Fakultät für Medizin

Neurologische Klinik und Poliklinik des Klinikums rechts der Isar

und

Institut für Humangenetik des Helmholtz Zentrums München

Genetic variants in restless legs syndrome and Parkinson's disease: the rare, the common and everything in between

Dr. med. Eva Christina Schulte

Vollständiger Abdruck der von der Fakultät für Medizin der Technischen Universität München zur Erlangung des akademischen Grades eines

Doctor of Philosophy (Ph.D.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Roland Schmid

Betreuerin: apl. Prof. Dr. Juliane Winkelmann

Prüfer der Dissertation:

1. Univ.-Prof. Dr. Bernhard Hemmer
2. Univ.-Prof. Dr. Thomas A. Meitinger
3. Univ.-Prof. Dr. Ortrud Steinlein, Ludwig-Maximilians-Universität München

Die Dissertation wurde am 01.08.2013 bei der Fakultät für Medizin der Technischen Universität München eingereicht und durch die Fakultät für Medizin am 14.11.2013 angenommen.

Für Johanna und Elisabeth

Table of Contents

Figures

Text Boxes

Tables

Abbreviations

AA	African American
ABI	Applied Biosystems, Inc.
AD	Alzheimer`s disease
APP	β amyloid precursor protein gene
ATP13A2	ATPase type 13A2 gene
auto dom	autosomal dominant
auto rec	autosomal recessive
BCA	balanced chromosomal abnormality
bp	base pairs
BST1	bone marrow stromal cell antigen 1 gene
BTBD9	BTB (POZ) domain containing 9 gene
Cas	CRISPR-associated
cf	confer
chr	chromosome
CNP	copy number polymorphism
CNS	central nervous system
CNV	copy number variant
CRISPR	clustered regularly interspaced short palindromic repeat
CSF	cerebrospinal fluid
ddNTPs	dideoxynucleotides (ddATP, ddCTP, ddGPT & ddTTP)
DET1	de-etiolated homolog 1 (*Arabidopsis*) gene
DMT1	divalent metal transporter 1 gene
DNA	desoxyribonucleic acid
DZ	dizygotic
e.g.	exempla gratia
EA	European American
EIF4G1	eukaryotic translation initiation factor 4G1 gene
ENCODE	Encyclopedia of DNA Elements
eQTL	expression quantitative trait locus
FBXO7	F-box protein 7 gene
FTLD	Fronto-temporal lobar degeneration
FUS	fused in sarcoma gene
GBA	glucocerebrosidase gene
GIGYF2	Grb10-interacting GYF protein 2 gene
GnRH	gonadotropin releasing hormone
GRN	progranulin gene
GTEx	genotype-tissue expression (NCBI project)
GWAS	genome-wide association study
HEK	human embryonic kidney
HLA	human leukocyte antigen
i.e.	id est

iPSCs	induced pluripotent stem cells
IRLSSG	International Restless Legs Syndrome Study Group
kb	kilobases
LBX1	ladybird homeobox 1 gene
LBXCOR1	ladybird homeobox corepressor 1 gene
LD	linkage disequilibrium
LDL	low-density lipoprotein
lincRNA	long intergenic non-coding RNA
LOD	logarithm of the odds
LRRK1	leucine-rich repeat kinase 1 gene
LRRK2	leucine-rich repeat kinase 2 gene
MAF	minor allele frequency
MALDI-TOF	matrix assisted laser desorption/ionization-time of flight
MAPT	microtubule-associated protein tau gene
Mb	megabases
MEIS1	myeloid ecotropic viral integration site 1 homolog (mouse) gene
miRNA	microRNA
MO	morpholino
MZ	monozygotic
NBIA	neurodegeneration with brain iron accumulation
ncRNA	non-coding RNA
NGS	next generation sequencing
NHLBI-ESP	National Heart, Lung and Blood Institute-Exome Sequencing Project
OR	odds ratio
PCR	polymerase chain reaction
PD	Parkinson`s disease
PINK1	PTEN-induced kinase gene
PLA2G6	phospholipase A2 group VI gene
PLMD	periodic limb movement disorder
PLMS	periodic limb movements in sleep
PLXNA4	plexin A4 gene
PRKN	parkin gene
PSEN1	presenilin 1 gene
PSEN2	presenilin 2 gene
PTEN	phosphatase and tensin homolog gene
PTPRD	protein tyrosine phosphatase receptor type delta gene
QTL	quantitative trait locus
ref.	reference
RLS	restless legs syndrome
RNA	ribonucleic acid
SCARB2	scavenger receptor class B member 2 gene
SKOR1	SKI family transcriptional corepressor 1 gene
SN	substantia nigra
SNCA	alpha-synuclein gene
SNP	single nucleotide polymorphism

SNpc	substantia nigra pars compacta
SNV	single nucleotide variant
TALE	three amino acid loop extension
TALEN	transcription activator-like effector nuclease
TARDBP	TAR DNA binding protein gene
TH	tyrosine hydroxylase
TOX3	TOX high mobility group box family member 3 gene
UCHL1	ubiquitin COOH-terminal hydrolase L1 gene
UK	United Kingdom
UTR	untranslated region
VPS35	vacuolar protein sorting-associated protein 35 gene
VUS	variant of unknown significance
wt	wildtype
ZFN	zinc finger nuclease

1 Introduction

1.1 Setting the Stage—Genetic Traits

Genetic disorders in the most classical sense, such as Huntington`s disease or cystic fibrosis, are inherited in a “Mendelian” (or monogenic) fashion, that is with only one major genetic alteration determining the trait of interest. However, with an overall prevalence of about 20/100,000 [1], monogenic diseases only account for a small portion of human traits believed to be influenced by genetic components. In many instances monogenic patterns of inheritance underlie rare diseases (prevalence less than 1/2000 per definition of the European Organisation for Rare Diseases[2]). In the far more prevalent portion of human diseases—the common diseases—multiple genetic and environmental factors collaborate in bringing about a specific phenotype (multifactorial diseases).[3]

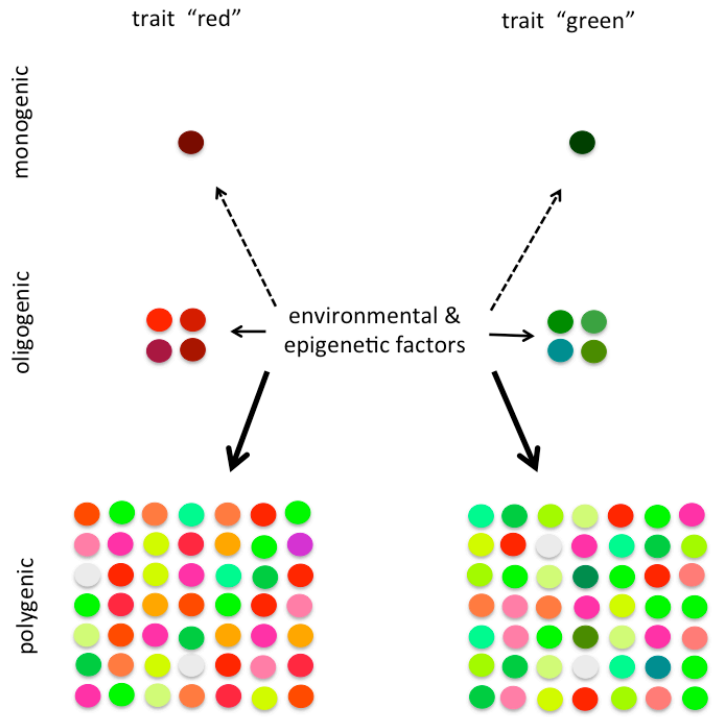


Figure 1.1: Schematic representation of the contribution of genetic variation as well as epigenetic and environmental factors to monogenic, oligogenic and polygenic diseases resulting in an overall phenotypic expression of either “red” or “green”. In complex polygenic diseases, the overall sum of genetic contribution results in the “red” or “green” trait although alleles predisposing for the opposite trait are also present. Lighter colors depict variants of smaller effect on a given phenotype and vice versa.

Such factors can include, but are not limited to, DNA modifications such as methylation or changes in chromatin conformation (epigenetics), aging or environmental influences (Figure 1.2).[4] Yet, this does not mean that in monogenic diseases additional genetic factors next to the causal variant, which is both necessary and sufficient to bring about the disease, do not act in phenotype modification.[5-7] Accordingly, monogenic and polygenic diseases are not likely to represent completely separate entities but rather two sides of a continuous spectrum (Figure 1.2).

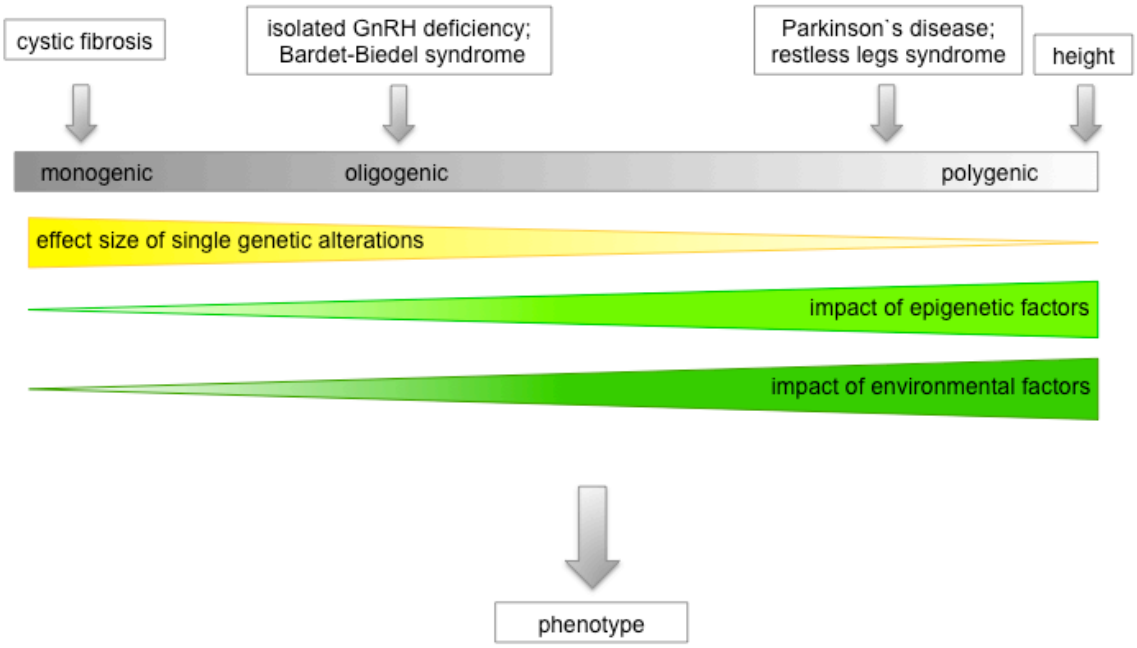


Figure 1.2: Spectrum of contribution of genetic and non-genetic factors to different forms of genetic disease. In complex genetic diseases such as most forms of Parkinson`s disease (or Parkinson disease, PD) or restless legs syndrome (RLS), many genetic variants alongside a significant contribution of epigenetic and environmental factors are supposed to bring about and also modify the phenotype. In monogenic diseases, the effect size of the causal variant is like to be much larger and the effects of external factors less pronounced. However, the exact contribution of the individual components involved in bringing about the phenotype are likely to differ from one phenotype to the next but often also within a given phenotype, creating a genetic framework unique to each (endo-) phenotype. Positioning of the different diseases on the spectrum of genetic complexity reflects the current best estimation of the genetic framework in the majority of individuals displaying a given trait and may vary for individual endophenotypes, such as in the case of monogenic familial PD.
GnRH=gonadotropin releasing hormone

1.2 Variance and Heritability

Especially with regard to non-monogenic diseases, in which several factors act in generating a given phenotype, it is relevant to understand the degree to which each factor contributes to the phenotype. Two central concepts in the study of trait-related genetic variation are heritability

and variance. The overall variance of a phenotype (V_P) is the sum of all independent genetic (V_G) and environmental (V_E) variation.

$$V_P = V_G + V_E$$

In other words, V_G is the contribution of genotypic differences among individuals to phenotypic variation. It can be broken down further to accommodate not only simple additive genetic effects (V_A) but also dominance effects at a given locus (V_D) and interaction effects (V_I).

Heritability (H^2, h^2), on the other hand, describes the proportion of V_P in a population that is due to either only additive (allelic, V_A) or all (V_G) genotypic differences among individuals.

$$H^2 = V_G/V_P \qquad \text{(broad-sense heritability)}$$
$$h^2 = V_A/V_P \qquad \text{(narrow-sense heritability)}$$

In the hypothetical scenario of a purely genetic disease, $V_P = V_G = H^2$. In reality, however, this scenario is unlikely, as is the assumption that the contributions of V_G and V_E to V_P are completely independent of each other. Heritability (h^2) estimates are usually derived from family or twin studies under the assumption of shared V_E and solely additive genetic variance.[4,8,9] It is important to note, however, that heritability describes the genetic contribution to a trait at the population and not the individual level. The contribution of a specific genetic variant to a phenotype is often given as a percent measure of h^2.

1.3 The Spectrum of Genetic Variation

Although nature has invested humans with a number of DNA repair and proofreading mechanisms[10,11], errors occur during DNA replication at a rate of approximately one every 10 to 100 million bases. Most sequence variation in humans can be attributed to such failures in DNA damage repair or the correction of replication errors.[8] Overall, the mutation rate has been estimated to be approximately 1.1×10^{-8} per site per generation[12-14] across the entire genome or 1.47×10^{-8} per person per generation[15] for non-synonymous changes. Within the protein-coding regions of the genome (exome), one single nucleotide variant (SNV) is found approximately every 52 base pairs (bp) when compared to the reference sequence.[16]

Known genetic variations in humans differ in size and composition (non-structural vs. structural). Other classifications include the location (within the coding vs. non-coding regions or intra- vs. intergenic regions), frequency and pathogenicity (also see discussion on the terminology in sections 1.3.2 and 1.3.3 below).

1.3.1 Different Classes of Genetic Variants

Genomic DNA can differ either with regard to its sequence (non-structural variation) or the number of copies that are present or an aberrant location or orientation (structural variation) in the form of so-called copy number variants (CNVs) and chromosomal abnormalities. Sizes can range from one nucleotide as in the form of most single nucleotide polymorphisms (SNPs) to CNVs of several thousand kb to duplications or deletions of entire chromosomes such as in Down or Turner syndrome.

Classically, the genome is divided into coding and non-coding parts—that is those that are translated into proteins and those that are not. Coding variants, in turn, can be either synonymous or non-synonymous depending on whether the amino acid sequence of the translated protein is changed by the variant or not (compare Figure 1.3).

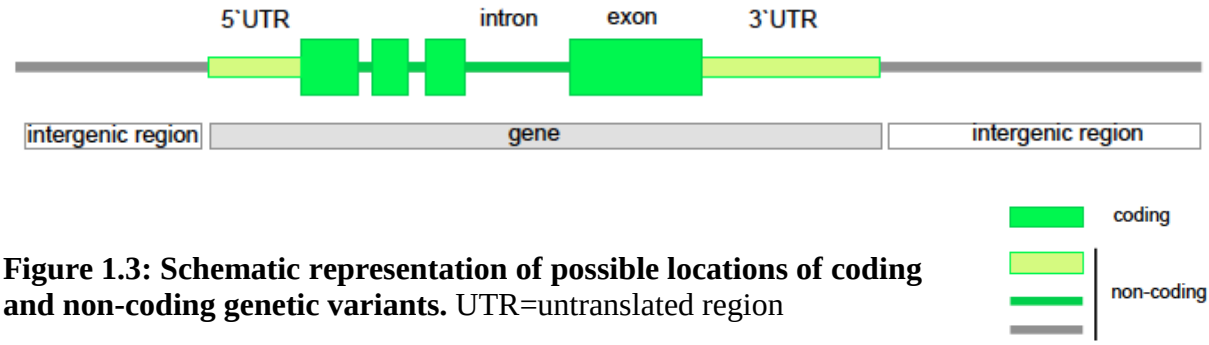


Figure 1.3: Schematic representation of possible locations of coding and non-coding genetic variants. UTR=untranslated region

Variants that alter the amino acid sequence can either result from single amino acid substitutions (missense variants), the introduction of a stop codon and a resulting truncation of the amino acid sequence (stop-loss variants) or the creation of a codon that does not naturally code for any amino acid (nonsense variants). Moreover, insertions or deletions (indels) of one to several hundred thousand nucleotides can also be found.[8]

Genetic variation is the norm and not the exception as demonstrated even by the first human genomes that were sequenced and published in 2001.[17,18] Compared to the reference sequence[19], the “Venter genome” supposed to belong to geneticist Craig Venter differed from the reference sequence in 12,290,978 nucleotide positions including 3.2 million SNVs, 849,000 heterozygous or homozygous indels (ranging from 1 to 82,711 bp), 90 large inversions and 62 large copy number variants.[8,17,18]

Furthermore, additional levels of genetic variation have been proposed such as the disruption of genomic DNA by retrotransposons (mobile genetic elements or ´jumping genes` belonging to the LINE1 and Alu families)[20-23] or the disputed extent of the introduction of sequence

changes at the RNA level (RNA editing)[24,25]. The notion is that we are currently just beginning to understand the factors involved in bringing about phenotypes and modifying them throughout an organism´s lifetime.

1.3.2 Genetic Variation by Frequency

The minor allele frequency (MAF) describes the frequency of the less common allele of two alleles at a given locus and is generally used to gage how frequently a variant occurs in a population. While strict definitions are lacking, if divided into two frequency categories—"common" and "rare"—variants in the first category usually have MAFs>5% whereas "rare" variants exhibit MAFs<5%. In recent years, an additional category of "low frequency" variants has been added, which most commonly comprises variants of 1%>MAF<5% or 0.5%>MAF<5%[26]. Accordingly, some publications only differentiate between common and rare variants, while others also describe low-frequency variants as a separate category. Throughout this work, wherever possible, all three frequency categories are addressed separately.

Although individually rare, variants with MAF ≤ 0.5% represent the most frequent variants while common variants with MAF>5% account for the largest portion of genetic differences between individuals[16,27] (Figure 1.4).

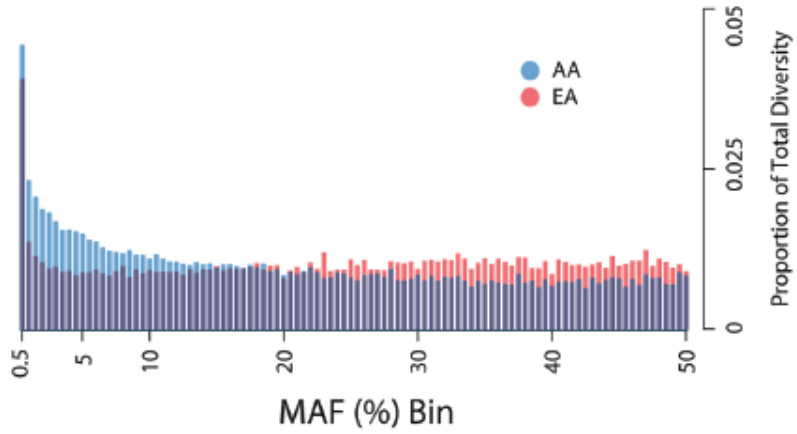


Figure 1.4: Rare variants with MAF ≤ 0.5% represent the proportionally largest contributor to genetic diversity in humans (taken from ref. [27]). AA=African American, EA=European American

It is important to realize that variant frequencies can differ between populations and that this can pose specific problems both in the analysis of common but also in the assessment of rare variants. These will be addressed in section 3.2.2 of the discussion.

1.3.2.1 Common Variants

Common variants are most frequently found as SNPs[28,29] but can also present as tandem or interspersed repeats or copy number polymorphisms (CNPs)[8]. While there are no strict definitions, SNPs with a MAF>5% are usually described as common. Throughout the human genome, one SNP is found approximately every 300 bp[8] and population geneticists project the existence of around 11 million SNPs with MAF>1% in the human genome[28,29]. Most SNPs have persisted from the earliest days of human evolution and most are found across all ethnic groups although often times at different allele frequencies.[8,16] They are frequently located in non-coding regions of the genome where some may serve gene regulatory functions.[30]
Disease associations are typically identified by means of genome-wide association studies (GWAS) in which the allele frequencies of currently more than one million SNPs across the genome are compared in case/control samples comprising tens to hundreds of thousands of individuals.[31]
With few exceptions[32], the disease-associated common variants identified thus far harbor only small effect sizes (odds ratio (OR) 1.1 to 1.5)[26,30,33] (Figure 1.6). According to the original "common disease, common variant" hypothesis[34-37], common, genetically complex diseases are caused by a collection of a few dozen loci of moderate effect. Yet, when looking at the hundreds of GWAS that have been published to date, this statement needs to be modified. When hundreds of thousands of individuals were examined for highly heritable traits such as height (heritability estimate = 0.8[38]), approximately 180 associated loci of small effect sizes were identified.[39,40] However, variants at these loci taken together only explain about 10.5 % of the total variance[40]. It is a common finding that the associated common variants identified as part of GWAS efforts across all different phenotypes rarely explain more than 5 to 10 % of the estimated heritability.[41] This dilemma has been termed "missing heritability"[33] and could be the result of a number of things: (1) the heritability estimates are inflated creating heritability that does not truly exist (so-called "phantom heritability")[9,42], (2) synthetic associations[43]—that is associations between common variants and a phenotype that in reality are not due to the common variant but due to (multiple) rare variants on the same haplotype—in certain constellations, (3) larger numbers of common variants of very small effects that the current GWAS are underpowered to detect[26,33,44,45], (4) rare variants of possibly stronger

effect that have escaped detection by GWAS because of their low MAFs[26,33,45], (5) structural variants[30,31,41], (6) epigenetic and environmental effects[30,31,41] (including parent-of-origin[46,47] and transgenerational effects), (7) (phase-dependent) interactions between genes[30,31,41]. Since the genetic architecture differs between diseases, it can be assumed that the "missing heritability" can be attributed to different causes in different diseases.

1.3.2.2 Low Frequency Variants

In between common and rare variants, low frequency variants, roughly defined by MAFs between 0.5 and 5 %[26] populate the middle ground. These variants have not received very much attention in the past because they were neither covered by the available SNP arrays used in GWAS (although some current generation arrays do cover variants down to MAFs of about 1% and variants of this frequency range can also be analyzed by imputation) nor were they detectable by linkage analyses[48-51]. Recently, newer technologies such as next generation sequencing (NGS, described in section 1.5.4 below) have been used to identify some low-frequency alleles associated with complex diseases[52,53]. These so-called "goldilocks" alleles[41] may actually prove very valuable to the study of genetic variation in complex diseases. Although this class of variants will often only account for a minor fraction of the population attributable risk, they are still frequent enough to be used in high-powered population-based studies (i.e. GWAS) while at the same time harboring effect sizes large enough to make them subject to purifying selection and, at the same time, suitable to functional analysis.[26,41,54-56] The effect of purifying selection is reflected in the fact that they were found to be 1.8 fold more likely predicted to be non-synonymous and deleterious than synonymous and benign.[57] In one study of 202 drug target genes sequenced in 14,002 people, goldilocks alleles (here defined as $0.5\% < MAF \geq 2\%$) were found in 73 genes and about half of the variants were predicted to be functional, compared to 31% of variants with MAF > 2% and 65% of singletons (also see section 1.3.2.3 below).[54]

1.3.2.3 Rare and Very Rare Variants

Variants with a MAF < 0.5% are generally referred to as rare variants. However, this category is broad, ranging from variants with MAF = 0.5% that can be found in approximately 400,000 German citizens to "singletons", that is variants that are only found in a single individual or family.

Generally, rare variants are expected to harbor larger effect sizes than more common variants (Figure 1.6) and are subject to purifying selection. Purifying selection describes nature´s

tendency to “purify” the genetic pool by selecting against variants that are directly disease-related and thus deleterious to an individual`s fitness to reproduce.[4] (Figure 1.5) Thus, comparatively deleterious variants stay rare within a population and reach MAFs near 1% only occasionally in recessive diseases.[4]

Overall, an excess of rare variants has been reported[27,54,58] with rare variants expected to effect approximately 1 out of every 2.5 intragenic bp in a linearly extrapolated sample of 1 million individuals.[54] This phenomenon has been ascribed to both rapid population growth over the past 1,400 years and relatively weak purifying selection in modern day societies (Figure 1.5).[27,54,58,59] In line with this, rare variants are also younger[16] and more likely to be non-synonymous and to harbor (predicted) functional effects[54,60] than more common alleles.

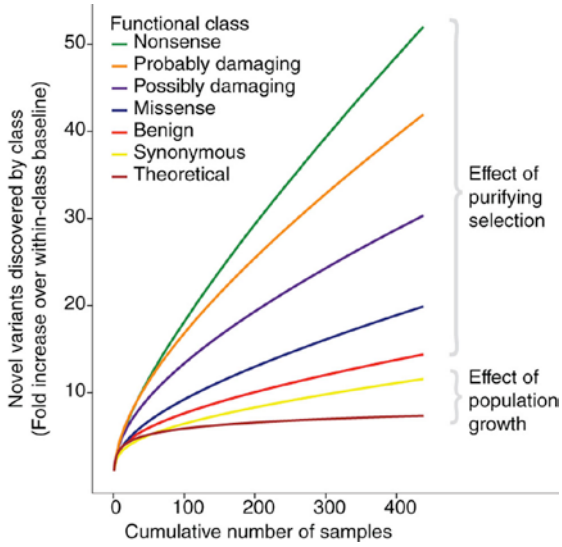


Figure 1.5: An excess of nonsense, missense and variants predicted to be damaging is found in human exomes due to the effect of purifying selection. To a lesser extent, explosive population growth has also sparked an increase in the number of (rare) variants found in human exomes across all functional classes. (taken from ref. [61])

Consequently, this argues for a distinctive value in studying rare variants in any disease context but also suggests that very large and homogeneous samples will need to be evaluated to statistically link rare variants to complex diseases.[54,61] Many presumably monogenic diseases are caused by rare variants of large effect. Traditionally, these were identified by linkage analysis and more recently also by whole-exome sequencing in families in which a single rare variant segregates with the disease.[62-64] Under the “common disease, rare variant” hypothesis[65-67], it has also been postulated that rare variants contribute to complex diseases. Several such rare variants have been identified via the re-analysis of GWAS data and next-generation sequencing.[68-75]

1.3.3 Variants vs. Mutations vs. Polymorphisms ?

Generally, the terms “variant”, “mutation”, and “polymorphism” can refer to an alteration in the nucleotide sequence of the DNA. The three terms are used by different branches of genetics in different ways. To molecular geneticist, “polymorphisms” are variants with MAF > 1%, while to population geneticists they are variants that stably coexist in a population at frequencies that cannot be explained by recurring mutations and to clinical geneticists they are simply non-pathogenic variants.[76] Along the same lines, “mutation” is sometimes used to describe variants of pathogenic effect but sometimes simply for variants of very low frequency.[76] For the sake of neutrality and lucidity, in this work, the term “variant” will be used to describe genetic alterations regardless of their frequency or supposed pathogenicity; “polymorphism” shall only be utilized in the context of SNPs, as defined above. In rare instances, variants that are accepted as being causal for a disease by the research community in the field will be referred to as “mutations”.

1.4 Different Effect Sizes

Next to the frequency, genetic variants also differ in the strength of the effect that they bring to the development of a given phenotype. This effect size is commonly measured as odds ratio (OR). An OR of 2, for example, would indicate that a carrier of the given variant is twice as likely to suffer from a given disease or display a given trait than one who does not harbor this variant.

Apart from a few exceptions[32], there appears to be an inverse correlation between the frequency of a given variant in the population and its effect size.[26] As explained above, this is mostly owed to the selective pressure that effects variants of large effect more than those of small effect. Today, there are strategies in place to examine all variants along this axis–money and sample number permitting[31] (also see Figure 1.6 below). In contrast, the discovery of rare variants of small effect (or reduced penetrance), which liably also exist and play a role in phenotype modification in complex diseases[77], will remain very difficult or impossible by genetic means.

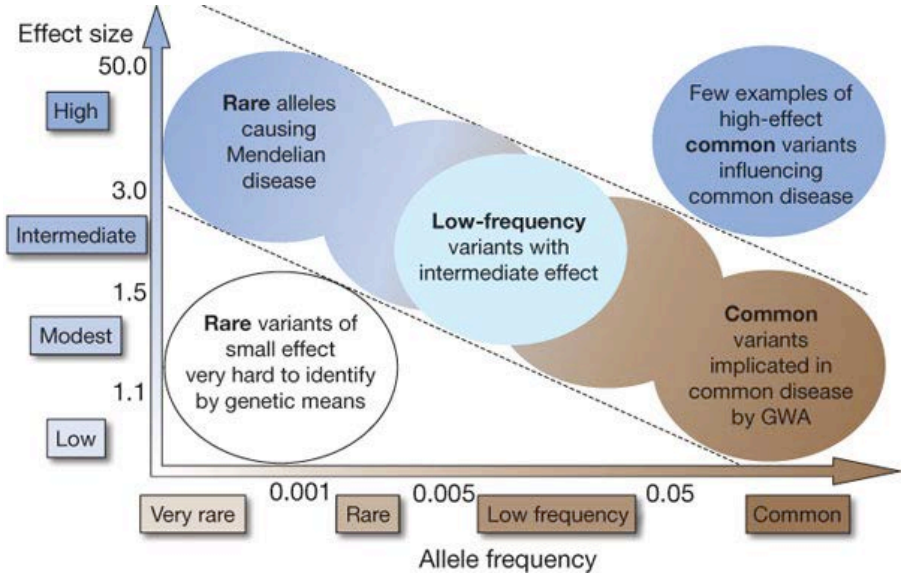


Figure 1.6: Relationship between allele frequency and effect size of a genetic variant. Generally, the rarer the variant, the stronger its effect on a given phenotype. (taken from ref. [26])

1.5 Methods of Investigating Genetic Variation

Approaches used in evaluating genetic variation vary depending on both the study sample to be used and the projected frequency and effect sizes of the variants to be analyzed. Also, these strategies are subject to a constant flux as new technologies become available[42] and especially in the field of human genetics, current "next generation" (or "second generation") sequencing technologies are vastly different now from what they were 10 years ago.[78] The past decade has seen a paradigm shift from sequence production to, currently, sequence analysis to soon sequence storage as the rate limiting factors in variant discovery.[41,78] While rare Mendelian diseases have proven well-suited to linkage analysis and especially exome sequencing, GWAS have contributed enormously to the identification of susceptibility alleles for common complex diseases.[79] (Figure 1.7)

Yet, as it becomes increasingly clear that not only common variants contribute to the genetic framework of complex genetic diseases, it has been argued that a "holistic" approach to the study of complex genetic diseases—combining family, case/control and extreme phenotype studies and all currently available technical approaches as illustrated in Figure 1.7 below—is

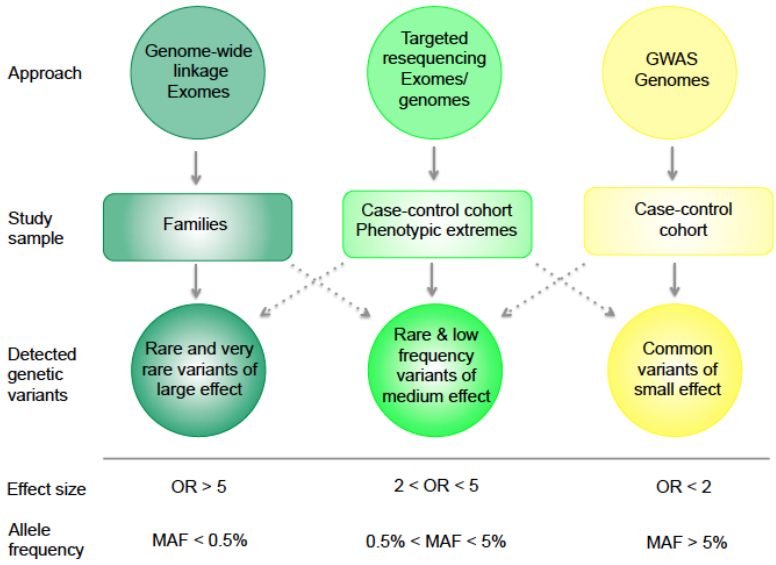


Figure 1.7: Strategies of analyzing genetic variation. (modified after ref. [31])

imperative.[41] This should become feasible within the next decade as tens to hundreds of thousands of samples become available through the collaborative efforts of consortia[41], the sequencing costs continue to drop and even newer, "third generation" sequencing technologies[80] become available.

In the dissertation work depicted herein, several of the above approaches were utilized to study the involvement of genetic variants of different frequency and effect sizes in two genetically complex neurogenetic diseases—Restless legs syndrome (RLS) and Parkinson`s disease (PD).

1.5.1 Association studies

The contribution of common variants to the genetic architecture of a disease is frequently assessed via association studies, which compare allele or genotype frequencies of genetic variants between unrelated affected (i.e. cases) and unaffected (i.e. controls) individuals within a given population. Controls can either be general population controls in which one can find the phenotype of interest at the same prevalence as in the general population or disease-free controls, which have been screened for the phenotype of interest (i.e. "super controls"). The controls used in all aspects of this work belong to the KORA (Kooperative Gesundheitsforschung in der Region Augsburg) general population cohort[81] or the KORA-AGE sub-survey of KORA, which specifically interrogates individuals born before 1944.[82]

A given variant is said to be associated with the phenotype if there is a statistically significant difference in variant frequency between cases and controls.[83,84] Accordingly, common variants are more suitable as genomic markers in association studies because they invest studies with much higher power to detect associations than rare variants.

Although current generation common variant genotyping arrays assess up to 4.3 million SNPs[85], this represents only a fraction of the genetic variation present in a given individual. However, alleles at genetic loci are not inherited independently but in blocks of 10 to 200kb, which are often times separated by recombination hotspots[86,87], and are, therefore, found more often together than expected by chance as reflected by the linkage disequilibrium (LD; r^2 ranging from 0 (no LD) to 1 (perfect LD)) between them[88,89]. Accordingly, tagging SNPs can be utilized to indirectly evaluate many more genetic variants known to be inherited together with the genotyped variant. Conversely, this also means that an association between a genotyped SNP and a phenotype is not necessarily caused by precisely that common variant but could be dependent upon any number of genetic variants—both common or rare—passed on together with the genotyped variant.

For genetically complex diseases, where additive effects are usually assumed[90], the Armitage trend test provides a powerful statistical tool which incorporates a suspected genotype-dependent ordering of effects[56,91,92] and was applied in the statistical analyses of the work illustrated below. Association studies can analyze either candidate genomic regions in a hypothesis-driven manner[93] or variants covering the entire genome in a hypothesis-free approach[94-96]. Association may also be analyzed between genotype and clinical phenotype directly or via an intermediate phenotype (i.e. a quantitative trait locus (QTL)) such as gene expression (eQTL) or metabolite levels[97-100].

In the work depicted herein, genotyping was performed either using genome-wide SNP arrays (Affymetrix® Genome-Wide Human SNP Arrays 5.0 and 6.0)[I,II] or on the Sequenom® MassARRAY system using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry with iPLEX Gold chemistry.[I-X]

1.5.2 Variant Screening

Next to the analysis of known variants by the genotyping methods outlined above, variant discovery also represents a central methodology in the study of genetic variants.

1.5.2.1 Sanger Sequencing

Since 1975, when Frederick Sanger first described a method of DNA sequencing by *in vitro* DNA replication using chain-terminating dideoxynucleotides (ddNTPs) by a DNA polymerase, "Sanger" sequencing has been the most widely used method in DNA sequence analysis.[101,102] In contemporary Sanger sequencing, an automated format using fluorescently labeled ddNTPs and polyacrylamide-gel-based size-fractionation in glass capillaries is employed.[8] Though arduous compared to other currently available sequencing technologies, it still counts as the most reliable method in DNA sequence detection and can be used to identify both known but more importantly also novel genetic variants. Sanger sequencing on an ABI Prism® 3730 automated sequencer using BigDye chemistry was used in a number of projects depicted in this work, either for variant discovery[VII,IX], the validation of variants discovered by medium- or high-throughput approaches such as high-resolution melting curve analysis or next generation sequencing[III-X], segregation testing in families[VII-X] or as fragment analysis in the evaluation of haplotypes[VI] or small indels[VII].

1.5.2.2 High-Resolution Melting Curve Analysis

High-resolution melting curve analysis on the LightScanner® system (Idaho Technology, Inc./BioFire Diagnostics, Inc.) represents an alternative method used in variant screening studies.[103-108] As it represents one of the central methods to the work depicted in this dissertation thesis, it shall be explained below in more detail.

When the melting process is visualized over time, each double-stranded DNA fragment has a characteristic melting temperature and pattern depending on its nucleotide composition. This can be utilized to identify DNA fragments with differences in the nucleotide composition by high-resolution melting curve analysis using the LightScanner® system. Here, PCR amplification of a genomic region of interest is performed in a 384-well format in the presence of a fluorescent dye (LCGreen; excitation 440 to 470 nm, emission 470 to 520 nm) that labels double-stranded DNA. In a second step, the melting of the amplified and labeled DNA is visualized by laser detection using the LightScanner®. After normalization of the recorded melting curves to wildtype DNA, DNAs with differing nucleotide composition can easily be recognized due to their altered melting profile (Figure 1.8).[103-108] In order to determine the exact underlying nucleotide alteration, the PCR product is, in a third step, sequenced using the Sanger method. Both heterozygous and homozygous variants as well as small indels can be identified using high resolution melting curve analysis. In a diagnostic context, where known heterozygous *BRCA1* mutations were detected within previously

optimized PCR amplicons, sensitivity was reported to be 100% with an average specificity of 98%. However, 3.4% of false positives where also seen in the same sample comprising 155 DNAs and only 93% of the known homozygous mutations were found even on repeated evaluation of multiple amplicons for the same stretch of DNA.[109] Moreover,

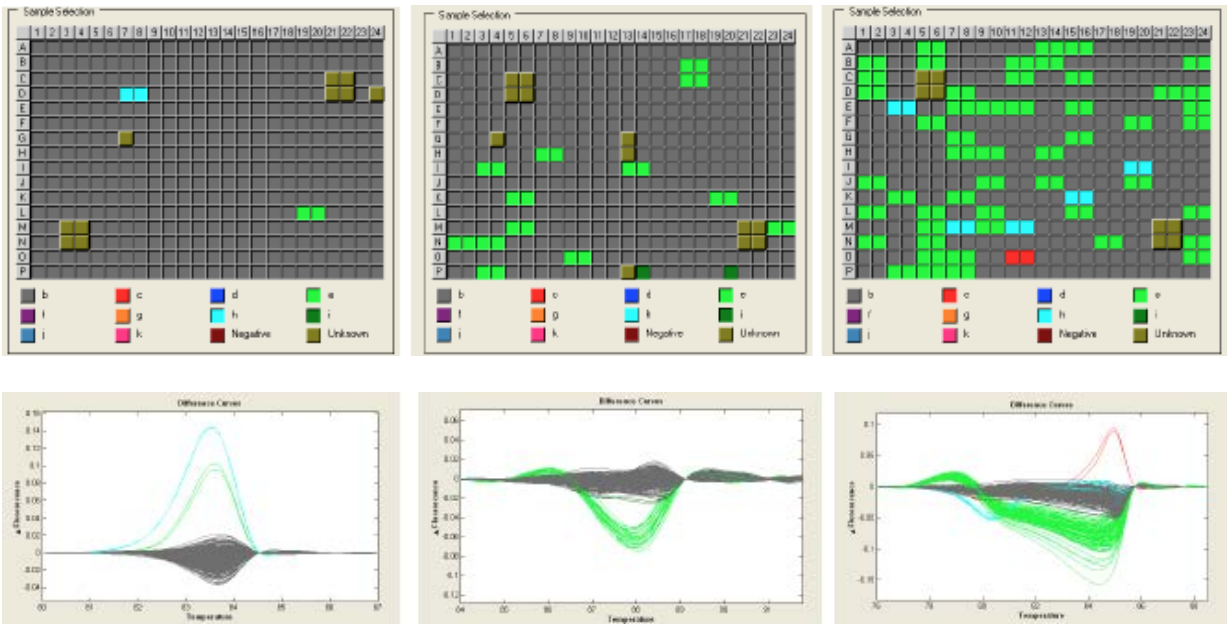


Figure 1.8: Examples of variant detection using high-resolution melting curve analysis (LightScanner®). The left panel depicts the position on the 384-well PCR plate (top) as well as the aberrant melting patterns (bottom) of two different rare variants. The central panel shows the same for a low-frequency variant present in heterozygous form in 13 out of 188 DNA samples (MAF = 3.5 %) tested. In the panel on the right, both a common variant present in the heterozygous (green) and homozygous (turquoise) state as well as an additional rare variant (red) is seen. All samples are run in duplicates for increased assay sensitivity.

additional predicaments specific to the identification of novel variants exist. For example, sensitivity varies with the location of the variant within the amplicon (peripheral vs. central), the overall variance within the amplicon, the GC-content (optimal range: 31% to 54%), the physical distance between the variants, the zygocity state (heterozygous variants, which alter the shape of the melting curve, are more easily detected than homozygous ones, which may only alter the melting temperature[110]) and the variant class (single bp exchanges are more readily detected than indels).[109, personal observation]

In the work described here, high-resolution melting curve analysis was used for variant screening entailing the coding regions and splice boundaries (± 10 bp) and—in some instance—the 5`and 3` untranslated regions (UTRs) of a total of 19 candidate genes for PD or RLS.[III-X]

1.5.3 Linkage Analysis

Linkage studies detect the co-segregation of genetic loci (as defined by specific genetic markers such as microsatellites or, more recently, SNPs) and a disease in a given family. A marker that is located in close physical proximity to a disease-causing genetic variant is inherited together with the causal variant more frequently than expected by chance because of the reduced meiotic recombination frequency between the two. Consequently, markers can be utilized to define candidate genomic regions expected to harbor the disease-causing variant.[8,111] Prerequisites for successful linkage studies include an unambiguous segregation patter in the family to be evaluated, a family structure resulting in sufficient number of meioses (i.e. several generations of affected individuals) and an underlying genetic variant of fairly large effect on disease risk ($OR \geq 4$)[37,48-50]. Most often, linkage analyses are encountered in a primary context in which they are used to highlight genomic regions suspected to be home to disease-causing genetic variants. In the context of whole exome or whole genome sequencing studies, however, it has also been shown that linkage signals below the generally accepted threshold for statistical significance (logarithm of the odds (LOD) score > 3.3) can be used as secondary analysis to prioritize candidate variants obtained independently[112,113]. As part of this work, parametric linkage analysis using a subset of 12,875 SNPs genotyped on the 500K oligonucleotide SNP array (Illumina) was used to prioritize variants identified by whole exome sequencing in six members of a family with autosomal dominant PD^X as part of a collaboration with Dr. Darina Czamara and Prof. Bertram Müller-Myhsok (both Max-Planck Institut für Psychiatrie, Munich, Germany).

1.5.4 Next-Generation Sequencing

Starting in 2005, the first instruments of a completely new generation of sequencing technology became available. The specifics of "massively parallel" or "next generation" sequencing differ by the company, which developed it (reviewed in ref. [114]). In the next generation sequencing (NGS) studies performed as part of this work, Illumina`s sequencing-by-synthesis technology was used. Here, in brief, a library of DNA fragments is prepared and via the addition of adaptors attached to a glass surface where it is amplified in an enzymatic reaction. A polymerase incorporates fluorescently labeled terminator nucleotides into growing DNA strands. After each cycle, the fluorescence signal is imaged and the fluorescent dye and the 3` blocking group that prevents elongation by more than one nucleotide at a time is chemically cleaved away. After a washing step, the cycle starts over.[78,115] DNA libraries were sequenced as paired-end reads of 100 bp length each, meaning that sequences were

determined from both ends of a linear DNA fragment. The three most basic applications of NGS are whole-genome, whole-exome and targeted resequencing of specific genes or genomic regions of interest. While per base, NGS is much cheaper and much faster than any previous sequencing approach, sequencing entire genomes is still costly and time-consuming.[78] Accordingly, the vast majority of NGS studies published to date have relied on whole exome sequencing, that is the analysis of solely those approximately 2% of the genome that are known to be protein-coding[e.g.3,62-64,116], and only few have assessed whole genomes[12,113,117]. Like targeted resequencing of genomic regions of interest, whole exome sequencing involves the targeted capture of predefined regions of the genome (i.e. all exons). In this work, Agilent`s SureSelect Human All Exon 38Mb and 50Mb kits were used for in-solution enrichment of coding sequences of DNA derived from peripheral blood lymphocytes. While initially the focus was placed on the study of familial diseases as here as little as one or a few exomes or genomes are sufficient to successfully identify disease genes, recently, as the cost of NGS continued to drop, first studies applying a case/control sample design known from GWAS to whole exome or targeted resequencing have been performed[6,70,73,74,118].
While the ability to sequence entire exomes or genomes is one thing, the ability to analyze the generated sequences is another. First, the generated sequences need to be aligned to a reference genome, then, deviations from the reference genome can be identified. In the exome sequencing projects described herein[VIII-X], read alignment was carried out with Burrows-Wheeler Aligner (BWA, version 0.5.8)[119] to the human genome assembly hg19. SNVs and indels were detected using SAMtools (version 0.1.7)[120]. Both exome sequencing and bioinformatics analysis were performed using the Institut für Humangenetik`s in-house exome sequencing pipeline established and run by PD Dr. Tim Strom (Institut für Humangenetik, Helmholtz Zentrum München, Munich, Germany).
Currently, 20,000 to 50,000 coding variants that differ from the human reference genome will be identified per individual in a typical exome.[3] Accordingly, the next challenge is to dissect out the one or a few disease-causing alleles from this large number of potential candidates.[121] A number of strategies based on (1) variant class[37,64], (2) variant frequency in public databases (dbSNP[122], 1000 genomes[123,124], NHLBI-ESP exomes[125]), the in house exome database or independent case/control samples, (3) conservation[126] and predicted functional relevance[121,127] and (4), in the case of family studies, segregation with the disease phenotype have been devised to address this challenge[3,121] and to reduce the number of potential variants to be followed up to a more manageable number. An example of the filtering algorithm used in the exome sequencing studies depicted in this work is given below (Figure 1.9).

Although NGS is most frequently encountered in the context of studies seeking trait-associated genetic variants, additional applications have been developed and are likely to gain increased importance in the future. These include but are not limited to studies of RNA and small RNA expression (RNA-Seq), the exact placement of regulatory DNA binding proteins (e.g. chromatin immunoprecipitation and sequencing (ChIP-Seq)) and genome-wide DNA methylation profiles.[78,128]

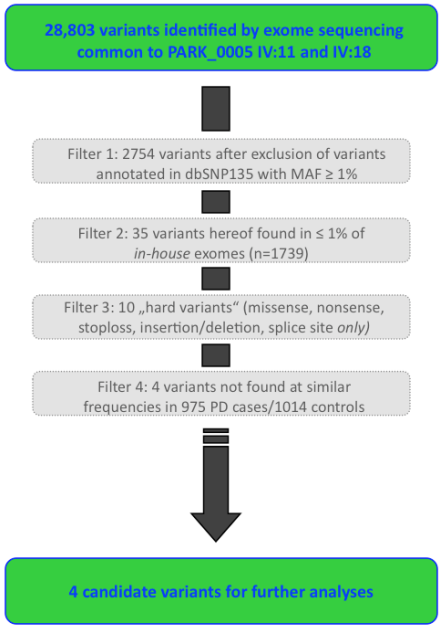


Figure 1.9: Filtering algorithm to prioritize variants identified by whole exome sequencing for follow up studies. First, variants common to two affected individuals from a family with autosomal dominant PD (PARK_0005) were filtered to remove common and low-frequency variants found in dbSNP135 and 1739 in house exomes. Then, variants were filtered by variant class to include only variant classes likely to comprise disease-causing variants. Lastly, variant frequency was assessed in a case/control sample and variants found at similar frequencies in both cases and controls were excluded, thus reducing the number of potential candidate variants from 28,803 to four.[X]

1.5.5 Rare and Low Frequency Variant Statistics

One of the most challenging aspects of all rare variant studies is the statistical evaluation of a potential link to a given phenotype, as the investigated variants are, by definition, rare and, accordingly, analyses are chronically underpowered[129-131]. According to current estimates, several tens of thousands of cases and controls—and likely even more—will be needed to confidently link rare variants to specific phenotypes[27,61]. The simplest way of statistically analyzing a possible contribution of rare variants to a specific phenotype is by simple group comparisons of the number of occurrences of a specific rare variant in cases vs. controls such as by Fisher`s exact or χ^2 tests.[61,132] However, in nearly all instances, this approach will be

unsuccessful due to a lack of power.[132] Consequently, alternate statistical analyses tools needed to be developed. The "collapsing" of variants provides one such tool[45,55,131-134]. It was first proposed in 2007 in its simplest form as the cohort allelic sum test (CAST)[135], which utilizes standard contingency table-based Fisher`s exact or χ^2 tests. To increase power, variants can be collapsed across a specific (1) genomic region, gene(s) or region of a gene, (2) variant class, (3) variant frequency[70], (4) predicted or experimentally ascertained function[62,63,70] or (5) pathway[136,137]. A number of extensions of the collapsing method integrating multivariate analysis (combined multivariate and collapsing method (CMC)[131]), the analysis of variants of different frequencies (weighted sum statistics (WSS)[138]) or different predicted functional impact (variable weight test (VWT)[55]) or different direction (adaptive sums test[139]) or size (kernel-based adaptive cluster test (KBAC)[140]) of effect have been published and it has been demonstrated that these invest the statistical analysis of rare variants with more power than CAST[132]. However, it has also been shown that for aggregate rare and low frequency variant analyses the genetic variance for a gene or a genomic region will always be underestimated because it is highly unlikely that all variants at a given locus will be disease-relevant and even the disease-related variants will have different effect sizes and directions.[141] With regard to performing rare variant statistics, it is also imperative to note, that the analysis of the most homogenous samples possible is imperative as population substructure and admixture play much larger roles in rare variant analyses as these variants show much larger population diversity than common variants.[45,142] Even within European populations, where common variation is fairly constant, rare variants show large differences[54]. In the work illustrated herein, CAST was used for aggregate analysis of rare and low frequency variants collapsed across genes, variant class, variant frequency, predicted functional impact and experimentally determined functionality.[IV,VII,IX,X]

1.5.6 Functional Assessment

NGS has opened the door to the discovery of innumerable genetic variants.[121] Now, it is no longer the rate of variant discovery that represents the limiting factor to uncovering the genetic factors underlying diseases but rather the development of bioinformatics capacities (Moore`s and Kryder`s Laws) and the functional analysis of variants.[41]

Evidence derived from variant frequency, segregation analyses, variant class, positional conservation, location directly or from *in silico* prediction of functionality (using prediction algorithms such as PolyPhen2[143], SIFT/PROVEAN[144,145] or MutationTaster[146] or a combination of a number of prediction algorithms such as described in[IX]) can provide some

idea of the likelihood with which a given variant will harbor a functional effect of pathophysiologic relevance.[121] Yet, prediction tools are notoriously inaccurate.[147] Accordingly, especially with regard to rare variants—where statistical evidence may not be very convincing due to impeded statistical power and where very rare variants or even singletons represent the largest fraction[16,27,54]—experimental functional analysis is vital to judge the nature of an identified variant.

In the work depicted herein, a number of standard experimental set-ups commonly employed in molecular biology were used to assess the functional effect of identified variants. These include the analysis of variant-dependent RNA and protein expression and localization by PCR, immunocytochemistry or Western blot in primary patient-derived fibroblasts or transfected HEK-T or SH-SY5Y cells as well as cell death detection by flow cytometry. However, the analysis of large numbers of variants by conventional cell biological and biochemical analyses is time consuming and more efficient strategies are needed to facilitate the evaluation of potentially pathophysiologically relevant variants and genes identified by various NGS approaches.[41] In this work, an *in vivo* complementation assay in zebrafish was employed as a medium-throughput way to evaluate the functionality of given coding variants[IV] and shall, therefore, be described in more depth.

In this assay, translation- or splice-blocking morpholinos (Gene Tools, Inc.) are used to ablate functional protein of a gene of interest in zebrafish embryos. Morpholinos are synthetic antisense RNA oligomers of approximately 25 bp in length that either prevent protein production or pre-mRNA splicing[148,149]. The resulting morphant phenotype is assessed and—where possible—analyzed quantitatively. To evaluate the effect of individual genetic variants, human cDNA of the gene of interest is cloned into the pCS2+ Gateway® destination expression vector. Subsequently, missense mutations are introduced into the open reading frame of the gene of interest using QuikChange® site-directed mutagenesis (Stratagene) and confirmed by Sanger sequencing. Morpholinos are injected into wildtype zebrafish embryos at the two-cell stage along with mRNA transcribed from linearized pCS2+-cDNA harboring the wildtype sequence or a specific variant. The wildtype mRNA should be able to rescue the morpholino-induced phenotype. With regard to the variants, however, this rescue capacity may be lacking (i.e. a null allele), attenuated (i.e. a hypomorphic allele) or uncompromised (i.e. a benign allele), thus providing information on the impact of a specific variant on the functionality of the gene of interest (also cf. Figure 1.10 panel A). Rescue capacity is usually evaluated in two rounds of approximately 100 injected zebrafish embryos each for each variant and either quantified—if the morphant phenotype allows this—or grouped into

different severity classes. The actual phenotype developed by the embryos is *not* necessarily related to the disease phenotype. Rather, the *in vivo* complementation assay is designed to test the impact of a genetic variant on the general function of a gene of interest in a disease-unspecific yet semi-high-throughput fashion. Figure 1.10 below gives an example of the *in vivo* complementation assay as used to assess the functionality of rare variants in the RLS-associated gene *MEIS1*. *In vivo* complementation assays described here were performed in conjunction with Dr. Maria Kousi and Prof. Dr. Nicholas Katsanis at the Center for Human Disease Modeling, Duke University, Durham, NC, USA.

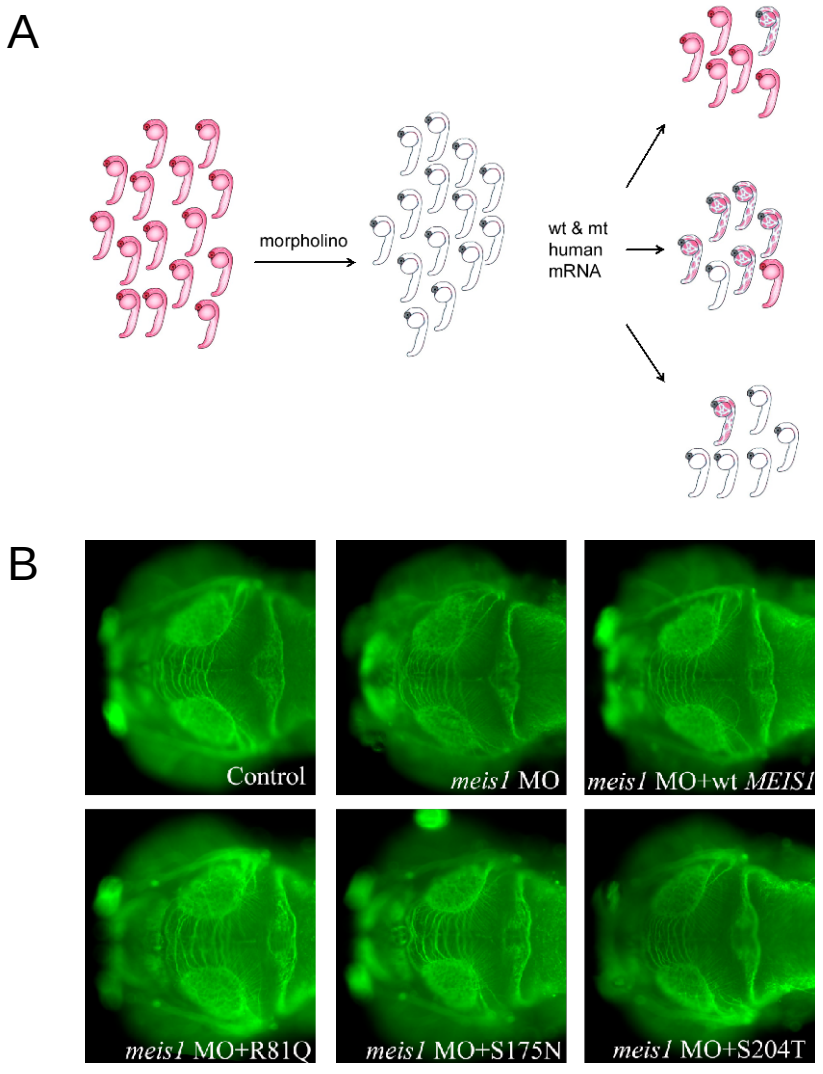


Figure 1.10: (A) Schematic depiction and (B) examples of the *in vivo* complementation assay in zebrafish. Injection of zebrafish eggs with a morpholino targeting *meis1* leads to a reduction in optic tectum size, which can be restored by co-injection with human wildtype *MEIS1* mRNA. (B, top row) Co-injection of *meis1* morpholino and *MEIS1* mRNA containing a number of different non-synonymous variants identified in patients with RLS leads to variably successful restoration of the optic tectum phenotype suggesting functional effects of differing severity on *MEIS1* (B, bottom row; benign, hypomorphic and null allele left to right).[IV]

1.6 Restless Legs Syndrome (adapted from book chapters published as ref. XI,XII)

1.6.1 Clinical Phenotype of RLS

1.6.1.1 Epidemiology

Age- and sex-dependent prevalences around 10% in adult populations of European descent render RLS one of the most common neurologic disorders overall[150,151]. In populations of Asian descent, prevelances are lower, ranging from around 1.57% in Taiwan[152] to 3.2% in Turkey[153]. However, a female preponderance (female to male ratio approx. 1.4:1.0[154]) is found in most populations examined so far. The cause for this overrepresentation of females in the RLS patient population is unclear to date. One study proposed that it could be related to the fact that pregnancies are associated with an increased risk of RLS.[155] The fact that brother-brother pairs have a higher RLS correlation than brother-sister or sister-sister pairs[156] also argues that non-genetic factors could contribute to the increased prevalence of RLS among women.

1.6.1.2 Definition

To date, there are no known biomarkers for RLS. During polysomnography, approximately 80% of individuals with RLS show an increased occurrence of periodic limb movements in sleep (PLMS)[157]. PLMS are defined as movements of >0.5s duration, occurring at 4 to 90 s intervals and of an amplitude of at least 25% of the calibration amplitude[158]. However, they are unspecific and also seen in other sleep disorders such as narcolepsy and rapid eye movement sleep behavior disorder (RBD) as well as in healthy individuals.[159] Accordingly, RLS remains a clinical diagnosis based largely on the patient`s account of his or her symptoms. The International RLS Study Group (IRLSSG) set forth four essential criteria as well as supportive and associated diagnostic features to be used in the diagnosis of RLS[160]. The essential criteria include the following: (1) dysesthesias affecting the legs, (2) triggered by periods of rest or inactivity, (3) relieved by movement and (4) occurring mostly during the evening and at night. Supportive criteria and associated features complete the current diagnostic criteria[160]. (Text Box 1.1) Specificity of the four essential diagnostic criteria has been estimated to be about 84%, with most of the diagnostic accuracy ascribed to the first and fourth criterion[161].

Essential criteria

1. an urge to move the legs, usually accompanied or caused by uncomfortable and unpleasant sensations in the legs

2. the urge to move or unpleasant sensations begin or worsen during periods of rest or inactivity such as lying or sitting

3. the urge to move or unpleasant sensations are partially or totally relieved by movement, at least as long as the activity continues

4. the urge to move or unpleasant sensations are worse in the evening or at night than during the day or only occur in the evening or at night

Supportive features

1. positive family history

2. positive response to levodopa or dopamine agonists

3. periodic limb movements in sleep

Associated features

1. a progressive and fluctuating natural clinical course of disease

2. sleep disturbances

3. normal physical examination findings in cases of primary RLS or findings of underlying conditions in cases of secondary RLS

Text Box 1.1: Essential criteria and supportive and associated features comprising the RLS diagnostic criteria set forth by the IRLSSG[160].

1.6.1.3 The Clinical Phenotype

Subjectively, patients often times find it very difficult to give an accurate rendition of the sensory phenomena they experience. The description of dysesthesia qualities is consequently broad and can vary from “tingeling” and “electrifying” sometimes “painful” sensations to “pulling”, “working”, “tensing” and “rumbeling” to “itching” and “pulsations” or “vibrations” which are classically felt deep inside the affected extremities. Dysesthesias typically begin after 10 to 20 minutes of rest but the time to the onset of symptoms can vary with severity and, in very severe cases, symptoms can be present near continuously. Emblematic accounts usually describe the occurrence of symptoms during periods of rest such as when falling asleep, during meetings in the afternoon or at night, during visits to the theater

or cinema or while watching TV at night, long bus, airplane or car rides or immobilization of the legs after surgery. There is a clear preponderance for the evening and night time, with the hours between midnight and 2 am representing the symptom maximum[162,163]. Moving the legs and feet or getting up and walking around usually provides prompt symptom relief. Patients have also reported to benefit from other sensory stimuli such as massaging their legs with cooling gels or cold showers. Moreover, >90% respond positively to treatment with dopaminergic drugs[164]. Generally, symptom intensity and frequency increase with disease duration and it is common to find symptom-free intervals of several weeks or months during the early stages of disease progressing to quotidian symptoms after a course of several years or decades.

Consequences of leg discomfort and restlessness are severe disturbances in sleep architecture and quality including a fragmented sleep profile with frequent changes in sleep stages and arousals as well increased stage 1 sleep, extended wake periods after sleep onset and increased latency to sleep onset[165-169]. Clinically, this results in difficulties in both falling and staying asleep as well as daytime sleepiness and fatigue[166]. Furthermore, RLS has also been linked to an increased incidence of depression, anxiety and possibly also increased cardiovascular diseases[170-173]. Diagnosis is made solely based on the medical history reported by the patient accompanied by an unremarkable examination finding, in the case of primary RLS, or an examination finding in line with the underlying condition in secondary RLS.

RLS is a disorder combining both sensory and motor symptoms on a spectrum that at its extremes consists of presentations with stark predominance of either one of the two features. Next to PLMS, paroxysmal myocloniformic hyperkinesias may also occur which can be severe enough to impair locomotion[174]. Sensory symptoms, on the other hand, can be very painful or cramp-like[175] rendering “leg pain” the chief complaint in some individuals with RLS.

Although the pathophysiologic factors underlying disease development are just starting to be uncovered, it is well established that sensory and motor symptoms can be alleviated by dopaminergic (levodopa and dopamine agonists[176,177]), opioidergic[178], and antiepileptic drugs such as alpha-2-delta calcium channel ligands[179]. Additionally, it has been shown that some individuals with RLS benefit from oral[180] and intravenous[181] iron substitution. At present, the mechanism of action of all of these treatments with regard to either the motor or the sensory components of the RLS phenotype remains unclear.

1.6.1.4 Endophenotypes—Primary and Secondary RLS

It is important to recognize that RLS can be a primary disorder or secondary to a number of other medical conditions. Primary RLS can be subdivided further into familial and idiopathic forms. A positive family history as defined by at least one affected 1st degree relative is reported by 40 to 90% of patients.[182-184] In 232 individuals with idiopathic or primary RLS, individuals with a positive family history had a younger age of onset (35.5 vs. 47.2 years, $p<0.05$)[183]. The age of onset shows a bimodal distribution with a larger suspected genetic contribution in those who start experiencing symptoms before the age of 30 and a larger contribution of environmental factors and secondary causes in those with an age of onset after 30 years[185].

Common causes of secondary RLS include pregnancy, iron deficiency and renal failure. Overall, about a third of all individuals with any one of these conditions also develop symptoms of RLS.[186-189]

A number of situations depleting body iron stores such as repeated blood donations[190] or disorders with increased blood loss such as severe myomas (personal observation) can lead to RLS. In 365 individuals between the age of 65 and 83 years of age, those with the lowest quantile of serum iron levels (odds ratio (OR)=3.08) and transferrin saturation (OR=5.68) were more likely to have RLS[191].

RLS during pregnancy is a common phenomenon reported by 28.9% of the total collective of 1079 women in three large studies from Norway, France, and Italy.[186-188] Symptoms are most severe during the third trimester. In most cases, RLS symptoms that first manifested during pregnancy disappear shortly after delivery but are believed to represent a predictor for the development of primary RLS later in life.

Approximately one third of hemodialysis patients suffer from RLS. Although, it has been reported that RLS symptoms in uremic patients positively correlate with serum kreatinin levels by some[189], a trustworthy serological correlate of RLS in hemodialysis patients is still lacking.

In the clinical setting, individuals with both RLS and polyneuropathy are also frequently encountered, however, it is unclear which percentage of these patients has RLS symptoms secondary to underlying polyneuropathy and in whom the two disease, which are both common in the elderly population, merely co-occur as no large studies have been conducted to date. Estimates of the prevalence of RLS in individuals with polyneuropathies vary widely (between 5.2% and 40%)[192-195].

When compared to its prevalence in the general population, RLS is also seen more frequently in a number of different conditions ranging from neurologic diseases such as amyotrophic lateral sclerosis[196], Parkinson`s disease[197] and multiple sclerosis[198] to rheumatoid arthritis[199] and celiac[200] as well as Crohn`s disease[201]. If the underlying condition can be treated, RLS symptoms also improve. Finally, a number of drugs affecting dopaminergic, serotoninergic, histaminergic or noradrenergic neurotransmission can also precipitate or worsen symptoms of RLS.[202-204]

1.6.1.5 Pathophysiologic Concepts

The pathophysiology and pathoanatomy underlying RLS are largely unclear. Although the age-dependent prevalence might suggest otherwise, there is no overt involvement of neurodegenerative processes.[205-207] Functional imaging studies and electrophysiological studies have seen both cortical and spinal hyperexcitability in individuals with RLS.[208-215] The therapeutic benefit derived from dopaminergics[176,177] and opioids[178,216] has implicated these neurotransmitter systems and the central nervous system (CNS) in the pathogenic framework of RLS. Yet, both the mechanism and the exact pathoanatomic location remain subject to speculation. The dopaminergic neurotransmitter system is intriguing because dopaminergic dysfunction could explain both motor and sensory symptoms[217] as well as the circadian predilection[218,219]. The nigrostriatal (A9) and diencephalospinal (A11) dopaminergic systems are most intensively discussed in the context of RLS.[217,220] A11 neurons inhibit both afferent sensory neurons and preganglionic sympathetic neurons and are involved in pain modulation and the control of autonomic and motor functions and could, in the case of dysfunction, amount in the spinal hyperexcitability seen in individuals with RLS.[221] Their location close to the suprachiasmatic nucleus, the “control hub” for circadian rhythms in the human body, further fuels speculations regarding a possible role in RLS.[222] Some clinical aspects of RLS, such as increased locomotion, which is attenuated by dopamine agonists and augmented by iron deficiency, can be reproduced in controversial mouse and rat models with A11 lesions.[223-225] However, in patients with RLS no signs of a neurodegenerative process could be detected in the A11 region[207]. With regard to the A9 region, the situation is equally as uncertain. A number of neuroimaging studies assessing both pre- and postsynaptic dopamine status in the nigrostriatal system have yielded contradicting results but insinuate a dys- and possible hypofunction in RLS.[226-231] However, an involvement of other structures of the CNS can also not be excluded as these have not been investigated in detail.

Clinical evidence also points to a role for iron in the pathogenesis of RLS (cf. section 1.7.1.3 above, reviewed in [232]). This is supported by the neuroimaging and transcranial sonography finding of reduced iron levels in the SN and multiple other brain regions in individuals with RLS.[233-240] Moreover, cerebrospinal fluid (CSF) studies revealed decreased ferritin and increased transferrin indicative of a depletion of body iron stores in the CSF of RLS patients when compared to controls.[241,242] The molecular mechanisms responsible for this iron depletion is not known. Attenuated uptake by dopaminergic neurons of the SN, the endothelial cells of the choroid plexus and the brain microvasculature or oligodendrocytes has been postulated, with a dysregulation of multiple players of the iron metabolism seen in these cells in *post mortem* neuropathological studies of a limited number of RLS brains.[206,243-248]
One possible link between the dopamine and iron pathways in RLS is that the rate-limiting enzyme in dopamine biosynthesis, tyrosine hydroxylase (TH), requires iron as a cofactor. Accordingly, decreased iron availability in dopaminergic neurons could amount in the dopaminergic hypofunction proposed to exist in RLS.[217] Yet, animal models of iron deficiency show increased extracellular dopamine and intracellular TH, while still recapitulating the dopaminergic alterations seen in RLS patients, at least in part.[248]
Overall, significant work remains on the road to fully understanding the pathogenic mechanisms involved in RLS. Recently, the discovery of the first genetic susceptibility factors for the disease have, moreover, implicated novel concepts, such as neurodevelopmental dysregulation, in the underpinnings of RLS.[I,249-252]

1.6.2 Genetics of RLS

1.6.2.1 RLS as a Genetic Disorder

During the first half of the 20th century, two of the earliest RLS researchers, Hermann Oppenheim and Karl Ekbom, already observed a familial aggregation of RLS cases. Ekbom estimated ‘one-third’ of all RLS cases to be hereditary and described families with an apparent autosomal-dominant pattern of inheritance.[253,254]
Next to an assessment of families, twin studies can be used to further investigate the heritable component of a disease and to evaluate the contribution of genotype and environment interactions to a phenotype. The larger the difference in concordance rates between monozygotic (MZ) and dizygotic (DZ) twins, the larger the genetic contribution to a given trait. With regard to RLS, three twin studies have been published with concordance rates ranging between 15% and 45% for DZ twin pairs[255,256] and 53% and 83% for MZ twin

pairs[255-257]. Under the assumption of shared environmental variation (V_E), the narrow-sense heritability (h^2) for RLS was estimated to be between 54% and 69%[255,256] and a collection of additive genetic effects combined with unique environmental influence proved to be the best approximation in multifactorial liability threshold modeling.[256]

Taken together the twin studies lend support to the perception of RLS as a highly heritable disease. At the same time, concordance rates among MZ twins fell short of 100% arguing for the existence of important individual epigenetic or environmental factors.

Genetic factors play a role in bringing about RLS in both primary and secondary cases—though likely to very different extends. Classically, RLS has been considered to be a complex genetic disorder. In secondary RLS, this may simply mean that individuals who develop RLS due to an underlying condition possess genetic variants conferring increased susceptibility to RLS but without an additional insult such as another predisposing medical condition these individuals would never develop symptoms of RLS. Some of these genetic risk factors may be identical to those seen in individuals with primary RLS (e.g. rs3923809 in *BTBD9* in hemodialysis patients with RLS[258]), however, it is also likely that genetic risk factors unique to secondary forms of RLS exist. On the other end of the spectrum, in familial RLS, mono- or oligogenic forms may exist in which, in the most extreme scenario, only one genetic alteration would be sufficient to cause disease. However, it is unclear whether this means that there is a single genetic variant in a single gene, different variants in a single gene or different variants in different genes in the affected individuals. Further, as discussed below in more detail, variable expressivity even within a single family, incomplete penetrance, the existence of phenocopies and genetic heterogeneity further characterize the genetics of RLS.

1.6.2.2 Family Studies of RLS

The large heritability estimates and the occurrence of large pedigrees with RLS prompted the first systematic family studies in the 1980s and 1990s[178,184,259,260]. Here, it was noted that in most pedigrees the recurring pattern of transmission seemed to be autosomal dominant.[259,260] This observation was later substantiated by the systematic evaluation of the pattern of inheritance in 300 individuals with RLS[183]. Under the assumption of a single causative gene playing a role in familial RLS, linkage analyses were used to identify genomic regions shared by affected individuals from a family.

To date, a total of seven such genomic loci have been identified (Table 1.1). In all but one, a model of autosomal dominant inheritance with reduced penetrance yielded the highest LOD scores.[156,261-267] For RLS-1, however, the first RLS linkage locus identified in a French-

Canadian family on chromosome 12q12-21, an autosomal-recessive model with a high allele frequency of 0.25, resulting in a pseudodominant mode of inheritance, represented the best fit[264,268].

Next to the seven linkage regions that were found to have genome-wide significant LOD scores above the conventional threshold of 3.3, a total of 21 linkage regions on 14 chromosomes have also been reported with LOD scores ranging between 1.00 and 2.61[156,262-265,269]. For a more in-depth discussion of the RLS linkage loci, please cf.[270].

Despite this plethora of evidence supporting the existence of single genetic variants of strong effect that play a role in familial RLS, it is also important to realize, that most of these loci where only found in single or—in the best case—a few families leaving many more families where the underlying genetic factors remain obscure.

The recurrent finding in the family studies was that of genetic heterogeneity and complexity in RLS. Interestingly, a large German RLS family in whom linkage analysis argued for the existence of two independent linkage loci on chromosomes 4 and 17 also exists, possibly reflecting an oligogenic mode of inheritance in this family. (Winkelmann et al., unpublished observation) Also, replication of the above loci has proven very difficult[266,268,269,271-274] and the maximum LOD scores found fall short of the maximum attainable scores projected by the pedigree structure. Overall, linkage studies in RLS have failed to the extent that no underlying genetic factor could be identified for any of the above loci, even when the most up-to-date technologies such as targeted next-generation sequencing were employed to resolve the regions[275].

Chr	Region (hg19)	Peak Marker	Size (Mb)	max LOD	Model	Replication Status	Reference
12q12-21	94176800-104264737	**D12S1044** 94176800	10.09	3.59	auto rec pseudodominant	+	[264]
14q13-21	34459194-47133518	**D14S288** 43171519	12.68 (1.3)	3.23	auto dom	+	[263]
9p24-22	516800-19680020	**D9S286** 8043378	19.18 (16.60)	3.9 3.22	auto dom model-free	+	[156]
9p21	22340644-ca. 3225000	**D9S147E** 31044744	9.9	3.6	auto dom	-	[266]
2q33	197566845-208825061	**D2S325** 207978881	11.26 (0.045)	5.5	auto dom reduced pen (0.7)	(+)	[265]
20p13	82754-5315186	**D20S849** 5142034	5.2 (4.5)	3.86	auto dom reduced pen (0.7)	(+)	[261]
16p12	22758479-23312075	several	1.18	3.5	auto dom reduced pen (0.8)	(+)	[262]
19p13	0-2518075	**D19S878** 2310697	2.5	3.59	auto dom	-	[267]

Table 1.1: Linkage regions in RLS. For the size of the linkage region, first the originally reported size is given and, secondly, if pertinent, the best current approximation after additional fine-mapping and replication studies. + = replicated with significant LOD score, (+) = replicated with LOD score suggestive of linkage, - = not replicated. (published in ref. XI,XII)

1.6.3 Genome-Wide Association Studies

To date, three genome-wide association studies (GWAS) have been carried out for RLS and one for RLS and periodic limb movements in sleep (PLMS) (Tables 1.3 and 1.4). The PLMS GWAS was performed out under the deCODE Genetics umbrella and included 306 cases with RLS and PLMS and 15,664 controls from Iceland in the genome-wide phase. An intronic variant in *BTBD9* within an LD block on chromosome 6p21.2 showed genome-wide significant association ($p_{nominal}=2x10^{-9}$, OR=1.8) and was replicated in a second Icelandic and a US-American sample (combined sample (617 cases/17,528 controls): $p_{nominal}=3x10^{-14}$, OR=1.7, lead SNP=rs3923809).[252]

Simultaneously, the first RLS GWAS, which included 401 German cases and 1,644 general population controls in the genome-wide phase as well as 903 German cases and 891 controls and 255 Canadian cases and 287 controls in the replication samples, also showed association to the same SNP and the same 115kb LD block on chromosome 6p containing intron 5 of *BTBD9*. However, on chromosome 2p, an association signal located within a 32 kb LD block containing intron 8 and exon 9 of *MEIS1* was more strongly associated with the RLS phenotype in all individuals included in the genome-wide phase as well as the combined sample (rs2300478, $p_{nominal}=3.41x10^{-28}$, OR=1.74). Fine mapping and haplotype analysis in the German replication sample revealed a haplotype associated with RLS with an increased OR of up to 2.75 (95% CI: 2.23-3.41) ($p_{nominal}=5.87x10^{-20}$, frequency in cases 0.231 vs. 0.102 in controls). A third association signal of genome-wide significance was located within a 48kb locus on chromosome 15q spanning the 3`end of *MAP2K5* as well as *SKOR1* (formerly called *LBXCOR1*) (combined $p_{nominal}=6.09x10^{-17}$).[249]

A GWAS-based analysis of the RLS-3 locus encompassing 31 Mb on chromosome 9p23-24 revealed and replicated two independent ($r^2=0$) SNPs within two independent LD blocks in intron 8 (rs4626664) and intron 10 (rs1975197) of the 5' UTR of *PTPRD*. In the combined analysis of discovery and replication samples, both SNPs surpassed thresholds for genome-wide significance (rs4626664: $p_{nominal}=5.91x10^{-10}$, OR=1.44; rs1975197: $p_{nominal}=5.81x10^{-9}$, OR=1.31). No variants in any of the 35 coding and 10 non-coding exons of *PTPRD* could be identified in nine affected individuals from an RLS-3 linked family and the common variants in *PTPRD* only explain a minor portion of the original RLS-3 linkage signal.[251]

An increased sample size of 922 cases and 1,526 controls in the genome-wide phase and a multi-national replication sample of 3,935 cases and 5,754 controls of European descent revealed two new loci of genome-wide significance: an intergenic region on chromosome 2p14 approximately 1.3 Mb downstream of *MEIS1* (rs6747972, $p_{nominal}=9.03x10^{-11}$, OR=1.23)

as well as a locus on chromosome 16q12.1 encompassing an LD block of 140 kb containing both the 5`-end of *TOX3* and the non-coding RNA *BC034767* (rs3104767, $p_{nominal}=9.4x10^{-19}$, OR=1.35).[I]

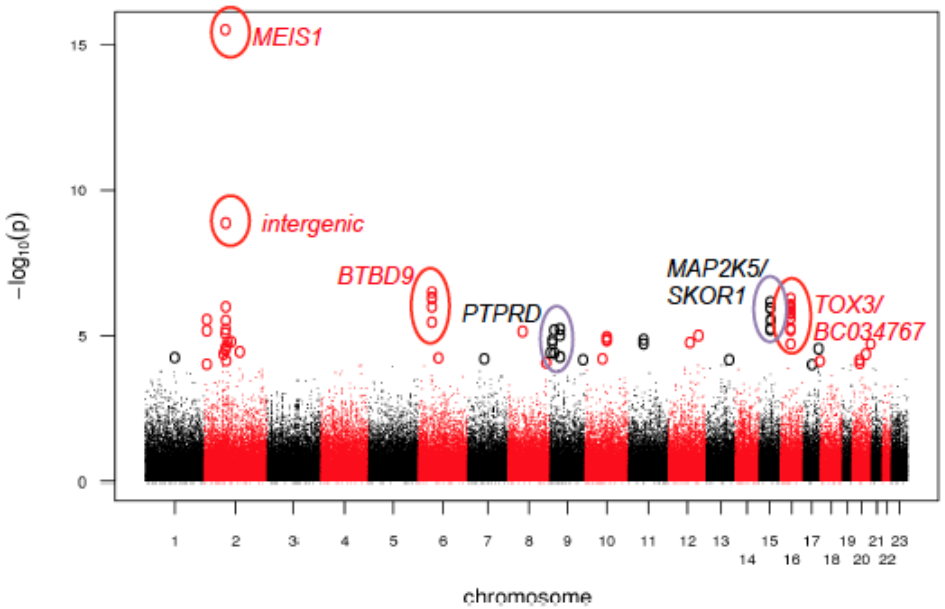


Figure 1.11: Manhattan plot showing RLS GWAS loci. By means of three GWAS, a total of six genomic loci of genome-wide significance were identified for RLS. (modified after ref. I)

While the two most recent loci still await replication in independent studies, the first four loci have been replicated in independent case/control samples[276-279] and the lead SNPs in *BTBD9* were also associated with increased susceptibility to RLS in European dialysis patients with RLS[258]. To date, no GWAS in non-European populations or considering specific endophenotypes have been performed for RLS.

Single SNPs at the RLS-associated loci identified by the above studies bear effect sizes between 1.23 and 1.68 and risk allele frequencies between 0.19 and 0.82.[I] (Tables 1.2 & 1.3) Although the conferred risk is large when compared to common variants associated with other complex traits, when taken together, the most significant SNPs at these loci only explain about 6.8% of the total heritability of RLS[I]. The significant portion of heritability that remains to be accounted for—the so-called "missing heritability"[33]—argues for—most likely—both the existence of additional independent RLS-related variants within these loci as well as a number of additional loci.

Genome-wide sample (cases/controls)	Origin	SNP Array	Replication sample(s) (cases/controls)	Origin	Lead SNPs	Candidate gene	Replication status	Reference
306/15,633	Iceland	Human Hap300 & Hap300-duo+ Bead, Illumina	123/1233 188/662	Iceland USA	rs3923809	*BTBD9*	+	[252]
401/1,644	Germany	500K, Affymetrix	903/891	Germany	rs2300478	*MEIS1*	+	[249]
			255/287	Canada	rs9296249	*BTBD9*	+	
					rs1026732	*MAP2K5/SKOR1*	+	
628/1,644	Germany	500K, Affymetrix (n=401+1,644)	1,271/1,901	Germany	rs4626664	*PTPRD*	+	[251]
		Genome-Wide Human SNP 5.0	279/368	Czech Republic	rs1975197	*PTPRD*	+	
		Array, Affymetrix (n=227)	285/842	Canada				
954/1,814	Germany	Genome-Wide Human SNP 5.0	1,236/1,471	Germany & Austria	rs2300478	*MEIS1*	+	[I]
	& Austria	Array, Affymetrix (cases)	1,104/1,065	Germany & Austria	rs9357271	*BTBD9*	+	
		Genome-Wide Human SNP 6.0	351/597	Czech Republic	rs1975197	*PTPRD*	+	
		Array, Affymetrix (controls)	141/360	Finland	rs12593813	*MAP2K5/SKOR1*	+	
			182/768	France	rs6747972	intergenic	-	
			285/285	Canada	rs3104767	*TOX3/BC034767*	-	
			556/1,208	USA				

Table 1.2: Summary of RLS GWAS. (published in ref. XI,XII)

Locus	Chr	LD Block (Mb)	Lead SNP	Risk allele	Risk allele freq cases/controls	P_{joint}	Odds ratio (95% CI)
MEIS1	2	66.57-66.64	rs2300478	G	0.35/0.24	3.40×10^{-49}	1.68 (1.57-1.81)
MAP2K5/SKOR1	15	65.25-65.94	rs12593813	G	0.75/0.68	1.37×10^{-22}	1.41 (1.32-1.52)
BTBD9	6	37.82-38.79	rs9357271	T	0.82/0.76	7.75×10^{-22}	1.47 (1.35-1.47)
TOX3/BC034767	16	51.07-51.21	rs3104767	G	0.65/0.58	9.40×10^{-19}	1.35 (1.27-1.43)
intergenic	2	67.88-68.00	rs6747972	A	0.47/0.44	9.03×10^{-11}	1.23 (1.16-1.31)
PTPRD	9	8.80-8.88	rs1975197	A	0.19/0.16	3.49×10^{-10}	1.29 (1.19-1.40)

Table 1.3: Summary of RLS GWAS loci. (published in ref. XI,XII)

1.6.1.3 Following-up on GWAS

The link between the most likely candidate genes at the associated GWAS loci and RLS is not readily apparent. It is also important to realize that the lead SNPs may not be identical with the causal genetic variants at these loci, which makes functional follow-up studies indispensible in order to utilize genetic variants to inform the pathophysiology of RLS. Functionally, most of the candidate genes highlighted by the GWAS are not well characterized. Transcriptional regulation especially in developmental processes in the nervous system seems to be the largest common denominator.

1.6.2.3.1 *MEIS1*

In the following, GWAS follow-up efforts regarding the *MEIS1* locus are portrayed in greater detail because it represents the currently only candidate gene at its locus, shows the most significant association with the RLS phenotype in the GWAS performed to date and because it is the most pertinent to the studies depicted in this dissertation. For a detailed review of the additional RLS GWAS loci (Table 1.3), please see ref. [270,280].

The transcription factor *MEIS1* (*myeloid ecotropic viral insertion site 1 homolog*) belongs to the family of highly conserved three-amino acid loop extension (TALE) homeobox (HOX) genes and interacts with PBX and HOX proteins to increase the affinity and specificity of HOX proteins[281] as well as CREB1[282] in DNA binding. In *Xenopus laevis*, *meis1* is known to be involved in neural crest development.[283] Murine Meis1 is essential for proximo-distal limb patterning[284] and plays a role in defining Hox transcriptional regulatory networks[285] that specify among others spinal motor neuron pool identity and connectivity[286]. In the CNS of adult mice, *Meis1* is expressed in cerebellar granule cells, the forebrain and the SN[287]. While *MEIS1* was initially identified in the context of acute myeloid leukemia[288,289], in recent years, a role in murine heart development has also been recognized[290] and SNPs in intron 8 (but in weak LD (r^2=0.3) with the known RLS SNPs) play a role in determining atrio-ventricular conduction velocity in both Europeans and African Americans[291,292]. *Meis1*$^{-/-}$ mice develop ocular and vascular defects, fail to produce megakaryocytes and display extensive hemorrhaging. They also die by embryonic day 14.5.[293]

Several rare non-synonymous variants in *MEIS1* have been identified in RLS patients.[294-296] However, coding variants in *MEIS1* are very rare in general (13 out of approximately 4300 individuals with a non-synonymous variant in the NHLBI-ESP exomes), possibly owing to the fact that *MEIS1* represents one of the most highly conserved genes in the human genome,

and, therefore, remain ambiguous with regard to possible causality of the RLS phenotype.[294,295,297]

Since the publication of the first GWAS, which identified SNPs in *MEIS1* as susceptibility factors for RLS, two studies have been reported which examine the functional differences brought about by the RLS-associated intronic variants. In the first, a significant decrease in *MEIS1* mRNA and protein expression was found in lymphoblastoid cell lines and brain tissue (pons and thalamus) from homozygous carriers of the risk haplotype when compared to homozygous carriers of the non-risk haplotype.[297] In a second study, knock-down of the *MEIS1* orthologue *unc-62* by RNA interference in *Caenorhabditis elegans* was related to increased ferritin expression and an extended lifespan. In thalamus but not in pons samples of RLS patients homozygous for the *MEIS1* risk haplotype (n=9), ferritin light and heavy chain as well as divalent metal transporter 1 (*DMT1*) mRNA and protein expression were significantly increased when compared to RLS patients carrying the protective haplotype (n=7).[298] The authors argue that these data are in support of a disruption of physiological iron transport into the brain and—in conjunction with the also observed decrease of MEIS1 expression in *in vitro* cell models of iron deprivation—provide a functional link between the RLS gene *MEIS1* and the iron metabolism.[298]

1.6.2.3.3 *BTBD9, PTPRD, MAP2K5/SKOR1 & TOX3/BC034767*

BTBD9, PTPRD, TOX3, MAP2K5, and *LBX1*, the transcriptional target of co-repressor *SKOR1*, have been reported to be—directly or indirectly—involved in transcription regulation and neuromuscular developmental processes.[299-304] All of them are expressed in a number of different cortical and subcortical brain regions.[305-308] Very little is known about the function of the non-coding RNA *BC034767*, the second candidate gene at the RLS-associated locus on chromosome 16. Common SNPs and structural variation in the above genes have also been related to a range of other neurologic—Tourette syndrome (*BTBD9*)[309], neuroblastoma (*PTPRD*)[310,311] and attention-deficit and hyperactivity disorder (*PTPRD*)[312]—and non-neurologic—type II diabetes (*PTPRD*)[313], coronary artery disease (*PTPRD*)[314] and breast cancer (*TOX3*)[315]—diseases. The only RLS-specific follow-up studies have been performed with regard to *BTBD9*. Here, in both drosophila and mice, the knock-down of *dBTBD9/Btbd9* leads to hyperlocomotion and changes in sleep architecture such as increased awake time and number of arousals as well as—in mice—decreased stage 3 and non-REM sleep reminiscent of RLS.[316-318] Although GWAS have written the first real “success stories” in RLS genetics, the identified genes only represent the most likely candidates at the given loci and one can

also not exclude that other genetic variants in high LD with the lead SNPs play a role or that the SNPs hold long-range regulatory function on other genes. Accordingly, ongoing efforts to functionally link the identified genes to the RLS phenotype will be important to better inform the nature of the observed associations.

1.7 Parkinson`s Disease

1.7.1 The Clinical Presentation of PD

1.7.1.1 Epidemiology

Next to Alzheimer`s disease (AD), PD represents the second most common neurodegenerative condition known today[319]. The point prevalence of PD in industrialized countries is estimated to be 0.3% with an age-dependent increase to 3 to 4% in those over the age of 80 years.[320-322] In Western Europe, the life time risk of developing PD currently ranges around 4% in European populations.[323] Men are more commonly affected than women (m:f = 1.46:1) in Caucasian but not in Asian populations, where men and women develop PD at equal frequencies.[324]

1.7.1.2 Definition and Diagnostics

Clinically, PD is characterized by a tetrad of motor symptoms consisting of (1) a low-frequency resting tremor (classically 4 to 6 Hz), (2) rigidity, (3) bradykinesia and (4) postural instability. Next to the motor impairments, non-motor features exemplified by neuropsychiatric symptoms—such as cognitive impairment[325,326], depression[327,328] and psychotic symptoms[329] —, dysautonomia[330,331] and sleep-wake[330,332] (reviewed in[333]) as well as pain disorders[334,335] also represent debilitating clinical aspects of PD[336] and can, in approximately 20% of cases, represent the initial presenting symptom in PD[337].
Internationally, the “UK Brain Bank Criteria”[338,339] are the most widely utilized diagnostic criteria for PD (Text Box 1.2) and also represent the criteria employed to ascertain PD cases in the studies depicted in this work.

Step 1—Diagnosis of a parkinsonian syndrome

Bradykinesia and at least one of the following:

1. Muscular rigidity
2. Resting tremor (4–6 Hz)
3. Postural instability unrelated to primary visual, cerebellar, vestibular or proprioceptive dysfunction

Step 2—Exclusion criteria for PD

1. Repeated strokes or head traumas
2. Dopamine depleting drugs
3. Encephalitis or oculogyric crisis
4. (More than one affected relative)
5. Sustained remission
6. Lacking response to large doses of levodopa
7. Strictly unilateral features after 3 years
8. Other neurological features: supranuclear gaze palsy, cerebellar signs, early severe autonomic involvement, Babinski sign, early severe dementia with disturbances of language, memory or praxis
9. Exposure to known neurotoxins
10. Presence of cerebral neoplasm or hydrocephalus on neuroimaging

Step 3—Supportive criteria for PD

Three or more required for the diagnosis of definite PD:

1. Unilateral onset
2. Excellent response to levodopa
3. Rest tremor
4. Severe levodopa-induced chorea
5. Progressive disorder
6. Levodopa response for over 5 years
7. Persistent asymmetry affecting the side of onset most
8. Clinical course of over 10 years

Text Box 1.2: The UK Brain Bank Criteria.[338,339]

Yet, it can be difficult to clinically differentiate PD from other disorders presenting with parkinsonism and generally a final diagnosis can only be established *post mortem*. In one study of 100 individuals prospectively diagnosed with PD, only 76 % to 82 % also showed neuropathologic features diagnostic of PD. The “misdiagnosed” cases displayed

neuropathologic characteristics of progressive supranuclear palsy, multiple system atrophy, AD or vascular disease affecting the basal ganglia.[338]

Neuropathologically, PD is characterized by the loss of dopaminergic neurons in the pars compacta of the SN and the presence of alpha-synuclein deposits in the form of Lewy bodies and Lewy neurites. In a quintessential study, Braak *et al.* defined the progression of PD neuropathology through six stages (I to VI)[340]. Under this hypothesis, the first alpha-synuclein deposits are described to appear in the olfactory bulb and the motor nuclei of the caudal cranial nerves followed on an ascending path by inclusions in the raphe nuclei, the gigantocellular reticular nucleus and the locus coeruleus before reaching the amygdala, the cholinergic nuclei of the basal forebrain and the SN. Stages V and VI see alpha-synuclein-positive deposits also in various sensory and motor regions of the cortex.[340] (Figure 1.12)

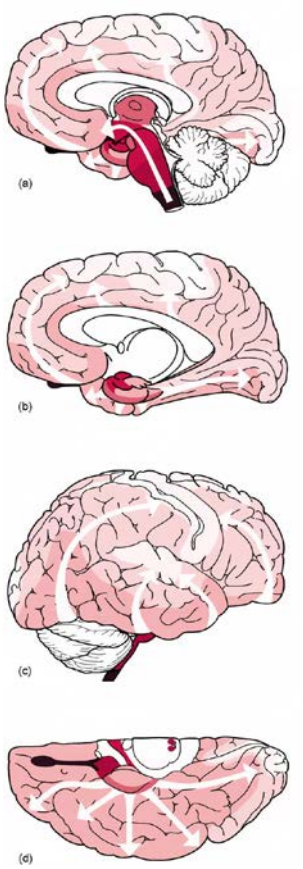


Figure 1.12: Braak Stages as applied in the neuropathological diagnosis of PD. (taken from ref. [340])

Because Lewy bodies are generally acknowledged as a neuropathologic hallmark of idiopathic PD[341,342], it was a striking finding that no Lewy bodies were found in a number of individuals who suffered from autosomal recessive PD harboring exonic deletions in the *parkin* gene[343,344] (for a detailed discussion of the genetic factors involved in autosomal

recessive PD, please cf. section 1.7.2.2 below). Overall, the neuropathologic findings in individuals with familial PD—sometimes even due to the same genetic alteration and within the same family—are very heterogeneous and not always reminiscent of the classical findings in idiopathic PD[344,345]. For example, four different neuropathologic profiles exist in the family in whom the p.R1441C missense variant in *LRRK2* was first identified, with 50 % of clinically affected carriers of the variant not showing any Lewy pathology[346,347].
Yet, despite this apparent discrepancy between the clinical and neuropathologic findings in idiopathic and familial PD[345], it was actually the discovery that Lewy bodies are comprised of alpha-synuclein[348,349]—encoded by *SNCA*, the first PD gene reported—that provided the first, lucid link between the neuropathology and genetics of PD.

1.7.1.3 Pathophysiologic Concepts

According to the current understanding, the loss of dopaminergic neurons in the pars compacta of the SN (SNpc) is central to the pathophysiology of PD. The loss of dopaminergeric projections from the SNpc to the striatum leads to a loss of inhibition of the globus pallidus internus (1) via the dopamine D1-receptor-dependent “direct” pathway and (2) by an excessive inhibition of the globus pallidus externus and subsequent loss of inhibition of the subthalamic nucleus and excessive activation of the globus pallidus internus via the D2-receptor-dependent “indirect” pathway. The result, in both cases, is an over-activation of inhibitory GABA-ergic projections from the globus pallidus internus to the ventrolateral thalamus and, thus, starkly decreased activation of glutamatergic neurons projecting to different cortical regions. (Figure 1.13) Overall, this dysfunction of the basal ganglial motor loop results in an inability to initiate automated movements. Goal-directed movements, on the other hand, remain unaffected.[350] On the contrary, the pathophysiology behind the resting tremor seen in PD is much less well defined. It has been postulated that the lack dopaminergic inhibition leads to dysfunction of the pallido-thalamic tracts and subsequently the cerebellar-thalamic innervation of the thalamus, where thalamic neurons turn into autonomous, synchronized rhythm generators, which further project to the motor cortex.[351,352]

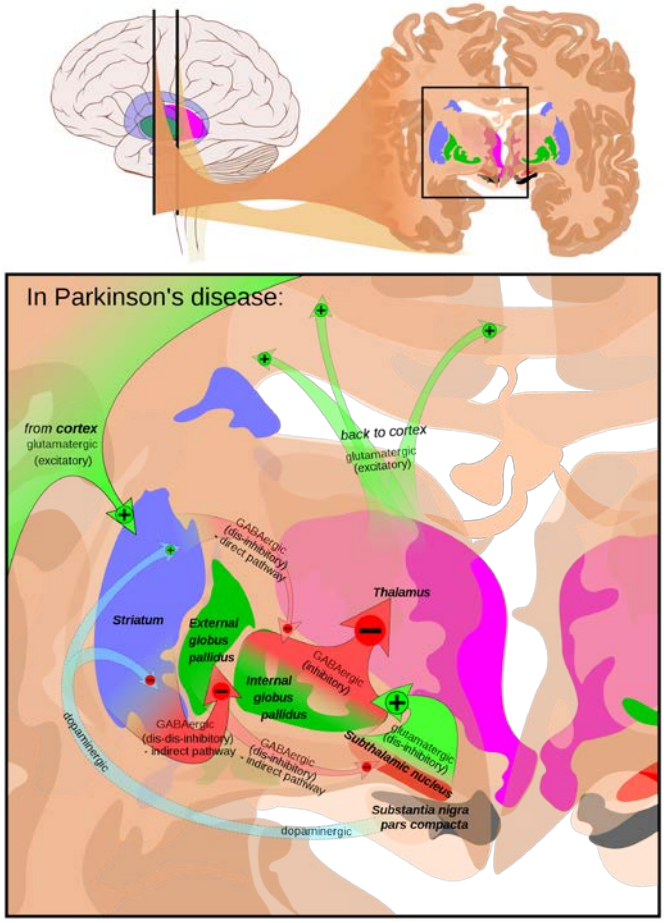


Figure 1.13: The neuroanatomy of PD.
(http://upload.wikimedia.org/wikipedia/commons/9/9e/Basal_ganglia_in_Parkinson%27s_disease.svg (accessed April 22, 2013; open source))

Yet, understanding the neuroanatomy of the cardinal symptoms seen in PD has far from explained its pathogenesis. The most central question that needs to be answered asks for the reason of neuronal cell death and the particular susceptibility of the dopaminergic neurons of the SNpc. The most widely accepted hypotheses revolve around—but are by no means limited to—oxidative stress and the mitochondria[e.g. 353-356], dysfunction of lysosomal pathways [e.g. 357-359] and the ubiquitin-proteasome system [e.g.360-362], aggregation and direct toxic effects of alpha-synuclein [e.g.348,363,364], neuroinflammation [e.g. 365-367] and environmental toxins [e.g.368-370]. Genetics (also see further discussion below) has played a large role in identifying most of these potential pathomechanisms.[371-379, reviewed in 345]

Overall, it seems fair to say that the picture remains very much obscure and the true cause of PD is likely to constitute a mix of a number of interdependent aspects—both genetic and environmental.

1.7.2 Genetics of PD

1.7.2.1 PD as a Genetic Disease

Although less than 15% of individuals with PD report a positive family history[345], it is likely that genetic factors contribute to PD development in nearly all cases—to a greater or lesser extend depending on the nature of the genetic variants involved. The heritability estimate derived from twin studies ranges around 34%[380] with MZ concordance rates in the two largest twin studies (n=193[60] and n=542[380]) between 11% and 15.5% and DZ concordance rates between 4% and 11.1%[60,380]. Although both heritability estimates and concordance rates are higher in RLS (compare section 1.6.3 above), in both diseases, in the large majority of cases, genetics are complex which a significant contribution of non-genetic factors to the phenotype. Among the much less common (near) monogenic forms of PD (summarized in Table 1.5), both autosomal dominant and autosomal recessive patterns of inheritance have been recognized (reviewed in ref. [345,381]). While in most of the autosomal dominant cases reported so far the phenotype is very similar to that of sporadic PD, autosomal recessive cases are characterized by a markedly earlier age of onset (mostly between the age of 20 and 30 years), slow disease progression, an excellent therapeutic response to levodopa but early fluctuations and dyskinesias, and additional clinical features such as dystonia or hyperreflexia[382].

1.7.2.2 Family Studies in PD

Only in 1996 and 1997, the first PD locus harboring the alpha-synuclein gene (*SNCA*) was identified in the large Italian-American Controusi kindred bearing the p.A53T substitution[371,383]. Additional missense variants[384,385] as well as duplications and triplications[372,386-389] of *SNCA* have also been shown to cause monogenetic forms of PD. A gene-dosage/disease severity relationship has been demonstrated for *SNCA*, in line with the hypothesis that increased deposition of abnormal proteins pathomechanistically contributes to the development of PD by a “toxic-gain-of-function” mechanism[381,389,390]. The second autosomal dominant PD gene encoding leucine-rich repeat kinase 2 (*LRRK2,* PARK8) was identified by two groups in 2004.[346,391] Missense variants in *LRRK2* represent the, to date, most common genetic factor in familial PD[346,391-393]. The most frequent missense variant p.G2019S[394-396] is estimated to be responsible for approximately 3.6 % of sporadic and approximately 10 % of autosomal dominant familial cases of PD in Europe[397] with reported frequencies of up to 20% in Ashkenazi Jewish[398] and up to 40% in the North African Berber Arab PD populations[399,400]. However, PD-linked variants are also found in 1.8 % of healthy

controls[397] and the penetrance of *LRRK2* p.G2019S is known to be incomplete with only 51 % of carriers showing clinical signs of PD by the age of 69[330]. Phenotypically, PD due to underlying variants in *LRRK2* shows a presentation very similar to idiopathic PD with symptom onset in the 6th or 7th decade but slightly slower progression and more prominent resting tremor and dystonia[330]. Interestingly, the neurologic phenotype of symptomatic carriers of *LRRK2* variants does not seem to be limited to PD as clinical and neuropathologic presentations reminiscent of multiple system atrophy or pure dementia have also been described[330,346]. The pathomechanism underlying *LRRK2* variants in PD has been subject of intense research efforts over the past decade but still remains very much unexplained. No gene dosage effect has been observed[400,401]. The fact that *LRRK2* holds both GTPase and kinase activity has sustained the notion that abnormal protein phosphorylation could play a role in PD pathogenesis. However, results regarding differential kinase activity in *LRRK2* mutants are conflicting[402,403] and several variants have also been linked to alterations in GTPase activity[404-406] or unspecific neurotoxic effects[407,408].

Linkage analyses have identified a total of four additional genes and two unresolved linkage regions segregating with the PD phenotype in an autosomal dominant fashion in single or a few families[409-415] (compare Table 1.5). Published in 2011, *eukaryotic translation initiation factor 4G1* (*EIF4G1*, PARK18) is the newest of these loci and was initially found in a family from Northern France[409]. Since variants in these genes in PD are extremely rare, many of these loci still await replication.

Most recently, family studies of familial PD have graduated from linkage analyses to the sequencing of entire exomes. Using whole exome sequencing, *vacuolar protein sorting-associated protein 35* (*VPS35*, PARK17) was identified simultaneously in a Swiss and an Austrian family[VIII,378,379]. Here, the pattern of inheritance of the index variant *VPS35* p.D620N was also autosomal dominant with near complete penetrance and the phenotype, as reported so far, was indistinguishable from "idiopathic" PD[VIII,378,379,416].

Next to autosomal dominant forms of familial PD, a number of genetic factors inherited in an autosomal recessive fashion have been identified. As a group, these lead to a PD phenotype characterized by an early age of onset and more benign course of disease than seen in "idiopathic" PD[345]. Shortly after the discovery of point mutations in *SNCA* in PD, homozygous multiple-exon deletions in *PARK2* encoding parkin were uncovered as the cause of juvenile-onset PD in a large Japanese family[374]. Today, homozygous and compound heterozygous missense variants as well as exonic deletions and rearrangements are estimated to account for up to 20 % of early-onset PD cases[382,417]. Functionally it has been postulated

that dysfunction of parkin, an E3 ubiquitin ligase, results in insufficient substrate clearance and, consequently, substrate aggregation[418]. Next to *PARK2* variants, homozygous and compound heterozygous loss-of-function variants (missense variants as well as small insertions or deletions) in *PTEN-induced kinase 1* (*PINK-1*, PARK6) are the second most common genetic factor in autosomal recessive PD identified thus far. Depending on the population, genetic variants have been reported to be the underlying cause in 0 % to 15 % of individuals with suspected autosomal recessive PD[419,420]. Clinically, PD related to *PINK-1* variants is very similar to the early-onset, levodopa-responsive, slowly progressive form with additional dystonic features and hyperreflexia, which is seen in carriers of *parkin* or *DJ-1* (see below) variants. Interestingly, in Asian populations, digenic inheritance of *PINK-1* and *parkin*[421] or *DJ-1*[422] variants has also been reported. The third uncontested genetic factor involved in early-onset autosomal recessive PD, *DJ-1* (PARK7), was discovered in the form of homozygous deletions and missense variants in two consanguineous families from the Netherlands and Italy[376]. A full array of point mutations in coding and promoter regions, frame-shift and splice-site mutations as well as exonic deletions in *DJ-1* have been identified in PD[345]. Mechanistically, both *PINK-1* and *DJ-1* have been linked to neuroprotection from oxidative stress and have highlighted a potential role of the mitochondria in PD pathogenesis[423,424]. In addition to these three “canonical” genetic factors in autosomal recessive PD, several other genes have also been reported to cause autosomal dominant or recessive parkinsonism (Table 1.5). Yet, in the large majority of cases, a number of other neurologic features such as ataxia, dementia, supranuclear gaze palsy, dysarthria, dystonia, or developmental delay differentiate these from “classical” autosomal recessive PD[381].

Controversy still abounds regarding the significance of heterozygous variants in the known autosomal recessive factors. Carriers of heterozygous variants in *parkin*, *PINK-1* or *DJ-1* show evidence of presynaptic dopamine deficits[425,426]. However, rare heterozygous variants in these three genes have been found in both PD cases and controls and variant rarity has so far precluded conclusive statistical evaluation. The current assumption is that these variants represent susceptibility factors for PD[427] and may modulate the phenotype by, for example, decreasing age of onset[428], but are not sufficient to cause overt PD on their own.

Despite large and successful strides that have been taken over the past two decades in identifying genes responsible for monogenic forms of PD, the currently identified genes only explain about 5 to 10 % of familial cases of PD.[345]

Locus	Gene	Chr	Inheritance	Begin	Clinical Phenotype	Reference
PARK1/PARK4	*SNCA*	4q21	auto dom	early onset	similar to sporadic PD, often cognitive impairment	371,372
PARK2	*PRKN*	6q25	auto rec	early onset	early accompanying dystonia, slow progression	374
PARK3	unknown	2p13	auto dom	late onset	similar to sporadic PD	429,430
PARK5	*UCHL1*/unknown	4p14	auto dom	late onset	similar to sporadic PD	382
PARK6	*PINK1*	1p35	auto rec	early onset	early accompanying dystonia, slow progression	373
PARK7	*DJ-1*	1p36	auto rec	early onset	slow progression	376
PARK8	*LRRK2*	12q12	auto dom	late onset	similar to sporadic PD	346,391
PARK9	*ATP13A2*	1p36	auto rec	early onset	atypical (Kufor-Rakeb syndrome)	377
PARK10	unknown	1p32	auto dom	late onset	similar to sporadic PD, to date only in Islandics	431,432
PARK11	*GIGYF2*/unknown	2q36	auto dom	late onset	similar to sporadic PD	411,412
PARK12	unknown	Xq	Unknown	late onset	unknown	433
PARK13	*OMI/HTRA2*	2p13	auto dom	late onset	similar to sporadic PD	413
PARK14	*PLA2G6*	22q13	auto rec	early onset	atypical (dystonia-parkinsonism, NBIA)	434
PARK15	*FBXO7*	22q12	auto rec	early onset	atypical	435
PARK17	*VPS35*	16q11	auto dom	late onset	similar to sporadic PD	VIII,378,379
PARK18	*EIF4G1*	3q27	auto dom	late onset	similar to sporadic PD, slow progression	409

Table 1.4: Genomic loci identified in family studies of PD.

1.7.2.3 Susceptibility Alleles of Intermediate Frequency in PD

The clinical observation that relatives of individuals with Gaucher`s disease, a lysosomal storage disorder where the dysfunction of glucocerebrosidase (*GBA*) leads to the accumulation of glucosylceramide, frequently suffer from PD led to the identification of heterozygous missense variants in *GBA* as a risk factor for PD. Initially, this was believed to be specific to the Ashkenazi Jewish population where the two most common variants, *GBA* p.N370S and p.L444P, are found in approximately 15 % of individuals with PD compared to 3 % of controls, resulting in an approximately sevenfold risk increase[436]. However, in a seminal study comprising a total of 5691 PD cases and 4898 controls from 16 centers across North America, Europe, Israel and Asia, the same two G*BA* variants were found in 3 % of PD patients but in less than 1 % of ethnically matched controls, accounting for a five-fold risk increase and establishing *GBA* variants as important, population-independent susceptibility factors for PD[375]. Overall, *GBA* is the quantitatively most significant genetic factor contributing to PD that has been identified to date. Clinically, symmetric onset and cognitive changes seem more frequent in *GBA* variant positive patients than in individuals with "idiopathic" PD[375].

Notably, low-frequency missense variants in *LRRK2* have also been reported as susceptibility factors for PD. In a large multi-center analysis of exonic variants in *LRRK2* in Caucasian (6995 PD cases/5595 controls) and Asian (1376 PD cases/962 controls) populations, *LRRK2* p.M1646T and p.A419V as well as p.G2385R were identified as low-frequency (MAF 1.6% to 3.3 %) risk alleles in Caucasians and Asians, respectively[437]. Additionally, a protective low-frequency haplotype (MAF 6.6 %) comprising p.N551K-p.R1398H-p.K1423K was identified[437]. These data along with the GWAS results depicted below argue for the existence of a full spectrum of risk and protective, rare and common genetic variants modulating the expression of the PD phenotype.

1.7.2.4 Genome-Wide Association Studies in PD

To date, a total of four large GWAS and at least four meta-analyses of GWAS have been performed for idiopathic PD[438-445] (Table 1.6). In these, a total of 21 loci surpassing thresholds for genome-wide significance in the joint analysis of discovery and replication samples were detected (Table 1.7, Figure 1.4). Among these are common variants in high LD with the already established PD genes *SNCA* and *LRRK2* but also with *microtubule-associated protein tau* (*MAPT*), a gene with a well-recognized role in other related neurodegenerative conditions such as progressive supranuclear palsy[446] or frontotemporal

lobar degeneration (FTLD)[447]. However, most of the genes implicated hold no easily discernible role in PD pathogenesis. Particularly intriguing, in this context, is the identification of an intronic variant (rs3129882) in *HLA-DRA* with known expression modulatory effects on HLA-DR and HLA-DQ identified in a GWAS of 2000 individuals with PD and 1986 unaffected controls[442]. The identification of this association signal has driven the afore-little-studied immune response and the major histocompatibility complex into the context of PD research[448].

In general, sample sizes are larger than for the RLS GWAS depicted above. The to date most extensive study combined genotype information on a total of 15,812 cases and 50,650 controls ascertained by academic institutions across North America and Europe as well as commercially through the California-based, direct-to-consumer genetic testing company 23andMe, Inc., and either newly identified or replicated a total of 16 loci of genome-wide significance[443]. This GWAS is also noteworthy because it represents one of the first ever performed using an internet-based study design for recruitment of both cases and controls. As such, it serves as a proof-of-principle that, at least with regard to PD, an individual`s self-reported phenotype is sufficiently accurate to both replicate known associations and to uncover new ones that can later be replicated in neurologist-ascertained case/control samples. This finding is important because it opens an avenue to multifold increases in the number of samples that can be analyzed and, thus, to the detection of variant alleles of even smaller effect sizes or of very low frequency.

The fact that one of the GWAS was conducted in an entirely Japanese sample (2,011 PD cases and 18,381 controls in genome-wide and replication phases combined)[439] appears important in light of the commonly reported differences in the genetic architecture of PD in different ethnic groups[449]. Here, interestingly, the *MAPT* locus, which is repeatedly one of the loci of most significant association in European populations, does not show any association with the PD phenotype[439]. This finding argues for the importance of carrying out both GWAS in samples of different ethnic background and of multi-ethnic GWAS in order to address the full spectrum of genetic factors important to PD pathogenesis.

Overall, these GWAS detected loci that account for a population attributable risk of more than 60 %[441], although this represents a likely overestimation due to inherent biases[381] and is in stark contrast to the low heritability estimates derived from twin studies (4 to 11 %)[60,380]. Still, as is the case for the vast majority of association signals identified across all phenotypes, for the largest number of association signals described in the PD GWAS—apart from those located in or close to genes already known to be involved in PD pathogenesis—, both the

truly causal genetic factor (or factors) generating the association signal as well as the functional relevance to the phenotype studied remain completely obscure.

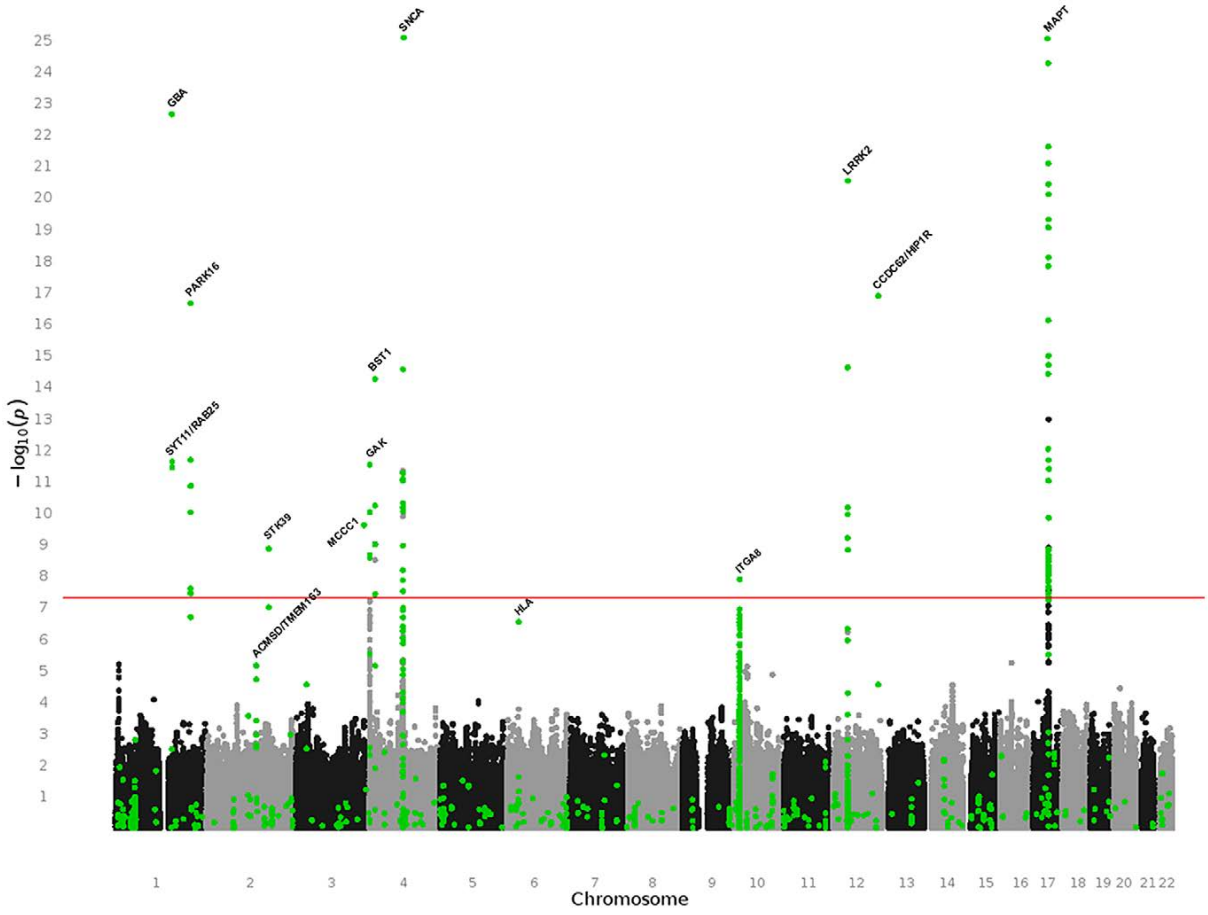


Figure 1.14: Manhattan plot showing PD GWAS loci as derived from the PDGene Database. (taken from ref. [445])

Genome-wide sample (cases/controls)	**Origin**	**SNP Array**	**Replication sample(s)** (cases/controls)	**Origin**	**Lead SNPs**	**Candidate gene** (newly identified)	**Replication status**	**Reference**
1,713/3,978	USA & Germany	HumanHap 550 beadchip, Illumina	1,528/2,044 1,100/2,168 824/544	USA Germany UK	rs2736990 rs393152	*SNCA* *MAPT*	+ +	[438]
1,078/2,628	Japan	HumanHap 550 beadchip, Illumina	612/14,139 321/1,614	Japan Japan	rs947211 rs1994090 rs4538475	*PARK16* *LRRK2* *BST1*	+ + +	[439]
2,000/1,986	USA	HumanOmni1-Quad, beadchip, Illlumina	none		rs3129882 rs11248051	*HLA-DRA* *GAK*	+ +	[442]
3,426/29,624	mostly USA (23andMe, Inc.)	HumanHap 550plus custom beadchip, Illumina	6,584/15,470	worldwide	rs1053789 rs6812193 rs11868035 rs823156	*MCCC1/LAMP3* *SCARB2* *SREBF1/RAI1* *SLC41A1*	+ + + +	[440]

Table 1.5: Summary of seminal GWAS performed for PD. In addition to these, four large meta-analyses jointly assessing the above studies or parts thereof have been conducted and any additionally associated loci are listed in Table 1.6 below.

Locus	Chr	Genomic position (hg19)	Lead SNP	Risk allele	Risk allele freq cases/controls	P_{joint}	Odds ratio (95% CI)	Reference
SNCA	4	90678291	rs2736990	T	0.51/0.46	$2.24x10^{-16}$	1.23	[438]
MAPT	17	43718893	rs393152	G	0.18/0.22	$1.95x10^{-16}$	0.77	[438]
PARK16/RAB7L1	1	205752415	rs947211	A	0.43/0.48	$1.75x10^{-12}$	1.30 (1.16-1.63)	[439]
BST1	4	15737687	rs4538475	A	0.41/0.36	$3.94x10^{-9}$	1.24 (1.16-1.34)	[439]
LRRK2	12	40428311	rs1994090	T	0.11/0.08	$2.72x10^{-8}$	1.39 (1.24-1.56)	[439]
HLA-DRA	6	32517508	rs3129882	G	0.45/0.40	$1.9x10^{-10}$	1.26 (1.17-1.35)	[442]
GAK	4	848332	rs11248051	T	0.12/0.09	$3.2x10^{-9}$	1.46 (1.29-1.65)	[442, 450]
SYT11	1	154105678	not named	T	not given	$1.02x10^{-8}$	1.67	[441]
ACMSD	2	135308851	rs6710823	A	not given	$1.35x10^{-9}$	1.38	[441]
STK39	2	168825271	rs2102808	T	not given	$3.31x10^{-11}$	1.28	[441]
MCCC1/LAMP3	3	184303969	rs11711441	G	not given	$2.10x10^{-8}$	0.82	[441]
CCDC62/HIP1R	12	121862247	rs12817488	A	not given	$4.43x10^{-9}$	1.16	[441]
SCARB2	4	77418010	rs6812193	T	not given	$7.55x10^{-10}$	0.84 (0.79-0.89)	[440]
SREBF/RAI	17	17655826	rs11868035	A	not given	$5.61x10^{-8}$	0.85 (0.80-0.90)	[440]
STBD1	4	77198736	rs6812193	T	not given	$1.17x10^{-17}$	0.88 (0.84-0.93)	[443]
GPNMB	7	23305770	rs156429	G	not given	$3.05x10^{-13}$	0.89 (0.85-0.94)	[443]
FGF20	8	16696841	rs591323	A	not given	$1.92x10^{-11}$	0.88 (0.84-0.94)	[443]
STX1B	16	30981975	rs4889603	A	not given	$6.98x10^{-13}$	1.12 (1.06-1.18)	[443]
RIT2	18	38927378	rs12456492	G	0.34	$2.0x10^{-10}$	1.19 (1.16-1.22)	[444]
ITGA8	10	15691549	rs7077361	T	0.12	$1.51x10^{-8}$	0.88 (0.84-0.92)	[445]
DGKQ	4	954359	rs11248060	T	0.12	$3.04x10^{-12}$	1.21 (1.15-1.27)	[445]

Table 1.6: Summary of the currently known PD GWAS loci of genome-wide significance. Associated loci are arranged in chronological publication order.

1.8 Aims

GWAS have been successful in identifying common variants associated with increased susceptibility to both RLS and PD. Moreover, for PD, linkage analyses have also identified rare variants of strong effect underlying familial forms, whereas for RLS, linkage analyses have not been equally successful—possibly due to the less intense research efforts in the field or a different underlying genetic architecture or variable phenotypic expressivity—and no variant of strong effect has been discovered to date. In both diseases, however, currently known genetic factors only explain a small percentage of the heritability and many more factors remain to be discovered. Some of this "missing heritability" could lie in a collection of many more common variants of relatively small effect such as those identified in GWAS but, in line with the "common disease, rare variant" hypothesis, rare variants are also likely to contribute to the genetic make-up of both diseases to a yet-unknown extent.

The aim of this thesis was the identification of genetic factors, which contribute to the genetic architecture of RLS and PD, two—for the most part—complex genetic neurologic diseases. Primarily, this was pursued via the study of rare genetic variants. To this end, with regard to PD, both a hypothesis-free family-based design employing whole exome sequencing[VIII-X] as well as candidate gene approaches analyzing genes known or projected to play a role in PD[V,VI] or other neurodegenerative conditions[VII] was applied. For RLS, on the other hand, the associated GWAS loci[I] were analyzed in depth for the existence of functionally relevant rare and low frequency variants[III,IV] in search of an allelic series of variants of the entire frequency spectrum in the same gene contributing to complex genetic diseases. The contribution of common variants to especially the RLS phenotype was assessed as part of on-going efforts in the host laboratory by GWAS[I] and GWAS follow-up studies using intermediate RNA expression phenotypes (eQTLs)[II] with the goal of identifying novel common variants as susceptibility factors for RLS.

2 Publications

2.1 Winkelmann *et al.*, Genome-Wide Association Study Identifies Novel Restless Legs Syndrome Susceptibility Loci on 2p14 and 16q12.1, *PLoS Genetics*, 2011[I]

Personal contributions: I participated in the Sequenom®-based genotyping and data analysis performed during the replication phase of the study as well as in the analysis of *cis*- and *trans*-eQTLs dependent on the lead SNPs at the newly identified loci. I also contributed to the writing and critique of the manuscript.

OPEN ACCESS Freely available online PLoS GENETICS

Genome-Wide Association Study Identifies Novel Restless Legs Syndrome Susceptibility Loci on 2p14 and 16q12.1

Juliane Winkelmann[1,2,3*], Darina Czamara[4], Barbara Schormair[1,3], Franziska Knauf[1,3], Eva C. Schulte[2], Claudia Trenkwalder[5], Yves Dauvilliers[6], Olli Polo[7,8], Birgit Högl[9], Klaus Berger[10], Andrea Fuhs[10], Nadine Gross[2], Karin Stiasny-Kolster[11,12], Wolfgang Oertel[12], Cornelius G. Bachmann[13], Walter Paulus[13], Lan Xiong[14], Jacques Montplaisir[15,16], Guy A. Rouleau[14], Ingo Fietze[17], Jana Vávrová[18], David Kemlink[18], Karel Sonka[18], Sona Nevsimalova[18], Siong-Chi Lin[19], Zbigniew Wszolek[19], Carles Vilariño-Güell[19], Matthew J. Farrer[19], Viola Gschliesser[9], Birgit Frauscher[9], Tina Falkenstetter[9], Werner Poewe[9], Richard P. Allen[20], Christopher J. Earley[20], William G. Ondo[21], Wei-Dong Le[21], Derek Spieler[1,3], Maria Kaffe[2,3], Alexander Zimprich[22], Johannes Kettunen[23,24], Markus Perola[23,24], Kaisa Silander[23,24], Isabelle Cournu-Rebeix[25,26,27], Marcella Francavilla[25,26,27], Claire Fontenille[25,26,27], Bertrand Fontaine[25,26,27], Pavel Vodicka[28], Holger Prokisch[1,3], Peter Lichtner[1,3], Paul Peppard[29], Juliette Faraco[30], Emmanuel Mignot[30], Christian Gieger[31], Thomas Illig[32], H.-Erich Wichmann[33,34,35], Bertram Muller-Myhsok[4], Thomas Meitinger[1,3]

1 Institute of Human Genetics, Technische Universität München, Munich, Germany, 2 Department of Neurology, Technische Universität München, Munich, Germany, 3 Institute of Human Genetics, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany, 4 Max Planck Institute of Psychiatry, Munich, Germany, 5 Paracelsus-Elena-Hospital, Kassel, Germany, 6 Unité du Sommeil, Service de Neurologie, Hôpital Gui-de-Chauliac, INSERM U1061, Montpellier, France, 7 Department of Pulmonary Medicine, Tampere University Hospital, Tampere, Finland, 8 Sleep Research Unit, University of Turku, Turku, Finland, 9 Department of Neurology, Innsbruck Medical University, Innsbruck, Austria, 10 Institute of Epidemiology and Social Medicine, University Münster, Münster, Germany, 11 Somnomar, Sleep Research Institute, Marburg, Germany, 12 Department of Neurology, Center of Nervous Diseases, Philipps University, Marburg, Germany, 13 Department of Clinical Neurophysiology, University of Göttingen, Göttingen, Germany, 14 Centre of Excellence in Neuromics, CHUM Research Centre and the Department of Medicine, University of Montreal, Montreal, Canada, 15 Laboratoire d'étude des maladies du cerveau, Centre de recherche du CHUM, Hôpital Notre-Dame, Université de Montréal, Montréal, Canada, 16 Centre d'étude du sommeil, Hôpital du Sacré-Coeur de Montréal, Montréal, Canada, 17 Charite – Universitätsmedizin Berlin Interdisciplinary Center of Sleep Medicine, Berlin, Germany, 18 Department of Neurology, 1st Faculty of Medicine, Charles University, Prague, Czech Republic, 19 Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, Canada, 20 Department of Neurology, Johns Hopkins University, Baltimore, Maryland, United States of America, 21 Department of Neurology, Baylor College of Medicine, Houston, Texas, United States of America, 22 Department of Neurology, Medical University of Vienna, Vienna, Austria, 23 Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland, 24 Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland, 25 INSERM, UMR_S975, Paris, France, 26 Centre de Recherche Institut du Cerveau et de la Moelle, CNRS 7225, Paris, France, 27 Fédération des maladies du système nerveux, Pitié – Salpêtrière Hospital, AP-HP, Paris, France, 28 Institute of Experimental Medicine, Czech Academy of Sciences, Prague, Czech Republic, 29 Department of Population Health Sciences, University of Wisconsin, Madison, Wisconsin, United States of America, 30 Center For Narcolepsy, Stanford University, Palo Alto, California, United States of America, 31 Institute of Genetic Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany, 32 Unit for Molecular Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany, 33 Institute of Epidemiology I, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany, 34 Institute of Medical Informatics, Biometry, and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany, 35 Klinikum Grosshadern, Munich, Germany

Abstract

Restless legs syndrome (RLS) is a sensorimotor disorder with an age-dependent prevalence of up to 10% in the general population above 65 years of age. Affected individuals suffer from uncomfortable sensations and an urge to move in the lower limbs that occurs mainly in resting situations during the evening or at night. Moving the legs or walking leads to an improvement of symptoms. Concomitantly, patients report sleep disturbances with consequences such as reduced daytime functioning. We conducted a genome-wide association study (GWA) for RLS in 922 cases and 1,526 controls (using 301,406 SNPs) followed by a replication of 76 candidate SNPs in 3,935 cases and 5,754 controls, all of European ancestry. Herein, we identified six RLS susceptibility loci of genome-wide significance, two of them novel: an intergenic region on chromosome 2p14 (rs6747972, P = 9.03 6 $10^{2\,11}$, OR = 1.23) and a locus on 16q12.1 (rs3104767, P = 9.4 6 $10^{2\,19}$, OR = 1.35) in a linkage disequilibrium block of 140 kb containing the 59-end of TOX3 and the adjacent non-coding RNA BC034767.

Citation: Winkelmann J, Czamara D, Schormair B, Knauf F, Schulte EC, et al. (2011) Genome-Wide Association Study Identifies Novel Restless Legs Syndrome Susceptibility Loci on 2p14 and 16q12.1. PLoS Genet 7(7): e1002171. doi:10.1371/journal.pgen.1002171

Editor: Mark I. McCarthy, University of Oxford, United Kingdom

Received December 7, 2010; **Accepted** May 24, 2011; **Published** July 14, 2011

Funding: The replication phase was supported by a grant from the US RLS Foundation. Part of this work was financed by the National Genome Research Network (NGFN). The KORA study group consists of H-E Wichmann (speaker), R Holle, J John, T Illig, C Meisinger, A Peters, and their coworkers, who are responsible for the design and conduction of the KORA studies. The KORA research platform (KORA, Cooperative Research in the Region of Augsburg) was initiated and financed by the Helmholtz Zentrum München, which is funded by the German Federal Ministry of Education and Research and by the State of Bavaria. The collection of sociodemographic and clinical data in the Dortmund Health Study was supported by the German Migraine & Headache Society (DMKG) and by unrestricted grants of equal share from Astra Zeneca, Berlin Chemie, Boots Healthcare, Glaxo-Smith-Kline, McNeil Pharma (former Woelm Pharma), MSD Sharp & Dohme, and Pfizer to the University of Muenster. Blood collection in the Dortmund Health Study was done through funds from the Institute of Epidemiology and Social Medicine, University of Muenster. Data collection in the COR-Study was supported by unrestricted grants of the German RLS Society (Deutsche Restless Legs Vereinigung e.V.) and Axxonis Pharma, Boehringer Ingelheim Pharma, Mundipharma Research, Roche Pharma, and UCB to the University of Muenster. CG Bachmann was supported by grants of the German RLS Society, Deutsche Restless Legs Vereinigung, e.V. H Prokisch and T Meitinger were supported by the German Federal Ministry of Education and Research (BMBF) project Systems Biology of Metabotypes (SysMBo #0315494A). RP Allen and CJ Earley were supported by the grant PO1-AG21190 National Institute of Health, USA; the Canadian part of the study was supported by a Canadian Institutes of Health Research (CIHR) grant to GA Rouleau and J Montplaisir. D Kemlink and S Nevsimalova were supported by an ESRS grant MSM0021620849; Jávrová and K Sonka were supported by grant MSM0021620816. Recruitment of Czech controls was funded by grant IGA NR 8563-5, Ministry of Health of the Czech Republic. B Frauscher, I Cournu-Rebeix, M Francavilla, and C Fontenille are co-authors on behalf of BRC-REFGENSEP, which is supported by INSERM, AFM (Généthon), ARSEP, and GIS-IBISA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: winkelmann@lrz.tu-muenchen.de

These authors contributed equally to this work.

Introduction

Restless legs syndrome (RLS) is a common neurological disorder with a prevalence of up to 10 %, which increases with age [1]. Affected individuals suffer from an urge to move due to uncomfortable sensations in the lower limbs present in the evening or at night. The symptoms occur during rest and relaxation, with walking or moving the extremity leading to prompt relief. Consequently, initiation and maintenance of sleep become defective [1]. RLS has been associated with iron deficiency, and is pharmacologically responsive to dopaminergic substitution. Increased cardiovascular events, depression, and anxiety count among the known co-morbidities [1].

Genome-wide association studies (GWAs) identified genetic risk factors within *MEIS1*, *BTBD9*, *PTPRD*, and a locus encompassing *MAP2K5* and *SKOR1* [2 4]. To identify additional RLS susceptibility loci, we undertook an enlarged GWA in a German case-control population, followed by replication in independent case-control samples originating from Europe, the United States of America, and Canada. In doing so, we identified six RLS susceptibility loci with genome-wide significance in the joint analysis, two of them novel: an intergenic region on chromosome 2p14 and a locus on 16q12.1 in close proximity to *TOX3* and the adjacent non-coding RNA *BC034767*.

Results/Discussion

We enlarged our previously reported [2,4] GWA sample to 954 German RLS cases and 1,814 German population-based controls from the KORA-S3/F3 survey and genotyped them on Affymetrix 5.0 (cases) and 6.0 (controls) arrays. To correct for population stratification, as a first step, we performed a multidimensional scaling (MDS) analysis, leading to the exclusion of 18 controls as outliers. In a second step, we conducted a variance components analysis to identify any residual substructure in the remaining samples, resulting in an inflation factor λ of 1.025 (Figures S1 and S2). The first four axes of variation from the MDS analysis were included as covariates in the association analysis of the genome-wide stage and all P-values were corrected for the observed λ.

Prior to statistical analysis, genotyping data was subjected to extensive quality control. We excluded a total of 302 DNA samples due to a genotyping call rate <98 %. For individual SNP quality control, we adopted a stringent protocol in order to account for the complexity of an analysis combining 5.0 and 6.0 arrays. We excluded SNPs with a minor allele frequency (MAF) <5%, a callrate <98%, or a significant deviation from Hardy-Weinberg Equilibrium (HWE) in controls ($P<0.00001$). In addition, we dropped SNPs likely to be false-positive associations due to differential clustering between 5.0 and 6.0 arrays by adding a second set of cases of an unrelated phenotype and discarding SNPs showing association in this setup (see Materials and Methods). Finally, we tested 301,406 SNPs for association in 922 cases and 1,526 controls. Based on a threshold level of a nominal λ-corrected $P_{GWA}<10^{-4}$, a total of 47 SNPs distributed over 26 loci were selected for follow-up in the replication study (Figure 1, Table S1).

We genotyped these 47 SNPs together with 29 adjacent SNPs in strong linkage disequilibrium (LD, $r^2=0.5$ 0.9) using the Sequenom iPLEX platform in seven case-control populations of European descent, comprising a total of 3,935 cases and 5,754 controls. Eleven SNPs with a call rate <95%, MAF<5%, and $P<0.00001$ for deviation from HWE in controls as well as 432 samples with a genotyping call rate <90% were excluded. A set of 47 SNPs, genotyped in 186 samples on both platforms (Affymetrix and Sequenom), was used to calculate an average concordance rate of 99.24 %.

The combined analysis of all replication samples confirmed the known four susceptibility loci and, in addition, identified two novel association signals on chromosomes 2p14 and 16q12.1 (Table 1). To address possible population stratification within the combined replication sample, we performed a fixed-effects meta-analysis. For four of the replication case-control populations, we included λ inflation factors which were available from a genomic controls experiment in a previous study in these populations [4]. These were used to correct the estimates for the standard error. Joint analysis of GWA and all replication samples showed genome-wide significance for these two novel loci as well as for the known RLS loci in *MEIS1*, *BTBD9*, *PTPRD*, and *MAP2K5/SKOR1* with a nominal λ -corrected $P_{JOINT}<5\times10^{-8}$ (Table 1). Depending on the variable power to detect the effects, the separate analyses of individual subsamples in the replication either confirmed the association after correction for multiple testing or yielded nominally significant results (Tables S2 and S3). The differing relevance of the risk loci in the individual samples is illustrated in

Author Summary

Restless legs syndrome (RLS) is one of the most common neurological disorders. Patients with RLS suffer from an urge to move the legs and unpleasant sensations located mostly deep in the calf. Symptoms mainly occur in resting situations in the evening or at night. As a consequence, initiation and maintenance of sleep become defective. Here, we performed a genome-wide association study to identify common genetic variants increasing the risk for disease. The genome-wide phase included 922 cases and 1,526 controls, and candidate SNPs were replicated in 3,935 cases and 5,754 controls, all of European ancestry. We identified two new RLS–associated loci: an intergenic region on chromosome 2p14 and a locus on 16q12.1 in a linkage disequilibrium block containing the 5′-end of *TOX3* and the adjacent non-coding RNA *BC034767*. *TOX3* has been implicated in the development of breast cancer. The physiologic role of TOX3 and BC034767 in the central nervous system and a possible involvement of these two genes in RLS pathogenesis remain to be established.

forest plots (Figure 2). There was no evidence of epistasis between any of the six risk loci ($P_{Bonferroni} > 0.45$).

The association signal on 2p14 (rs6747972: nominal λ-corrected $P_{JOINT} = 9.03 \times 10^{-11}$, odds ratio (OR) = 1.23) is located in an LD block of 120 kb within an intergenic region 1.3 Mb downstream of *MEIS1* (Figure 3). Assuming a long-range regulatory function of the SNP-containing region, *in silico* analysis for clusters of highly conserved non-coding elements using the ANCORA browser (http://ancora.genereg.net) identified *MEIS1* as well as *ETAA1* as potential target genes [5,6].

The second locus on chromosome 16q12.1 (rs3104767: nominal λ-corrected $P_{JOINT} = 9.4 \times 10^{-19}$, OR = 1.35) is located within an LD block of 140 kb (Figure 3), which contains the 5′UTR of *TOX3* (synonyms *TNRC9* and *CAGF9*) and the non-coding RNA *BC034767* (synonym *LOC643714*). *TOX3* is a member of the high mobility box group family of non-histone chromatin proteins which interacts with *CREB* and *CBP* and plays a critical role in mediating calcium-dependent transcription in neurons [7]. GWAs have identified susceptibility variants for breast cancer in the identical region [8]. The best-associated breast cancer SNP, rs3803662, is in low LD ($r^2 \sim 0.1$, HapMap CEU data) with rs3104767, but showed association to RLS (λ-corrected nominal $P_{GWA} = 7.29 \times 10^{-7}$). However, logistic regression analysis conditioned on rs3104767 demonstrated that this association is dependent on rs3104767 (rs3803662: $P_{GWA/conditioned} = 0.2883$).

BC034767 is represented in GenBank by two identical mRNA transcripts, BC034767 and BC029912. According to the gene model information of the UCSC and Ensembl genome browsers (http://genome.ucsc.edu and http://www.ensembl.org/index.html), these mRNAs are predicted to be non-coding. Additional *in silico* analysis using the Coding Potential Calculator (http://cpc.cbi.pku.edu.cn) supported this by attributing only a weak coding potential to this RNA, suggesting a regulatory function instead [9]. We also searched for rare alleles with strong effects and performed a mutation screening by sequencing all coding and non-coding

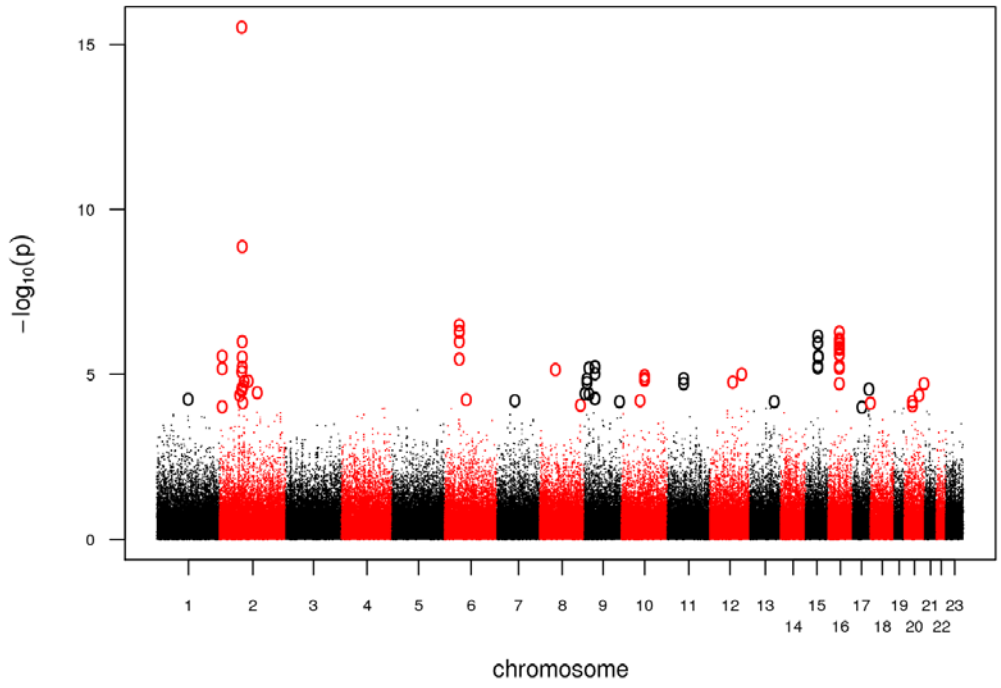


Figure 1. Manhattan plot of the GWA. Association results of the GWA stage. The x-axis represents genomic position along the 22 autosomes and the x-chromosome, the y-axis shows -log10(P) for each SNP assayed. SNPs with a nominal λ-corrected $P < 10^{-4}$ are highlighted as circles.
doi:10.1371/journal.pgen.1002171.g001

exons of *TOX3* and *BC034767* in 188 German RLS cases (Table S4). In *TOX3*, a total of nine variants not listed in dbSNP (Build 130) were found, three of which are non-synonymous. Only one of these is also annotated in the 1000 Genomes project (November 2010 data release). Three additional new variants were located in putative exons 1 and 2 of *BC034767*. Analysis of the frequency of these variants as well as all known non-synonymous, frameshift, and splice-site coding SNPs in *TOX3* in a subset of one of the replication samples (726 cases and 735 controls from the GER1 sample) did not reveal any association to RLS. For a power of >80%, however, variants with an OR above 4.5 and a MAF ≥0.01 would be required. For even lower MAFs, ORs ≥10 would be necessary for sufficient power. Furthermore, the described CAG repeat within exon 7 of *TOX3* was not polymorphic as shown by fragment analysis in 100 population-based controls.

According to publicly available expression data (http://genome.ucsc.edu), in humans, *BC034767* is expressed in the testes only, while *TOX3* expression has been shown in the salivary glands, the trachea, and in the CNS. Detailed in-depth real time PCR profiling of *TOX3* showed high expression levels in the frontal and occipital cortex, the cerebellum, and the retina [10]. To assess a putative eQTL function of rs6747972 or rs3104767, we studied the SNP-genotype-dependent expression of *TOX3* and *BC034767* as well as of genes known to directly interact with *TOX3* (*CREB-1*/*CREBBP*/*CITED1*) and potential target genes of long-range regulatory elements at the locus on chromosome 2 (*MEIS1*/*ETAA1*) in RNA expression microarray data from peripheral blood in 323 general population controls [11]. No differential genotype-dependent expression variation was found.

To assess the potential for genetic risk prediction, we split our GWA sample in a training and a test set and determined classifiers for case-control status in the training set to predict case-control status in the test set. Training and test set were independent of each other not only with respect to included individuals but also with respect to the genotyping procedure as we used genotypes generated on different genotyping platforms. As training set, we used those cases of the current GWA which had been genotyped on 500K arrays in a previous GWA and the corresponding control set [2], in total, 326 cases and 1,498 controls. The test set comprised 583 cases and 1,526 controls, genotyped on 5.0/6.0 arrays as part of the current study. Prior to the analysis, we removed the six known risk loci and performed LD-pruning to limit the analysis to SNPs not in LD with each other. In the end, a total of 76,532 SNPs were included in the pruned dataset. We conducted logistic regression with age and sex as covariates. Based on these association results, the sum score of SNPs showing the most significant effects (i.e. the number of risk alleles over all SNPs) weighted by the ln(OR) of these effects was chosen as predictor variable in the test set. We then varied the P-value threshold for SNPs included in the sum score. For a P-value <0.6, we observed a maximum area under the curve (AUC) of 63.9% and an explained genetic variance of 6.6% (Nagelkerke's R), values comparable to estimates obtained for other complex diseases such as breast cancer or diabetes (Table S5) [12 14]. Inclusion of the six known risk loci in this analysis resulted in a maximum AUC of 64.2% and an explained genetic variance of 6.8%.

Additionally, we performed risk prediction in the combined GWA and replication sample including only the six established RLS risk loci. For this purpose, we used the weighted risk allele score resulting in ORs of up to 8.6 (95% CI: 2.46 46.25) and an AUC of 65.1% (Figures S3 and S4).

By increasing the size of our discovery sample, we have identified two new RLS susceptibility loci. The top six loci show effect sizes between 1.22 and 1.77 and risk allele frequencies between 19 and 82%, and reveal genes in neuronal transcription pathways not previously suspected to be involved in the disorder.

Materials and Methods

Study population and phenotype assessment

Ethics statement. Written informed consent was obtained from each participant in the respective language. The study has

Table 1. Association results of GWA and joint analysis of GWA and replication.

Chr	Locus	LD block (Mb)	SNP	Position (bp)	Risk allele	Risk allele frequency cases/controls	P_{GWA}	$P_{REPLICATION}$	P_{JOINT}	Odds ratio (95% CI)
Known risk loci (1 SNP per locus)										
2	MEIS1	66.57–66.64	rs2300478	66634957	G	0.35/0.24	7.77×10^{-16}	4.39×10^{-35}	3.40×10^{-49}	1.68 (1.57–1.81)
6	BTBD9	37.82–38.79	rs9357271	38473851	T	0.82/0.76	6.74×10^{-7}	2.01×10^{-16}	7.75×10^{-22}	1.47 (1.35–1.47)
9	PTPRD	8.80–8.88	rs1975197	8836955	A	0.19/0.16	4.94×10^{-5}	1.07×10^{-6}	3.49×10^{-10}	1.29 (1.19–1.40)
15	MAP2K5/ SKOR1	65.25–65.94	rs12593813	65823906	G	0.75/0.68	1.49×10^{-6}	1.54×10^{-17}	1.37×10^{-22}	1.41 (1.32–1.52)
New genome-wide significant loci ($P_{JOINT} < 5.2\times10^{-8}$)										
2	intergenic region	67.88–68.00	rs6747972	67923729	A	0.47/0.44	1.37×10^{-6}	3.73×10^{-6}	9.03×10^{-11}	1.23 (1.16–1.31)
			rs2116050	67926267	G	0.49/0.47	7.84×10^{-6}	4.85×10^{-6}	4.83×10^{-10}	1.22 (1.15–1.30)
16	TOX3/ BC034767	51.07–51.21	rs3104767	51182239	G	0.65/0.58	7.38×10^{-7}	2.16×10^{-13}	9.40×10^{-19}	1.35 (1.27–1.43)
			rs3104788	51196004	T	0.65/0.58	1.19×10^{-6}	2.42×10^{-13}	1.63×10^{-18}	1.33 (1.25–1.43)

RLS-associated SNPs with genome-wide significance. P_{GWA}, λ-corrected nominal P-value of GWA stage. $P_{REPLICATION}$, nominal P-value obtained from meta-analysis of the replication stage samples. P_{JOINT}, nominal P-value of the joint meta-analysis of GWA and replication stage, λ-corrected in samples where λ-values were available. Nominal P-values in GWA were calculated using logistic regression with sex, age, and the first four components from the MDS analysis of the IBS matrix as covariates. For nominal $P_{REPLICATION}$ and P_{JOINT} -values, a fixed-effects inverse-variance meta-analysis was performed. Risk allele frequencies and odds ratios were calculated in the joint sample. LD blocks were defined by D' using Haploview 4.2 based on HapMap CEU population data from HapMap release #27. CI, 95% confidence interval. Genome positions refer to the Human March 2006 (hg18) assembly.

doi:10.1371/journal.pgen.1002171.t001

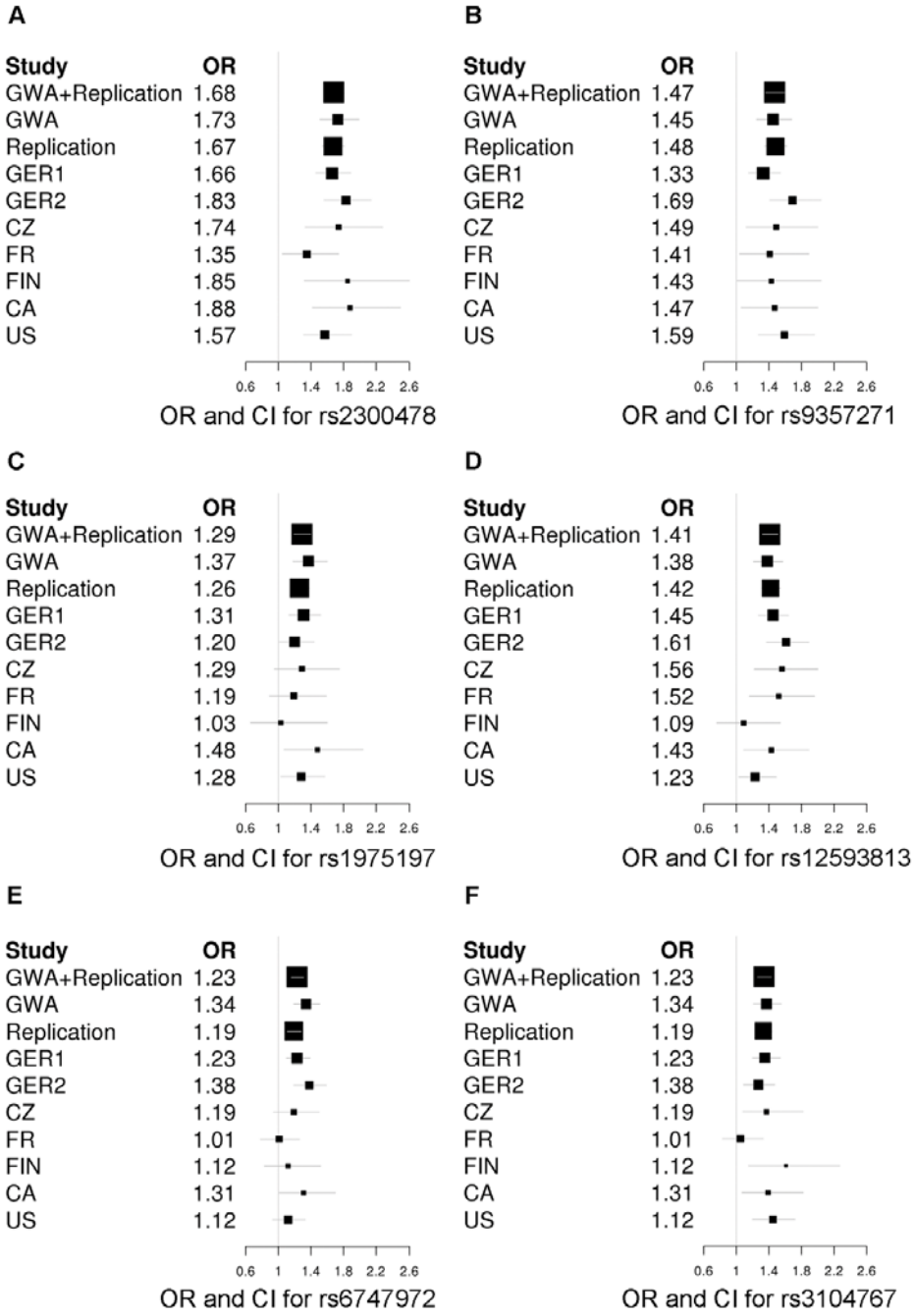

A
Study OR
GWA+Replication 1.68
GWA 1.73
Replication 1.67
GER1 1.66
GER2 1.83
CZ 1.74
FR 1.35
FIN 1.85
CA 1.88
US 1.57
0.6 1 1.4 1.8 2.2 2.6
OR and CI for rs2300478
B
Study OR
GWA+Replication 1.47
GWA 1.45
Replication 1.48
GER1 1.33
GER2 1.69
CZ 1.49
FR 1.41
FIN 1.43
CA 1.47
US 1.59
0.6 1 1.4 1.8 2.2 2.6
OR and CI for rs9357271
C
Study OR
GWA+Replication 1.29
GWA 1.37
Replication 1.26
GER1 1.31
GER2 1.20
CZ 1.29
FR 1.19
FIN 1.03
CA 1.48
US 1.28
0.6 1 1.4 1.8 2.2 2.6
OR and CI for rs1975197
D
Study OR
GWA+Replication 1.41
GWA 1.38
Replication 1.42
GER1 1.45
GER2 1.61
CZ 1.56
FR 1.52
FIN 1.09
CA 1.43
US 1.23
0.6 1 1.4 1.8 2.2 2.6
OR and CI for rs12593813
E
Study OR
GWA+Replication 1.23
GWA 1.34
Replication 1.19
GER1 1.23
GER2 1.38
CZ 1.19
FR 1.01
FIN 1.12
CA 1.31
US 1.12
0.6 1 1.4 1.8 2.2 2.6
OR and CI for rs6747972
F
Study OR
GWA+Replication 1.23
GWA 1.34
Replication 1.19
GER1 1.23
GER2 1.38
CZ 1.19
FR 1.01
FIN 1.12
CA 1.31
US 1.12
0.6 1 1.4 1.8 2.2 2.6
OR and CI for rs3104767

Figure 2. Forest plots of the RLS risk loci (1 SNP per locus). OR and corresponding confidence interval for the GWA sample, all individual replication samples, the combined replication sample as well as the combined GWA and replication sample are depicted. ORs are indicated by squares with the size of the square corresponding to the sample size for the individual populations. (A) rs2300478 in *MEIS1*; (B) rs9357271 in *BTBD9*; (C) rs1975197 in *PTPRD*; (D) rs12593813 in *MAP2K5/SKOR1*; (E) rs6747972 in intergenic region on chromosome 2; (F) rs3104767 in *TOX3/BC034767*.
doi:10.1371/journal.pgen.1002171.g002

been approved by the institutional review boards of the contributing authors. The primary review board was located in Munich, Bayerische Ärztekammer and Technische Universität München.

RLS patients (GWA and replication phase). A total of 2,944 cases (GWA = 954, replication = 1,990) of European descent were recruited in two cycles via specialized outpatient clinics for RLS. German and Austrian cases for the GWA (GWA) and the replication sample (GER1) were recruited in Munich, Marburg, Kassel, Göttingen, Berlin (Germany, n in GWA = 830, n in GER1 = 1,028), Vienna, and Innsbruck (Austria, n in GWA = 124, n in GER1 = 288). The additional replication samples originated from Prag (Czech Republic (CZ), n = 351), Montpellier (France (FR), n = 182), and Turku (Finland (FIN), n = 141). In all patients, diagnosis was based upon the diagnostic criteria of the International RLS Study Group [1] as assessed in a personal interview conducted by an RLS expert. A positive family history was based on the report of at least one additional family member affected by RLS. We excluded patients with secondary RLS due to uremia, dialysis, or anemia due to iron deficiency. The presence of secondary RLS was determined by clinical interview, physical and neurological examination, blood chemistry, and nerve conduction studies whenever deemed clinically necessary.

In addition, 1,104 participants (GER2) of the "Course of RLS (COR-) Study", a prospective cohort study on the natural course of disease in members of the German RLS patient organizations, were included as an additional replication sample. After providing informed consent, study participants sent their blood for DNA extraction to the Institute of Human Genetics, Munich, Germany. A limited validation of the RLS diagnosis among the majority of members was achieved through a diagnostic questionnaire. Five percent had also received a standardized physical examination and interview in one of the specialized RLS centers in Germany prior to recruitment. To avoid doublets, we checked these subjects against those recruited through other German RLS centers and excluded samples with identical birth date and sex.

556 cases (US) were recruited in the United States at Departments of Neurology at Universities in Baltimore, Miami, Houston, and Palo Alto. Diagnosis of RLS was made as mentioned above.

285 cases (CA) were recruited and diagnosed as above in Montréal, Canada. All subjects were exclusively of French-Canadian ancestry as defined by having four grandparents of French-Canadian origin.

Detailed demographic data of all samples are provided in Table S6.

Control populations (GWA and replication phase). Controls for German and Austrian cases were of European descent and recruited from the KORA S3/F3 and S4 surveys, general population-based controls from southern Germany. KORA procedures and samples have been described [15]. For the GWA phase, we included 1,814 subjects from S3/F3, and, for the replication stage, 1,471 subjects from S4.

For replication of the GER2 sample, we used controls from the Dortmund Health Study (DHS), a population-based survey conducted in the city of Dortmund with the aim of determining the prevalence of chronic diseases and their risk factors in the general population. Sampling for the study was done randomly from the city's population register stratified by five-year age group and gender [16]. 597 subjects selected at random from the Czech blood and bone marrow donor registry served as Czech controls [17]. French controls included 768 parents of multiple sclerosis patients recruited from the French Group of Multiple Sclerosis Genetics Study (REFGENSEP) [18]. Finnish controls comprised 360 participants of the National FINRISK Study, a cross-sectional population survey on coronary risk factors collected every five years. The current study contains individuals recruited in 2002. Detailed description of the FINRISK cohorts can be found at www.nationalbiobanks.fi.

French-Canadian controls were 285 unrelated individuals recruited at the same hospital as the cases.

1,200 participants of the Wisconsin Sleep Cohort (WSC), an ongoing longitudinal study on the causes, consequences, and natural course of disease of sleep disorders, functioned as US controls [19].

None of the controls were phenotyped for RLS. All studies were approved by the institutional review boards in Germany, Austria, Czech Republic, France, Finland, the US, and Canada. Written informed consent was obtained from each participant. Detailed demographic data of all samples are provided in Table S6.

Genotyping

GWA. Genotyping was performed on Affymetrix Genome-Wide Human SNP Arrays 5.0 (cases) and 6.0 (controls) following the manufacturer's protocol. The case sample included 628 cases from previous GWAs [2,4] and 326 new cases. After genotype-calling using the BRLMM-P clustering algorithm [20], a total of 475,976 overlapping SNPs on both Affymetrix arrays were subjected to quality control. We added 655 cases of a different phenotype unrelated to RLS, genotyped on 5.0 arrays, to the analysis and excluded those SNPs which showed a significant difference of allele frequencies in cases (RLS and unrelated phenotype on 5.0) and controls (6.0) (n = 92). Thereby, we filtered out SNPs likely to be false-positive associations. We excluded SNPs with a minor allele frequency (MAF) <5% (n = 88,582), a callrate <98% (n = 65,906) or a significant deviation from Hardy-Weinberg Equilibrium (HWE) in controls (P<0.00001) (n = 20,060). Cluster plots of the GWA genotyping data for the best-associated SNPs in Table 1 are shown in Figure S5. Genotypes of these SNPs are available in Table S7.

Replication. We selected all SNPs with a λ-corrected $P_{nominal}<10^{-4}$ in the GWA for replication. These SNPs clustered in 26 loci (defined as the best associated SNP ±150 kb of flanking sequence). We genotyped a total of three SNPs in each of the 26 regions. These were either further associated neighbouring SNPs with a λ-corrected $P_{nominal}<10^{-3}$ or, in case of singleton SNPs, additional neighbouring SNPs from HapMap with the highest possible r^2 (at least >0.5) with the best-associated SNP. We also genotyped the best-associated SNPs identified in the previous GWAs [2,4].

Genotyping was performed on the MassARRAY system using MALDI-TOF mass spectrometry with the iPLEX Gold chemistry (Sequenom Inc, San Diego, CA, USA). Primers were designed using AssayDesign 3.1.2.2 with iPLEX Gold default parameters. Automated genotype calling was done with SpectroTYPER 3.4. Genotype clustering was visually checked by an experienced evaluator.

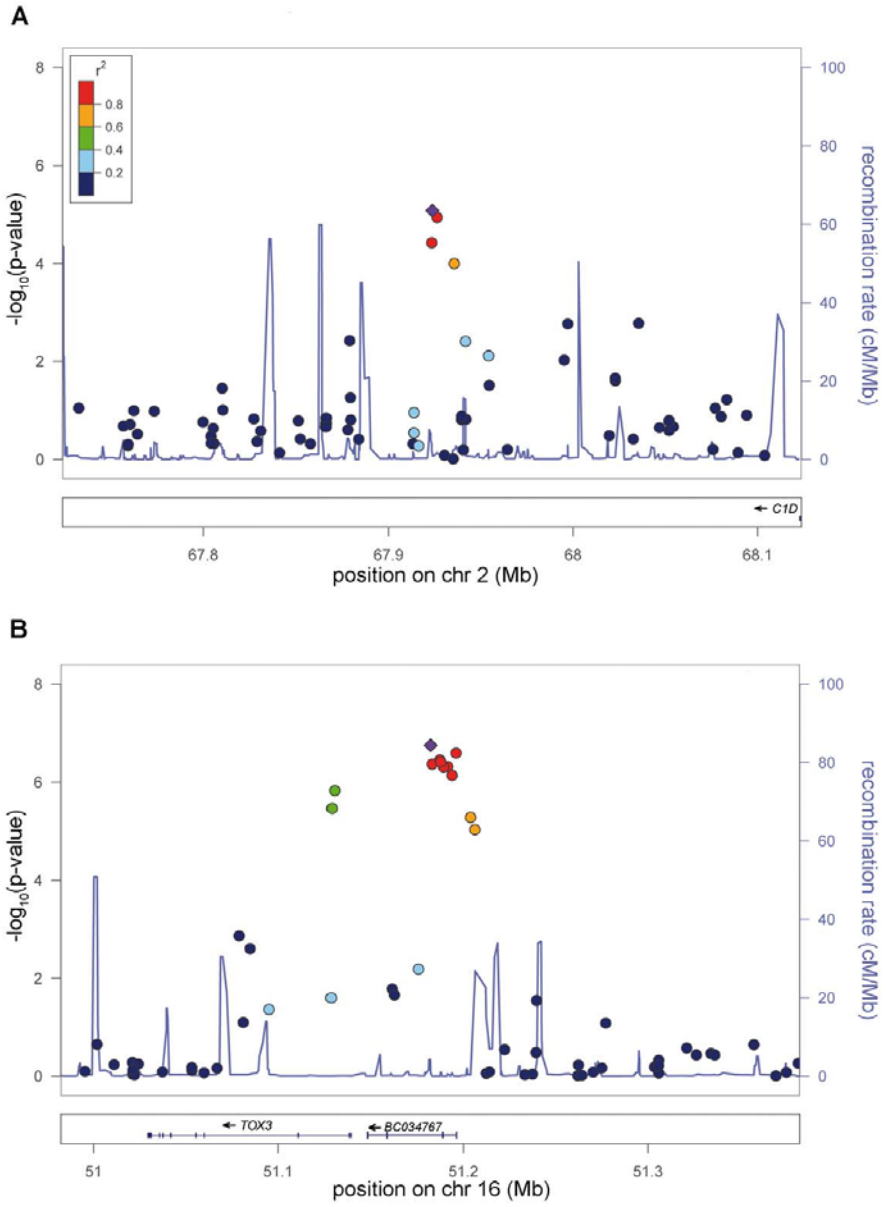


Figure 3. New genome-wide significant RLS loci. a) Risk locus on chromosome 2p14, showing the best-associated SNP rs6747972 and ±200 kb of surrounding sequence. b) Risk locus on chromosome 16p21, showing the best-associated SNP rs3104767 and ±200 kb of surrounding sequence. The left-hand x-axis shows the negative log10 of the nominal λ-corrected P-values of the GWA stage for all SNPs genotyped in the respective region. The right-hand x-axis shows the recombination frequency in cM/Mb. The y-axis shows the genomic position in Mb based on the hg18 assembly. The r^2-based LD between SNPs is colour-coded, ranging from red (r^2>0.8) to dark blue (r^2<0.2) and uses the best-associated SNP as reference. This SNP is depicted as a violet diamond. Recombination frequency and r^2 values are calculated from the HapMap II (release 22) CEU population. Plots were generated with LocusZoom 1.1 (http://csg.sph.umich.edu/locuszoom/).
doi:10.1371/journal.pgen.1002171.g003

SNPs with a call rate<95%, MAF<5%, and P<0.00001 for deviations from HWE in controls were excluded. DNA samples with a call rate<90% were also excluded.

Population stratification analysis

GWA. To identify and correct for population stratification, we performed an MDS analysis as implemented in PLINK 1.07 (http://pngu.mgh.harvard.edu/~purcell/plink, [21]) on the IBS matrix of our discovery sample. After excluding outliers by plotting the main axes of variation against each other, we performed logistic regression with age, sex, and the values of the MDS components as covariates. Using the Genomic Control approach [22], we obtained an inflation factor λ of 1.11.

Additionally, we performed a variance components analysis using the EMMAX software (http://genetics.cs.ucla.edu/emmax, [23]) and, again, calculated the inflation factor with Genomic Control, now resulting in a λ of 1.025. EMMAX uses a mixed linear model and does not only correct for population stratification but also for hidden relatedness. We, therefore, decided to base correction for population substructure on the EMMAX results.

Replication. Correction for population stratification was performed for the German, Czech, and the Canadian subsamples. The λ-values of 1.1032, 1.2286, and 1.2637 were derived from a previous Genomic Control experiment within the same samples using 176 intergenic or intronic SNPs [4]. Here, we had applied the expanded Genomic Control method GCF developed by Devlin and Roeder [24]. In the meta-analysis of all replication samples, the λ-corrected standard errors were included for the German, Czech, and Canadian samples. For the other replication samples from France, Finland, and the USA, no such data was available and, therefore, no correction factor was included in the analysis.

Statistical analysis

Statistical analysis was performed using PLINK 1.07 (http://pngu.mgh.harvard.edu/~purcell/plink, [21]). In the GWA sample, we applied logistic regression with age, sex, and the first four axes of variation resulting from an MDS analysis as covariates.

P-values were λ-corrected with the λ of 1.025 from the EMMAX analysis. In the individual analysis of the single replication samples, we tested for association using logistic regression and correcting for gender and age as well as for population stratification where possible (see Population Stratification). Each replication sample was Bonferroni-corrected using the number of SNPs which passed quality control for the respective sample.

For the combined analysis of all replication samples, we performed a fixed-effects inverse-variance meta-analysis. Where available, we used λ-corrected standard errors in this analysis. Bonferroni-correction was performed for 74 SNPs, i.e. the number of SNPs which passed quality control in at least one replication sample.

For the joint analysis of the GWA and the replication samples, we also used a fixed-effects inverse-variance meta-analysis and again included λ-corrected values as far as possible. For the conditioned analysis, the SNP to be conditioned on was included as an additional covariate in the logistic regression analysis as implemented in PLINK.

Interaction analysis was performed using the epistasis option in PLINK. Significance was determined via Bonferroni-correction (i.e. 0.05/28, as 28 SNP combinations were tested for interaction).

Power calculation

Power calculation was performed using the CaTS power calculator [25] using a prevalence set of 0.08 and an additive genetic model (Table S3). The significance level was set at 0.05/74 for replication stage analysis and at 0.05/301,406 for genome-wide significance in the joint analysis of GWA and replication. For the rare variants association study, the significance level was set at 0.05/12.

Mutation screening of *TOX3* and *BC034767*

All coding and non-coding exons including adjacent splice sites of *TOX3* (reference sequence NM_001146188) and *BC034767* (reference sequence IMAGE 5172237) were screened for mutations in 188 German RLS cases.

Mutation screening was performed with high resolution melting curve analysis using the LightScanner technology and standard protocols (IDAHO Technology Inc.). DNAs were analyzed in doublets. Samples with aberrant melting pattern were sequenced using BigDyeTerminator chemistry 3.1 (ABI) on an ABI 3730 sequencer. Sequence analysis was performed with the Staden package [26]. Primers were designed using ExonPrimer (http://ihg.gsf.de) or Primer3plus (www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). All identified variants were then genotyped in 735 RLS cases and 735 controls of the general population (KORA cohort) on the MassARRAY system, as described above.

In addition, fragment analysis of exon 7 of *TOX3* was performed to screen for polymorphic CAG trinucleotide repeats. DNA of 100 controls (50 females, 50 males) was pooled and analyzed on an ABI 3730 sequencer with LIZ-500 (ABI) as a standard. Primers were designed using Primer3plus, the forward Primer contains FAM for detection. Analysis was performed using GeneMapper v3.5.

Expression analyses

Associations between *MEIS1/ETAA1* RNA expression and rs6747972 and between *TOX3/BC034767/CREB-1/CREBBP/CITED1* expression and rs3104767 were assessed using genome-wide SNP data (Affymetrix 6.0 chip) in conjunction with microarray data for human blood samples (n = 323 general population controls from the KORA cohort, Illumina Human WG6 v2 Expression BeadChip) [11]. A linear regression model conditioned on expression and controlling for age and sex was used to test for association.

Prediction of genetic risk

Based on the performance of P-value-threshold selected SNPs in a training and a test sample. As training sample, we used those GWA-cases which had also been genotyped for our previous study [2]. We also included the control samples from this study. As a first quality control step, we carried out an association analysis comparing the Affymetrix 500K genotypes of these GWA-cases to the Affymetrix 5.0 genotypes of the same cases. Significant P-values would indicate systematic differences in the genotyping between the different chips. For further analysis, we only used those 259,302 SNPs with P-values >0.10. We performed a second quality control step in which IDs with a callrate below 98% and SNPs with a callrate below 98%, a MAF lower than 5%, or a P-value for deviation from HWE<0.00001 were removed.

Further, we excluded the four already known risk loci as well as the two newly identified loci and performed LD-pruning to limit the analysis to SNPs not in LD with each other. This was performed using a window-size of 50 SNPs. In each step, this window was shifted 5 SNPs. We used a threshold of 2 for the VIF (variance inflation factor). 76,532 SNPs, 326 cases, and 1,498 controls were included in the final training dataset. We conducted logistic regression with age and sex as covariates. Based on these association results, the sum score of SNPs showing the most

significant effects (i.e. the number of risk alleles over all SNPs) weighted by the ln(OR) of these effects was chosen as predictor variable in the test set, comprising the remaining 583 cases of the GWA sample and 1,526 controls. None of these cases/controls were included in the training-sample, i.e. the test-sample constitutes a completely independent sample. Based on this sum score, we calculated the ROC curve and Nagelkerke's R to measure the explained variance.

Based on a weighted risk allele score. To evaluate the predictive value in our sample, we calculated a weighted sum score of risk alleles in the combined GWA and replication sample. To this end, we used one SNP from each RLS risk region and also included markers from the two newly identified regions on chromosome 16q12 and 2p14 (*MEIS1*: rs2300478, *2p14*: rs6747972, *BTBD9*: rs9296249, *PTPRD*: rs1975197, *MAP2K5*: rs11635424, *TOX3/BC034767*: rs3104767). At each SNP, the number of risk alleles was weighted with the corresponding ln(OR) for this SNP. The corresponding distribution of the score in cases and controls is illustrated in Figure S3. Employing this score for risk prediction resulted in an AUC of 0.651 (Figure S4).

Supporting Information

Figure S1 MDS analysis plot for GWA. Distribution of cases (red) and controls (black) along the two main axes of variation identified in the MDS analysis. The three visible clouds are due to a common 3.8 Mb inversion polymorphism on chromosome 8 (described in: Tian C, Plenge RM, Ransom M, Lee A, Villoslada P, et al. (2008) Analysis and Application of European Genetic Substructure Using 300 K SNP Information. PLoS Genet 4: e4. doi:10.1371/journal.pgen.0040004).
(TIFF)

Figure S2 QQ-plot of GWA results. QQ-plot showing the P-value distribution before (red) and after (blue) correction for population stratification using Genomic Control.
(TIFF)

Figure S3 Weighted risk allele score analysis. Histogram of the weighted risk allele scores for cases and controls. The corresponding OR and CI for each category against the median category is depicted in green. The left y-axis refers to the number of individuals (in %), the right-axis refers to the OR values.
(TIFF)

Figure S4 ROC curve for weighted risk score analysis. Receiver operating characteristic (ROC) curve for the weighted risk allele score approach of risk prediction. The area under the curve (AUC) is 65.1%.
(TIFF)

Figure S5 Cluster plots of GWA genotyping for the six risk loci. For the best-associated SNPs at each risk locus, clusterplots were generated for cases and controls. Intensities of the A and B allele (based on the Affymetrix annotation of the SNPs) are given on the x- and y-axes and the respective genotypes are indicated in blue, green, and orange.
(PDF)

Table S1 GWA results for SNPs with λ-corrected $P_{GWA}<10$ 4 and additional SNPs selected for replication. A star (*) indicates SNPs which had been identified in previous RLS GWAs [2 4]. P-values of the GWA phase are given as λ-corrected nominal P-values. Two different methods for λ correction were applied, multi-dimensional-scaling (MDS)-analysis using PLINK and variance components (VC)-analysis using the EMMAX software with the P-values listed in the respective columns "MDS λ-corrected P_{GWA}" and "VC λ-corrected P_{GWA}". The selection of SNPs for replication was based on the MDS λ-corrected P-values. r^2-values based on Hapmap CEU data are given for those SNPs which were selected for replication based on their LD with the best-associated SNP in each region. Genomic position and gene annotation refer to the hg18 genome.
(DOC)

Table S2 Replication stage association results for individual replication samples. P-values are derived from logistic regression and correcting for gender and age as well as for population stratification where possible (see Materials and Methods). Each replication sample was Bonferroni-corrected using the number of SNPs which passed quality control for the respective sample. The OR refers to the minor allele. NA; SNP could not be analysed due to failing quality control in the respective sample.
(DOC)

Table S3 Power analysis for GWA, replication and joint analysis of GWA and replication. Power calculation was performed using the CaTS power calculator [25] using a prevalence set of 0.08 and an additive genetic model. The significance level α was set at 0.05/74 for replication stage analysis and at 0.05/301,406 for genome-wide significance in the joint analysis of GWA and replication.
(DOC)

Table S4 Results of TOX3 and BC034767 mutation screening. * "A" refers to the mutant allele, "B" to the reference allele. Position refers to hg18 genome annotation. Codon numbering refers to the reference sequence NM_001146188. Data of the 1000 genomes project was obtained from the November 2010 release via the 1000 genomes browser (http://browser.1000genomes.org/index.html).
(DOC)

Table S5 Prediction of genetic risk; training- and test-set approach. Inclusion threshold P-values were derived from a logistic regression with age and sex as covariates in the training sample. # SNPs indicates the number of SNPs passing the inclusion threshold. Based on these association results, the sum score of SNPs showing the most significant effects (i.e. the number of risk alleles over all SNPs) weighted by the ln(OR) of these effects was chosen as predictor variable in the test set. Based on this sum score, an AUC and Nagelkerke's R were calculated.
(DOC)

Table S6 Demographic data of GWA and replication samples. Mean age, mean age of onset and respective standard deviations and ranges are given in years. N: number of individuals; SD: standard deviation; AAO: age of onset. GWA: Genome-wide association study; CZ: Czechia; FR: France; FIN: Finland; CA: Canada; US: United States. - indicates that this information is not applicable for the respective sample.
(DOC)

Table S7 Genotype data of GWA samples. Genotypes of the GWA samples are given for the eight best-associated SNPs (see Table 1). SNP alleles are ACGT-coded. Phenotype information includes gender (1 = male, 2 = female) and disease status (1 = unaffected, 2 = affected).
(XLS)

Acknowledgments

We are grateful to all patients who participated in this study. We thank Jelena Golic, Regina Feldmann, Sibylle Frischholz, Susanne Lindhof,

Katja Junghans, Milena Radivojkov-Blagojevic, and Bianca Schmick for excellent technical assistance.

Author Contributions

Study design: J Winkelmann, B Müller-Myhsok, T Meitinger. Recruitment and biobanking of German/Austrian RLS cases: J Winkelmann, C Trenkwalder, B Högl, K Berger, N Gross, K Stiasny-Kolster, W Oertel, CG Bachmann, W Paulus, I Fietze, V Gschliesser, B Frauscher, T Falkenstetter, W Poewe, D Spieler, M Kaffe, A Zimprich, T Meitinger. Recruitment and biobanking of KORA controls: C Gieger, T Illig, H-E Wichmann. Recruitment and biobanking of Canadian RLS cases and controls: L Xiong, J Montplaisir, GA Rouleau. Czech RLS cases and controls: Jávrová, D Kemlink, K Sonka, S Nevsimalova, P Vodicka. US cases and controls: S-C Lin, Z Wszolek, C Vilariño-Güell, MJ Farrer, RP Allen, CJ Earley, WG Ondo, W-D Le, P Peppard, J Faraco, E Mignot. Finnish cases and controls: O Polo, J Kettunen, M Perola, K Silander. French cases and controls: Y Dauvilliers, I Cournu-Rebeix, M Francavilla, C Fontenille, B Fontaine. Affymetrix genotyping: B Schormair, P Lichtner. Sequenom genotyping: B Schormair, F Knauf, EC Schulte, P Lichtner. Sequencing and Fragment analysis: F Knauf. Expression analysis: EC Schulte, H Prokisch. Supervision of all markers typed: J Winkelmann, P Lichtner. Statistical analysis: D Czamara, B Müller-Myhsok. Clustering of Affymetrix genotypes: D Czamara, B Müller-Myhsok. Wrote the manuscript: J Winkelmann, D Czamara, B Schormair, B Müller-Myhsok, T Meitinger.

References

1. Allen RP, Picchietti D, Hening WA, Trenkwalder C, Walters AS, et al. (2003) Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology. A report from the restless legs syndrome diagnosis and epidemiology workshop at the National Institutes of Health. Sleep Med 4: 101 119.
2. Winkelmann J, Schormair B, Lichtner P, Ripke S, Xiong L, et al. (2007) Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. Nat Genet 39: 1000 1006.
3. Stefansson H, Rye DB, Hicks A, Petursson H, Ingason A, et al. (2007) A genetic risk factor for periodic limb movements in sleep. N Engl J Med 357: 639 647.
4. Schormair B, Kemlink D, Roeske D, Eckstein G, Xiong L, et al. (2008) PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome. Nat Genet 40: 946 948.
5. Engstrom PG, Fredman D, Lenhard B (2008) Ancora: a web resource for exploring highly conserved noncoding elements and their association with developmental regulatory genes. Genome Biol 9: R34.
6. Kikuta H, Laplante M, Navratilova P, Komisarczuk AZ, Engstrom PG, et al. (2007) Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. Genome Res 17: 545 555.
7. Yuan SH, Qju Z, Ghosh A (2009) TOX3 regulates calcium-dependent transcription in neurons. Proc Natl Acad Sci U S A 106: 2909 2914.
8. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, et al. (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447: 1087 1093.
9. Mercer TR, Dinger ME, Mattick JS (2009) Long non-coding RNAs: insights into functions. Nat Rev Genet 10: 155 159.
10. Dittmer S, Kovacs Z, Yuan SH, Siszler G, Kögl M, et al. (2011) TOX3 is a neuronal survival factor that induces transcription depending on the presence of CITED1 or phosphorylated CREB in the transcriptionally active complex. J Cell Sci 124: 252 60.
11. Meisinger C, Prokisch H, Gieger C, Soranzo N, Mehta D, et al. (2009) A genome-wide association study identifies three loci associated with mean platelet volume. Am J Hum Genet 84(1): 66 71.
12. Wacholder S, Hartge P, Prentice R, Garcia-Closas M, Feigelson HS, et al. (2010) Performance of common genetic variants in breast-cancer risk models. N Engl J Med 362: 986 993.
13. Lango H, Palmer CN, Morris AD, Zeggini E, Hattersley AT, et al. (2008) Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. Diabetes 57: 3129 3135.
14. van Hoek M, Dehghan A, Witteman JC, van Duijn CM, Uitterlinden AG, et al. (2008) Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study. Diabetes 57: 3122 3128.
15. Wichmann HE, Gieger C, Illig T (2005) KORA-gen resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen 67 Suppl 1: S26 30.
16. Happe S, Vennemann M, Evers S, Berger K (2008) Treatment wish of individuals with known and unknown restless legs syndrome in the community. J Neurol 255: 1365 1371.
17. Pardini B, Naccarati A, Polakova V, Smerhovsky Z, Hlavata I, et al. (2009) NBN 657del5 heterozygous mutations and colorectal cancer risk in the Czech Republic. Mutat Res 666: 64 67.
18. Cournu-Rebeix I, Genin E, Leray E, Babron MC, Cohen J, et al. (2008) HLA-DRB1*15 allele influences the later course of relapsing remitting multiple sclerosis. Genes Immun 9: 570 574.
19. Young T, Palta M, Dempsey J, Peppard PE, Nieto FJ, et al. (2009) Burden of sleep apnea: rationale, design, and major findings of the Wisconsin Sleep Cohort study. Wmj 108: 246 249.
20. Affymetrix Inc. (2007) BRLMM-P: a Genotype Calling Method for the SNP 5.0 Array. http://www.affymetrix.com/support/technical/whitepapers.affx. Accessed 03. December 2010.
21. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559 575.
22. Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55: 997 1004.
23. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, et al. (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42: 348 54.
24. Devlin B, Bacanu SA, Roeder K (2004) Genomic controls to the extreme. Nat Genet 36: 1129 1130.
25. Skol AD, Scott LJ, Abecasis GR, Boehnke M (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet 38: 209 213.
26. Staden R, Beal KF, Bonfield JK (2000) The Staden package, 1998. Methods Mol Biol 132: 115 130.

2.2 Schulte *et al.*, Blood cis-eQTLs in Prioritizing Sub-Threshold Association Signals from Genome-Wide Association Studies in Restless Legs Syndrome for Follow-Up Analysis, submitted to *PLoS ONE*, 2013[II]

Personal contributions: I designed the study, selected the subthreshold SNPs to be analyzed, analyzed *cis*- and *trans*-eQTLs derived from existing genotyping and expression data in KORA and SHIP-TREND, designed the multiplex PCR for Sequenom®-based genotyping and analyzed genotyping data. I wrote the manuscript and designed all figures and tables.

Blood *cis*-eQTLs in Prioritizing Sub-Threshold Association Signals from Genome-Wide Association Studies in Restless Legs Syndrome for Follow-Up Analysis

Eva C. Schulte[1,2], Katharina Schramm[2,3], Claudia Schurmann[4], Peter Lichtner[3], Christian Herder[5], Michael Roden[5,6], Christian Gieger[7], Annette Peters[8], Claudia Trenkwalder[9,10], Birgit Högl[11], Birgit Frauscher[11], Klaus Berger[12], Ingo Fietze[13], Nadine Gross[1], Karin Stiasny-Kolster[14,15], Wolfgang Oertel[14], Cornelius G. Bachmann[16], Walter Paulus[17], Alexander Zimprich[18], Henry Völzke[19], Ulf Schminke [20], Matthias Nauck[21], Thomas Illig[22,23], Bertram Müller-Myhsok[24], Holger Prokisch[2,3], Juliane Winkelmann[1,2,3,25]*

[1] Neurologische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität München, Munich, Germany
[2] Institut für Humangenetik, Helmholtz Zentrum München, Munich, Germany
[3] Institut für Humangenetik, Klinikum rechts der Isar, Technische Universität München, Munich, Germany
[4] Interfaculty Institute for Genetics and Functional Genomics, Ernst-Moritz-Arndt Universität Greifswald, Greifswald, Germany
[5] Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany
[6] University Clinics of Endocrinology and Diabetology, University Hospital Düsseldorf, Düsseldorf, Germany
[7] Institute for Genetic Epidemiology, Helmholtz Zentrum München, Munich, Germany
[8] Institute for Epidemiology II, Helmholtz Zentrum München, Munich, Germany
[9] Paracelsus Elena Klinik, Kassel, Germany
[10] Department of Neurosurgery, University Medical Center, Georg August Universität Göttingen, Göttingen, Germany
[11] Neurologische Klinik, Medizinische Universität Innsbruck, Innsbruck, Austria
[12] Institut für Epidemiologie und Sozialmedizin, Westfälische Wilhelms Universität Münster, Münster, Germany
[13] Zentrum für Schlafmedizin, Charite Universitätsmedizin, Berlin, Germany
[14] Neurologische Klinik, Philips Universität Marburg, Marburg, Germany
[15] Somnomar Institut für Medizinische Forschung und Schlafmedizin, Marburg, Germany
[16] Abteilung für Neurologie, Paracelsus Klinik Osnabrück, Osnabrück, Germany
[17] Department of Clinical Neurophysiology, University Medical Center, Georg August Universität Göttingen, Göttingen, Germany
[18] Neurologische Klinik, Medizinische Universität Wien, Vienna, Austria
[19] Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany
[20] Institute of Neurology, University Medicine Greifswald, Greifswald, Germany
[21] Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany
[22] Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Munich, Germany
[23] Hannover Unified Biobank, Hannover Medical School, Hannover, Germany
[24] Max-Planck Institute for Psychiatry, Munich, Germany
[25] Munich Cluster for Systems Neurology (SyNergy), Munich, Germany

*address correspondence to
Prof. Juliane Winkelmann, Neurologische Klinik und Poliklinik, Klinikum rechts der Isar, Ismaningerstr. 22, 81675 München, Germany
Phone: +49 89 4140 4688, Fax: +49 89 4140, Email: winkelmann@lrz.tum.de

Number of characters in title: 151 | Number of references: 32
Number of words in abstract: 217 | Number of figures: 1
Number of words in text body: 2886 | Number of tables: 2
Keywords: restless legs syndrome, *cis*-eQTL, peripheral blood

1

ABSTRACT

Restless legs syndrome (RLS) is a common neurologic disorder characterized by nightly dysesthesias affecting the legs primarily during periods of rest and relieved by movement. Genetically, RLS is a complex genetic disease and susceptibility factors in six genomic regions have been identified by means of genome-wide association studies (GWAS). For some complex genetic traits, expression quantitative trait loci (eQTLs) are enriched among trait-associated single nucleotide polymorphisms (SNPs). With the aim of identifying new genetic susceptibility factors for RLS, we assessed the 332 best-associated SNPs from the genome-wide phase of the to date largest RLS GWAS for *cis*-eQTL effects in peripheral blood from individuals of European descent. In 740 individuals belonging to the KORA general population cohort, 53 *cis*-eQTL with $p_{nominal}<10^{-3}$ were identified. Twenty-three hereof were replicated at $p_{nominal}<10^{-3}$ in 976 individuals belonging to the SHIP-TREND general population study. Subsequently, the nine of the 23 *cis*-eQTL SNPs, which were not located at an already published RLS-associated locus, were tested for association in 2449 RLS cases and 1462 controls. The top SNP, located in the *DET1* gene, was nominally significant (p=0.038) but did not withstand correction for multiple testing (p=0.418). Although no new genetic susceptibility factor for RLS was identified, whole blood *cis*-eQTL analysis of known and potential RLS-associated SNPs adds a novel level of functional assessment to RLS GWAS data and suggests new potential candidate genes, although it is limited by the lack of tissue-specificity for RLS.

INTRODUCTION

Restless legs syndrome (RLS) is a common sensory-motor disorder characterized by dysesthesias affecting the legs, triggered by periods of rest, relieved by movement and occurring mostly during the evening and at night. [1] Consequences are severe sleep disturbances, depression, anxiety and possibly also increased cardiovascular risk. [2,3] With age-dependent prevalences of up to 10% in the adult European population, it is one of the most common neurologic diseases. [4] RLS is a complex polygenic phenotype and genome-wide association studies (GWAS) have already identified a total of six RLS-associated genomic loci associated with the disease. [5-8] Known susceptibility alleles confer relatively high risk increases when compared to other complex genetic traits (odds ratios (OR) of up to 2.75) [5]. It is likely that additional risk loci of weaker effect sizes exist that have not yet been ascertained in the GWAS carried out so far. It has been shown that single nucleotide polymorphisms (SNPs) associated with complex genetic traits are more likely to have an effect on gene expression and, thus, represent expression quantitative trait loci (eQTLs).

[9,10] The use of *cis*-eQTL analyses in prioritizing sub-threshold association signals for GWAS follow-up, has already been successfully employed with regard to several complex diseases such as Crohn's disease [11], asthma [12], or schizophrenia [13]. Accordingly, we sought to prioritize sub-threshold RLS association signals from an RLS GWAS [7] via *cis*-eQTLs in the human blood for follow-up association study seeking to highlight additional genetic factors involved in RLS.

MATERIALS & METHODS

Ethics Statement

The KORA and SHIP-TREND studies as well as the recruitment of the RLS case/control sample was carried out in accordance with the recommendations of the Declaration of Helsinki and was approved by ethics committee of the "Bayerische Landesärztekammer" and the Technische Universität München (for KORA and the RLS case/control sample) and the University of Greifswald (for SHIP-TREND). Written informed consent was obtained from each of the study participants.

Study Design and SNP selection

The objective of the study was to use blood-based *cis*-eQTL analysis as a filter in the identification of new RLS susceptibility factors from sub-threshold association signals from a previously-published GWAS. We selected all SNPs with an association signal of $p_{nominal}<1x10^{-3}$ (λ-corrected, n=332) in the genome-wide phase (i.e. discovery phase) of a recently published RLS GWAS (n=922 cases and 1526 controls) [7] for *cis*-eQTL analysis to identify SNPs linked to differential mRNA expression (*cis*-eSNPs). These 332 signals represent 197 loci containing a single SNP and 101 loci with two or more SNPs in very high linkage disequilibrium (LD; $r^2\geq0.8$). All 332 SNPs with $p_{nominal}<1x10^{-3}$ were analyzed for *cis*-eQTLs first in 740 individuals belonging to the KORA general population-based study and subsequently replicated in 976 individuals belonging to the SHIP-TREND general population-based studies. The *cis*-eSNPs with $p_{nominal}<1x10^{-3}$ not located at loci of published association with RLS [5-8] were replicated in an independent case/control sample (Figure 1) with the objective of identifying new SNPs associated with the RLS phenotype.

Cohorts and Case/Control Samples

KORA F4 cohort (discovery sample)

Based in southwestern Germany, KORA (Cooperative Health Research in the Region of Augsburg) is a regional research platform for population-based surveys and follow-up studies. Four cross-sectional health surveys have been performed at five-year intervals between 1984 and 2001 with more than 18,000 subjects' participating. Each of the four survey contained independent random samples of individuals of German nationality resident in the city of Augsburg or one of sixteen adjacent communities. Blood samples were collected from 3080 subjects as part of the KORA F4 examination between 2006 and 2008 [14] . The expression analysis in this study was based on whole-blood samples of 740 KORA F4 participants aged 62 to 81 years [15] and represents the discovery sample.

SHIP-TREND cohort (replication sample)

SHIP (Study of Health in Pomerania in northeastern Germany) is a population-based project consisting of two independent cohorts, SHIP and SHIP-TREND. SHIP has been described in detail. [16] For this eQTL analysis, the SHIP-TREND cohort was used where a total of 976 samples, from individuals aged 20 to 81 years, with both imputed genotypes and whole-blood gene expression data were available. [17] SHIP-TREND was used as a replication sample for *cis*-eSNPs identified in the KORA discovery sample.

Case/control sample for eSNP association study

The replication sample comprised 2449 German and Austrian individuals with RLS (average age 48.0±34.7 years, 70.7% female) and 1462 individuals belonging to the S4 survey of the KORA general population cohort [18] (average age 49.9±13.4 years, 51.7% female), who were not genotyped in the genome-wide phase of the GWAS [7]. Both case and control populations were entirely of European descent. In all patients, diagnosis was based on the diagnostic criteria of the International RLS Study Group [1] as assessed in a personal interview conducted by an RLS expert. We excluded patients with secondary RLS due to uremia, dialysis, or anemia due to iron deficiency. The presence of secondary RLS was determined by clinical interview, physical and neurological examination, blood chemistry, and nerve conduction studies whenever deemed clinically necessary.

Genome-wide genotyping

As described previously [17,20], genome-wide genotyping of the KORA sample was performed on Affymetrix Genome-Wide Human SNP Arrays 6.0 following the

manufacturer`s protocol. SNPs with minor allele frequency (MAF) <5%, a call rate <98% or a significant deviation from Hardy-Weinberg Equilibrium (HWE) ($p<1x10^{-5}$) were excluded. The SHIP-TREND probands (n=986) were genotyped using the Illumina HumanOmni2.5-Quad arrays. Genotypes were imputed to HapMap v22 using IMPUTE [19].

Determination of cis- and trans-eQTLs

The eQTL analyses were performed based on gene expression data derived from whole-blood samples as described earlier [20]. Briefly, in both studies blood was taken and stored in PAXgene blood RNA tubes, RNA was isolated using the PAXgene Blood miRNA Kit (Qiagen, Hilden, Germany) and reverse transcribed using the Illumina TotalPrep-96 RNA Amp Kit (Ambion, Darmstadt, Germany). The labelled cRNA was hybridized to the Illumina HumanHT-12 v3 Expression BeadChip. After quality control there were altogether 740 KORA F4 and 976 SHIP-TREND samples with gene expression and as well as genotype data. For KORA F4 and SHIP-TREND raw gene expression intensities were exported from Illumina's 'GenomeStudio' software to the R environment. Data were log2-transformed and quantile normalized. The eQTL analysis was carried out for all 332 selected sub-threshold SNPs. For *cis*-eQTL analyses, all probes less than 500 kilobases (kb) away from the 332 selected SNP were used. Associations between the SNP and the respective mRNA probes were analyzed with a linear model adjusted for age and sex. P-values were corrected using the Benjamini and Hochberg procedure. For SHIP-TREND, 324 out of the 332 selected SNPs were analyzed directly. For eight SNPs (rs141616460, rs158019569, rs29574363, rs42812468, rs5679963, rs78275963, rs89086653 and rs9920066), proxy SNPs with $r^2=1$ (SNP Annotation and Proxy Search (SNAP), HapMap release 22, CEU panel [21]) were used.

Genotyping of replicated cis-eQTL SNPs in case/control replication sample

Genotyping was performed on the MassARRAY system using MALDI-TOF mass spectrometry with iPLEX Gold chemistry (Sequenom Inc, San Diego, CA, USA). Primers were designed using AssayDesign 3.1.2.2 with iPLEX Gold default parameters. Automated genotype calling was performed with SpectroTYPER 3.4. Genotype clustering was visually checked by an experienced evaluator. SNPs with a call rate<95%, MAF<5%, and Hardy-Weinberg p-value < $1x10^{-5}$ in controls were excluded. Known RLS-associated SNPs were not followed up in the replication. Association was tested using the allelic test as implemented in PLINK [22].

RESULTS

RLS-associated SNPs are more likely to be cis-eQTLs

To test whether RLS-associated SNPs are more commonly *cis*-eQTLs than those not associated, we compared the number of *cis*-eQTLs among the 332 most significantly associated SNPs (all with $p_{nominal}<1x10^{-3}$) from the latest RLS GWAS [7] to the 332 with the worst association p-values. Minor allele frequency (MAF) distribution when MAFs were divided into 5% bins was similar in both groups. Among the associated 332 SNPs, 52 *cis*-eQTLs ($p_{nominal}<1x10^{-3}$) were found while 37 *cis*-eQTLs were present among the 332 not associated SNPs. After very stringent LD pruning (setting a threshold of $r^2 <= 0.5$), which was necessary as there was significantly higher LD among the potentially associated SNPs compared to the not associated SNPs, we found evidence for an enrichment of *cis*-eQTLs ($p_{nominal}<1x10^{-3}$) in the associated vs. the not associated SNPs (34 *cis*-eQTLs among 246 SNPs harboring the most significant association signals vs. 28 *cis*-eQTLs among 313 SNPs showing the least significant association signals; Fisher's exact test, one-sided, $p < 0.05$, OR = 1.63).

Analysis of specific cis-eQTLs

Among the 332 best-associated SNPs, 52 *cis*-eSNPs with $p_{nominal}<1x10^{-3}$ resulting in 45 independent *cis*-eQTLs were found when assessed in whole-blood samples from 740 KORA controls. These 45 *cis*-eQTLs represent 33 loci with LD<0.8. Four SNPs (intronic rs17487827 in *BARD1* as well as intronic rs6714954, rs7592599 and rs13387588 in *SLC4A5*) represented *cis*-eQTLs of transcriptome-wide significance ($p_{nominal}<1x10^{-8}$) (Table 1). Also, of the six known RLS loci [5,6,7,8], only SNPs at loci on chromosome 6p and 15q are *cis*-SNPs with $p_{nominal}<1x10^{-3}$ (Table 1). In both cases, the expression change seen, however, did not primarily affect the candidate genes at these loci but rather genes in the vicinity (Table 2). Of the 52 *cis*-eQTLs identified in the KORA sample, 23 (44.2%) could be replicated in peripheral blood mRNA profiles of 976 SHIP-TREND samples with $p_{nominal}<1x10^{-3}$ and eleven of these reached transcriptome-wide significance when using a 5% Benjamini-Hochberg false discovery rate for multiple testing correction. rs17487827 in *BARD1* was the only *cis*-eSNP that reached transcriptome-wide significance independently in both cohorts. The 23 replicated *cis*-eSNPs contained nine that were dependent upon SNPs at the known RLS-associated locus on chromosome *15q* (*MAP2K5/SKOR1*) (Table 2). Interestingly, none of these were dependent upon expression changes of *SKOR1* and only two (rs4489954 and rs28670272) affected the expression of *MAP2K5*, the two candidate genes underlying the

GWAS association signal at this locus. Instead, seven *cis*-eSNPs coincided with differential expression of *CALML4*, located approximately 400 kb upstream of the known locus. The remaining 14 *cis*-eSNPs represented twelve individual loci as three SNPs (rs7592599, rs6714954 and rs13387588) located in *SLC4A5* all associated with decreased expression levels of two neighboring genes, *AUP1* and *MRPL53*. Directions of differential expression only concurred in 42.9% (6 out of 14) of replicated *cis*-eSNPs (Table 1).

Trans-eQTLs linked to RLS-associated SNPs

In order to examine possible *trans*-effects of SNPs of known association with the RLS phenotype, we also assessed transcriptome-wide *trans*-eQTLs in the whole-blood samples for 13 SNPs known to be associated with RLS [5-8]. However, none of the *trans*-eQTLs identified in KORA F4 or SHIP-TREND could be replicated at $p_{nominal}<1x10^{-3}$ in the respective other cohort (data not shown).

Replication of sub-threshold SNPs representing cis-eQTLs

Twelve *cis*-eSNPs with $p_{nominal}<10^{-3}$ in both the KORA and the SHIP-TREND study were selected for replication in an independent sample comprising 2449 German and Austrian RLS cases and 1462 general population-based controls belonging to the KORA cohort. Due to technical reasons, intergenic SNP rs6746899 could not be included in the replication. One SNP in *DET1* (rs9920066) showed nominally significant association ($p_{nominal}=0.03$) but did not withstand Bonferroni correction ($p_{corrected}=0.41$) while the other ten SNPs were not associated with the RLS phenotype in the replication sample (Table1).

Expression in Brain

The relevance of blood *cis*-eQTLs or *cis*-eSNPs —that is *cis*-eQTLs reflecting SNP-dependent mRNA expression changes in whole blood samples—to neurologic and psychiatric diseases has been shown. [9] However, differences between blood eQTLs or eSNPs and brain eQTLs or eSNPs —that is those *cis*-eQTLs or eSNPs showing SNP-dependent mRNA expression changes in brain tissue samples— have also been demonstrated. [23] Therefore, we analyzed the brain *cis*-eQTL status of all 23 replicated *cis*-eSNPs using the NCBI GTEx eQTL browser (http://www.ncbi.nlm.nih.gov/gtex/GTEX2/gtex.cgi, accessed August 5, 2012) and the seeQTL browser (http://gbrowse.csbio.unc.edu/cgi-bin/gb2/gbrowse/seeqtl/, accessed August 5, 2012). None of the 23 blood *cis*-eSNPs were also *cis*-eSNPs with $p_{nominal}<1x10^{-3}$ in the cerebellum, frontal and temporal cortex or pons (n=142 to 144) [24] or in whole brain (n=193). [25]

8

DISCUSSION

Blood *cis*-eQTL analysis has been successfully used in enhancing output from GWAS. [11, 12, 13] Here, we evaluated *cis*-eSNPs and *cis*-eQTLslinked to (potential) RLS susceptibility genes identified in previous RLS GWAS in order to prioritize sub-threshold candidates for follow-up evaluation. In doing so, we identified one SNP in the de-etiolated 1 encoding gene *DET1* (rs9920066) that reached nominal significance in the replication phase but did not withstand correction for multiple testing. Re-evaluation of the *DET1* locus in data from the genome-wide phase of the last RLS GWAS [7] uncovered three additional SNPs—one in high LD (r^2=0.96) and two in lower LD ($r^2 \leq 0.4$) with rs9920066—marginally associated with the RLS phenotype at $p_{nominal}<0.05$. A direct functional link between *DET1* and the RLS phenotype is not apparent. *DET1* is known to endorse degradation of the protooncogenic transcription factor c-Jun via a multi-subunit ubiquitin ligase. [26] It is much better characterized in the plant *Arabidopsis thaliana* where it acts in light-dependent transcriptional repression in the circadian clock. [27] Even if rs9920066 were a true susceptibility factor, risk increases conferred by the A allele of rs9920066 would be small (OR = 1.11 (95% confidence interval: 1.00-1.22)). Statistical power calculation using the Purcell Power Calculator [28] revealed that in order to replicate an association for one SNP with MAF = 0.30 at $\alpha = 0.05$, one would need a minimum of 5,767 cases and 5,767 controls to achieve 80% power for detecting effects with an OR of 1.11, sample sizes which are currently unattained even by the world-wide RLS genetics consortium.

Moreover, when drawing conclusions on RLS-specific functions, one caveat has to be that *cis*-eQTLs employed in selecting SNPs for replication were evaluated in peripheral blood and not a more disease-specific tissue. Although the underlying pathophysiology is not entirely clear, an involvement of the central nervous system in RLS pathophysiology seems most likely. Evaluation of the 23 replicated *cis*-eSNPs in two human brain expression data sets showed that none of the blood *cis*-eQTLs were also found in the brain samples. One reason could be that the number of brain samples in both studies was smaller than the number of blood samples evaluated in our study (347 brain samples vs. 1716 blood samples) and the analysis could, therefore, be underpowered to detect the same effects. Also, it is possible that *cis*-eQTLs dependent on (potentially) RLS-associated SNPs in the peripheral blood do not overlap with those in the brain and are not functionally relevant for disease pathogenesis. Lastly, it is known that eQTLs can be specific to developmental time points [24,25] and brain regions [23] and that they were, therefore, not detected in the available data. However, at the

moment, neither the time point nor the brain region most relevant to RLS pathophysiology has been identified and can be evaluated specifically.

Although none of the RLS-associated SNPs selected for follow-up could be replicated, two additional interesting aspects emerged. For one, one of the known RLS susceptibility loci on chromosome 15q [5-7] comprising RLS candidate genes *MAP2K5* and *SKOR1* harbored nine *cis*-eQTLs with $p_{nominal}<1x10^{-3}$. Two of these showed altered *MAP2K5* expression dependent on the RLS-risk allele though in different directions in KORA and SHIP-TREND, while none were related to altered expression of *SKOR1*. Interestingly, seven RLS-linked SNPs in *MAP2K5* were further related to altered expression of calmodulin-like 4 (*CALML4*), a gene located approximately 400 kb upstream of the RLS-associated *MAP2K5/SKOR1* locus. However, here, too, the direction of differential expression was not the same in both cohorts. In one study, Calml4 was linked to murine chondrogenesis [29], however, beyond this, data are lacking. Despite the fact that several studies have been successful in using *cis*-eQTLs to fine-map or provide functional support for specific genes at a GWAS locus [12,30,31], here, the situation is not as clear. It is possible that potential RLS-associated expression changes in *CALML4* are due to SNPs in *CALML4,* which are in high LD with RLS-associated SNPs at the *MAP2K5/SKOR1* locus. Alternatively, it cannot be excluded that variation in *CALML4* instead of, or in addition to, *MAP2K5/SKOR1* could play a role in RLS pathogenesis, as has been postulated with regard to other complex traits such as the body mass index [32] or that these expression changes are artificial as they do not concur in the two cohorts.

Lastly, one *cis*-eSNP in *BRCA1*-associated ring domain, *BARD1*,was found with transcriptome-wide significance independently in both KORA and SHIP-TREND. However, the SNP`s association with the RLS phenotype could not be replicated in our second case/control sample and is, therefore, unclear.

Overall, we were unable to establish a new genetic susceptibility factor for RLS, although, at least in the case of *DET1*, this may be due to the lack of power to replicate alleles conferring only a small risk increase. However, our analysis of peripheral blood *cis*-eQTLs has added a new level to the functional assessment of GWAS results for RLS and the fact that the best associated SNPs harbor more blood *cis*-eQTLs than the worst associated SNPs justifies this approach specifically with regard to RLS. DET1 emerged as a new candidate gene of potential relevance to RLS although our study is challenged by the fact that *cis*-eQTLs were evaluated in peripheral blood and not a tissue of more pathophysiologic relevance to RLS. In the future, as the neuroanatomic correlates of RLS become more defined and more expression profiles of different brain regions become available, it will be interesting to assess whether

these also play a role in brain-region-specific, RLS-allele-dependent eQTLs and in disease development.

ACKNOWLEDGEMENTS

We are very grateful to Jelena Golic, Susanne Lindhof, Katja Junghans, Regina Feldmann and Sybille Frischholz at Institute for Human Genetics at the Helmholtz Zentrum München for expert technical assistance. Recruitment of the KORA cohort was supported by institutional (Helmholtz Zentrum München, Munich, Germany) and government funding from the German Bundesministerium für Bildung und Forschung (03.2007-02.2011 FKZ 01ET0713). The German Diabetes Center is funded by the German Federal Ministry of Health (Berlin, Germany) and the Ministry of Innovation, Science and Research of the State of North Rhine-Westphalia (Düsseldorf, Germany). This study was supported in part by a grant from the German Federal Ministry of Education and Research (BMBF) to the German Center for Diabetes Research (DZD e.V.). SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Deutsche Forschungsgemeinschaft (DFG GRK840-D2), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania. This work is also part of the research project Greifswald Approach to Individualized Medicine (GANI_MED), which is funded by the Federal Ministry of Education and Research and the Ministry of Cultural Affairs of the Federal State of Mecklenburg–West Pomerania (03IS2061A). Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg, West Pomerania. Whole-body MR imaging was supported by a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg West Pomerania. The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG and the Caché Campus program of the InterSystems GmbH. The SHIP authors are grateful to Mario Stanke for the opportunity to use his Server Cluster for the SNP imputation.

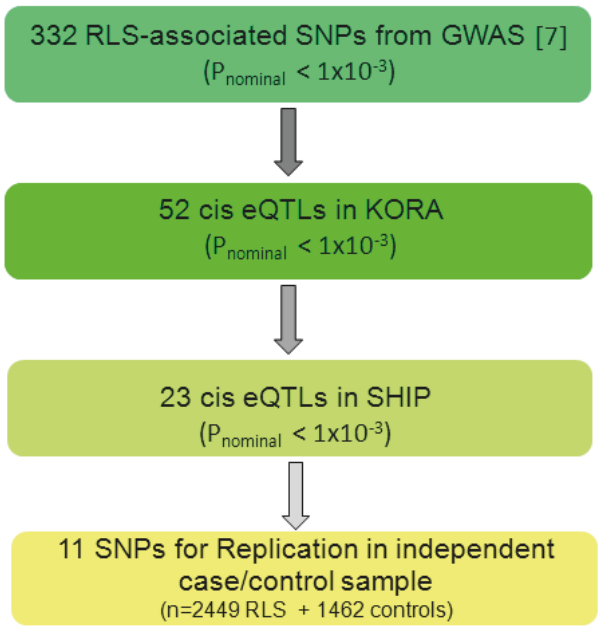
REFERENCES

1. Allen RP, Picchietti D, Hening WA, Trenkwalder C, Walters AS, et al. (2003) Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology. A report from the restless legs syndrome diagnosis and epidemiology workshop at the National Institutes of Health. Sleep Med 4:101-119.
2. Winkelmann J, Prager M, Lieb R, Pfister H, Spiegel B, et al. (2005) "Anxietas tibiarum". Depression and anxiety disorders in patients with restless legs syndrome. J Neurol 252: 67-71.
3. Winkelman JW, Shahar E, Sharief I, Gottlieb DJ (2008) Association of restless legs syndrome and cardiovascular disease in the Sleep Heart Health Study. Neurology 70: 35-42.
4. Allen RP, Walters AS, Montplaisir J, Hening W, Myers A, Bell TJ, Ferini-Strambi L (2005) Restless legs syndrome prevalence and impact: REST general population study. Arch Intern Med 165: 1286-1292.
5. Winkelmann J, Schormair B, Lichtner P, Ripke S, Xiong L, et al. (2007) Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. Nat Genet 39: 1000-1006.
6. Schormair B, Kemlink D, Roeske, D, Eckstein G, Xiong L, et al. (2008) PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome. Nat Genet 40: 946-948.
7. Winkelmann J, Czamara D, Schormair B, Knauf F, Schulte EC, et al. (2011) Genome-Wide association Study identifies novel restless legs syndrome susceptibility loci on 2p14 and 16q12.1. PLoS Genet 7: e1002171.
8. Stefansson H, Rye DB, Hicks A, Petursson H, Ingason A, et al. (2007) A genetic risk factor for periodic limb movements in sleep. N Engl J Med 357: 639-647.
9. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, et al. (2010) Trait-Associated SNPs Are More Likely to Be eQTLs: Annoation to Enhance Discovery from GWAS. PLoS Genet 6: e1000888.
10. Richards AL, Jones L, Moskvina V, Kirov G, Gejman PV, et al. (2012) Schizophrenia susceptibility alleles are enriched for alleles that affect gene expression in adult human brain. Molec Psych 17:193-201.
11. Fransen K, Visschedijk MC, van Sommeren S, Fu JY, Franke L, et al. (2010) Analysis of SNPs with an effect on gene expression identifies UBE2L3 and BCL3 as potential new risk genes for Crohn's disease. Hum Mol Genet 19: 3482-3488.
12. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, et al. (2007) Genetic variants regulating childhood asthma. Nature 448: 470-473.
13. Bray NJ, Preece A, Williams NM, Moskvina V, Buckland PR, et al. (2005) Haplotypes at the dystrobrevin binding protein 1 (DTNBP1) gene locus mediate risk for schizophrenia through reduced DTNBP1 expression. Hum Mol Genet 14: 1947-1954.
14. Holle R, Happich M, Löwel H, Wichman HE, MONICA/KORA Study Group (2005) KORA—a research platform for population based health research. Gesundheitswesen 67: S19-25.
15. Rathmann W, Strassburger K, Heier M, Holle R, Thorand B, et al. (2009) Incidence of Type 2 diabetes in the elderly German population and the effect of clinical and lifestyle risk factors: KORA S4/F4 cohort study. Diabet Med 26:1212-1219.
16. Völzke H, Alte D, Schmidt CO, Radke D, Lorbeer R, et al. (2011) Cohort profile: the study of health in Pomerania. Int J Epidemiol: 40, 294-307.
17. Schurmann C, Heim K, Schillert A, Blankenberg S, Carstensen M, et al. (2012) Analyzing Illumina Gene Expression Microarray Data from Different Tissues:

Methodological Aspects of Data Analysis in the MetaXpress Consortium. PLoS One 7: e50938.
18. Wichmann HE, Gieger C, Illig T (2005) KORA-gen–resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen 67: Suppl 1S26–30.
19. Howie BN, Donnelly P, Marchini JA (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5: e1000529.
20. Mehta D, Heim K, Herder C, Carstensen M, Homuth G, et al. (2012) Impact of common regulatory single-nucleotide variants on gene expression profiles in whole blood. Eur J Hum Genet: doi:10.1038.
21. Johnson AD, Handsaker RE, Pulit S, Nizzari MM, O'Donnell CJ, de Bakker PIW (2008) SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24:2938-2939.
22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al. (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics 81: 559-575.
23. Hernandez DG, Nalls MA, Moore M, Chong S, Dillman A, et al. (2012) Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain. Neurobiol Dis 47: 20-28.
24. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, et al. (2010) Abundant quantitative trait loci exist for DNA methylation and gene expression in the human brain. PLoS Genet 6:e1000952.
25. Myers AJ, Gibbs JR, Webster JA, Rohrer K, Zhao A, et al. (2007) A survey of genetic human cortical gene expression. Nat Genet 39: 1494-1499.
26. Wertz IE, O`Rourke KM, Zhang Z, Dornan D, Arnott D, et al. (2004) Human de-etiolated-1 regulates c-Jun by assembling a CUL4A ubiquitin ligase. Science 303: 1371-1374.
27. McClung CR (2011) The photomorphogenic protein, DE-ETIOLATED 1, is a critical transcriptional corepressor in the central loop of the Arabidopsis circadian clock. Mol Cell 43: 693-694.
28. Purcell S, Cherny SS, Sham PC (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics 19:149-150.
29. Chen L, Fink T, Zhang XY, Ebbesen P, Zachar V (2005) Quantitative transcriptional profiling of ATDC5 mouse progenitor cells during chondrogenesis. Differentiation 73:350-363.
30. Okada Y, Shimane K, Kochi Y, Tahira T, Suzuki A, et al. (2012) A Genome-Wide Association Study Identified AFF1 as a Susceptibility Locus for Systemic Lupus Erythematosus Japanese. PLoS Genet 8: e1002455.
31. Wild PS, Zeller T, Schillert A, Szymczak S, Sinning CR, et al. (2011) A Genome-Wide Association Study Identifies LIPA as a Susceptibility Gene for Coronary Artery Disease. Circ Cardiovasc Genet 4: 403-412.
32. Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, et al. (2009) Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. Nat Genet 41: 25-34.

FIGURE LEGEND

Figure 1: Study Design.



TABLES

Table 1: RLS-associated SNPs representing *cis*-eQTLs in peripheral blood.

14

SNP	Locted in	Gene expression altered	Association ($p_{nominal}$ from past GWAS) [7]	KORA *cis*-eQTL ($p_{nominal}$)	Major Allele KORA	KORA beta	SHIP-TREND *cis*-eQTL ($p_{nominal}$)	SHIP-TREND major allele	SHIP-TREND beta	Association Replication ($p_{nominal}$)	Association Replication ($p_{corrected}$)
rs17487827	***BARD1***	***BARD1***	**0.00021**	**$1.02x10^{-13}$**	C	-0.171	**$9.20x10^{-17}$**	C	-0.110	0.255	NS
rs6746899	***intergenic***	***intergenic***	**0.00089**	**$1.53x10^{-8}$**	A	-0.172	**$1.61x10^{-8}$**	A	-0.130	no assay	
rs6714954	***NBC4***	***MRPL53/AUP1***	**0.00021**	**$2.22x10^{-8}$**	G	-0.145	**$8.50x10^{-7}$**	G	0.104	0.148	NS
rs9920066	***DET1***	***DET1***	**0.00053**	**$2.26x10^{-8}$**	T	-0.071	**$2.31x10^{-8}$**	T	-0.054	0.038	NS (0.418)
rs9354792	***CR595314***	***CR595314***	**0.00078**	**$3.20x10^{-8}$**	A	0.086	**$3.36x10^{-10}$**	A	0.048	0.521	NS
rs6696524	*SCMH1*	*SCMH1*	0.00084	$1.23x10^{-7}$	G	-0.106					
rs2250205	*EIF6/ITGB4BP*	*EIF6*	0.00033	$5.00x10^{-7}$	G	0.141					
rs17125761	***intergenic***	***ERO1L***	**0.00054**	**$9.40x10^{-7}$**	T	-0.091	**$4.85x10^{-5}$**	T	0.066	0.965	NS
rs1076094	*GRB2*	*GRB2/NUP85*	0.0002	$1.87x10^{-6}$	C	-0.142					
rs4646861	*ALDH6H1*	*LIN52*	0.00071	$3.88x10^{-6}$	A	0.071					
rs11024433	***SERGEF***	***SAAL1***	**0.00088**	**$5.58x10^{-6}$**	G	-0.074	**$1.35x10^{-6}$**	G	0.051	0.557	NS
rs7670748	***intergenic***	***intergenic***	**0.00013**	**$7.86x10^{-6}$**	C	0.134	**$8.78x10^{-12}$**	C	0.125	0.499	NS
rs2250205	*EIF6/ITGB4BP*	*EIF6*	0.00033	$9.26x10^{-6}$	G	0.115					
rs2029361	***intergenic***	***SMC4***	**0.00095**	**$9.95x10^{-6}$**	C	0.099	**0.0003**	C	0.063	0.385	NS
rs738415	***intergenic***	***BC033837***	**0.00062**	**$1.04x10^{-5}$**	G	0.142	**$1.49x10^{-7}$**	G	-0.119	0.937	NS
rs4388643	***ZNF364***	***ANKRD35***	**0.00047**	**$1.51x10^{-5}$**	G	0.072	**$4.41x10^{-6}$**	G	-0.055	0.533	NS
rs12600464	*SKAP1*	*CDK5RAP3*	0.00012	$2.30x10^{-5}$	A	0.127					
rs11182236	*PUS7L*	*IRAK4*	0.0007	$3.22x10^{-5}$	G	-0.124					
rs3127065	*intergenic*	*L3MBTL*	0.00062	0.00011	A	0.047					
rs28670272	***MAP2K5***	***CALML4***	**0.00025**	**0.00019**	A	-0.064	**$4.62x10^{-6}$**	A	0.053		
rs12593813	***MAP2K5***	***CALML4/MAP2K5***	**1.49E-06**	**0.00022**	G	-0.059	**$3.65x10^{-6}$**	G	0.048		
rs2814888	*BTBD9*	*GLO1*	0.00028	0.0003	T	-0.096					
rs1822736	*intergenic*	*C12orf23*	1.28E-05	0.00034	T	0.051					
rs683856	***KCNC3***	***NAPSB***	**0.00018**	**0.00034**	T	-0.125	**$1.64x10^{-7}$**	T	0.116	0.618	NS
rs9380748	*BTBD9*	*GLO1*	0.00054	0.00035	T	-0.095					
rs2306219	*SLC25A19*	*GRB2*	0.0009	0.00058	T	-0.108					
rs11231	*RCAN3*	*SRRM1*	0.00041	0.00058	C	-0.065					
rs7602776	*intergenic*	*LOXL3*	2.14E-05	0.00061	T	-0.049					
rs3799979	*SUPT3H*	*intergenic*	0.00081	0.00063	T	0.241					
rs2025682	*intergenic*	*intergenic*	0.00019	0.0008	A	0.028					
rs10972919	*intergenic*	*MELK*	0.00094	0.00081	A	0.027					
rs1555322	*MMP24*	*CPNE1*	0.00056	0.00086	G	-0.116					
rs7916801	*XPNPEP1*	*XPNPEP1*	0.00075	0.0009	G	0.049					

A total of 52 *cis*-eQTLs dependent upon 45 RLS-associated SNPs were found in KORA ($p_{nominal}<1x10^{-3}$). These 45 SNPs represent 33 independent loci. The SNP showing the best association with the RLS phenotype is listed. SNPs for which the *cis*-eQTL was replicated in the SHIP-TREND cohort and which were carried further into the replication phase are printed in bold. NS=not significant

15

Table 2: cis-eQTLs at known RLS-associated loci.

SNP	Located in	Gene expression altered	Association ($p_{nominal}$ from past GWAS)	KORA cis-eQTL ($p_{nominal}$)	KORA major allele	KORA beta	SHIP-TREND cis-eQTL ($p_{nominal}$)	SHIP-TREND beta	SHIP-TREND major allele
rs4591850	*BTBD9*	*GLO1*	0.000377322	0.000283749	T	-0.09549	2.73E-01	0.0183	T
rs2814888	*BTBD9*	*GLO1*	0.000284845	0.000306096	T	-0.09579	1.89E-01	0.0221	T
rs9380748	*BTBD9*	*GLO1*	0.000541914	0.000358813	T	0.09466	2.61E-01	0.0188	T
rs12593813	*MAP2K5*	*CALML4*	1.49E-06	0.00022361	G	-0.05867	3.65E-06	0.0484	G
rs11635424	*MAP2K5*	*CALML4*	1.52E-06	0.000226735	G	-0.05867	3.59E-06	0.0485	G
rs868037	*MAP2K5*	*CALML4*	9.51E-07	0.000267826	G	-0.05848	2.62E-06	0.0491	G
rs4489954	*MAP2K5*	*CALML4*	8.26E-06	0.000388586	G	-0.05767	1.32E-06	-0.0519	G
rs1026732	*MAP2K5*	*CALML4*	4.01E-06	0.000515309	G	-0.05569	3.56E-06	0.0484	G
rs6494696	*MAP2K5/LBXCOR1*	*CALML4*	3.70E-06	0.000424605	G	-0.05669	3.23E-06	0.0487	G
rs28670272	*MAP2K5/LBXCOR1*	*CALML4*	0.0002593	1.93E-04	A	-0.06438	4.62E-06	-0.0531	A
rs4489954	*MAP2K5*	*MAP2K5*	8.26E-06	0.0004486	G	-0.05767	0.000837691	-0.0351	G
rs28670272	*MAP2K5/LBXCOR1*	*MAP2K5*	0.0002593	0.000980145	A	-0.04949	7.13E-05	-0.0450	A

2.3 Schulte *et al.*, Variant screening of the coding regions of *MEIS1* in patients with restless legs syndrome, *Neurology*, 2011[III]

Personal contributions: I performed LightScanner® high-resolution melting curve analysis and follow-up Sanger sequencing of the coding regions of *MEIS1*, analyzed the LightScanner® data, participated in Sequenom®-based genotyping and data analysis and recruited part of the family described. I wrote the manuscript and designed all figures and tables.

Clinical/Scientific Notes

E.C. Schulte, MD
F. Knauf, MSc
D. Kemlink, MD
B. Schormair, PhD
P. Lichtner, PhD
C. Gieger, PhD
T. Meitinger, MD
J. Winkelmann, MD

VARIANT SCREENING OF THE CODING REGIONS OF *MEIS1* IN PATIENTS WITH RESTLESS LEGS SYNDROME

Restless legs syndrome (RLS) is a common and genetically complex neurologic disease presenting with an urge to move the legs and dysesthesias in the evening and at times of rest. Genome-wide association studies have linked single nucleotide polymorphisms in *MEIS1* and 3 other loci to an increased susceptibility to RLS.[1-3] However, to date, only one potentially causal variant has been reported.[4] Therefore, we screened the coding regions and exon-intron boundaries of *MEIS1* for variants, which by exerting a strong phenotypic effect could provide a basis for assessing the function of the gene in RLS.

Methods. Using Idaho LightScanner high-resolution melting curve analysis, we screened DNA of 188 patients with RLS of a first discovery sample (72.8% female, mean age 60.0 ± 11.2 years), all harboring RLS risk alleles of *MEIS1* (G/T or G/G for rs2300478), for aberrant melting patterns (e-Methods on the *Neurology*® Web site at www.neurology.org). Exons showing changes suggestive of variants were sequenced on an ABI Prism 3730 sequencer, and variants identified were subsequently genotyped in an independent German sample (henceforth termed "second sample") consisting of 735 patients with RLS (70.8% female, mean age 61.5 ± 14.2 years) and 735 unrelated control subjects (74.5% female, mean age 59.8 ± 11.3 years) by matrix-assisted laser desorption ionization/time-of-flight mass spectrometry. Disease segregation was evaluated in one family with the p.R272H mutation of exon 8 of *MEIS1*, previously related to RLS.[4] RLS in all patients was diagnosed in accordance with standard diagnostic criteria[5] (e-Methods).

Standard protocol approvals, registrations, and patient consents. Ethics review board approval and participants' written informed consent were obtained.

Results. In the discovery sample, we identified 3 novel nonsynonymous nucleotide substitutions in exons 3 (p.H81Q) and 6 (p.S204T) and in one transcript containing exon 13 (p.M453T) of *MEIS1* in one patient each. In addition, 2 patients showed the p.R272H variant.

Genotyping in the second sample revealed p.H81Q in 2 control subjects and p.M453T in 2 case patients and 3 control subjects, whereas p.S204T was observed in one case patient. p.R272H was not found in any additional case patients or control subjects (figure, A).

Segregation analysis performed on 7 family members revealed 4 affected individuals who presented the p.R272H variant (figure, B). One family member (III-1) and one married-in individual (III-3) were affected but did not show the p.R272H variant. p.R272H was not found in one unaffected individual (II-1).

Because all p.R272H patients were of Czech heritage, we further evaluated the presence of a putative p.R272H founder mutation in 279 Czech patients with RLS (63.1% female, mean age 55.8 ± 14.9 years), but did not find any additional carriers.

Discussion. Screening of the coding regions of *MEIS1* in patients with RLS revealed 3 novel variants. In all cases, patients reported one or more relatives with at least a suspicion of RLS; however, families were small and individuals in many instances not available for further study so that pathogenicity of the variants was not evaluated. We also confirmed the p.R272H variant of *MEIS1*[4] in 2 patients with RLS, one belonging to a family in which the RLS trait seems to be inherited in an autosomal dominant fashion. As observed previously, the p.R272H variant is located within the highly conserved TALE homeobox domain, which is essential for dimerization and transcription activation and disruption of which is known to be detrimental.[4] Therefore, this variant is considered the most likely candidate for an RLS-linked pathogenic mutation to date. In our sample, it was only present in RLS-affected individuals. However, segregation analysis also revealed affected individuals without the variant, suggesting that because RLS is a common disease, these cases could represent phenocopies, that is, similar phenotypes due to different genetic alterations. As opposed to the phenotype described in the North American p.R272H family, disease representation was rather homogeneous in our family with severe and early onset of symptoms. The second patient with p.R272H reported only a daughter

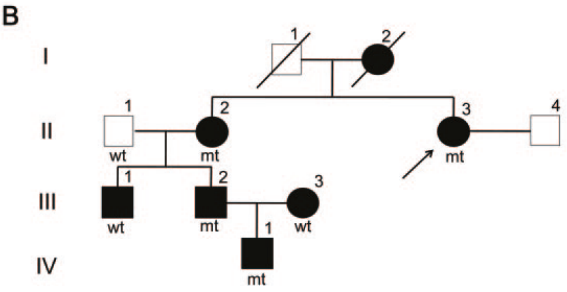
Supplemental data at www.neurology.org

Figure Variants of the coding regions of *MEIS1*

A

Region		Genomic position (chr. 2; NCBI 36/hg18)	Nucleotide/amino acid substitution		First sample	Second sample		
						Cases	Controls	MAF
MEIS1 exon 3	novel	66,520,482	CA[C/G] - His>Gln	H81Q	1/186	0/0/722	0/2/717	0.001
MEIS1 exon 6	novel	66,523,664	[T/A]CA - Ser>Thr	S204T	1/188	0/1/723	0/0/723	<0.001
MEIS1 exon 8	Ref. (4)	66,592,857	C[G/A]T - Arg>His	R272H	2/365	0/1/723	0/0/725	<0.001
MEIS1 exon 13	novel	66,652,131	A[T/C]G - Met>Thr	M453T	1/178	0/2/716	0/3/722	0.002

B



(A) Table illustrating the nucleotide and amino acid substitutions of novel coding variants and p.R272H of *MEIS1* as well as their frequency in the first (mutation carriers/total number of patients with restless legs syndrome [RLS] tested) and the second sample (homozygotes/heterozygotes/homozygotes in case patients and control subjects). (B) Segregation analysis of the p.R272H variant of *MEIS1* in an RLS family. Men are represented by squares and women by circles; a diagonal line indicates a deceased individual. For individuals sequenced, mt indicates those carrying the p.R272H variant and wt indicates noncarriers. The arrow denotes the index patient. MAF = minor allele frequency.

with possible RLS during pregnancy, which could be suggestive of variable expressivity of the p.R272H variant as well as the presence of additional modifying factors. Overall, segregation analyses remain inconclusive because of the small size of pedigrees, yet support the notion that p.R272H *MEIS1* could be causally related to RLS.

All additional variants affect amino acids highly conserved in vertebrates. Although not located within a known functional protein domain, bioinformatics algorithms[6] predict p.S204T and p.H81Q to be disease-causing, whereas p.M453T is likely to be functionally neutral. However, the fact that, as opposed to p.H81Q and p.M453T, p.S204T was only found in RLS-affected individuals renders this variant a second potential candidate for a disease-causing genetic alteration.

However, one limitation of this study is the fact that the control subjects used are general population control subjects and thus we cannot exclude the possibility that variants also found in control subjects (p.H81Q and p.M453T) are not related to RLS. Accordingly, these warrant further replication in an independent dataset.

Our results show that exonic variants in *MEIS1* are not common in RLS. However, it is still possible that rare exonic variants of strong effect could play a causative role in RLS in rare cases, as is known for other complex diseases,[7] and their study is important because they could provide significant clues toward understanding of the disease mechanism.

From the Neurologische Klinik und Poliklinik (E.C.S., J.W.) and Institut für Humangenetik (T.M., J.W.), Klinikum rechts der Isar, Technische Universität München, Munich; Institut für Humangenetik (F.K., B.S., P.L., T.M., J.W.) and Institut für Epidemiologie (C.G.), Helmholtz Zentrum München, Munich, Germany; and Department of Neurology (D.K.), Charles University, 1st Faculty of Medicine and General Teaching Hospital, Prague, Czech Republic.

Study funding: Supported by in-house institutional funding from the Helmholtz Zentrum München, Munich, Germany.

Disclosure: Dr. Schulte receives a postdoctoral fellowship from the Technische Universität München, Munich, Germany. F. Knauf reports no disclosures. Dr. Kemlink receives research support from the Czech Ministry of Education. Dr. Schormair, Dr. Lichtner, and Dr. Gieger report no disclosures. Dr. Meitinger serves on scientific advisory boards for Estonian Biobank and the European Genetics Foundation; serves on the editorial board of the European Journal of Human Genetics*; receives license fee payments for a patent from KIRIN breweries; and has served as a legal consultant. Dr. Winkelmann serves on a scientific advisory board for UCB; has received speaker honoraria from UCB and Boehringer Ingelheim; has filed a patent re: Winkelmann et al. Nat Genet 2007; and receives research support from the German RLS Foundation and the Fritz Thyssen Foundation.*

Received July 8, 2010. Accepted in final form October 10, 2010.

Address correspondence and reprint requests to Dr. Juliane Winkelmann, Neurologische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität München, Ismaningerstr. 22, 81675 Munich, Germany; winkelmann@lrz.tum.de

ACKNOWLEDGMENT

The authors thank all individuals who participated in this study and Jelena Golic for technical assistance in performing Sequenom genotyping.

1. Winkelmann J, Schormair B, Lichtner P, et al. Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. Nat Genet 2007;39:1000–1006.
2. Stefansson H, Rye DB, Hicks A, et al. A genetic risk factor for periodic limb movements in sleep. N Engl J Med 2007; 357:639–647.
3. Schormair B, Kemlink D, Roeske D, et al. PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome. Nat Genet 2008; 40:946–948.
4. Vilariño-Güell C, Chai H, Keeling BH, et al. MEIS1 p.R272H in familial restless legs syndrome. Neurology 2009;73:243–245.
5. Allen RP, Picchietti D, Hening WA, et al. Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology: a report from the restless legs syndrome diagnosis and epidemiology workshop at the National Institutes of Health. Sleep Med 2003;4:101–119.
6. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 2010;7:575–576.
7. McClellan J, King MC. Genetic heterogeneity in human disease. Cell 2010;141:210–216.

Andrew McKeon, MD
Sean J. Pittock, MD
Vanda A. Lennon, MD, PhD

CSF COMPLEMENTS SERUM FOR EVALUATING PARANEOPLASTIC ANTIBODIES AND NMO-IgG

The detection of neural-reactive immunoglobulin G (IgG) autoantibodies aids the diagnosis of organ-specific autoimmune neurologic disorders. Many paraneoplastic autoantibodies reliably predict a particular cancer type and are accompanied by varied neurologic presentations of subacute onset.[1] The detection of neuromyelitis optica (NMO)–IgG predicts a relapsing inflammatory demyelinating disorder predominated by optic neuritis and transverse myelitis.[2] When an autoimmune neurologic disorder is suspected, serologic testing of serum is frequently undertaken before more invasive CSF evaluation. However, CSF evaluation can complement testing of serum when suspicion for an autoimmune etiology persists despite a negative serum result. Here we report, for a 25-year period of testing by standardized indirect immunofluorescence protocols, the frequency of neural autoantibody detection in serum and CSF.

Methods. The immunofluorescence protocols we used were validated in this laboratory for detection of paraneoplastic antibodies (anti-neuronal nuclear antibody [ANNA]-1; ANNA-2; ANNA-3; Purkinje cell cytoplasmic antibody [PCA]-1; PCA-2; PCA-Tr; collapsin response-mediator protein [CRMP]-5-IgG; amphiphysin antibody; antiglial/neuronal nuclear antibody [AGNA]-1; NMDA receptor antibody) and NMO-IgG. We searched the Mayo Clinic Neuroimmunology Laboratory database (January 1986 to March 2010) for all patient samples submitted for service evaluation of paraneoplastic or NMO-IgG. We included both Mayo Clinic and non-Mayo patients for whom both serum and CSF were submitted, and reviewed available oncologic data for patients with antibodies identified by CSF testing.

Results. Testing was performed on a clinical service basis for a median of 12 years (range 2–25 years). The antibody detection rate in all specimens ranged from 0.01% for PCA-Tr to 7% for NMO-IgG (table).

In patients for whom paired serum and CSF samples were tested, the antibody detection rate ranged from 0.08% (PCA-Tr) to 9% (NMDA receptor antibody). One or more neural autoantibodies were detected in 462 patients (497 antibodies detected). In 405 of those 462 patients, both serum and CSF yielded a positive result (88%). In 57 patients, serum or CSF alone was positive (12%). Among those patients, serum alone yielded a positive result in 31 (54%) and CSF alone in 26 (46%). For classic paraneoplastic antibodies, CSF alone yielded a positive result in 20 patients, twice as commonly as serum alone (10 patients). For NMO-IgG, serum alone yielded a positive result in 21 patients 3.5 times more commonly than CSF alone (6 patients).

Discussion. From our review of a 25-year experience with immunofluorescence testing on a service basis in the Mayo Clinic Neuroimmunology Laboratory, we found that the rate of clinically pertinent autoantibody detection was highest when both serum and CSF were tested. It is plausible that this finding may reflect a greater likelihood of physicians deciding to test both serum and CSF in patients with the highest index of clinical suspicion.

When both serum and CSF were tested, CSF was more commonly informative than serum for paraneoplastic antibody detection. This raises concern that clinically important neural antibodies may be missed when only serum is tested. This finding was most prominent for NMDA receptor-specific IgG. Consistent with this finding, Kumar et al.[3] recently reported 3 patients, each of whom had NMDA receptor IgG detected in CSF but not in serum. Where there is a high suspicion for

2.4 Schulte *et al.*, An excess of rare loss-of-function alleles substantiates MEIS1 as a genetic factor in restless legs syndrome, *in preparation*[IV]

Personal contributions: I designed the study and performed LightScanner® high-resolution melting curve analysis of the coding regions of all seven genes as well as of *MEIS1* in the large case/control sample, follow-up Sanger sequencing and all data analyses. I designed the multiplex PCRs for Sequenom®-based genotyping and analyzed all genotyping data. I carried out part of the statistical analyses used in burden testing. Moreover, at the laboratory of Prof. Nicholas Katsanis at the Center for Human Disease Modeling, Department of Cell Biology, Duke University, Durham, NC, USA, I cloned all mutagenized constructs used in the *in vivo* complementation assay in zebrafish, made all mRNA for injections, participated in the injections of morpholino and mRNA into zebrafish embryos and performed most of the staining, imaging and image analysis as well as the statistics. I wrote the manuscript and designed all figues and tables except for Supplementary Figure 2 and Figure 3 B.

An excess of rare loss-of-function alleles substantiates *MEIS1* as a genetic factor in restless legs syndrome

Eva C. Schulte[1,2], Maria Kousi[3], Percillaz Tan[3], Barbara Schormair[2,4], Franziska Knauf[2], Peter Lichtner[2], Claudia Trenkwalder[5,6], Birgit Högl[7], Birgit Frauscher[7], Klaus Berger[8], Ingo Fietze[9], Nadine Gross[1], Magdolna Hornyak[1,10,11], Karin Stiasny-Kolster[12,13], Wolfgang Oertel[13], Cornelius G. Bachmann[14], Walter Paulus[15], Alexander Zimprich[16], Annette Peters[17], Christian Gieger[18], Thomas Meitinger[2,4], Bertram Müller-Myshok[19], Nicholas Katsanis[3], Juliane Winkelmann[1,2,4*]

[1]Neurologische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität München, Munich, Germany
[2]Institut für Humangenetik, Helmholtz Zentrum München, Munich, Germany
[3]Center for Human Disease Modelling, Department of Cell Biology, Duke University, Durham, NC, USA
[4]Institut für Humangenetik, Technische Universität München, Munich, Germany
[5]Paracelsus Elena Klinik, Kassel, Germany
[6]Klinik für Neurochirurgie, Georg August Universität, Göttingen, Germany
[7]Department of Neurology, Medizinische Universität Innsbruck, Innsbruck, Austria
[8]Institut für Epidemiologie und Schlafmedizin, Westfälische Wilhelms Universität Münster, Münster, Germany
[9]Zentrum für Schlafmedizin, Charité Universitätsmedizin, Berlin, Germany
[10]Interdisziplinäres Schmerzzentrum, Albert-Ludwigs Universität Freiburg, Freiburg, Germany
[11]Algesiologikum München, Munich, Germany
[12]Schlaflabor Marburg, Marburg, Germany
[13]Klinik für Neurologie, Philipps Universität Marburg, Marburg, Germany
[14]Abteilung für Neurologie, Paracelsus Klinikum Osnabrück, Osnabrück, Germany
[15]Abteilung für Neurophysiologie, Georg August Universität, Göttingen, Germany
[16]Department of Neurology, Medizinische Universität Wien, Vienna, Austria
[17]Institute of Epidemiology II, Helmholtz Zentrum München, Munich, Germany
[18]Institute of Genetic Epidemiology, Helmholtz Zentrum München, Munich Germany
[19]Max-Planck Institut für Psychiatrie München, Munich, Germany

[*]address correspondence to:
Juliane Winkelmann
Neurologische Klinik und Poliklinik
Klinikum rechts der Isar
Technische Universität München
Ismaningerstr. 22
81675 Munich
Germany
Phone: +49 89 4140 4688
Fax: +49 89 4140 4867
Email: winkelmann@lrz.tum.de

INTRODUCTORY PARAGRAPH

Restless legs syndrome (RLS) is a common neurologic disorder characterized by nocturnal dysesthesias and an urge to move affecting the legs. Genetically, it is a complex disease and genome-wide association studies (GWAS) have identified common susceptibility alleles at six genomic loci. Performing a fine-mapping analysis, we identified low-frequency and rare coding variants at these six risk loci and observed an excess of rare variants both across all loci but also for *MEIS1* alone. The rarer the variants, the more likely they were found in individuals with RLS. In *MEIS1*, functional assessment in zebrafish embryos revealed an excess of loss-of-function alleles among the rare non-synonymous variants. At the *MEIS1* locus, the whole spectrum of genetic variation with regard to frequency and effect sizes contributes to the genetic framework of a complex genetic phenotype such as RLS.

MANUSCRIPT BODY

Restless legs syndrome (RLS, OMIM*102300) is one of the most common neurologic disorders with an age-dependent prevalence of up to 5-10% in Europe and North America[1]. RLS is characterized by an irresistible urge to move the legs accompanied by disagreeable, often painful sensations in the lower limbs at night. Moving the affected legs or walking leads to a prompt but only temporary relief[1]. As a consequence, patients suffer from insomnia leading to an impairment of quality of life and mental health. RLS is genetically complex and a highly familial trait with estimates of heredity between 54% and 69% as derived from twin studies[2,3].

For RLS, susceptibility alleles at loci identified by current genome-wide association studies (GWAS) explain only about 6.8% of the estimated heritability[4]. Some of the „missing heritability“[5] in RLS and other complex genetic diseases may lie hidden in a collection of rare and very rare variants of strong effect, which cannot be identified by means of GWAS[6,7]. For a few other complex genetic diseases such as diabetes or chronic inflammatory bowel disease, rare variants contributing to disease development have recently been identified within known GWAS loci[8,9], thus supporting the concept of allelic series in complex genetic disorders. Although one potentially causal rare variant has been described in *MEIS1*[10,11], to date, no variants of strong effect are established for RLS.

In order to assess low-frequency (1% < minor allele frequency (MAF) < 5%) coding variation at the known RLS susceptibility loci entailing candidate genes *MEIS1*, *BTBD9*, *PTPRD*, *MAP2K5*, *SKOR1*, *TOX3* and the non-coding RNA *BC034767*[4,12-14], we screened the coding regions and exon-intron boundaries (± 10 bp) using high-resolution melting curve analysis (LightScanner, Idaho Technology, Inc.) and Sanger sequencing in 188 German individuals

with RLS (demographics to be added, also see Online Methods) and 188 controls belonging to the KORA general population cohort[15] based in Southern Germany (demographics to be added). A total of 50 variants with MAF < 5% were identified in both cases and controls (Supplementary Table 1). When collapsed across all seven loci, low-frequency synonymous and non-synonymous variants combined occurred more frequently in cases than in controls (78 cases with a variant vs. 46 controls; p=0.004, McNemar test; odds ratio (OR)=2.19 (95% confidence interval (CI): 1.41-3.40)). Non-synonymous variants alone also showed a trend towards being more frequent in cases than in controls (40 in cases vs. 24 in controls; p=0.045, McNemar; OR=1.85 (95% CI: 1.06-3.21)). Within *MEIS1*, synonymous or non-synonymous low-frequency variants were present in 9 cases but only 1 control (p=0.011, McNemar; OR=9.4 (95% CI: 1.18-74.97)) while no difference in the amount of low-frequency variation was obvious in any of the other genes (Supplementary Table 1). Variants did not seem to cluster within specific regions of the examined genes (Figure 1). To assess a possible association with the RLS phenotype, we next genotyped 40 of the 50 identified variants in 3265 German and Austrian RLS cases (65.2±11.3yrs; 29.3% male) and 2944 KORA controls (56.1±13.3yrs; 48.7% male). 10 of the 50 variants could not be included for technical reasons. While low-frequency variants—either all together or only non-synonymous ones—with MAF < 5% where not significantly more frequent in RLS cases than controls, there was a distinct excess of rare variants with a KORA-derived MAF < 1% (total: 200 vs. 136 individuals with variants, p=0.008, χ^2 test; OR=1.35 (95% CI: 1.08-1.65); non-synonymous only: 156 vs. 48 individuals with variants, $p=3.71\times10^{-12}$, χ^2 test; OR=3.02 (95% CI: 2.18-4.20)) and even more so with MAF < 0.1% (total: 51 vs. 12 individuals with variants, $p=5.84\times10^{-6}$, χ^2 test; OR=3.88 (95% CI: 2.06-7.28); non-synonymous only: 35 vs. 5 individuals with variants, $p=9.16\times10^{-6}$, χ^2 test; OR=6.36 (95% CI: 2.49-16.27)) in individuals with RLS compared to controls (Figure 1). At the *MEIS1* locus alone but not at any of the other six loci, the same held true. Here, too, rare variants with MAF<1% (total: 116 vs. 67 individuals with variants, p=0.002, χ^2 test; OR=1.58 (95% CI: 1.17-2.15); non-synonymous only: 39 vs. 14 individuals with variants, p=0.002, χ^2 test; OR=2.53 (95% CI: 1.37-4.67)) and MAF<0.1% (both total and non-synonymous only: 9 vs. 1 individuals with variants, p=0.01, χ^2 test; OR=8.15 (95% CI: 1.03-64.25)) were seen more frequently in cases than in controls. Permutation analysis (n=5000) substantiated this finding (total: p=0.0037, OR=18.62 (95% CI: 2.48-139.44); non-synonymous only: p=0.0072, OR=30.48 (95% CI: 2.34-395.87)) (Figure 1). No low-frequency coding variants with 1%<MAF<5% were found in *MEIS1*. Several individual rare variants in *MEIS1* (p.R272H and p.M453T), *TOX3* (p.A233A), and *PTPRD* (c.551-4 C>G,

p.P278P and R1898C) were nominally associated with RLS in the large case/control sample but associations did not withstand correction for multiple testing (Supplementary Table 1). The excess of low-frequency and rare variants at RLS-associated GWAS loci was most pronounced for *MEIS1*. Accordingly, we sought to expand our analysis to a more comprehensive investigation of genetic variation both with regard to frequency and location within *MEIS1* by screening the coding regions ± 10 bp as well as the 5' and 3' untranslated regions (UTRs) ± 100 bp for variants with MAF<5% in 3760 German and Austrian RLS cases of German and Austrian origin (demographics to be added) and 3542 KORA controls (demographics to be added) by means of LightScanner® high-resolution melting curve analysis (Online Methods). We identified a total of 89 such variants (Supplementary Tables 2 and 3), 22 of these coding in either isoform 1 (ENSP00000272369) or isoform 2 (ENSP00000381518) of *MEIS1*. Overall, there was an excess of variants with MAF<5% across all examined regions of *MEIS1* (1451 occurrences of a variant in cases vs. 649 in controls, $p=3.22x10^{-67}$, χ^2 test; OR=2.35 (95% CI: 2.13-2.60)) although this was largely dependent on rs11693221 in the post-3' UTR region ($MAF_{cases}=13.55\%$, $MAF_{controls}=3.58\%$; $p=8.79x10^{-99}$, χ^2 test; OR=4.16 (95% CI: 3.61-4.80)) (Supplementary Table 3). Without this variant, the overall numbers of individuals with low-frequency and rare variants across all regions of *MEIS1* were similar in cases and controls (432 vs. 396; p=0.68, χ^2 test). However, stratification of variants according to their localization showed an excess of rare variants with MAF<1% in the 5' UTR (16 cases vs. 2 controls; $p=6.2x10^{-4}$, logistic regression of the phenotype on number of variants of each type; OR=7.62 (95% CI: 1.75-33.19)) and among non-synonymous coding variants in isoform 2 (34 vs. 15; p=0.005, logistic regression; OR=2.31 (95% CI: 1.24-4.32)) in individuals with RLS compared to controls (Figure 2). Non-synonymous coding variants in isoform 1 also showed a trend towards being more frequent in the cases than in the controls (18 vs. 9; p=0.11, logistic regression; OR=1.88 (95% CI: 0.85-4.21)). Low-frequency and rare variants located in the 3' UTR collectively showed a trend toward being protective (240 vs. 275; p=0.03, logistic regression; OR=0.84 (95% CI:0.71-0.99)) (Figure 2).

Because it is likely that a locus associated with a complex genetic phenotype will harbor not only disease-linked variants of different frequency and statistical effect sizes, we next investigated the functional effects of the identified non-synonymous coding variants of *MEIS1*. The morpholino-based knock-down of *meis1* is known to result in a reduction of optic tectum size in zebrafish embryos 72 hours post fertilization (hpf)[17]. Co-injection with human *MEIS1* wildtype mRNA can rescue this phenotype (Figure 3) while co-injection with *MEIS1* mRNA harboring a variant predicted to ablate DNA binding ability

(p.R276A+N325T) serving as a positive control cannot (Supplementary Figure 2). Of note, the morpholino-dependent knock-down of *map2k5,* another putative RLS gene identified in the GWAS (also see above), yields a very similar phenotype with regard to the size of the optic tectum, which is likely to be specific because it can be rescued by co-injection with human *MAP2K5* mRNA (Figure 3). The fact that the knock-down of other genes of known involvement in neurodevelopment, such as, for example, *myst3/MYST3*, which showed no association with the RLS phenotype in previous GWAS (data not shown), supports the relevance of this phenotype to genes believed to be involved in RLS and not merely neurodevelopment as the whole.

All 13 non-synonymous coding variants identified in isoform 1 as well as the four non-synonymous variants, which are coding solely in isoform 2, observed in the 3760 RLS cases and 3542 controls (Figure 3 and Supplementary Table 3) were analyzed using this *in vivo* complementation assay in zebrafish embryos (n = at least 150 per variant). Incomplete rescue by human mRNA harboring non-synonymous variants is indicative of dysfunctional hypomorphic or null alleles (Online Methods). Among the 13 alleles of isoform 1, we identified two benign, six hypomorphic and five null alleles (Figure 3). There was an excess of complete loss-of-function alleles (i.e. null alleles) in individuals with RLS compared to those without (13 in cases vs. 2 in controls; p=0.006, χ^2 test; OR=6.14 (95% CI: 1.38-27.23)) (Figure 3). When assayed in isoform 2, however, none of the four non-synonymous variants coding only in isoform 2 nor two variants shown to be functional nulls in isoform 1 (p.S204T and p.R272H) showed any effect on tectum size in isoform 2 (Supplementary Figure 2). As overexpression of *MEIS1* mRNA harboring the 13 variants identified in isoform 1 also had no effect on optic tectum size (Supplementary Figure 2), loss-of-function represents the likely mechanism underlying the involvement of rare coding variants in *MEIS1* in RLS.

Previous GWAS have established the genomic locus encompassing *MEIS1* as the most significant susceptibility region for RLS[4,14]. The most likely candidate gene in this region is *MEIS1*, a TALE homeobox transcription factor known to be involved in specifying spinal motor neuron pool identity and connectivity[18] as well as proximo-distal limb patterning[19] during embryonic development. Recent studies in the context of RLS have also suggested a link to the iron metabolism in the central nervous system[20,21].

The excess of rare alleles of functional effect in RLS patients compared to controls shown here substantiates *MEIS1* as the causal genetic factor underlying the observed associations. Moreover, it implicates loss-of-function as the mechanism underlying this association, at least with regard to rare variants. Interestingly, we also observed a new association with a low-frequency variant (rs1169322) located approximately 70 bp downstream of the 3' UTR of *MEIS1* in its current annotation, that represents the largest genetic risk factor for RLS

identified to date. Its haplotypic relation to known common variants associated with RLS is currently unknown (linkage disequilibrum (r^2<0.6)).

Nonetheless, this finding in conjunction with the observed excess of rare in the 5' UTR (and to a much lesser extend also the decreased amount of rare variation in the 3' UTR) of *MEIS1* in individuals with RLS implicates the UTRs in disease pathogenesis. The excess of rare non-coding variants in the 100 bp surrounding the exons of nine genes associated with asthma[23] and the fact that fine-mapping studies located about 22 % of 36 GWAS association signals for celiac disease to either the 5` or the 3` non-coding regions (UTRs and several kb up- or downstream)[24] could indicate that these regions are indeed important in the context of complex genetic diseases and might be overlooked by the current surge of whole-exome sequencing studies.

Recent studies have implicated allelic series of variants of different frequency and effect sizes at loci identified in the context of GWAS in complex genetic diseases. In single cases, individual rare variants were shown to be associated with the phenotype[9,25,26] while in other cases it was either the collective of rare variation within a single gene[8] or across a number of GWAS-identified loci[27,28]. Our data substantiate this role of the whole of genetic variation from common to low-frequency to rare and very rare variation at a GWAS-identified locus in the genetic architecture of a complex genetic disease, such as RLS.

We were unable to establish a single rare variant of large effect involved in RLS at the examined loci possibly due to a lack of power in sight of the large amount of rare and very rare genetic variation and the excess of singletons known to exist in the human genome[29-31]. Alternatively, it is, of course, also possible that no such variant exists within the coding regions of *MEIS1* or any of the other RLS-associated genes that were investigated. Still, we observed a clear tendency across all investigated loci that the rarer the genetic variant, the more likely it was found in an individual with RLS. This extends the previous general observation that the rarer a genetic variant, the more likely it is functionally relevant[32] to a disease context. Moreover, the follow-up of two null alleles which primarily or only in cases (*MEIS1* p.R272H and p.Q353H) and are located in the homeobox and transcription activation domains, respectively, in murine models could prove worthwhile to establish their relevance to the RLS phenotype and to further explore the pathophysiology underlying the disease.

REFERENCES

1 Allen, R. P. *et al.* Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology. A report from the restless legs syndrome diagnosis and epidemiology workshop at the National Institutes of Health. *Sleep medicine* **4**, 101-119 (2003).

2 Desai, A. V., Cherkas, L. F., Spector, T. D. & Williams, A. J. Genetic influences in self-reported symptoms of obstructive sleep apnoea and restless legs: a twin study. *Twin research : the official journal of the International Society for Twin Studies* **7**, 589-595, doi:10.1375/1369052042663841 (2004).

3 Xiong, L. *et al.* Canadian restless legs syndrome twin study. *Neurology* **68**, 1631-1633, doi:10.1212/01.wnl.0000261016.90374.fd (2007).

4 Winkelmann, J. *et al.* Genome-wide association study identifies novel restless legs syndrome susceptibility loci on 2p14 and 16q12.1. *PLoS genetics* **7**, e1002171, doi:10.1371/journal.pgen.1002171 (2011).

5 Maher, B. Personal genomes: The case of the missing heritability. *Nature* **456**, 18-21, doi:10.1038/456018a (2008).

6 Gibson, G. Rare and common variants: twenty arguments. *Nature reviews. Genetics* **13**, 135-145, doi:10.1038/nrg3118 (2011).

7 Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753, doi:10.1038/nature08494 (2009).

8 Bonnefond, A. *et al.* Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. *Nature genetics* **44**, 297-301, doi:10.1038/ng.1053 (2012).

9 Rivas, M. A. *et al.* Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nature genetics* **43**, 1066-1073, doi:10.1038/ng.952 (2011).

10 Schulte, E. C. *et al.* Variant screening of the coding regions of MEIS1 in patients with restless legs syndrome. *Neurology* **76**, 1106-1108, doi:10.1212/WNL.0b013e318211c366 (2011).

11 Vilarino-Guell, C. *et al.* MEIS1 p.R272H in familial restless legs syndrome. *Neurology* **73**, 243-245, doi:10.1212/WNL.0b013e3181ae7c79 (2009).

12 Schormair, B. *et al.* PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome. *Nature genetics* **40**, 946-948, doi:10.1038/ng.190 (2008).

13 Stefansson, H. *et al.* A genetic risk factor for periodic limb movements in sleep. *The New England journal of medicine* **357**, 639-647, doi:10.1056/NEJMoa072743 (2007).

14 Winkelmann, J. *et al.* Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. *Nature genetics* **39**, 1000-1006, doi:10.1038/ng2099 (2007).

15 Wichmann, H. E., Gieger, C., Illig, T. & Group, M. K. S. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67 Suppl 1**, S26-30, doi:10.1055/s-2005-858226 (2005).

16 Jonsson, T. *et al.* A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* **488**, 96-99, doi:10.1038/nature11283 (2012).

17 Erickson, T., French, C. R. & Waskiewicz, A. J. Meis1 specifies positional information in the retina and tectum to organize the zebrafish visual system. *Neural development* **5**, 22, doi:10.1186/1749-8104-5-22 (2010).

18 Dasen, J. S., Tice, B. C., Brenner-Morton, S. & Jessell, T. M. A Hox regulatory network establishes motor neuron pool identity and target-muscle connectivity. *Cell* **123**, 477-491, doi:10.1016/j.cell.2005.09.009 (2005).

19 Mercader, N. *et al.* Conserved regulation of proximodistal limb axis development by Meis1/Hth. *Nature* **402**, 425-429, doi:10.1038/46580 (1999).

20 Catoire, H. *et al.* Restless legs syndrome-associated MEIS1 risk variant influences iron homeostasis. *Annals of neurology* **70**, 170-175, doi:10.1002/ana.22435 (2011).

21 Xiong, L. *et al.* MEIS1 intronic risk haplotype associated with restless legs syndrome affects its mRNA and protein expression levels. *Human molecular genetics* **18**, 1065-1074, doi:10.1093/hmg/ddn443 (2009).

22 Schulte, E. C. K., M.; Schormair, B.; Winkelmann, J. Iron in Restless Legs Syndrome. *International Review of Neurobiology* **in press** (2013).

23 Torgerson, D. G. *et al.* Resequencing candidate genes implicates rare variants in asthma susceptibility. *American journal of human genetics* **90**, 273-281, doi:10.1016/j.ajhg.2012.01.008 (2012).

24 Trynka, G. *et al.* Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nature genetics* **43**, 1193-1201, doi:10.1038/ng.998 (2011).

25 Holm, H. *et al.* A rare variant in MYH6 is associated with high risk of sick sinus syndrome. *Nature genetics* **43**, 316-320, doi:10.1038/ng.781 (2011).

26 Raychaudhuri, S. *et al.* A rare penetrant mutation in CFH confers high risk of age-related macular degeneration. *Nature genetics* **43**, 1232-1236, doi:10.1038/ng.976 (2011).

27 Diogo, D. *et al.* Rare, low-frequency, and common variants in the protein-coding sequence of biological candidate genes from GWASs contribute to risk of rheumatoid arthritis. *American journal of human genetics* **92**, 15-27, doi:10.1016/j.ajhg.2012.11.012 (2013).

28 Johansen, C. T. *et al.* Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nature genetics* **42**, 684-687, doi:10.1038/ng.628 (2010).

29 Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216-220, doi:10.1038/nature11690 (2013).

30 Kiezun, A. *et al.* Exome sequencing and the genetic basis of complex traits. *Nature genetics* **44**, 623-630, doi:10.1038/ng.2303 (2012).

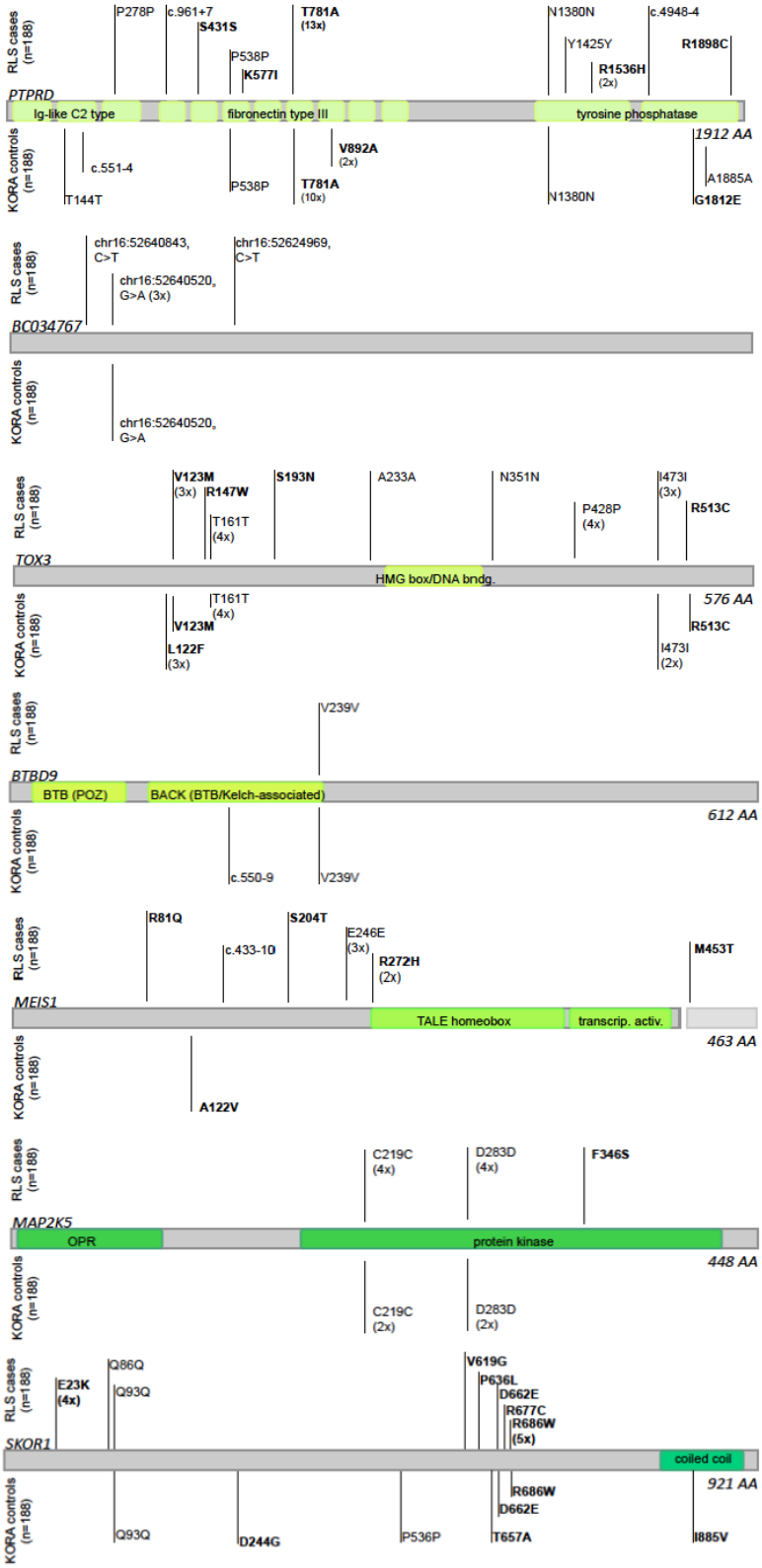
31 Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64-69, doi:10.1126/science.1219240 (2012).

32 Nelson, M. R. *et al.* An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* **337**, 100-104, doi:10.1126/science.1217876 (2012).

33 Project, N. G. O. E. S. *The NHLBI GO ESP Exome Variant Server*, <http://evs.gs.washington.edu/EVS/> (2013).

FIGURES

A


RLS cases (n=188)
P278P
c.961+7
S431S
T781A (13x)
P538P
K577I
N1380N
c.4948-4
Y1425Y
R1898C
R1536H (2x)
PTPRD
Ig-like C2 type
fibronectin type III
tyrosine phosphatase
KORA controls (n=188)
c.551-4
T144T
P538P
V892A (2x)
T781A (10x)
N1380N
1912 AA
A1885A
G1812E
RLS cases (n=188)
chr16:52640843, C>T
chr16:52640520, G>A (3x)
chr16:52624969, C>T
BC034767
KORA controls (n=188)
chr16:52640520, G>A
RLS cases (n=188)
V123M (3x)
R147W
T161T (4x)
S193N
A233A
N351N
P428P (4x)
I473I (3x)
R513C
TOX3
HMG box/DNA bndg.
KORA controls (n=188)
T161T (4x)
V123M
L122F (3x)
576 AA
R513C
I473I (2x)
RLS cases (n=188)
V239V
BTBD9
BTB (POZ)
BACK (BTB/Kelch-associated)
612 AA
KORA controls (n=188)
c.550-9
V239V
RLS cases (n=188)
R81Q
c.433-10
S204T
E246E (3x)
R272H (2x)
M453T
MEIS1
TALE homeobox
transcrip. activ.
463 AA
KORA controls (n=188)
A122V
RLS cases (n=188)
C219C (4x)
D283D (4x)
F346S
MAP2K5
OPR
protein kinase
448 AA
KORA controls (n=188)
C219C (2x)
D283D (2x)
RLS cases (n=188)
Q86Q
E23K (4x)
Q93Q
V619G
P636L
D662E
R677C
R686W (5x)
SKOR1
coiled coil
921 AA
KORA controls (n=188)
Q93Q
D244G
P536P
R686W
D662E
T657A
I885V

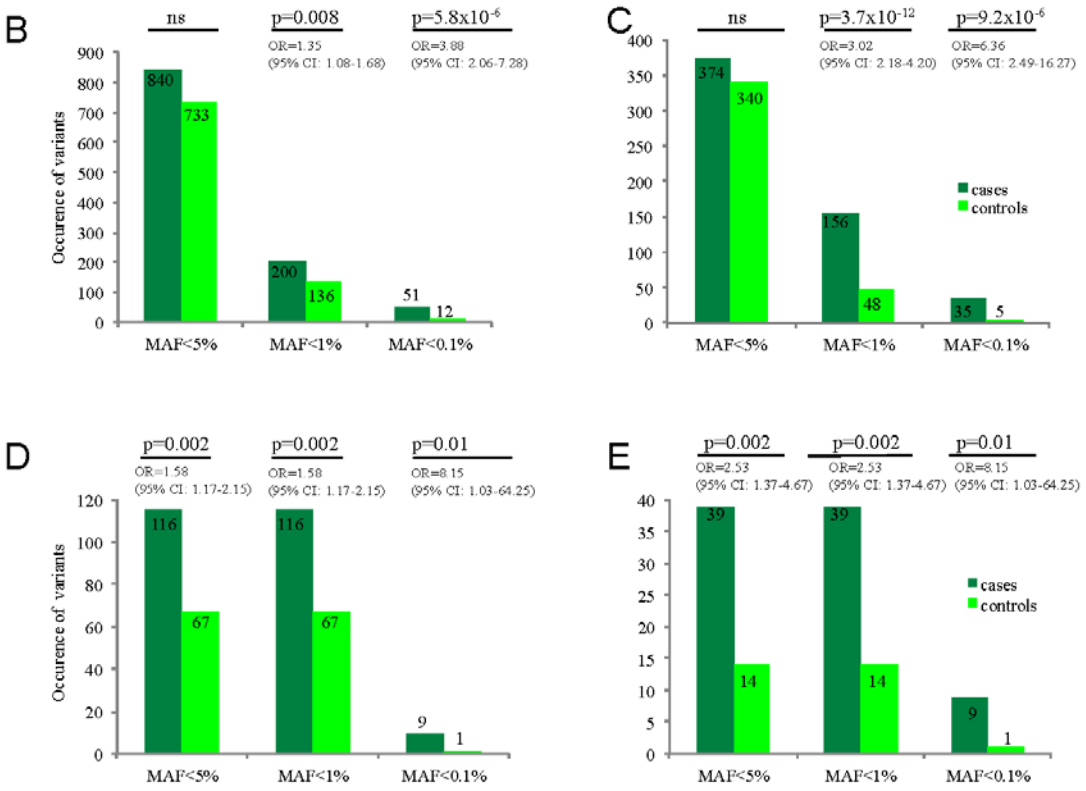


Figure 1: Low-frequency and rare variants identified in the coding sequences of seven genes associated with RLS in GWAS[4,12-14] in 188 cases and 188 controls. (A) Schematic representation of identified low-frequency variants in *PTPRD*, *BC034767*, *TOX3*, *MEIS1*, *BTBD9*, *MAP2K5* and *SKOR1*. When assessed in 3265 cases and 2944 controls, there was an excess of both overall (B) and non-synonymous (C) variants with MAF<1% and MAF<0.1% across all examined loci. The same held true for the combined (D) and solely the non-synonymous (E) variants at the *MEIS1* locus individually.

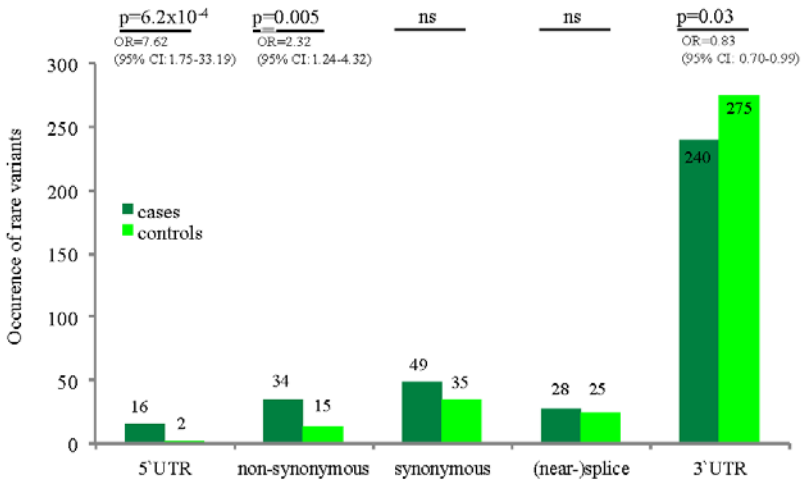
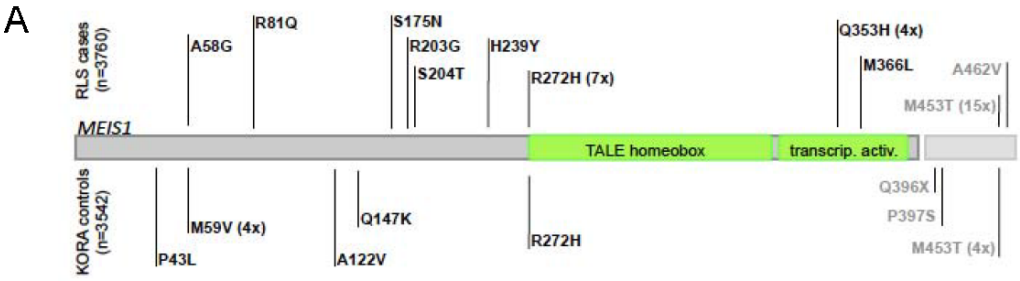
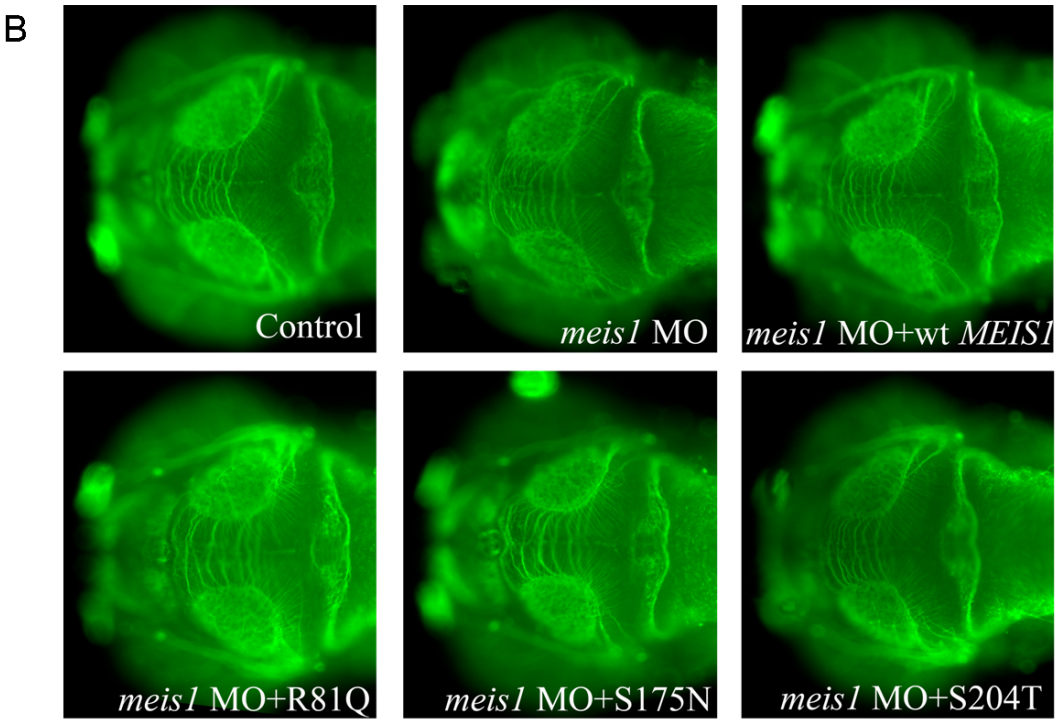
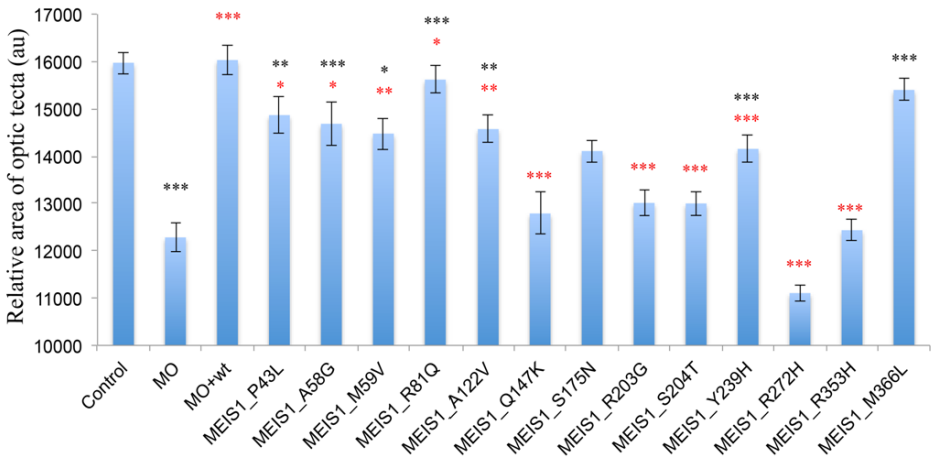


Figure 2: Variant screening of the coding regions and UTRs of *MEIS1* in 3760 individuals with RLS and 3542 KORA controls. Stratification according to variant localisation shows an excess of rare variants in both the 5' UTR and among non-synonymous coding variants. Low-frequency and rare variants in the 3' UTR showed a trend towards being protective. No difference was observed in the number of individuals carrying synonymous coding or (near-) splice variants between cases and controls.


A
RLS cases (n=3760)
A58G
R81Q
S175N
R203G
S204T
H239Y
R272H (7x)
Q353H (4x)
M366L
A462V
M453T (15x)
MEIS1
TALE homeobox
transcrip. activ.
KORA controls (n=3542)
P43L
M59V (4x)
A122V
Q147K
R272H
Q396X
P397S
M453T (4x)


B
Control
meis1 MO
meis1 MO+wt MEIS1
meis1 MO+R81Q
meis1 MO+S175N
meis1 MO+S204T


Relative area of optic tecta (au)
17000
16000
15000
14000
13000
12000
11000
10000
Control
MO
MO+wt
MEIS1_P43L
MEIS1_A58G
MEIS1_M59V
MEIS1_R81Q
MEIS1_A122V
MEIS1_Q147K
MEIS1_S175N
MEIS1_R203G
MEIS1_S204T
MEIS1_Y239H
MEIS1_R272H
MEIS1_R353H
MEIS1_M366L

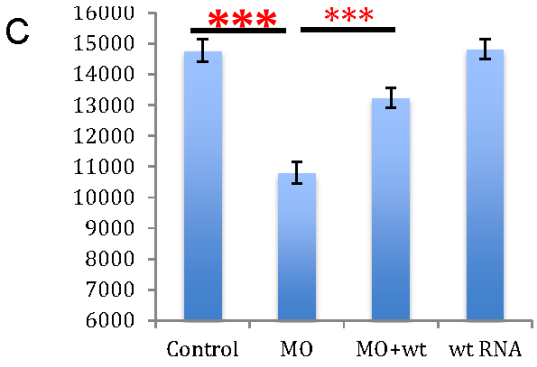


Figure 3: Functional assessment of rare non-synonymous variants in *MEIS1* by in vivo complementation in zebrafish embryos. (A) Location and frequency of non-synonymous *MEIS1* variants examined by in zebrafish. Variants found in cases are given above the gene, those found in controls below. (B) At 72 hpf, zebrafish larvae were stained as whole mounts using an antibody against acetylated tubulin and the size of the optic tecta was measures for phenotypic read out. Control, morpholino injection and rescue by human wt mRNA are shown in the upper panels. The lower panels illustrate the effects of different alleles tested. (C) Quantification of optic tectum area in zebrafish larvae at 72 hpf (n=at least 150 per genotype). Benign alleles show a significant difference with regard to the MO injection, hypomorphic alleles a significant difference with regard to both the MO injection and the rescue (MO plus wt) injection and null alleles are significantly different from the rescue only. Asterisks denote significance sevels as determined by student`s t-test: black=allele vs. MO, red=allele vs. rescue.

SUPPLEMENTARY MATERIALS

An excess of rare loss-of-function alleles substantiates *MEIS1* as a genetic factor in restless legs syndrome

Eva C. Schulte[1,2], Maria Kousi[3], Barbara Schormair[2,4], Franziska Knauf[2], Peter Lichtner[2], Claudia Trenkwalder[5,6], Birgit Högl[7], Birgit Frauscher[7], Klaus Berger[8], Ingo Fietze[9], Nadine Gross[1], Magdolna Hornyak[1,10,11], Karin Stiasny-Kolster[12,13], Wolfgang Oertel[13], Cornelius G. Bachmann[14], Walter Paulus[15], Alexander Zimprich[16], Annette Peters[17], Christian Gieger[18], Thomas Meitinger[2,4], Bertram Müller-Myshok[19], Nicholas Katsanis[3], Juliane Winkelmann[1,2,4*]

[1]Neurologische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität München, Munich, Germany
[2]Institut für Humangenetik, Helmholtz Zentrum München, Munich, Germany
[3]Center for Human Disease Modelling, Department of Cell Biology, Duke University, Durham, NC, USA
[4]Institut für Humangenetik, Technische Universität München, Munich, Germany
[5]Paracelsus Elena Klinik, Kassel, Germany
[6]Klinik für Neurochirurgie, Georg August Universität, Göttingen, Germany
[7]Department of Neurology, Medizinische Universität Innsbruck, Innsbruck, Austria
[8]Institut für Epidemiologie und Schlafmedizin, Westfälische Wilhelms Universität Münster, Münster, Germany
[9]Zentrum für Schlafmedizin, Charité Universitätsmedizin, Berlin, Germany
[10]Interdisziplinäres Schmerzzentrum, Albert-Ludwigs Universität Freiburg, Freiburg, Germany
[11]Algesiologikum München, Munich, Germany
[12]Schlaflabor Marburg, Marburg, Germany
[13]Klinik für Neurologie, Philipps Universität Marburg, Marburg, Germany
[14]Abteilung für Neurologie, Paracelsus Klinikum Osnabrück, Osnabrück, Germany
[15]Abteilung für Neurophysiologie, Georg August Universität, Göttingen, Germany
[16]Department of Neurology, Medizinische Universität Wien, Vienna, Austria
[17]Institute of Epidemiology II, Helmholtz Zentrum München, Munich, Germany
[18]Institute of Genetic Epidemiology, Helmholtz Zentrum München, Munich Germany
[19]Max-Planck Institut für Psychiatrie München, Munich, Germany

[*]address correspondence to:

Juliane Winkelmann
Neurologische Klinik und Poliklinik
Klinikum rechts der Isar
Technische Universität München
Ismaningerstr. 22
81675 Munich
Germany
Phone: +49 89 4140 4688
Fax: +49 89 4140 4867
Email: winkelmann@lrz.tum.de

ONLINE METHODS

Participants

Both case and control populations were entirely of German and Austrian descent. In all patients, diagnosis was based on the diagnostic criteria of the International RLS Study Group[1] as assessed in a personal interview conducted by an RLS expert. We excluded patients with secondary RLS due to uremia, dialysis, or anemia due to iron deficiency. The presence of secondary RLS was determined by clinical interview, physical and neurological examination, blood chemistry, and nerve conduction studies whenever deemed clinically necessary. Participants' written informed consent were obtained prior to the initiation of the study. Written informed consent was obtained from each participant in the respective language. The study was approved by the institutional review boards of the contributing authors. The primary review board was located in Munich, Bayerische Ärztekammer and Technische Universität München.

Variant screening

In a first step, we used Idaho LightScanner® high-resolution melting curve analysis (Biofire, Inc.) to screen the coding regions and exon/intron boundaries of *PTPRD*, *BTBD9*, *TOX3*, *BC034767*, *MAP2K5* and *MEIS1* for variants. Due to the very high GC-content, the coding regions ± 10bp of *SKOR1* could not be subjected to LightScanner® analysis and were Sanger sequenced instead.188 German RLS cases and 188 general population controls belonging to the KORA cohort[2] based in Southern Germany were included in the screening. Of the188 cases used to screen *MEIS1* (60.0±11.2 years, 72.8% female) and *BTBD9* (demographics to be added) half were homozygous and half were heterozygous for the published risk alleles (rs2300478 and rs3923809)[3,4]. The same set of case (demographics to be added) and controls (demographics to be added) was used for all other screening experiments. In the case of an altered melting pattern suggestive of variants, Sanger sequencing ensued to identify the underlying variant. The same method was used to screen the coding regions of *MEIS1* isoform 1 and 2 ± 10 bp as well as the 5' and 3' UTRs ± 100 bp in 3760 RLS cases of German and Austrian descent (demographics to be added) and 3542 general-population controls (demographics to be added) belonging to the S4 and F4 surveys of KORA[2]. Group comparisons between cases and controls were performed for each gene and each variant separately using McNemar, χ^2 tests and logistic regressions as appropriate.

Genotyping

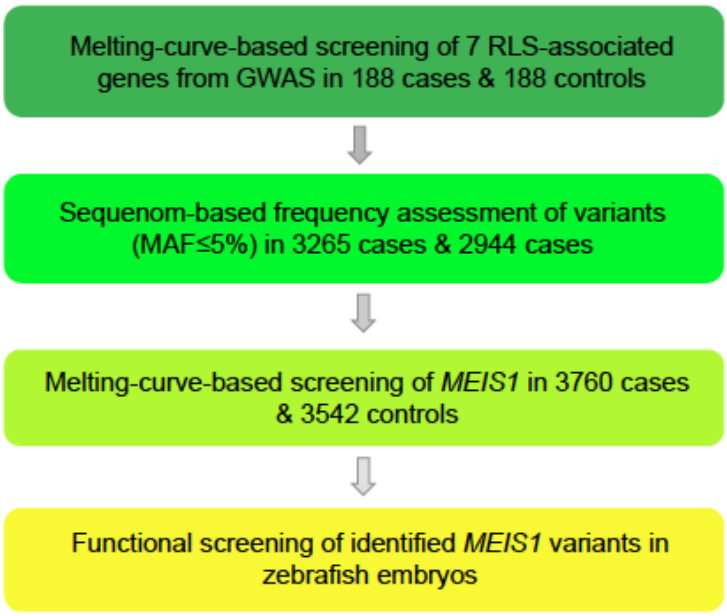
Genotyping was performed on the MassARRAY system using MALDI-TOF mass spectrometry with the iPLEX Gold chemistry (Sequenom Inc, San Diego, CA, USA). Primers were designed using AssayDesign 3.1.2.2 with iPLEX Gold default parameters. No assay could be designed for seven variants, largely those located in the extremely GC-rich gene *SKOR1*. Three assays failed two or more times and were, therefore, not pursued further. Automated genotype calling was done with SpectroTYPER 3.4. Genotype clustering was visually checked by an experienced evaluator. SNPs and DNA samples with a call rate<90% were excluded. The genotyping sample consisted of 3265 cases auf German and Austrian descent (65.2±11.3yrs; 29.3% male) and 2944 general population controls (56.1±13.3yrs; 48.7% male) from the KORA F4 survey[2].

***In vivo* complementation in zebrafish embryos and whole-mount immunostaining**

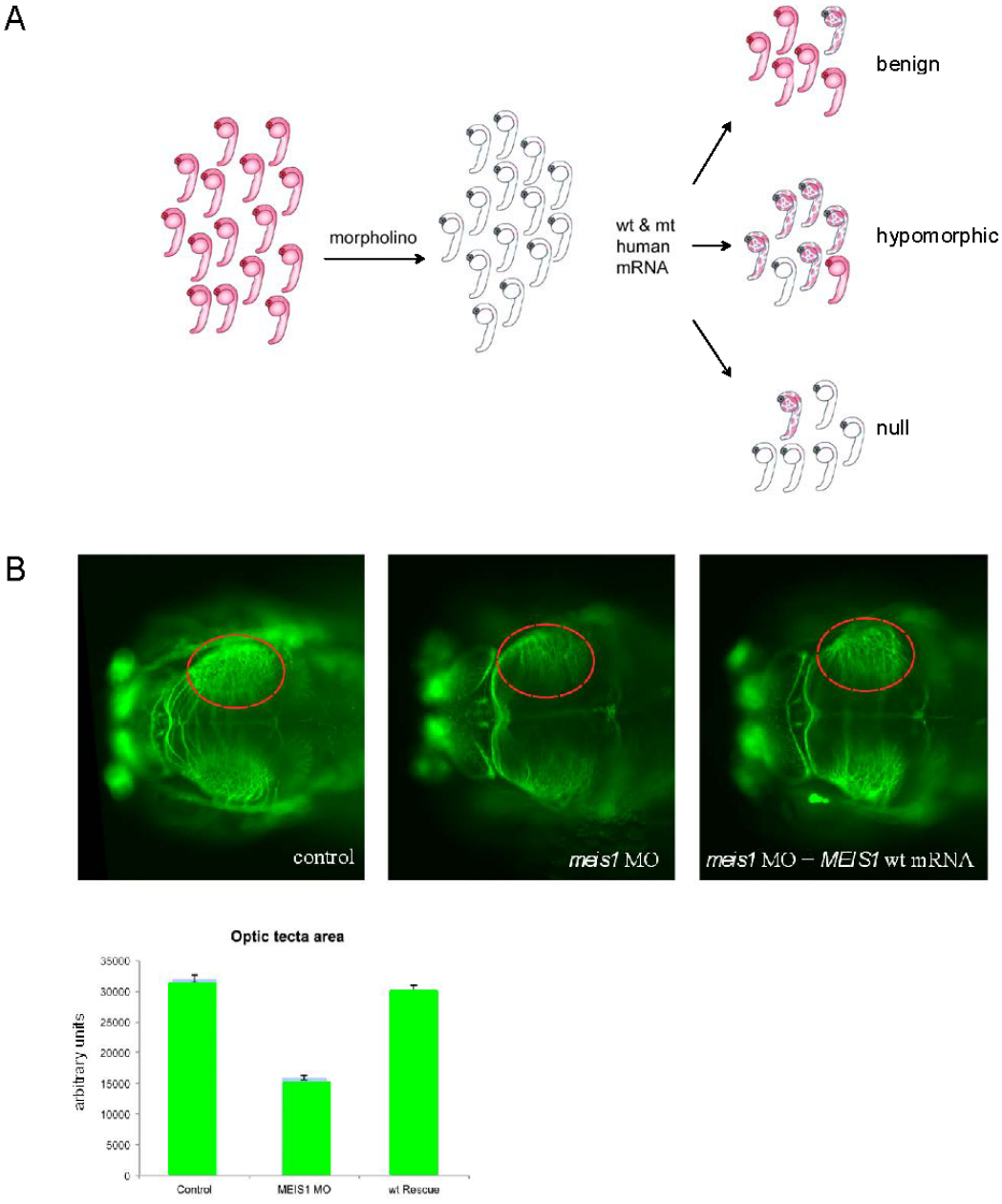
A translation-blocking morpholino (MO) against meis1 was designed and obtained from Gene Tools, LLC. We injected 1nl of diluted MO (4 ng) and/or RNA (75 pg) into wildtype zebrafish embryos at the 1-to 2-cell stage. Injected embryos were fixed overnight at 72 hpf in 4% paraformaldehyde (PFA) and stored in 100% methanol at −20°C. For acetylated tubulin staining embryos were fixed in Dent's fixative (80% methanol, 20% DMSO) overnight at 4°C. The embryos were permeabilized with proteinase K followed by post-fixation with 4% PFA, washed in PBSTX (PBS+0.5% Triton X-100). After rehydration in PBS, PFA-fixed embryos were washed in IF buffer (0.1% Tween-20, 1% BSA in PBS 1X) for 10 minutes at room temperature. The embryos were incubated in the blocking buffer (10% FBS, 1% BSA in PBS1X) for 1hr at room temperature. After two washes in IF Buffer for 10 minutes each, embryos were incubated in the primary antibody solution (anti-acetylated tubulin (T7451, mouse, Sigma-Aldrich), 1:1000) in blocking solution, overnight at 4°C. After two washes in IF Buffer for 10 minutes each, embryos were incubated in the secondary antibody solution, 1:1000 Alexa Fluor rabbit anti-goat IgG (A21207, Invitrogen) in blocking solution, for 1hr at room temperature.

For RNA rescue and overexpression experiments, the human wild type mRNAs was cloned into the pCS2 vector and transcribed *in vitro* using the SP6 Message Machine kit (Ambion). All the experiments were repeated three times and significance of the morphant phenotype was judged using student`s t-test.

SUPPLEMENTARY FIGURES



Supplementary Figure 1: Study Design.



Supplementary Figure 2: Rescue of the *meis1* morpholino knock-down phenotype by co-injection with human *MEIS1* wildtype mRNA. (A) Schematic depiction of the in vivo complementation assay in zebrafish embryos. (B) 50 zebrafish embryos from pooled batches were injected per condition and immunostained using an antibody raised against acetylated tubulin at 72 hpf. MO=morpholino, wt=wildtype

SUPPLEMENTARY TABLES

gene	exon	genomic position (hg19)	variant	class	reference	Frequency		total (all/non-syn)		frequency		p-value	MAF	
						188 cases	188 controls	cases	controls	3265 cases	2944 controls	nominal	dbSNP137	NHLBI-ESP[5]
PTPRD	4	chr9:8,528,700, C>T	T144T	syn	novel		1			no assay	no assay			not found
PTPRD	6	chr9:8,526,648, C>G	c.551-4	nearsplice	rs142871508		1			1/12/3196	0/4/2923	0.03	no data	21/8165
PTPRD	9	chr9:8,521,404, T>G	P278P	syn	novel	1				0/5/3210	0/0/2922	0.03		not found
PTPRD	9	chr9:8,521,270, T>G	c.961+7	nearsplice	novel	1				0/4/3210	0/0/2853	0.05		not found
PTPRD	**10**	**chr9:8,518,098, G>T**	**S431R**	**non-syn**	**rs145325302**	**2**				**0/6/3185**	**0/10/2924**	**0.24**	**no data**	**7/8599**
PTPRD	11	chr9:8,507,364, T>C	P538P	syn	rs147878183	2	2			0/16/3173	0/20/2914	0.35	no data	15/8598
PTPRD	**12**	**chr9:8,504,356, A>T**	**K577I**	**non-syn**	**novel**	**1**				**0/1/3210**	**0/0/2927**	**0.33**		**1/8599**
PTPRD	**15**	**chr9:8,497,250, T>C**	**T781A**	**non-syn**	**rs72694737**	**13**	**10**			**4/180/2985**	**5/175/2737**	**0.51**	**1.90%**	**256/8344**
PTPRD	**17**	**chr9:8,486,142, T>C**	**V892A**	**non-syn**	**rs151005956**		**2**			**0/30/3165**	**0/38/2896**	**0.18**	**no data**	**36/8564**
PTPRD	25	chr9:8,404,607, C>T	N1380N	syn	rs138755135	1	1			assay failed	assay failed		no data	1/8599
PTPRD	26	chr9:8,389,343, A>G	Y1425Y	syn	rs75966719	1				**0/0/3192**	**0/0/2934**	n/a	1.50%	not found
PTPRD	**28**	**chr9:8,375,990, G>A**	**R1536H**	**non-syn**	**rs142960593**	**2**				**0/4/3188**	**0/0/2934**	**0.05**	**no data**	**6/8592**
PTPRD	30	chr9:8,341,272, A>G	c.4948-4	nearsplice	novel	1				0/5/3201	0/3/2924	0.56		9/8591
PTPRD	**33**	**chr9:8,331,681, G>A**	**G1812E**	**non-syn**	**novel**		**1**			**0/0/3212**	**0/1/2928**	**0.29**		**1/8599**
PTPRD	34	chr9:8,319,846, T>G	A1885A	syn	novel		1			0/0/3215	0/1/2926	0.29		not found
PTPRD	**35**	**chr9:8,317,921, C>T**	**R1898C**	**non-syn**	**rs150063446**	**1**		**26/19**	**19/13**	**0/8/3205**	**0/0/2930**	**0.006**	**no data**	**1/8597**
BC034767	1	chr16:52,640,843, C>T		non-coding	novel	1				no assay	no assay			
BC034767	1	chr16:52,640,520, G>A		non-coding	rs78030692	3	1			**0/27/3183**	**0/21/2905**	**0.58**	0.70%	
BC034767	2	chr16:52,624,969, C>T		non-coding	novel	1		**5**/-	**1**/-	assay failed	assay failed			
TOX3	**3**	**chr16:52,497,872, G>A**	**V123M**	**non-syn**	**rs16951186**	**3**	**1**			**0/60/3155**	**0/69/2858**	**0.18**	**2.70%**	**80/8166**
TOX3	**3**	**chr16:52,497,869, C>T**	**L122F**	**non-syn**	**novel**		**3**			**0/5/3167**	**0/9/2921**	**0.22**		**10/8226**
TOX3	**4**	**chr16:52,484,428, C>T**	**R147W**	**non-syn**	**novel**	**1**				**0/2/3214**	**0/0/2925**	**0.17**		**not found**
TOX3	4	chr16:52,484,384, C>T	T161T	syn	rs147678931	4	4			0/49/3144	1/42/2878	0.92	no data	53/8411
TOX3	**4**	**chr16:52,484,286, G>A**	**S193N**	**non-syn**	**novel**	**1**				**0/1/3195**	**0/0/2933**	**0.33**		**not found**
TOX3	5	chr16:52,480,113, T>C	A233A	syn	novel	1				1/3/3186	**0/0/2934**	**0.03**		not found
TOX3	7	chr16:52,473,815, T>C	N351N	syn	novel	1				0/1/3207	0/0/2926	0.22		not found
TOX3	7	chr16:52,473,584, C>T	P428P	syn	rs148327560	4				0/61/3126	0/59/2872	0.78	no data	84/8502
TOX3	7	chr16:52,473,449, G>T	I473I	syn	rs117772184	3	2			1/65/3148	1/55/2872	0.7	0.27%	92/8444
TOX3	**7**	**chr16:52,473,331, C>T**	**R513C**	**non-syn**	**novel**	**1**	**1**	**19/6**	**11/5**	**no assay**	**no assay**			**55/8343**
BTBD9	5	chr6:38,560,469, C>T	V239V	syn	rs41303370	1	1			0/22/3177	0/19/2904	0.85	0.30%	not found
BTBD9	5	chr6:38,560,625, C>T	c.550-9	nearsplice	novel		1	**1/0**	**2/0**	0/0/3192	0/3/2931	0.07		not found
MEIS1	**3**	**chr2:66,666,978, C>G**	**H81Q**	**non-syn**	**novel**	**1**				**0/1/3218**	**0/0/3218**	**0.34**		**not found**
MEIS1	**6**	**chr2:66,670,160, T>A**	**S204T**	**non-syn**	**novel**	**1**				**0/1/3210**	**0/02918**	**0.34**		**not found**
MEIS1	**8**	**chr2:66,739,353, G>A**	**R272H**	**non-syn**	**rs61752693**	**2**				**0/7/3187**	**0/1/2933**	**0.04**	**no data**	**4/8520**
MEIS1	**13**	**chr2:66,798,627, T>C**	**M453T**	**non-syn**	**novel**	**1**				**1/28/61**	**0/13/2921**	**0.02**		**n/a**
MEIS1	5	chr2:66,668,536, A>G	c.433-10	nearsplice	rs41285949	1				0/21/3180	0/16/2906	0.58	no data	not found
MEIS1	7	chr2:66,691,348, G>A	E246E	syn	rs1135875	3				0/56/3136	0/37/2897	**0.11**	no data	89/8279
MEIS1	**3**	**chr2:66,667,100, C>T**	**A122V**	**non-syn**	**novel**		**1**	**9/5**	**1/1**	**0/0/3192**	**0/0/2934**	**n/a**		**not found**
MAP2K5	11	chr15:67,938,728	C219C	syn	rs56254481	4	2			0/50/3147	0/51/2864	0.57	1.40%	101/8469
MAP2K5	14	chr15:67,984,818	D283D	syn	rs55966838	4	4			1/83/3133	1/67/2860	0.47	0.30%	80/8516
MAP2K5	**16**	**chr15:67,995,739, T>C**	**F346S**	**non-syn**	**novel**	**1**		**9/1**	**6/0**	**0/1/3190**	**0/0/2934**	**0.33**		**not found**

SKOR1	**2**	**chr15:68,118,924, A>G**	**D244G**	**non-syn**	**novel**		**1**			**0/0/3215**	**0/0/2929**	**n/a**		**not found**
SKOR1	3	chr15:68,119,906, G>A	P536P	syn	novel		1			no assay	no assay			not found
SKOR1	**3**	**chr15:68,120,154, T>G**	**V619G**	**non-syn**	**novel**	**1**				**0/1/3191**	**0/0/2920**	**0.34**		**not found**
SKOR1	**3**	**chr15:68,120,205, C>T**	**P636L**	**non-syn**	**novel**	**1**				**no assay**	**no assay**			**not found**
SKOR1	**3**	**chr15:68,120,267, A>G**	**T657A**	**non-syn**	**novel**		**1**			**0/0/3205**	**0/1/2921**	**0.29**		**not found**
SKOR1	**3**	**chr15:68,120,284, C>A**	**D662E**	**non-syn**	**novel**	**1**	**1**			**no assay**	**no assay**			**6/8250**
SKOR1	**3**	**chr15:68,120,329, C>T**	**R677C**	**non-syn**	**novel**	**1**				**no assay**	**no assay**			**not found**
SKOR1	**3**	**chr15:68,120,354, C>T**	**R686W**	**non-syn**	**novel**	**5**	**1**			**no assay**	**no assay**			**not found**
SKOR1	**6**	**chr15:68,125,577, A>G**	**I885V**	**non-syn**	**rs143419968**		**1**	**9/9**	**6/5**	**1/3/3186**	**0/2/2919**	**0.47**	**no data**	**7/8589**

Supplementary Table 1: Variants identified within the coding regions ± 10 bp in seven genes that have been associated with RLS via GWAS[3,6-8]. The table also displays the detailed results of the frequency assessment of the identified variants in a case/control sample consisting of 3265 individuals with RLS and 2944 control individuals belonging to the KORA general population cohort. Bold print denotes non-synonymous variants. n/a=not applicable

region	*number of unique variants*	*occurrence of variants in cases (n=3760)*	*occurrence of variants in controls (n=3542)*
coding	22	67	41
non-synonymous	13	18	9
synonymous	9	49	32
nearsplice (±10bp)	6	29	25
5`UTR	7	16	2
3`UTR	36	253	287
intergenic (UTRs ±100bp)	1	1019	253
all	**89**	**1451**	**649**
NHLBI-ESP exomes (EA only)			***n=4300***
coding	12		112
non-synonymous	7		13
synonymous	5		99

Supplementary Table 2: Variants with MAF < 5% identified in the coding regions ± 10bp as well as the UTRs ± 100bp of *MEIS1*. The overall number of variants detected by LightScanner high-resolution melting curve analysis is comparable to the genetic variance in *MEIS1* detected by whole-exome sequencing and annotated in the NHLBI-ESP exome variant server[5]. EA=European American

gene	*location*	*genom. pos.*	*variant*	*AA substitution*	*frequency*		*dbSNP137*	*category*	*PolyPhen2*	*NHLBI-ESP (n=4300)*
					cases	*controls*				
MEIS1	5'UTR	chr2:66662563	c.1-426 G>C		1		novel	5'UTR		
MEIS1	5'UTR	chr2:66662566	c.1-423 G>A			1	novel	5'UTR		
MEIS1	5'UTR	chr2:66662686	c.1-302 C>T		1		novel	5'UTR		
MEIS1	5'UTR	chr2:66662695	c.1-294 T>C		3		novel	5'UTR		
MEIS1	5'UTR	chr2:66662902	c.1-87 G>A		1		novel	5'UTR		
MEIS1	5'UTR	chr2:66662932	c.1-57 A>T		1		novel	5'UTR		
MEIS1	5'UTR	chr2:66662948	c.1-41 A>G		9	1	rs71411941 (0.10%)	5'UTR		
MEIS1	intron 5	chr2:66668536	c.433-10 A>G		23	25	rs41285949 (n/d)	nearsplice		
MEIS1	intron 6	chr2:66691237	c.631-4 A>G		1		novel	nearsplice		
MEIS1	intron 7	chr2:66739271	c.743-10 A>G		1		novel	nearsplice		
MEIS1	intron 9	chr2:66794579	c.966-6 T>C		1		novel	nearsplice		
MEIS1	intron 10	chr2:66795790	c.1025-9 C>T		2		novel	nearsplice		
MEIS1	intron 11	chr2:66796175	c.1115-7 C>A		1		novel	nearsplice		
MEIS1	exon 2	chr2:66664947	c.91 A>C	R31R	1		novel	syn		6/8438
MEIS1	exon 2	chr2:66664984	c.128 C>T	P43L		1	novel	non-syn	benign	not found
MEIS1	exon 2	chr2:66665029	c.173 C>G	A58G	1		novel	non-syn	benign	not found
MEIS1	exon 2	chr2:66665031	c.175 A>G	M59V		4	novel	non-syn	benign	not found
MEIS1	exon 2	chr2:66665063	c.207 C>T	D69D	1	2	novel	syn		not found
MEIS1	exon 3	chr2:66666978	c.243 C>G	H81Q	1		novel	non-syn	prob dam	not found
MEIS1	exon 3	chr2:66666994	c.259 T>C	L87L		1	novel	syn		not found
MEIS1	exon 3	chr2:66667100	c.365 C>T	A122V		1	novel	non-syn	pos dam	not found
MEIS1	exon 5	chr2:66668552	c.439 C>A	Q147K		1	novel	non-syn	prob dam	not found
MEIS1	exon 6	chr2:66670039	c.489 C>T	H163H	1		novel	syn		not found
MEIS1	exon 6	chr2:66670074	c.524 G>A	S175N	1		novel	non-syn	benign	not found
MEIS1	exon 6	chr2:66670157	c.607 A>G	R203G	1		novel	non-syn	benign	1/8241
MEIS1	exon 6	chr2:66670160	c.610 T>A	S204T	1		novel	non-syn	pos dam	not found
MEIS1	exon 7	chr2:66691270	c.660 G>A	T220T	2		novel	syn		not found
MEIS1	exon 7	chr2:66691325	c.715 C>T	H239Y	1		novel	non-syn	benign	not found
MEIS1	exon 7	chr2:66691333	c.723 G>A	G241G	1		novel	syn		not found
MEIS1	exon 7	chr2:66691348	c.738 G>A	E246E	43	27	rs13005707 (0.50%)	syn		89/8279
MEIS1	exon 8	chr2:66739300	c.762 T>C	S254S		1	novel	syn		not found
MEIS1	exon 8	chr2:66739353	c.815 G>A	R272H	7	1	rs61752693 (n/d)	non-syn	benign	4/8520
MEIS1	exon 8	chr2:66739405	c.867 G>A	A289A		1	novel	syn		not found
MEIS1	exon 11	chr2:66795833	c.1059 G>C	Q353H	4		novel	non-syn	prob dam	not found
MEIS1	exon 11	chr2:66795870	c.1096 A>T	M366L	1		novel	non-syn	pos dam	not found
MEIS1	exon 13,3'UTR	chr2:66798455	c.*115 C>T, c.1186 C>T	Q398X		1	novel	non-syn/stop, 3'UTR		
MEIS1	exon 13,3'UTR	chr2:66798458	c.*118 C>T, c.1189 C>T	P399S		1	novel	non-syn, 3'UTR		
MEIS1	exon 13,3'UTR	chr2:66798472	c.*132 C>G, c.1204 C>G	P403P		2	novel	syn, 3'UTR		
MEIS1	exon 13,3'UTR	chr2:66798496	c.*156 G>A, c.1227 G>A	G411G		1	novel	syn, 3'UTR		
MEIS1	exon 13,3'UTR	chr2:66798627	c.*287 T>C, c.1359 T>C	M455T	15	4	novel	non-syn, 3'UTR		
MEIS1	exon 13,3'UTR	chr2:66798654	c.*314 C>T, c.1385 C>T	A464V	1		rs76963732 (1.83%)	non-syn, 3'UTR		
MEIS1	3'UTR	chr2:66798698			1		novel	3'UTR		
MEIS1	3'UTR	chr2:66798701			1	1	novel	3'UTR		
MEIS1	3'UTR	chr2:66798743				2	novel	3'UTR		
MEIS1	3'UTR	chr2:66798789				1	novel	3'UTR		
MEIS1	3'UTR	chr2:66798823	c.*483 T>C		8	13	rs111248476 (0.05%)	3'UTR		
MEIS1	3'UTR	chr2:66798844				5	novel	3'UTR		
MEIS1	3'UTR	chr2:66798902				1	novel	3'UTR		
MEIS1	3'UTR	chr2:66798910	c.*570 C>T			1	novel	3'UTR		
MEIS1	3'UTR	chr2:66798924			1		novel	3'UTR		

MEIS1	3'UTR	chr2:66798932	c.*592 A>T	20	13	rs140870220 (0.02%)	3'UTR
MEIS1	3'UTR	chr2:66798936			1	novel	3'UTR
MEIS1	3'UTR	chr2:66798939_940			2	novel	3'UTR
MEIS1	3'UTR	chr2:66799129		1	1	rs149727570 (n/d)	3'UTR
MEIS1	3'UTR	chr2:66799130			2	novel	3'UTR
MEIS1	3'UTR	chr2:66799132			1	novel	3'UTR
MEIS1	3'UTR	chr2:66799143			1	novel	3'UTR
MEIS1	3'UTR	chr2:66799145_154		1		novel	3'UTR
MEIS1	3'UTR	chr2:66799162			1	novel	3'UTR
MEIS1	3'UTR	chr2:66799193			1	novel	3'UTR
MEIS1	3'UTR	chr2:66799195			1	novel	3'UTR
MEIS1	3'UTR	chr2:66799394	c.*1054 A>G	3/184	1/214	rs72824830 (1.95%)	3'UTR
MEIS1	3'UTR	chr2:66799420		1		novel	3'UTR
MEIS1	3'UTR	chr2:66799458		1		novel	3'UTR
MEIS1	3'UTR	chr2:66799471		1		novel	3'UTR
MEIS1	3'UTR	chr2:66799499		2	2	novel	3'UTR
MEIS1	3'UTR	chr2:66799613		1		novel	3'UTR
MEIS1	3'UTR	chr2:66799682			4	novel	3'UTR
MEIS1	3'UTR	chr2:66799782		1		novel	3'UTR
MEIS1	3'UTR	chr2:66799788			3	novel	3'UTR
MEIS1	3'UTR	chr2:66799828	c.*1488 T>G	10	6	novel	3'UTR
MEIS1	post-3'UTR	chr2:66799986		44/931	5/243	rs11693221 (2.09%)	post-3'UTR, intergenic

Supplementary Table 3: Rare and low-frequency variants indentified by screening the coding regions ± 10 bp and the 5' and 3' UTRs ± 100 bp of *MEIS1* in 3760 individuals with RLS and 3542 general population controls. The coding variants below the dashed line are only coding in isoform 2 of *MEIS1* (ENSP00000381518). AA=amino acid, n/d=no data.

SUPPLEMENTARY REFERENCES

1 Allen, R. P. *et al.* Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology. A report from the restless legs syndrome diagnosis and epidemiology workshop at the National Institutes of Health. *Sleep medicine* **4**, 101-119 (2003).

2 Wichmann, H. E., Gieger, C., Illig, T. & Group, M. K. S. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67 Suppl 1**, S26-30, doi:10.1055/s-2005-858226 (2005).

3 Winkelmann, J. *et al.* Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. *Nature genetics* **39**, 1000-1006, doi:10.1038/ng2099 (2007).

4 Schulte, E. C. *et al.* Variant screening of the coding regions of MEIS1 in patients with restless legs syndrome. *Neurology* **76**, 1106-1108, doi:10.1212/WNL.0b013e318211c366 (2011).

5 Project, N. G. O. E. S. *The NHLBI GO ESP Exome Variant Server*, <http://evs.gs.washington.edu/EVS/> (2013).

6 Schormair, B. *et al.* PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome. *Nature genetics* **40**, 946-948, doi:10.1038/ng.190 (2008).

7 Stefansson, H. *et al.* A genetic risk factor for periodic limb movements in sleep. *The New England journal of medicine* **357**, 639-647, doi:10.1056/NEJMoa072743 (2007).

8 Winkelmann, J. *et al.* Genome-wide association study identifies novel restless legs syndrome susceptibility loci on 2p14 and 16q12.1. *PLoS genetics* **7**, e1002171, doi:10.1371/journal.pgen.1002171 (2011).

2.5 Hopfner *et al.*, The Role of SCARB2 as Susceptibility Factor in Parkinson`s Disease, *Movement Disorders*, 2013[V]

Personal contributions: I participated in LightScanner® high-resolution melting curve analysis and follow-up Sanger sequencing as well as Sequenom®-based genotyping. Further, I also contributed to the writing of the manuscript and the design of the table.

The Role of *SCARB2* as Susceptibility Factor in Parkinson's Disease

Franziska Hopfner, MD,[1,2,3]* Eva C. Schulte, MD,[1,2,4] Brit Mollenhauer, MD,[5,6] Benjamin Bereznai, MD,[7] Franziska Knauf, MSc,[1] Peter Lichtner, PhD,[1,2] Alexander Zimprich, MD,[8] Dietrich Haubenberger, MD,[8] Walter Pirker, MD,[8] Thomas Brücke, MD,[9] Annette Peters, PhD,[10] Christian Gieger, PhD,[11] Gregor Kuhlenbäumer, MD, PhD,[12] Claudia Trenkwalder, MD[5,6] and Juliane Winkelmann, MD[1,2,4]

[1]*Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany* [2]*Institute of Human Genetics, Klinikum rechts der Isar, Technische Universität München, Munich, Germany* [3]*Department of Neurology, Christian-Albrechts-University Kiel, Kiel, Germany* [4]*Department of Neurology, Klinikum rechts der Isar, Technische Universität München, Munich, Germany* [5]*Paracelsus-Elena-Klinik, Center of Parkinsonism and Movement Disorders, Kassel, Germany* [6]*Georg-August-University, Department of Clinical Neurophysiology, Göttingen, Germany* [7]*Clinical and Research Center of Molecular Neurology, Semmelweis University, Budapest, Hungary* [8]*Department of Neurology, Medizinische Universität Wien, Vienna, Austria* [9]*Neurological Department, Wilhelminenspital, Vienna, Austria* [10]*Institut of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany* [11]*Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany* [12]*Institute of Experimental Medicine, Christian-Albrechts-University Kiel, Kiel, Germany*

Correspondence to: Franziska Hopfner, Klinik für Neurologie, Arnold-Heller-Straße 3, Haus 41, Neurozentrum, 24105 Kiel, Germany; f.hopfner@neurologie.uni-kiel.de

Funding agencies: The study was funded and sponsored by institutional funding from Helmholtz Zentrum München, Munich, Germany. Recruitment of case and control cohorts was supported by institutional (Helmholtz Zentrum München, Munich, Germany) and government funding from the German Bundesministerium für Bildung und Forschung (03.2007-02.2011 FKZ 01ET0713) and the Hungarian National Innovation Office (TAMOP-4-2-1/B-03/1/KMR-2010-001). Gregor Kuhlenbäumer is a member of the DFG-funded Cluster of Excellence "Inflammation at Interfaces."

Relevant conflicts of interest/financial disclosures: Nothing to report. Full financial disclosures and author roles may be found in the online version of this article.

Received: 23 May 2012; **Revised:** 6 November 2012; **Accepted:** 3 December 2012

Published online 13 February 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.25349

ABSTRACT

Background: Genetic variation in the glucocerebrosidase (*GBA*) gene is strongly associated with Parkinson's disease (PD). Transport of glucocerebrosidase to the lysosome involves the protein encoded by the *SCARB2* gene. An association between the common SNP rs6812193, upstream of *SCARB2*, and PD has been reported previously. The role of exonic variants in the *SCARB2* gene in PD has not been examined.

Methods: We studied the role of exonic variants in *SCARB2* and tried to replicate the association between the SNP rs6812193 and PD in a German and Austrian sample. Screening of all *SCARB2* exons by high-resolution melting curve analysis was performed in 376 German PD patients. The SNP rs6812193 was analyzed in 984 PD patients and 1014 general population controls.

Results: We identified no novel exonic variants in *SCARB2* but confirmed the association between SNP rs6812193 and PD (OR, 0.86; *P*=.02). © 2013 *Movement* Disorder Society

Key Words: genetics; Parkinson's disease

Homozygous mutations in the glucocerebrosidase (*GBA*) gene cause Gaucher disease (GD), whereas heterozygous mutations have been identified as genetic risk factors for PD.[1] Lysosomal membrane protein 2 (LIMP2) encoded by the *SCARB2* gene mediates GBA trafficking from the endoplasmic reticulum to the lysosome. α-synuclein, a key molecule in PD pathogenesis, is degraded in lysosomes. In model systems, impaired GBA function leads to reduced lysosomal protein degradation and accumulation as well as aggregation of α-synuclein. Furthermore, it has been shown that α-synuclein inhibits the lysosomal activity of normal GBA in neurons and brain tissue of patients with idiopathic PD, suggesting that GBA depletion contributes to the pathogenesis of sporadic synucleinopathies.[2,3] Genomewide association studies have identified an association between the single-nucleotide polymorphism (SNP) rs6812193, upstream of *SCARB2*, and PD.[4,5] Whether rare exonic variants in *SCARB2* are associated with PD has not been investigated. The aim of the current study was to assess the genetic contribution of exonic variants in *SCARB2* to PD and to replicate the recently reported association between the SNP rs6812193, upstream of *SCARB2* in an intron of *FAM47E* and coding for a transcript with unknown function, and PD in a German/Austrian PD sample.

Patients and Methods

Patient and Control Samples

Variant screening of the 12 exons and the exon–intron boundaries of *SCARB2* was performed in 376 unrelated German PD patients (mean age, 71.1±9.4 years; 68.4% male). The case–control sample used to replicate the association between rs6812193 and PD comprised 984 German and Austrian patients (63.7±10.5 years; 34.6% female) including the 376 patients screened for mutations in the *SCARB2* gene and 1014 controls (76±6.6 years; 50.1% female).[6] Controls were of European descent and were recruited in the KORA-AGE surveys from the general population living in or near the city of Augsburg, Germany.[7] All PD patients were diagnosed according to the United Kingdom Brain Bank criteria. Ethics review board approval and participants' written consent were obtained.

Mutation Screening of *SCARB2*

PCR primers were designed using ExonPrimer (http://ihg.gsf.de) or Primer3plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). The mutation screening was performed using high-resolution melting curve analysis (HRM) employing LightScanner technology and standard protocols (IDAHO Technology Inc., Salt Lake City, UT).[8] Samples with aberrant melting patterns were sequenced using BigDye Terminator Chemistry 3.1 on an ABI 3730 sequencer (Life Technologies, Carlsbad, CA).

Analysis of SNP rs6812193

Genotyping was performed by matrix-assisted laser desorption ionization/time-of-flight mass spectrometry using iPLEX chemistry on the Sequenom platform (Sequenom, San Diego, CA). Statistical analysis using logistic regression was performed using PLINK v1.07 with correction for age and sex.[9]

Results

The mutation screening revealed no novel unknown variants in the exonic regions of *SCARB2*. Already known annotated variants (dbSNP [vs.135] rs894253, rs3853 189, rs61598131, rs116738729, rs17001633, rs2289512) were detected in the patients but not followed up any further because the allele frequencies were comparable to those found in dbSNP for white populations.[10]

There was no deviation from Hardy–Weinberg equilibrium for SNP rs6812193 in either patients or controls. We found a significant association between SNP rs6812193 and PD in the logistic regression analysis with correction for sex and age ($P_{corr.}$=.02; $MAF[T]_{cases}$, 0.33; $MAF[T]_{controls}$, 0.36; OR, 0.86; 95% CI, 0.75–0.98; Table 1).

TABLE 1. Association results of rs6812193 in our current study compared with 2 previous studies[4,5]

Study	Odds ratio (95% CI)	Minor allele frequency$_{(controls)}$[a]	Cases, n	Controls, n
Current study	0.86 (0.75–0.98)	0.36	984	1014
IPDGC[4]	0.90 (0.86–0.94)	0.36	12,386	21,026
Do et al, 23andMe[5]	0.84 (0.79–0.89)	0.37	3426	29,624

[a]Minor allele frequencies in controls were not available for all studies.

Discussion

Our findings indicate that exonic variants in *SCARB2* do not play a major role in the development of PD in our German PD sample. A validation study of HRM showed an average specificity of 98%, indicating a low incidence of false-positives. Nevertheless, uncertainty remains about whether all variants present in the patients were detected.[8] We confirmed the association between the common variant rs6812193, located upstream of *SCARB2*, and PD in our German and Austrian sample. The minor allele frequencies in controls were comparable to previous studies (Table 1).[4,5] In conclusion, our results do not support a major role of rare exonic variants in *SCARB2* in the pathogenesis of PD in Germany but confirm the association between SNP rs6812193, upstream of *SCARB2*, and PD. However, it remains unclear if and how this SNP influences the *SCARB2* gene. After completion of our study, an association between SNP rs6825004, in an intron of *SCARB2*, and PD has been reported, supporting the notion that common genetic variants in the genomic region of *SCARB2* are associated with PD.[11]

References

1. Sidransky E, Nalls MA, Aasly JO, et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N Engl J Med. 2009;361:1651–1661.
2. Mazzulli JR, Xu YH, Sun Y, et al. Gaucher disease glucocerebrosidase and alpha-synuclein form a bidirectional pathogenic loop in synucleinopathies. Cell. 2011;146:37–52.
3. Cuervo AM, Stefanis L, Fredenburg R, Lansbury PT, Sulzer D. Impaired degradation of mutant alpha-synuclein by chaperone-mediated autophagy. Science. 2004;305:1292–1295.
4. Consortium IPsDG, 2 WTCCC. A two-stage meta-analysis identifies several new loci for Parkinson's disease. PLoS Genet. 2011;7:e1002142.
5. Do CB, Tung JY, Dorfman E, et al. Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. PLoS Genet. 2011;7:e1002141.
6. Zimprich A, Benet-Pages A, Struhal W, et al. A Mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. Am J Hum Genet. 2011;89:168–175.
7. Wichmann HE, Gieger C, Illig T. KORA-gen—resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen. 2005;67(Suppl 1):S26–S30.

8. van der Stoep N, van Paridon CD, Janssens T, et al. Diagnostic guidelines for high-resolution melting curve (HRM) analysis: an interlaboratory validation of BRCA1 mutation scanning using the 96-well LightScanner. Hum Mutat. 2009;30:899–909.
9. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–575.
10. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29:308–311.
11. Michelakakis H, Xiromerisiou G, Dardiotis E, et al. Evidence of an association between the scavenger receptor class B member 2 gene and Parkinson's disease. Mov Disord. 2012;27:400–405.

Functional Movement Disorders Are Not Uncommon in the Elderly

Amit Batla, MD DM (Neuro),[1] Maria Stamelou, MD, PhD,[1] Mark J. Edwards, MD,[1] Isabel Pareés, MD,[1] Tabish A. Saifee, MRCP,[1] Zoe Fox, PhD[2] and Kailash P. Bhatia, MD FRCP[1]*

[1]*Sobell Department of Motor Neuroscience and Movement Disorders, UCL Institute of Neurology, University College London, London, United Kingdom;* [2]*Department of Biostatistics, UCL Institute of Neurology, University College London, London, United Kingdom*

ABSTRACT

Background: Functional movement disorders (FMDs) are thought to be rare in the elderly. Clinical characteristics of the elderly people who develop FMDs are rarely reported. The objective of this study was to highlight the clinical characteristics of FMD in the elderly and compared these with a cohort of patients with a younger age of onset.

Methods: The authors performed a retrospective review of the clinical records of patients with FMD who were seen at their center in the last 5 years and had consented to be included in research studies. Patients fulfilling currently accepted diagnostic criteria for FMD as documented, clinically established, or probable were included.

Results: Of 151 patients with FMD who were identified and had sufficient information, 21.0% (n=33) had an onset after age 60 years (elderly group). The mean age of onset of FMD was 63.5 years (standard deviation, 5.2 years) in the elderly group and 35.5 years (standard deviation, 12.6 years) in the younger group. Tremor was the most common movement disorder in both groups (elderly group, 33.3%; younger group: 38.9%). Fixed dystonia was not observed in any patient who had an FMD onset after age 60 years. Gait abnormalities were significantly more common in the elderly group (69.7%) than in younger patients (23.5%; $P<0.001$). Associated psychogenic nonepileptic seizures tended to be more common in elderly patients (18.2%) compared with younger patients (13%; $P=0.06$).

Conclusions: Contrary to common perceptions, FMDs are not uncommon in the elderly, and 1 in 5 patients in the current cohort, onset of FMD occurred after age 60 years. Gait abnormalities and psychogenic nonepileptic seizures may be more common in older patients.

Key Words: functional movement disorders; elderly; psychogenic; gait; psychogenic non epileptic seizures

Functional movement disorders (FMDs) are clinically challenging and phenotypically diverse.[1] There are internationally accepted diagnostic criteria for FMD,[2] highlighting the importance of positive clinical signs in making a diagnosis.[1] FMDs often are not suspected in the elderly, and previous studies have confirmed that most patients are young.[2–5] The mean age at FMD onset ranges from 36.9 to 50.0 years in large cohort studies.[3,4,6,7] Although some patients with an FMD onset after age 60 years have been included in previous cohorts,[4,5,8,9] specific evaluation of an older onset group or comparison in detail with a younger onset group have not been performed. We retrospectively reviewed a large cohort of patients with FMD at our center to assess how commonly FMD begins after age 60 years and whether the clinical features of these patients differ from those who have FMD onset at a younger age.

Patients and Methods

We retrospectively reviewed the clinical records of patients with FMD who had given their consent for inclusion in clinical research and were recruited from the movement disorders clinics of K.P.B. from 2006 to 2012 and M.J.E. from 2009 to 2012. Patients who fulfilled the currently accepted diagnostic criteria for FMD[4] as documented, clinically established, or probable were included. Patients with pending investigations or evidence of any significant imaging or laboratory abnormality were excluded. A pro forma procedure was used to record clinical characteristics. For comparison, the patients were divided into 2 groups based on age at onset before or after 60 years, in line with other studies

Correspondence to: Prof. K. P. Bhatia, Sobell Department of Motor Neuroscience and Movement Disorders, UCL Institute of Neurology, Queen Square, London, WC1N 3BG, United Kingdom; k.bhatia@ion.ucl.ac.uk

Relevant conflicts of interest/financial disclosures: Nothing to report. Full financial disclosures and author roles may be found in the online version of this article.

Received: 31 October 2012; **Revised:** 30 November 2012; **Accepted:** 16 December 2012

Published online 15 February 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.25350

2.6 Schulte *et al.*, Variants in *eukaryotic translation initiation factor 4G1* in sporadic Parkinson`s disease, *Neurogenetics*, 2012[VI]

Personal contributions: I participated in designing the study, performed LightScanner® high-resolution melting curve analysis and follow-up Sanger sequencing of the coding regions of *EIF4G1* and analyzed the LightScanner® data. I designed the multiplex PCR for Sequenom®-based genotyping, analyzed genotyping data and participated in the fragment analysis used for haplotype determination and performed *in silico* predictions of variant pathogenicity. I wrote the manuscript and designed all figures and tables.

Neurogenetics (2012) 13:281–285
DOI 10.1007/s10048-012-0334-9

SHORT COMMUNICATION

Variants in *eukaryotic translation initiation factor 4G1* in sporadic Parkinson's disease

Eva C. Schulte · Brit Mollenhauer · Alexander Zimprich · Benjamin Bereznai · Peter Lichtner · Dietrich Haubenberger · Walter Pirker · Thomas Brücke · Maria J. Molnar · Annette Peters · Christian Gieger · Claudia Trenkwalder · Juliane Winkelmann

Received: 5 March 2012 / Accepted: 30 May 2012 / Published online: 16 June 2012

Abstract Recently, mutations in *eukaryotic translation initiation factor 4G1* (*EIF4G1*) were reported as a rare cause of familial Parkinson's disease (PD). We screened the 33 exons of *EIF4G1* by high-resolution melting curve analysis for variants in our Central European cohort of 376 PD cases. Variant frequency was assessed in a total of 975 PD cases and 1,014 general population controls. Eight novel nonsynonymous and four synonymous variants were identified. In our cohort, novel and previously identified nonsynonymous variants were very rare. Although it is possible that our general population controls also comprise individuals who have or could develop PD in the future, the presence of the original mutation (*EIF4G1* p.Arg1205 His) in three controls only, raises questions about the causality of this variant with regard to PD.

Keywords Genetics · Rare variants · Parkinson's disease · *EIF4G1*

Introduction

Genome-wide association studies and linkage analyses have identified at least 19 genes associated with idiopathic Parkinson's disease (PD). Most recently, variants in *eukaryotic translation initiation factor 4G1* (*EIF4G1*) were implicated in familial PD, linking dysfunctional mRNA translation initiation to PD pathogenesis. [1] Here, we assess the role of *EIF4G1* variants in our Central European PD cohort.

E. C. Schulte · J. Winkelmann (✉)
Neurologische Klinik und Poliklinik, Klinikum rechts der Isar,
Technische Universität München,
Ismaningerstr. 22,
81675 Munich, Germany
e-mail: winkelmann@lrz.tum.de

E. C. Schulte · P. Lichtner · J. Winkelmann
Institut für Humangenetik, Helmholtz Zentrum München,
Munich, Germany

B. Mollenhauer · C. Trenkwalder
Paracelsus Elena Klinik,
Kassel, Germany

B. Mollenhauer · C. Trenkwalder
Neurologische Klinik, Georg August Universität,
Göttingen, Germany

A. Zimprich · D. Haubenberger · W. Pirker
Department of Neurology, Medical University of Vienna,
Vienna, Austria

B. Bereznai · M. J. Molnar
Center for Molecular Neurology, Department of Neurology,
Semmelweis University,
Budapest, Hungary

P. Lichtner · J. Winkelmann
Institut für Humangenetik, Technische Universität München,
Munich, Germany

T. Brücke
Department of Neurology, Wilhelminenspital,
Vienna, Austria

A. Peters
Institute of Epidemiology II, Helmholtz Zentrum München,
Munich, Germany

C. Gieger
Institute of Genetic Epidemiology, Helmholtz Zentrum München,
Munich, Germany

Methods

Using Idaho® melting curve analysis, we screened the 33 exons and exon intron boundaries of *EIF4G1* in a discovery sample of 376 German PD patients (71.1±9.4 years, 31.6 % female). When altered melting patterns suggested variants, Sanger sequencing ensued. To assess variant frequency, we genotyped the novel as well as four of the five variants previously described in PD (*EIF4G1* c.1505C > T (p.Ala502Val), c.2056G > T (p.Gly686Cys), c.3490A > C (p.Ser1164Arg), and c.3614G > A (p.Arg1205His)) [1] in 975 familial and sporadic PD cases from Austria (n=486, 58.7±11.3 years, 35.4 % female, family history known in n=413, 33.4 % thereof positive for PD in a first or second degree relative), Germany (n=450, 376 of which comprised the discovery sample, 70.2±9.7 years, 32.2 % female, family history known in n=105, 24.7 % thereof positive for PD in a first or second degree relative), and Hungary (n=39, 50.4±10.8 years, 53.9 % female, family history known in n=39, 28.2 % thereof positive for PD in a first- or second-degree relative) and 1,014 general population controls belonging to the KORA-AGE cohort (76.0±6.6 years, 50.1 % female) [2] by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry on the Sequenom platform. The KORA-AGE cohort is a follow-up study of the initial surveys, enriched for older individuals. Individuals known to take dopaminergic medication where excluded from the control sample. All individuals included in this study were Caucasian. German and Austrian PD samples and KORA controls originate from the same geographic region. The small number of Hungarian patients either have an early age of onset or are index patients of larger PD families and were, therefore, genotyped as well. For technical reasons, four novel variants could not be included in the genotyping assay. Haplotype analysis in carriers of the original c. 3614G > A (p.Arg1205His) variant was performed using haplotype-tagging SNPs rs4912537, rs2178403, rs2293605, rs1879244, and rs2230571 and polymorphic markers D3S3609, D3S3578, and D3S3583 by Sanger sequencing. All subjects were diagnosed according to the UK Brain Bank criteria by a senior neurologist specializing in movement disorders. Ethics review board approval and participants' written informed consent were obtained.

Results

In addition to several common and rare synonymous variants, we identified seven nonsynonymous variants, not previously reported in PD, in six individuals. These include c.47C > T (p.Pro16Leu), c.211C > T (p.Pro71Ser, rs113810947), c.953C > T (p.Thr318Ile),c.1622T > G (p.Val541Gly), c.1648G > C (p.Ala550Pro, rs111924994), c.2093G > C (p.Gly698Ala), and c.2149G > C (p.Ala717Pro, rs11396765) as well as c.1456C > T (p.Pro486Ser, rs112545306) previously reported in two individuals suffering from PD [3] (Fig. 1). Similar to the phenotype described [1], all individuals presented with classic PD with an age of onset at 64.5±5.5 years and positive response to dopaminergic therapy. Where available, family history was negative (Table 1).

Overall, the identified variants were very rare in our population. Four c.47C > T (p.Pro16Leu), c.953C > T (p.Thr318Ile), c.1622T > G (p.Val541Gly), and c.2093G > C (p.Gly698Ala) were validated in the PD individual in whom they were first identified but were not found in any additional PD subjects. Of these, c.953C > T (p.Thr318Ile), c.1622T > G (p.Val541Gly), and c.2093G > C (p.Gly698Ala) were not present in controls, while c.47C > T (p.Pro16Leu) was identified in three controls. Of the previously reported [1] variants, c.1505C > T (p.Ala502Val) and c.3490A > C (p.Ser1164Arg) were not seen in the 1989 individuals assessed. Five cases and three controls, on the other hand, were heterozygous for c.2056G > T (p.Gly686Cys). Surprisingly, the original mutation, c.3614G > A (p.Arg1205His), which had, so far, only been identified in PD cases [1], was only present in three controls. In the original publication, all eight PD probands heterozygous for c. 3614G > A (p.Arg1205His; out of 4,708 cases and 4,576 controls) shared the same minimal haplotype [1]. Genotyping of five haplotype-tagging SNPs and three

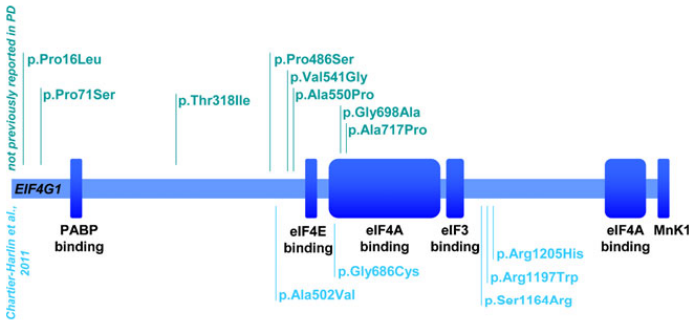


Fig. 1 *EIF4G1* sheme depicting novel and previously described [1] missense variants in individuals with PD and their relative location in relation to known and predicted functional domains. *PABP* polyadenylate binding protein, *eIF* eukaryotic translation initiation factor

Table 1 Rare *EIF4G1* variants identified in individuals with PD, clinical phenotype, and variant frequencies

EIF4G1 variant	Genomic position (hg19)	Exon	Family history	AoO	DD	IS	B	R	RT	PI	L-Dopa/DA	Frequency		Frequency NHLBI-ESP (EA)	Mutation Taster	PolyPhen2
												Cases	Controls			
p.Pro16Leu (c.47C>T)	chr3:184,033,631	1	n/a	n/a	n/a	n/a	+	+	–	++	+	1/975	3/1014	Not found	dc	n/s
p.Pro71Leu (c.211C>T)	chr3:184,035,172	4	n/a	59	4	RT	+	+	+	+	+	n/a	n/a	0.000285	Poly	prob.dam.
p.Ala239Ala (c.717A>G)	chr3:184,039,089	9										1/975	0/1014	0.000142		
p.Thr318Ile (c.953C>T)	chr3:184,039,325	9	neg	62	3	B	+	+	–	+	+	1/975	0/1014	Not found	poly	prob.dam.
p.Pro486Ser (c.1456C>T)	chr3:184,039,828	9	n/a	70	1	B	+	+	+	+	+	n/a	n/a	0.00057	poly	benign
p.Val541Gly (c.1622T>G)	chr3:184,040,345	11	n/a	72	3	TR	+	+	+	+	+	1/975	0/1014	Not found	poly	poss. dam.
p.Ala550Pro (c.1648G>C)	chr3:184,040,371	11	n/a	72	3	RT	+	+	+	+	+	n/a	n/a	0.001994	poly	benign
p.Gly698Ala (c.2093G>C)	chr3:184,041,200	14	neg	65	4	B	+	+	–	+	+	1/975	0/1014	Not found	dc	prob. dam.
p.Ala717Pro (c.2149G>C)	chr3:184,041,256	14	neg	59	9	B	+	+	–	+	+	n/a	n/a	Not found	poly	benign
p.Pro992Pro (c.2971A>G)	chr3:184,043,282	20										14/975	10/1014	0.00616		
p.Val141Val (c.4251C>T)	chr3:184,049,143	29										n/a	n/a	0.0245		
p.Ala1517Ala (c.4551C>T)	chr3:184,049,807	32										n/a	n/a	0.00818		
Chartier_A502V		10										0/975	0/1014	0.000285		
Chartier_G686C		14										5/975	3/1014	Not found		
Chartier_S1164R		24										0/975	0/1014	No found		
Chartier_R1197W		24										n/a	n/a	Not found		
Chartier_R1205H		24										0/975	3/1014	Not found		

Four nonsynonymous and two synonymous variants present in our sample in addition to the *EIF4G1* variants previously identified in familial PD [1] were genotyped in 975 cases and 1,014 controls. Additional clinical information and in silico predictions of the damaging potential of the amino acid exchange as assessed by MutationTaster [5] and PolyPhen2 [6] are presented for the newly identified missense variants. Additionally, variant frequencies as found in the approximately 3,500 European American exomes found in the NHLBI exome sequencing project (NHLBI-ESP) are noted for all newly identified and previously reported [1, 3, 4]

n/a Not available, *neg* negative, *AoO* age of onset, *DD* disease duration, *IS* initial symptom, *B* bradykinesia, *R* rigor, *RT* resting tremor, *PI* postural instability, *D* dementia, *DA* dopamine agonist, *dc* disease causing, *poly* polymorphism, *n/s* not scored, *prob. dam.* probably damaging, *poss. dam.* possibly damaging

microsatellite markers indicated that two of our three c.3614G > A (p.Arg1205His) controls could share this minimal haplotype (Table 2). Seven out of the 12 variants identified in our PD cohort were also found in the approximately 3,500 European American exomes pertaining to the NHLBI exome sequencing project [4] (Table 1).

Discussion

Of the newly identified variants, c.47C > T (p.Pro16Leu), c.211C > T (p.Pro71Ser), c.953C > T (p.Thr318Ile), c.1622T > G (p.Val541Gly), and c.2093G > C (p.Gly698Ala) are predicted to damage protein structure, while c.1456C > T (p.Pro486Ser) and c.2149G > C (p.Ala717Pro) are likely functionally neutral [5, 6] (Table 1). Of these, c.2093G > C (p.Gly698Ala) emerges as the best potentially pathogenic candidate. Contrary to most other amino acids affected, the glycine in position 698 is conserved in all vertebrates. The variant, moreover, was ranked most likely to be damaging by two prediction algorithms [5, 6] and is located in the eIF3/eIF4A binding domain necessary for formation of the translation initiation complex (Fig. 1). However, caution is mandated as a nearby variant (c.2056G > T (p.Gly686Cys)), previously only found in two individuals with PD [1], was present in five cases and three controls in our much smaller sample, suggesting that population-specific effects can misconstrue frequency assessment especially with regard to rare genetic variation. Consequently, further assessment of the role of *EIF4G1* variants in PD is warranted.

Haplotype analysis in the three control subjects harboring c. 3614G > A (p.Arg1205His) supports the idea of an ancestral founder mutation. Linkage analysis and segregation in the original family [1] back pathogenicity of this variant and this is not necessarily disparaged by the presence of the variant in our controls. First, we used general population controls and it is not unlikely that some of the controls may have or may develop PD. Second, it is possible that this mutation shows incomplete penetrance or that other protective factors exist. However, the presence of c. 3614G > A (p.Arg1205His) in our control cohort could also indicate that its role in PD pathogenesis is questionable as has just now also been suggested for the *EIF4G1* p.Ala502Val variant initially also reported by Chartier-Harlin et al. [1, 3]. Overall, the *EIF4G1* locus naturally holds a lot of genetic variance [1, 3]. Accordingly, much larger case–control samples than those used in either the original [1], a follow-up [3], or our study will be necessary to answer this question.

Although not common, it still cannot be excluded that rare exonic *EIF4G1* variants of strong effect could play a causative role in PD in rare cases. And their study is important as they can provide significant clues in understanding disease mechanism. This idea is supported by the fact that *LRRK2*, which harbors both rare and common genetic variation contributing to PD development [7, 8], has recently also been implicated in dysfunctional mRNA translation initiation [9].

Table 2 Haplotype of *EIF4G1* p.Arg1205His carriers

Marker ID	KORA_315	KORA_330	KORA_944
D3S3609	163/179	163/167	163/165
rs4912537	***T***	***T*****/***C*	***T*****/***C*
rs2178403	***G***	***G*****/***A*	***G***
rs2293605	***T*****/***C*	***T*****/***C*	***T*****/***C*
p.Arg1205His	***A*****/***G*	***A*****/***G*	***A*****/***G*
rs1879244	***T/T***	***T*****/***C*	***T***
rs2230571	***C***	***C***	***C/T***
D3S3578	*240/240*	***230*****/***240*	***230*****/***240*
D3S3583	262/**272**	268/270	268/270

Since phase is unknown for all three individuals, where necessary, both alleles are given (with the one pertaining to the described haplotype in bold). Variants comprising the reported minimal haplotype [1] are in italics

Accession numbers

NCBI accessions NM_198241.2 and NP_937884.1 were used to number all variants within the *EIF4G1* gene and eIF4G1 protein. Functional domains were assessed using UniProtKB/Swiss-Prot Q04637 (accessed January 24, 2012).

Acknowledgments We are very grateful to Jelena Golic, Susanne Lindhof, Katja Junghans, Regina Feldmann, and Sybille Frischholz for expert technical assistance.

Competing interests The authors declare that they have no conflict of interest with regard to the above study. Full financial disclosures are listed below.

Dr. Schulte received a postdoctoral fellowship from Technische Universität München, Munich, Germany. Dr. Mollenhauer has received speaker honoraria from Orion Corporation and GlaxoSmithKline; serves as an Associate Editor for the *Journal of Alzheimer Disease*; holds or has pending patents re: Method of differentially diagnosing dementias; Novel ELISA-based quantification of alpha-synuclein proteins in cerebrospinal fluid and peripheral blood products using 384-well plates; and MicroRNA expression profiling of cerebrospinal fluid; serves as a consultant for Bayer Schering Pharma AG; and receives research support from Teva Pharmaceutical Industries Ltd., Desitin Pharmaceuticals, GmbH, Boehringer Ingelheim, GE Healthcare, the Michael J. Fox Foundation for Parkinson's Research, the American Parkinson's Disease Association, and the Stifterverband für die Deutsche Wissenschaft (Dr. Werner Jackstädt-Stipend). Dr. Zimprich reports no disclosures. Dr. Bereznai receives research support from the Hungarian National Innovation Office (TÁMOP-4-2-1/B-03/1/KMR-2010-001). Dr. Lichtner reports no disclosures. Dr. Haubenberger received an NINDS Intramural Competitive Fellowship and research report from the Austrian Science Fund (Erwin Schroedinger Fellowship, project# J2783-B09) and the NINDS Intramural Research Program. Dr. Pirker has received speaker honoraria and travel compensation from Boehringer Ingelheim, Novartis, Abbott Pharmaceuticals, Medtronic, and UCB. Dr.

Brücke has received honoraria for lecturing and travel compensation from CSC, USB, Boehringer Ingelheim, Novartis, Aventis, GE Healthcare, Lundbeck, Merz, GlaxoSmithKline, and Pfizer. Dr. Molnar serves/has served on scientific advisory boards for Genzyme Europe B.V., received speaker honoria from Roche, serves as the Editor-in-Chief of the Hungarian edition of *Neurology*, and receives research support from the Hungarian National Innovation Office (TÁMOP-4-2-1/B-03/1/KMR-2010-001). Dr. Peters and Dr. Gieger report no disclosures. Dr. Trenkwalder serves on scientific advisory boards for Boehringer Ingelheim, Cephalon, Inc., UCB, Novartis, Mundipharma International Limited, and Solvay Pharmaceuticals, Inc.; has received speaker honoraria from Boehringer Ingelheim, Cephalon, Inc., UCB, Novartis, Pfizer Inc, and GlaxoSmithKline; serves on the editorial boards of *Sleep Medicine* and *Movement Disorders* and as an Associate Editor for *Focus on Parkinson Disease*. Dr. Winkelmann serves on a scientific advisory board for UCB; has received speaker honoraria from UCB and Boehringer Ingelheim; has filed a patent re: Winkelmann et al. Nat Genet 2007; and receives research support from the German RLS foundation, the Deutsche Forschungsgemeinschaft (DFG) and the Fritz Thyssen Foundation.

Funding The study depicted herein was funded by institutional funding from Helmholtz Zentrum München, Munich, Germany. Recruitment of case and control cohorts was supported by institutional (Helmholtz Zentrum München, Munich, Germany) and government funding from the German Bundesministerium für Bildung und Forschung (03.2007-02.2011 FKZ 01ET0713), and the Hungarian National Innovation Office (TAMOP-4-2-1/B-03/1/KMR-2010-001).

References

1. Chartier-Harlin MC, Dachsel JC, Vilariño-Güell C et al (2011) Translation initiator EIF4G1 in familial Parkinson disease. Am J Hum Genet 89:398–406
2. Wichmann HE, Gieger C, Illig T (2005) KORA-gen—resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen 67(Suppl 1):26–30
3. Tucci A, Charlesworth G, Sheerin UM et al (2012) Study of the genetic variability in a Parkinson's disease gene: EIF4G1. Neurosci Lett (in press)
4. Exome Variant Server, NHLBI Exome Sequencing Project (ESP), Seattle, WA. http://evs.gs.washington.edu/EVS/. Accessed 14 May 2012.
5. Schwarz JM, Rödelsperger C, Schuelke M et al (2010) MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 8:575–576
6. Adzhubei IA, Schmidt S, Peshkin L et al (2010) A method and server for predicting damaging missense mutations. Nat Methods 7:248–249
7. Satake W, Nakabayashi Y, Mizuta I et al (2009) Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson`s disease. Nat Genet 41:1303–1307
8. Zimprich A, Biskup S, Leitner P et al (2004) Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron 44:601–607
9. Gehrke S, Imai Y, Sokol N, Lu B (2010) Pathogenic LRRK2 negatively regulates microRNA-mediated translational repression. Nature 466:637–641

2.7 Schulte *et al.*, Rare Variants in *β-Amyloid Precursor Protein (APP)* and Parkinson`s Disease, *in preparation*[VII]

Personal contributions: I participated in designing the study, performed the LightScanner® high-resolution melting curve analysis of *APP*, *PSEN1*, *PSEN2*, *FUS*, *GRN*, *TARDBP* and *MAPT* as well as follow-up Sanger sequencing and analyzed the LightScanner® data. I designed the multiplex PCRs for Sequenom®-based genotyping, analyzed genotyping data, performed *in silico* predictions of variant pathogenicity and participated in fragment analysis used for frequency assessment of small deletions in *APP*. Moreover, I analyzed the CSF data and performed genotype-phenotype correlations. I wrote the manuscript and designed all tables and figures except for Figures 1B and 2B.

Rare Variants in *β-Amyloid Precursor Protein (APP)* and Parkinson's Disease

Eva C. Schulte, MD[1,2]; Akio Fukumori, MD, PhD[3,4]; Brit Mollenhauer, MD[5,6]; Hyun Hor, PhD[7]; Thomas Arzberger, MD[8]; Robert Perneczky, MD[9]; Alexander Kurz, MD[9]; Michael Hüll, MD[10]; Peter Lichtner, PhD[2,11]; Gertrud Eckstein, PhD[2]; Alexander Zimprich, MD[12]; Dietrich Haubenberger, MD[12]; Walter Pirker, MD[12]; Thomas Brücke, MD[13]; Benjamin Bereznai, MD, PhD[14]; Maria J. Molnar, MD, PhD[14]; Oswaldo Lorenzo-Betancor, MD[15,16,17]; Pau Pastor, MD,PhD[15,16,17]; Annette Peters, PhD[18]; Christian Gieger, PhD[19]; Xavier Estivill, PhD[7]; Thomas Meitinger, MD[2,11]; Hans A. Kretzschmar, MD[10]; Claudia Trenkwalder, MD[5,20]; Christian Haass, PhD[3,4]; Juliane Winkelmann, MD[1,2,11,]*

[1] Klinik und Poliklinik für Neurologie, Klinikum rechts der Isar, Technische Universität München, Munich, Germany
[2] Institut für Humangenetik, Helmholtz Zentrum München, Munich, Germany
[3] Adolf-Butenandt-Institut, Department of Biochemistry, Ludwig-Maximillians Universität München, Munich, Germany
[4] German Center for Neurodegenerative Diseases (DZNE), Munich, Germany
[5] Paracelsus Elena Klinik, Kassel, Germany
[6] Neurochirurgische Klinik, Georg August Universität Göttingen, Göttingen, Germany
[7] Genes and Disease Program, Centre for Genomic Regulation and UPF and Centro de Investigación Biomédica en Red en Epidemiología y Salud Pública, Barcelona, Catalonia, Spain
[8] Institut für Neuropathologie, Ludwig-Maximillians Universität München, Munich, Germany
[9] Psychiatrische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität München, Munich, Germany
[10] Psychiatrische Universitätsklinik, Albert Ludwigs Universität, Freiburg, Germany
[11] Institut für Humangenetik, Technische Universität München, Munich, Germany
[12] Department of Neurology, Medical University of Vienna, Vienna, Austria
[13] Neurologische Klinik, Wilhelminenspital, Vienna, Austria
[14] Center for Molecular Neurology, Department of Neurology, Semmelweis University, Budapest, Hungary
[15] Neurogenetics Laboratory, Division of Neurosciences, Center for Applied Medical Research, University of Navarra, Pamplona, Spain
[16] Department of Neurology, Clinica Universidad de Navarra, University of Navarra School of Medicine, Pamplona, Spain
[17] CIBERNED, Centro de Investigacion Biomedica en Red en Enfermedades Neurodegenerativas, Instituto de Salud Carlos III, Spain
[18] Institut für Epidemiologie II, Helmholtz Zentrum München, Munich, Germany
[19] Institut für Genetische Epidemiologie, Helmholtz Zentrum München, Munich, Germany
[20] Neurologische Klinik, Georg August Universität Göttingen, Göttingen, Germany

*corresponding author:
Prof. Dr. Juliane Winkelmann, Klinik und Poliklinik für Neurologie, Klinikum rechts der Isar, Technische Universität München, Ismaningerstr. 22, 81675 München, Germany
Phone: +49 89 4140 4688, Fax: +49 89 4140 7681, Email: winkelmann@lrz.tum.de

Character count title: 74
Character count running head:
Word count abstract: 104
Word count text (manuscript body): 1764
Figures: 2
Color figures: 2
Tables: 1
References: 18
Supplementary figures: 1
Supplementary tables: 1
Supplementary references: 18

ABSTRACT

Many individuals with Parkinson`s disease (PD) develop cognitive deficits and phenotypic and molecular overlap between neurodegenerative conditions exists. We investigated the contribution of rare variants in 7 dementia genes (*APP*, *PSEN1/2*, *MAPT*, *FUS*, *GRN*, *TDP-43*) to PD and PD plus dementia (PD+D). PD+D cases harbored more rare variants across all 7 genes than PD individuals and rare variants in *APP* were more common in PD cases than either Alzheimer`s disease cases or controls. CSF parameters differed for *MAPT* and *APP* variants and one variant altered the Aβ spectrum. Our data suggest a possible role for *APP* in bringing about or modifying the PD phenotype.

INTRODUCTION

Linkage analyses as well as genome-wide association and exome sequencing studies have uncovered at least 20 genes associated with idiopathic Parkinson's disease (PD). Still, to date, the identified genes only explain a small portion of the genetic burden in PD. It is likely that genetic factors involved in bringing about the PD phenotype comprise both genetic variants of strong effect, which alone are causative, as well as variants of weaker effect, which contribute to disease risk or phenotypic modification.

Significant overlap between different neurodegenerative diseases has been described on the neuropathologic, the genetic and the phenotypic level. Neuropathologically, the neurodegenerative overlap is exemplified by the coexistence of hallmark features of both Alzheimer`s disease (AD) and PD in individuals with dementia with Lewy bodies.[1] On the genetic level, common genetic variants in myelin-associated protein tau (*MAPT*) represent risk factors for PD[2] while, at the same time, rare variants of strong effect in *MAPT* have long been recognized as a cause of frontotemporal dementia (FTD)[3]. Phenotypically, it is known that at least 30 % of individuals with PD develop dementia[4]. Accordingly, we sought to assess the contribution of genetic factors known to be involved in dementias such as AD[5-8] or FTD[3,9-11] to the PD phenotype.

METHODS

Participants, Variant Screening and Genotyping

We used Idaho®'s LightScanner melting curve analysis to screen the coding regions and exon-intron boundaries of β-amyloid precursor protein (*APP*), presenilin 1 and 2 (*PSEN1*, *PSEN2*), tau (*MAPT*), TAR DNA binding protein 43 (*TDP-43*), granulin (*GRN*) and fused in sarcoma (*FUS*) in 376 individuals with PD (188 with solely PD, 188 with PD plus dementia and 376 KORA-AGE controls (*APP* and *MAPT* only). In the case of altered melting patterns suggestive of variants,

Sanger sequencing ensued. Variants indentified during the screening phase were genotyped in 975 PD cases, 93 neuropathologically confirmed cases of Lewy body disease, 613 AD, 182 FTD cases and 1014 controls using Sequenom® MALDI-TOF mass spectrometry. For technical reasons, *MAPT* p.R546H and *PSEN2* p.R71W were not included. Two 3-base pair (bp) deletions in *APP* were assessed by fragment analysis as described previously[12]. One variant (*APP* p.E599K) which showed significant association in the first sample was also assessed in a second sample of 715 PD cases and 948 controls. Significance was judged using the χ^2 test. Ethics review board approval and participants' written informed consent were obtained. For a detailed description, please cf. supplement.

Aß40, Aß42, Tau and phospho-Tau Determination in Cerebrospinal Fluid (CSF)

Total tau, phospho-tau, β-amyloid (Aβ)1–40 and Aβ1–42 in CSF were measured by commercially available validated ELISAs and protein detection kits as described in the supplement.

Immunohistochemistry

Cortical and midbrain sections of the individual harboring the *APP* p.E599K variant were stained for Aβ and alpha-synuclein. Staining procedure and antibodies can be found in the supplement.

Cloning, Transfections and Analysis of Aß-Spectrum

cDNA of the pCMV/betaAPP695sw vector containing all identified *APP* variants were transiently transfected into HEK293 cells and the Aβ spectrum was analyzed by mass spectrometry as depicted in the supplement.

RESULTS

Variant screening of "dementia genes" in individuals with PD

Within the coding regions and exon-intron boundaries (±10 bp) of *APP*, *PSEN1*, *PSEN2*, *MAPT*, *FUS*, *TDP-43* and *GRN*, we identified a total of 27 rare variants with minor allele frequency (MAF) < 5 % in 376 individuals with PD (n=188) or PD+D (n=188). Interestingly, more individuals with PD+D (10.11%) than solely PD (4.26%) harbored a rare variant with MAF < 5 % in any of the seven "dementia genes" (19 PD+D individuals with a variant vs. 8 PD individuals with a variant; $p<0.05$, χ^2 test). Four individuals harbored the *GRN* p.R433W (rs63750412) variant and one *GRN* p.G35R. One novel variant in *PSEN1* (p.I148W) within two amino acids of known pathogenic mutations as well as three previously reported variants in *PSEN2* (p.R62H (rs58973334), p.R71W (rs140501902), p.S130L (rs63750197)) were found. No variants were identified in either *TDP-43* or

FUS. 9 were also found by the NHLBI-GO exome sequencing project[13]. (Tab 1) For a detailed discussion of the phenotype of variant carriers, please cf. supplement. (Suppl Tab 2)
For *APP* and *MAPT*, the screening was performed in the above 376 PD cases and 376 KORA-AGE controls. In *APP*, 11 rare variants with MAF < 5 % (7 missense, two 3-bp deletions, 2 nearsplice variants) were seen. In total, 10 cases but only 4 controls carried a rare *APP* variant. None of these variants have previously been reported in individuals with a neurodegenerative condition. In *MAPT*, we identified a total of 10 rare variants (9 missense, 1 stop). Overall, 7 cases and 5 controls harbored a rare *MAPT* variant. (Tab 1, Suppl Tab 1). As is frequently the case, analysis by common prediction algorithms yielded contradicting results for most variants (Tab 1), thus warranting additional frequency assessment and functional study.

Frequency Assessment in Individuals with PD, AD and FTD

Frequency assessment for 25 of the 27 variants identified in the screening phase was carried out in a sample consisting of 975 PD patients (including the 376 used above), 613 AD patients, 182 FTD patients, 93 neuropathologically confirmed cases of Lewy body disease and 1014 controls (also including the 376 used above). 68.0% of the variants were very rare with MAF < 0.1 % in the control sample. When compared to controls, the *APP* p.E599K variant was significantly more frequent in the PD phenotype than in controls ($p<0.01$, χ^2 test) prior to correction for multiple testing. When data from the NHLBI-ESP exomes were added as controls, the finding remained significant even after correction for multiple testing (14 out of 1068 cases vs. 12 out of 5310 controls; $p_{nominal}=3.8x10^{-7}$, $p_{corrected}=9.5x10^{-6}$, χ^2 test). However, when trying to replicate this finding in a Spanish PD case/control sample, we did not find any *APP* p.E599K carriers in either cases or controls, possibly suggesting a population-specific finding. *APP* p.E599K was the only variant identified in the 93 Lewy body disease cases. Clinically, this individual had suffered from classical, levodopa-responsive PD with an age of onset at 60 years. Her mother had also had PD. Histology revealed both Lewy bodies in the substantia nigra and some amyloid plaques in the frontal and parietal cortex and the hippocampus, in line with a diagnosis of idiopathic PD.(Fig 1)
Burden tests analyzing the load of rare variants were performed for both *APP* and *MAPT*. This revealed an access of rare variants with MAF < 5 % in *APP* in PD (27 individuals with a variant out of 975) when compared to either controls alone (13 out of 1014, $p<0.02$, χ^2 test), AD cases alone (4 out of 613, $p<0.01$, χ^2 test) or the combined sample of controls, AD and FTD cases ($p<0.001$, χ^2 test). The frequency of rare variants in *MAPT* was similar is all groups.

Rare Variants in *MAPT* and *APP* Influence CSF Levels of Aβ and Tau

In order to evaluate a possible functional effect of the identified variants, we analyzed neurodegeneration markers in the CSF of the PD individuals in whom the variants were identified. Interestingly, comparison between those carrying rare *MAPT* variants (n=4) and those carrying rare *APP* variants (n=5) uncovered comparatively elevated total tau (311.2±122.1 vs. 114.5±44.1 pg/ml; $p=1.0x10^{-4}$) and pathologically elevated (normal <75 pg/ml) mean phosphorylated tau (78.3±23.9 vs. 40.6±5.0 pg/ml; p=0.005) levels in those with *MAPT* variants. Carriers of rare *APP* variants, on the other hand, showed a comparative increase in Aβ1-42 load (982.2±170.0 vs. 704.0±65.1 pg/ml p=0.009) and Aβ42/40 ratio (1.32±0.18 vs. 0.68±0.19, $p=5.9x10^{-4}$) and attenuation of Aβ1-40 levels (7754.0±1585.0 vs. 10967.3±3047.2 pg/ml; p=0.03, all Student`s t-test) when compared to *MAPT* variant carriers.(Fig 2) There was no difference in the four CSF parameters when carriers of rare *APP* or *MAPT* variants were compared to PD individuals of the screening sample without a rare variant (n=333) or between the 188 individuals with PD and PD+D. (data not shown)

Impact of Rare Variants in *APP* on Aβ Processing

Aβ spectral analysis was performed to further evaluate a potential functional effect of the identified coding variants in *APP*. In all but one, the Aβ spectrum reflected the wildtype situation. However, *APP* p.G702S, located within the Aβ domain, shifted the spectrum from Aβ40 as the main species to Aβ39 and—to a lesser extend—Aβ37 (Fig 2, Suppl Fig 1).

DISCUSSION

Screening of seven genes known to be strong genetic factors in AD or FTD in a sample comprising both individuals with PD and PD+D revealed a number of rare variants not previously described. Interestingly, identified variants in *APP* were more common in PD than in either controls or AD. Next to a mere chance occurrence, there are several possible explanations for this finding. For one, rare variants in known dementia genes could represent phenotype modifiers in PD. This is supported by the fact that in the screening sample, rare variants were more frequent in the PD+D group than in the PD group when all seven genes were analyzed together. Also, the "dementia gene" variants could contribute to the overall "neurodegenerative burden" that an individual carries which reflects an increased susceptibility for neurodegenerative conditions in general. In this scenario, an access of genetic alterations in a specific pathway plus additional non-genetic factors could then tip the balance towards one neurodegenerative phenotype or the other or create phenotypes in which features of multiple neurodegenerative diseases coexist. Alternatively, this

could also mean that the phenotypic spectrum of AD or FTD is broader than previously recognized and could include PD-like aspects.

The boldest proposal would be that rare variants of strong effect in *APP* or *MAPT* alone could cause PD. *Mapt*$^{-/-}$ mice have recently been shown to develop not only memory deficits but also PD-specific features such as loss of neurons in the substantia nigra and reduced locomotion.[14] Common variants in *MAPT* are an established risk factor for PD[2] and the relevance of allelic series—that is both common variants of weak effect and rare variants of strong effect in one gene—to PD has already be shown[15]. Yet, in our sample, rare variants in *APP*, not *MAPT*, were enriched in PD. However, since a physical interaction between MAPT and APP and a role of MAPT in trafficking APP to the cell membrane has been reported[14,16], rare variants in APP could have a similar effect with regard to PD as MAPT variants.

Based to the CSF findings and the Aβ spectra, some of the variants we identified appear to harbor functional effects on the protein. Yet, from our data we cannot conclude that this function is truly relevant to PD nor which function of the proteins would be relevant to a potential (modifying) role in PD. Next to its role in amyloid production, APP`s ferroxidase[17] and amine oxidase[18] activities could even more plausibly fit a potential role in PD pathogenesis or phenotype modification and could be explored further.

ACKNOWLEDGEMENTS

We are gratefully indebted to Katja Junghans, Susanne Lindhof, Jelena Golic, Sybille Frischholz und Regina Feldmann at the Institut für Humangenetik, Helmholtz Zentrum München, Munich, Germany, for their expert technical assistance in performing Sequenom genotyping and fragment analyses. This study was funded exclusively by in-house institutional funding from the Technische Universität München and the Helmholtz Zentrum München, Munich, Germany.

AUTHOR CONTRIBUTIONS

Eva C. Schulte-study design, figures, literature search, data collection, data analysis, data interpretation, manuscript writing
Akio Fukumori-figures, data collection, data analysis, data interpretation, manuscript writing
Brit Mollenhauer-data collection, data analysis, data interpretation, manuscript writing
Hyun Hor-data collection, data analysis, manuscript writing
Thomas Arzberger-figure, data collection, data analysis, data interpretation, manuscript writing
Robert Perneczky-data collection, data analysis, data interpretation, manuscript writing
Alexander Kurz-data collection, data analysis, data interpretation, manuscript writing
Michael Hüll-data collection, data analysis, data interpretation, manuscript writing
Peter Lichtner-data collection, data analysis, data interpretation
Gertrud Eckstein-data collection, data analysis, data interpretation

Alexander Zimprich-data collection
Dietrich Haubenberger-data collection
Walter Pirker-data collection
Thomas Brücke-data collection
Benjamin Bereznai-data collection
Maria J. Molnar-data collection
Oswaldo Lorenzo-Betancor-data collection
Pau Pastor-data collection
Annette Peters-data collection
Christian Gieger-data collection
Xavier Estivill-data collection, data analysis, data interpretation
Thomas Meitinger-data interpretation
Hans A. Kretzschmar-data collection, data analysis, data interpretation
Claudia Trenkwalder-data collection, data analysis, data interpretation
Christian Haass-data collection, data analysis, data interpretation
Juliane Winkelmann-study design, data collection, data analysis, data interpretation, manuscript writing

REFERENCES

1. Deramecourt V, Bombois S, Maurage CA, Ghestem A, et al. Biochemical staging of synucleinopathy and amyloid deposition in dementia with Lewy bodies. J Neuropathol Exp Neurol 2006;65:278-288.

2. Simon-Sanchez J, Schulte C, Bras JM, Sharma M, et al. Genome-wide association study reveals genetic risk underlying Parkinson`s disease. Nat Genet 2009;41:1308-1312.

3. Hutton M, Lendon CL, Rizzu P, Baker M, et al. Association of missense and 5`-splice-site mutations with the inherited dementia FTDP-17. Nature 1998;393:702-705.

4. Aarsland D, Andersen K, Larsen JP, Lokl A, et al. Prevalence and characteristics of dementia in Parkinson disease: an 8-year prospective study. Arch Neurol 2003;60:387-392.

5. Goate A, Chartier-Harlin MC, Mullan M, Brown J, et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature 1991;349: 704–706.

6. Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature 1995;375: 754–760.

7. Rogaev EI, Sherrington R, Rogaeva EA, Levesque G, et al. Familial Alzheimer`s disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer`s disease type 3 gene. Nature 1995;376:775-778.

8. Levy-Lahad E, Wasco W, Poorkaj P, Romano DM, et al. Candidate gene for the chromosome 1 familial Alzheimer's disease locus. Science 1995;269:973-977.

9. Cruts M, Gijselinck I, Van Der Zee J, Engelborghs S, et al. Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. Nature 2006;442: 920-925.

10. Neumann M, Rademakers R, Roeber S, Baker M, et al. A new subtype of frontotemporal lobar degeneration with FUS pathology. Brain 2009;132: 2922–2931

11. Neumann M, Sampathu, DM, Kwong, LK, Truax, AC, et al. Ubiquitinated TDP-43 in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis". Science 2006;314:130–133.

12. Winkelmann J, Czamara D, Schormair B, Knauf F, et al. Genome-wide association study identifies novel restless legs syndrome susceptibility loci on 2p14 and 16q12.1. PLOS Genet 2011;7:e1002171.

13. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/) [accessed December 1, 2012]

14. Lei P, Ayton S, Finkelstein DI, Spoerri L, et al. Tau deficiency induces parkinsonism with dementia by impairing APP-medicated iron export. Nat Med 2012;18:291-295.

15. Ross OA, Soto-Ortolaza AI, Heckman MG, Aasly JO, et al. Association of LRRK2 exonic variants with susceptibility to Parkinson`s disease: a case-control study. Lancet Neurol 2011;10:898-908.

16. Guo JP, Arai T, Miklossy J, McGeer PL. Aβ and tau form soluble complexes that may promote self aggregation of both into the insoluble forms observed in Alzheimer`s disease. Proc Natl Acad Sci 2006;103:1953-1958.

17. Duce JA, Tsatsanis A, Cater MA, James SA, et al. Iron-export ferroxidase activity of β-amyloid precursor protein is inhibited by zinc in Alzheimer`s disease. Cell 2010;142:857-867.

18. Duce JA, Ayton S, Miller AA, Tsatsanis A, et al. Amine oxidase activity of β-amyloid precursor protein modulates systemic and local catecholamine levels. Mol Psychiatry 2012; doi: 10.1038, Epub ahead of print Jan 3, 2012.

TABLES

Gene	Genomic position (hg19)	dbSNP137	Variant	Cases		Controls	Frequency Assessment (975PD:613AD:182FTD:1014KORA)	NHLBI-ESP[1] (EA only)	Predictions		
				PD (n=188)	PD+D (n=188)	(n=376)			PolyPhen2	Mutation Taster	SIFT
APP	chr21:27,423,376, C>T	rs149995576	A201V		1		2:0:0:1	A=3/G=8597			N/S
APP	chr21:27,394,297-299, delCTT	N/A	c.722_724delCTT		1		4:1:0:4	not found	n/a		n/a
APP	chr21:27,354,793, A>G	N/A	c.1091-3	1			1:0:0:0	not found	n/a		n/a
APP	chr21:27,347,438, C>T	N/A	R468H			1	0:0:0:1	not found			
APP	chr21:27,328,030, C>T	rs201547994	A500T			1	0:0:0:2	not found			
APP	chr21:27,284,167, G>A	rs140304729	E599K	1	2	1	13:3:0:3	T=9/C=8591			N/S
APP	chr21:27,284,163, G>A	rs200088099	T600M			1	0:0:0:1	not found			
APP	chr21:27,269,955-957,delTCC	N/A	c.1992_1994delGGA		1		4:0:1:0	not found	n/a		n/a
APP	chr21:27,269,961, G>A	rs200260102	T663M		1		1:0:0:0	not found			
APP	chr21:27,264,120, C>T	rs201269325	G709S		1		1:0:0:0	not found			N/S
APP	chr21:27,254,092-093, delAA	N/A	c.2212-10_-11delTT		1		1:0:0:1	not found	n/a		n/a
GRN	chr17:42,426,635, G>A	N/A	G35R		1		1:0:0:0	not found			
GRN	chr17:42,429,500, C>T	rs63750412	R433W	1	3		2:0:0:0 homo &.6:5:2:6 hetero	T=24/C=8576			N/S
MAPT	chr17:44,039,716, C>T	N/A	R5C		1		1:0:0:0	not found			
MAPT	chr17:44,039,717, G>A	rs63750959	R5H			1	0:0:0:1	not found			
MAPT	chr17:44,039,824, G>A	rs115239819	A41T	1			1:0:0:0	not found			
MAPT	chr17:44,067,341, C>T	rs143956882	S427F		1	1	2:4:0:3	T=18/C=8582			
MAPT	chr17:44,067,403, C>T	rs200099007	R448X	1			2:4:0:3	T=2/C=8598	n/a		N/S
MAPT	chr17:44,068,850, G>A	rs143624519	A469T	1	1	2	3:4:0:5	A=23/G=8577			
MAPT	chr17:44,071,314, C>G	N/A	P511R			1	1:0:0:2	not found			
MAPT	chr17:44,073,840, G>A	N/A	R546H			1	n/a	not found			
MAPT	chr17:44,096,064, A>G	N/A	I695V	1			1:0:0:0	not found			
MAPT	chr17:44,101,491, C>T	rs63750991	T762M	1			1:0:0:0	not found			
PSEN1	chr14:73,640,377, A>G	N/A	I148W		1		1:0:0:0	not found			
PSEN2	chr1:227,071,449, G>A	rs58973334	R62H		1		4:3:0:5	A=22/G=8578			N/S
PSEN2	chr1:227,071,475, C>T	rs140501902	R71W		1		n/a	T=32/C=8568			
PSEN2	chr1:227,073,271, C>T	rs63750197	S130L		1		2:2:0:4	T=9/C=8591			

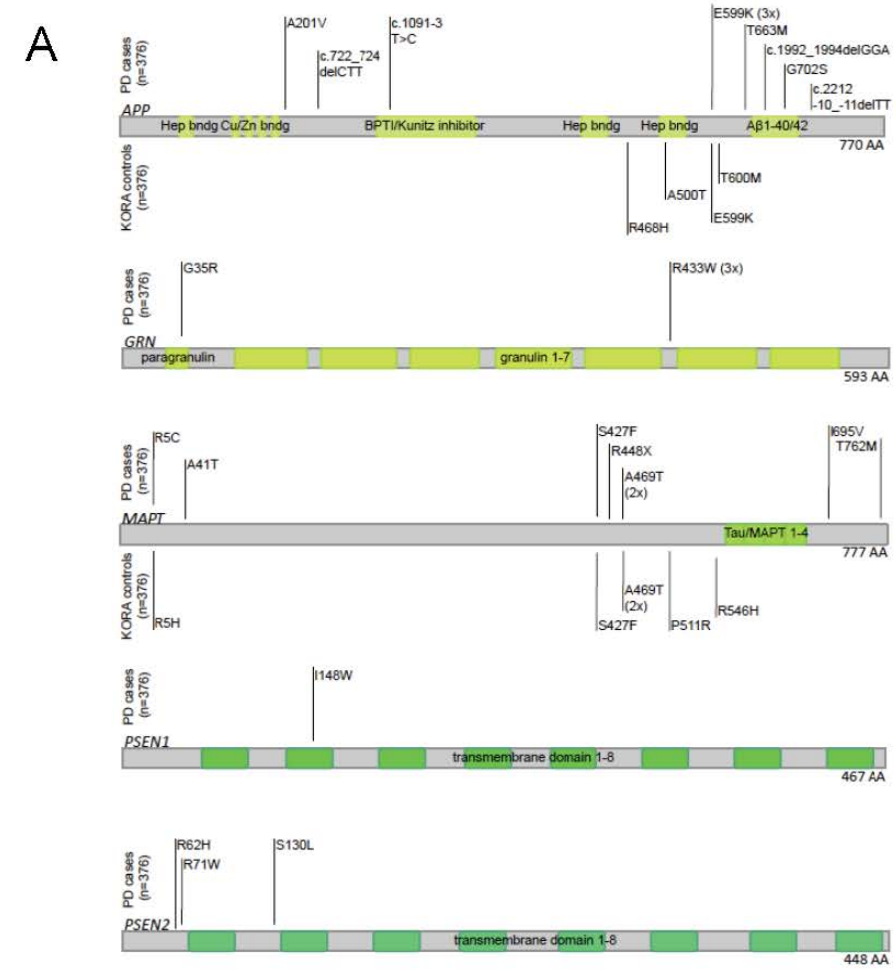
Table 1
Rare variants in "dementia genes" identified in variant screening. Gray boxes in the controls column symbolize that the given gene was not screened in controls. Orange = probably damaging (PolyPhen2), disease causing with a score ≥ 0.9 (Mutation Taster) or damaging (SIFT); yellow = possibly damaging (PolyPhen2), disease causing with a score < 0.9 (Mutation Taster) or damaging with low confidence (SIFT); light green = polymorphism with a score < 0.9 (Mutation Taster); dark green = benign (PolyPhen2), polymorphism with a score ≥ 0.9 (Mutation Taster) or tolerated (SIFT). N/A = not annotated, N/S = not scored, n/a = not available because this class of variants cannot be tested using the given prediction algorithm. [1]Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/) [accessed December 1, 2012]

FIGURES

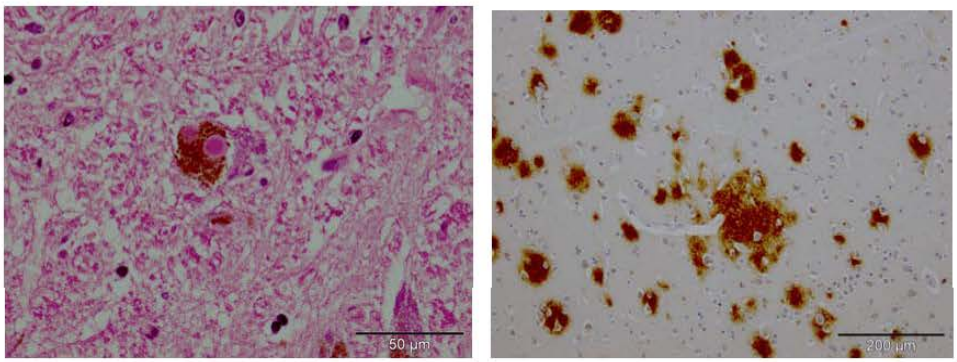
Figure 1

Location of rare variants in *APP*, *GRN*, *MAPT*, *PSEN1* and *PSEN2* and histological features of an individual harboring the p.E599K variant of *APP*. (A) Variants with minor allele frequency (MAF) < 5% found in PD cases are depicted above the schematic illustration of each gene, those found in controls—if the gene was analyzed in controls—below the gene. If variants were present more than once in the discovery sample, the number of occurrences is given in parentheses. Domain annotations were taken from Uniprot (accessed December 12, 2012). PD = Parkinson`s disease, Hep = heparin, AA = amino acids. (B) Depiction of a classical nigral Lewy body (left, antibody: anti-alpha synuclein KM51, 1:1000, Novocastra/NCL-ASYN, counter stain: hematoxylin-eosine) and cortical Aβ plaques (right, antibody: 4G8, 1:2000, Signet) found in an individual with classical idiopathic PD and the *APP* p.E599K variant.

12


A
PD cases (n=376)
A201V
c.722_724 delCTT
c.1091-3 T>C
E599K (3x)
T663M
c.1992_1994delGGA
G702S
c.2212 -10_-11delTT
APP
Hep bndg
Cu/Zn bndg
BPTI/Kunitz inhibitor
Hep bndg
Hep bndg
Aβ1-40/42
770 AA
KORA controls (n=376)
R468H
A500T
T600M
E599K
PD cases (n=376)
G35R
R433W (3x)
GRN
paragranulin
granulin 1-7
593 AA
PD cases (n=376)
R5C
A41T
S427F
R448X
A469T (2x)
I695V
T762M
MAPT
Tau/MAPT 1-4
777 AA
KORA controls (n=376)
R5H
S427F
A469T (2x)
P511R
R546H
PD cases (n=376)
I148W
PSEN1
transmembrane domain 1-8
467 AA
PD cases (n=376)
R62H
R71W
S130L
PSEN2
transmembrane domain 1-8
448 AA

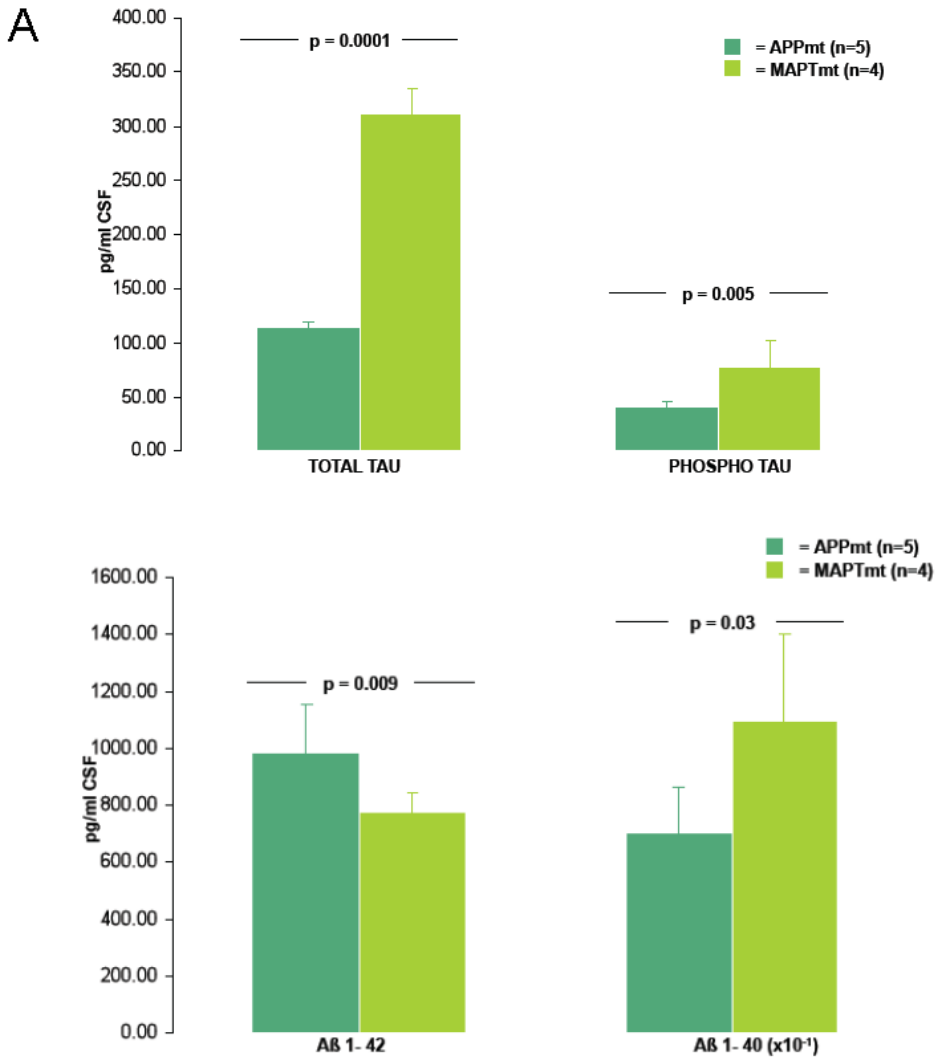
B


50 µm
200 µm

13

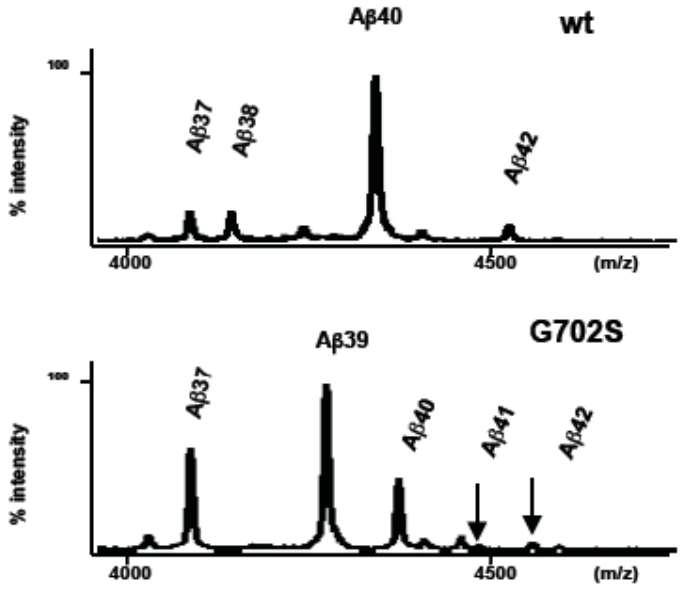
Figure 2

Functional effect of rare *APP* variants. (A) Differences in CSF total tau levels and tau phosphorylation as well as Aβ1-42 and Aβ1-40 levels between PD individuals carrying rare variants in either *APP* or *MAPT*. (B) When overexpressed in HEK293 cells, the *APP* G702S variant shifts the Aβ spectrum from Aβ40 to Aβ39 and Aβ37.



14

B


Aβ40
wt
100
% intensity
Aβ37
Aβ38
Aβ42
4000
4500
(m/z)
Aβ39
G702S
100
% intensity
Aβ37
Aβ40
Aβ41
Aβ42
4000
4500
(m/z)

15

SUPPLEMENTARY MATERIALS

Eva C. Schulte, MD[1,2]; Akio Fukumori, MD, PhD[3,4]; Brit Mollenhauer, MD[5,6]; Hyun Hor, PhD[7]; Thomas Arzberger, MD[8]; Robert Perneczky, MD[9]; Alexander Kurz, MD[9]; Michael Hüll, MD[10]; Peter Lichtner, PhD[2,11]; Gertrud Eckstein, PhD[2]; Alexander Zimprich, MD[12]; Dietrich Haubenberger, MD[12]; Walter Pirker, MD[12]; Thomas Brücke, MD[13]; Benjamin Bereznai, MD, PhD[14]; Maria J. Molnar, MD, PhD[14]; Oswaldo Lorenzo-Betancor, MD[15,16,17]; Pau Pastor, MD,PhD[15,16,17]; Annette Peters, PhD[18]; Christian Gieger, PhD[19]; Xavier Estivill, PhD[7]; Thomas Meitinger, MD[2,11]; Hans A. Kretzschmar, MD[10]; Claudia Trenkwalder, MD[5,20]; Christian Haass, PhD[3,4]; Juliane Winkelmann, MD[1,2,11,]*

[1] Klinik und Poliklinik für Neurologie, Klinikum rechts der Isar, Technische Universität München, Munich, Germany
[2] Institut für Humangenetik, Helmholtz Zentrum München, Munich, Germany
[3] Adolf-Butenandt-Institut, Department of Biochemistry, Ludwig-Maximillians Universität München, Munich, Germany
[4] German Center for Neurodegenerative Diseases (DZNE), Munich, Germany
[5] Paracelsus Elena Klinik, Kassel, Germany
[6] Neurochirurgische Klinik, Georg August Universität Göttingen, Göttingen, Germany
[7] Genes and Disease Program, Centre for Genomic Regulation and UPF and Centro de Investigación Biomédica en Red en Epidemiología y Salud Pública, Barcelona, Catalonia, Spain
[8] Institut für Neuropathologie, Ludwig-Maximillians Universität München, Munich, Germany
[9] Psychiatrische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität München, Munich, Germany
[10] Psychiatrische Universitätsklinik, Albert Ludwigs Universität, Freiburg, Germany
[11] Institut für Humangenetik, Technische Universität München, Munich, Germany
[12] Department of Neurology, Medical University of Vienna, Vienna, Austria
[13] Neurologische Klinik, Wilhelminenspital, Vienna, Austria
[14] Center for Molecular Neurology, Department of Neurology, Semmelweis University, Budapest, Hungary
[15] Neurogenetics Laboratory, Division of Neurosciences, Center for Applied Medical Research, University of Navarra, Pamplona, Spain
[16] Department of Neurology, Clinica Universidad de Navarra, University of Navarra School of Medicine, Pamplona, Spain
[17] CIBERNED, Centro de Investigacion Biomedica en Red en Enfermedades Neurodegenerativas, Instituto de Salud Carlos III, Spain
[18] Institut für Epidemiologie II, Helmholtz Zentrum München, Munich, Germany
[19] Institut für Genetische Epidemiologie, Helmholtz Zentrum München, Munich, Germany
[20] Neurologische Klinik, Georg August Universität Göttingen, Göttingen, Germany

*corresponding author:
Prof. Dr. Juliane Winkelmann, Klinik und Poliklinik für Neurologie, Klinikum rechts der Isar, Technische Universität München, Ismaningerstr. 22, 81675 München, Germany
Phone: +49 89 4140 4688, Fax: +49 89 4140 7681, Email: winkelmann@lrz.tum.de

SUPPLEMENTARY METHODS

Description of case/control samples

All 376 PD cases (71.1±9.4 years, 31.6% female; 188 PD and 188 PD+D) used in variant screening were recruited at Paracelsus-Elena Klinik, a hospital specializing in Parkinson's disease (PD), in Kassel, Germany. Cases used in genotyping included a total of 450 cases from Kassel (age 70.2 ± 9.7yrs, 32.2% female, including the 376 cases using in variant screening), 486 cases (age 58.7 ±11.3yrs, 35.4% female) collected at the Departments of Neurology at Wilhelminenspital and Allgemeines Krankenhaus in Vienna, Austria, and 39 cases (age 50.4 ±10.8yrs, 53.8% female) collected at the Department of Neurology, Semmelweis University, Budapest, Hungary, for a total of 975 cases. The Hungarian cases were included because they are either individuals with a very early age of onset or probands of large PD families. All PD cases used in variant screening and genotyping have been described previously[1]. PD diagnosis was made in accordance with the UK Brain Bank Criteria by a senior neurologist specializing in movement disorders.

The Spanish replication sample consisted of 715 individuals (72.6 ± 11.2 years, 40% female) diagnosed with PD originating from the Spanish regions of Catalonia and Navarra and 948 general population controls (38.6 ± 11.5 years, 55% female).

Samples of 93 neuropathologically diagnosed cases of Lewy body disease were collected by BrainNet Europe and archived at the Department of Neuropathology, Ludwig Maximillians Universität, Munich, Germany.

613 individuals diagnosed with Alzheimer`s disease (AD) were recruited at the department of Psychiatry, Klinkum rechts der Isar, Technische Universität München, Munich, Germany, and were included in the genotyping.

The frontotemporal dementia (FTD) sample included 146 cases from the department of Psychiatry, Klinikum rechts der Isar, Technische Universität München, Munich, Germany, as well as 36 cases from the department of Psychiatry, Albert-Ludwigs Universität, Freiburg, Germany.

The 1014 controls used in genotyping (age 76 ± 6.6yrs, 50.1% female) belong to a large general population cohort (KORA) based in the region around Augsburg in Southern Germany and have been described previously.[1,2] KORA-AGE represents a subset of the KORA cohort collected in 2009 as a gender- and age-stratified subsample of the KORA[3] S1-S4 studies comprising participants born before 1944. All individuals taking dopaminergic medication were excluded from the control sample.

Variant Screening

The following transcripts were used in primer design and variant annotation: *APP*—NM_000484.3, *PSEN1*—000021.3, *PSEN2*—NM_000447.2, *FUS*—NM_004960.3, *GRN*—NM_002087.2, *MAPT*—NM_001123066.3, *TDP-43*—NM_007375.3. Primer sequences are available upon request.

Immunohistochemistry

Cortical and midbrain sections of the individual carrying the *APP* p.E599K variant were rehydrated and pre-treated with 80% formic acid for 1 h. Subsequently, sections were incubated over night at 4°C with a monoclonal primary antibody directed against Aβ (Signet, clone 4G8, dilution 1:2000) or a monoclonal primary antibody directed against alpha-synuclein (Novocastra/NCL-ASYN, clone KM51, dilution 1:1,000). The reaction product was visualized using the Zymed Lab-SA detection system (Zymed, San Fransisco, CA, USA) with the use of Biosource Romulin AEC as chromogen (Biocare medical, Walnut Creek, CA, USA).[4,5]

Determination of Aß40, Aß42, Tau and phosphorylated Concentrations in Cerebrospinal Fluid (CSF)

Commercially available assays from Innogenetics, Ghent, Belgium, were used to quantify total tau protein, phospho-tau 181 protein as well as Aβ1-42.[6] Aβ1-40 was measured by a commercially available assay from ABETA GmbH, Heidelberg, Germany, as described previously[6]. The assays were performed at the CSF diagnostics lab of the Department of Neurology, Georg August Universität, Göttingen, Germany, where they are routinely used in clinical practice.

Cloning and Transfections

All identified coding variants in *APP* were introduced into the pCMV/betaAPP695sw vector[7] by QuikChange mutagenesis (Stratagene) using oligonucleotide primers encoding the respective point mutations. Human embryonic kidney (HEK) 293 cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum and penicillin/streptomycin on poly-L-lysine-coated plates. Cells were plated at a density of 1,000,000 cells/6-well plate, and the following day, cells were transiently transfected with the indicated *APP* cDNAs using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions.

Mass Spectrometry Analysis of Aβ

As described previously[8], Aβ species were immunoprecipitated from conditioned medium from each cell line with antibody 82E1 for 4 h at 4 °C. Immunoprecipitates were washed four times with immunoprecipitation-MS buffer (0.1% N-octylglucoside, 140 mM NaCl, 10 mM Tris, pH 8.0) and two times with distilled water. Immunoprecipitated peptides were eluted with 0.3% trifluoroacetic acid in 40% acetonitrile saturated with a-cyano-4-hydroxy cinnamic acid. The dissolved samples were dried on a stainless plate and subjected to MALDI-TOF MS analysis using Voyager DE STR (Applied Biosystems).

SUPPLEMENTARY RESULTS and DISCUSSION

Clinical phenotypes of variant carriers

Clinical details of variant carriers are depicted in Suppl Table 1. Most individuals suffered from classic idiopathic PD with or without dementia with an age of onset at 64.1±7.9 years and an average disease duration of 9.0±6.0 years. Occuring in 65% of patients with a rare variant (minor allele frequency (MAF) < 5%), unilateral resting tremor was the most common initial symptom. In a few individuals diagnoses other than idiopathic PD had been discussed in the past (Suppl Table 1) because of the presence of additional symptoms such as prominent dementia or vertical gaze palsy. However, in the end, in all cases idiopathic PD was considered the most likely diagnosis by a senior movement disorders specialist. Upon administration of 200mg levodopa, no improvement of symptoms was seen in one carrier of an *APP* and one carrier of a *MAPT* variant and two carriers of *GRN* as well as one carrier of a *PSEN2* variant showed less than 20 % improvement in the motor section score of the United Parkinson`s Disease Rating Scale (UPDRS III). This could be attributed to either advanced disease or an underlying diagnosis other than idiopathic PD. Transcranial sonography of the substantia nigra was performed in six out of 26 cases and demonstrated either uni- or bilateral nigral hyperechogenicity in all cases. Seven out of 26 cases underwent polysomnography. In six, results suggested REM-sleep behavior disorder (RBD) and in three of these, an increase in periodic limb

movements in sleep (PLMSs) was also present suggesting the presence of restless legs syndrome (RLS). One individual suffered from RLS but not RBD. Magnetic resonance imaging (MRI) was unremarkable or showed only unspecific microangiopathic changes in most individuals (9 of 22). However, in 7 out of 22 cases, signal alterations classified by the attending neuroradiologists as likely calcifications or iron depositions were present in the basal ganglia. In four cases, these changes were accompanied by some degree of temporal, hippocampal and/or mesencephalic atrophy (compare Suppl Tab 1). Further, the phenotypes of carriers of some previously described variants are noteworthy in light of what can be currently found in the literature and are described below in more detail.

MAPT p.R5C and p.R5H

One individual with idiopathic PD carried the *MAPT* p.R5C variant. His dizygotic twin brother was reported to have suffered from essential tremor all his life but was not available for detailed clinical examination or genotyping. The *MAPT* p.R5C variant was not previously descripted in the literature. However, two different variants affecting the same codon (*MAPT* p.R5H and p.R5L) are listed in the AD-FTD mutation database (www.molgen.ua.ac.be/ADMutations, accessed Dec 1, 2012) as pathogenic. These were linked to two different neuropathologically confirmed phenotypes—FTD for *MAPT* p.R5H and progressive supranuclear palsy (PSP) for *MAPT* p.R5L—and both impair microtubule assembly in vitro[9,10]. Resting tremor and bradykinesia as the presenting symptoms as well as a relatively long disease course of currently 12 years substantiate idiopathic PD as the primary diagnosis in our patient. Accordingly, one could speculate about a potential role of the p.R5C variant with regard to the PD phenotype. However, this remains a speculation as long as there is not functional or statistical genetic evidence to support it. Interstingly, however, we also identified the p.R5H variant of *MAPT* but in a control individual. We cannot exclude that the control individual belonging to our general population cohort has or could develop a neurodegenerative phenotype, yet, this could also suggest reduced penetrance or variable expressivity of this variant.

GRN p.G35R

The PD individual who harbored the *GRN* p.G35R variant reported the onset of a Parkinson syndrome at the age of 56 with bradykinesia. Resting tremor was never present and levodopa response was limited. Dementia was present at the time of inclusion into the study (MMSE 19/30) and LBD and MSA-P were considered as differential diagnoses during the 18-year course of disease but the final

diagnosis was idiopathic PD. The *GRN* p.G35R variant was previously observed in an individual with late-onset sporadic AD[11] but its pathogenic nature is unclear.

PSEN2 p.R62H

PSEN2 p.R62H was identified in one case of idiopathic PD with good response to levodopa and anosmia. However, a connection with the PD phenotype is questionable because when analyzed by a recently suggested algorithm to be used in the analysis of AD variants, *PSEN2* p.R62H is not pathogenic[12]. This is further supported by the fact that is was found with a MAF of approximately 9% in a cohort of 130 individuals from seven African populations[12] and 22 of approximately 4300 individuals in the NHLBI-ESP exomes[13] and it does not alter the production of Aβ1-42[14].

PSEN2 p.R71W

The individual carrying the *PSEN2* p.R71W variant presented with resting tremor and developed idiopathic PD with only moderate responsiveness to levodopa as well as cognitive impairment (mini mental state exam (MMSE) 24/30). Polysomnography was in line with REM sleep behavior disorder (RBD) and restless legs syndrome (RLS). The *PSEN2* p.R71W variant has previously been reported in one individuals with idiopathic AD[12] and one individual with LBD and his brother who also suffered from dementia[15]. According to the proposed pathogenicity algorithm[12], it was classified as possibly pathogenic and it has been recognized to affect PSEN2 protein stability and its function in Notch signaling[16]. Yet, the fact that ist is also found in 32 out of approximately 4300 individuals examined as part of the NHLBI-ESP[13], makes a causal role unlikely.

PSEN2 p.S130L

An 85-year-old individual in whom symptoms had begun at age 73 with bradykinesia carried the *PSEN2* p.S130L variant. At presentation, the phenotype consisted of bradykinesia, resting tremor, rigor, postural instability, and pronounced dementia (MMSE 10/30). In CSF, Aβ42/40 ratio was within normal limits (0.9). *PSEN2* p.S130L has previously been reported in one family with AD[17] and one individual with sporadic AD[18] but also in two families with dilated cardiomyopathy and heart failure[19]. Interestingly, in the family of the „sporadic“ AD patient with the *PSEN2* p.S130L genotype, the father and sister suffered from PD but were not available for genotyping[18]. *PSEN2* p.S130L does not change Aβ1-42 production[14] and was also found in 9 individuals out of approximately 4300 evaluated as part of the NHLBI-ESP[13].

SUPPLEMENTARY REFERENCES

1. Schulte EC, Mollenhauer B, Zimprich A, Bereznai B, et al. Variants in *eukaryotic translation initiation factor 4G1* in sporadic PD. Neurogenetics 2012;13:281-285.

2. Zimprich A, Benet-Pages A, Struhal W, Graf E, et al. A Mutation in VPS35, Encoding a Subunit of the Retromer Complex, Causes Late-Onset Parkinson Disease. Am J Hum Genet 2011;89:168-175.

3. Wichmann HE, Gieger C, Illig T. KORAgen—Resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen 67 Suppl 1:26-30.

4. Alafuzoff I, Ince PG, Arzberger T, Al-Sarraj S, Bell J, Bodi I, Bogdanovic N, Bugiani O, Ferrer I, et al. Staging/typing of Lewy body related alpha-synuclein pathology: a study of the BrainNet Europe Consortium. Acta Neuropathol 2009;117:635-652.

5. Alafuzoff I, Thal DR, Arzberger T, Bogdanovic N, Al-Sarraj S, Bodi I, Boluda S, Bugiani O, Duyckaerts C, et al. Assessment of beta-amyloid deposits in human brain: a study of the BrainNet Europe Consortium.Acta Neuropathol. 2009;117:309-20.

6. Mollenhauer B, Bibl M, Wiltfang J, Steinacker P, et al. Total tau protein, phosphorylated tau (181p) protein, beta-amyloid(1-42), and beta-amyloid(1-40) in cerebrospinal fluid of patients with dementia with Lewy bodies. Clin Chem Lab Med 2006;44:192-195.

7. Haass C, Koo EH, Capell A, Teplow DB, et al. Polarized sorting of beta-amyloid precursor protein and its proteolytic products in MDCK cells is regulated by two independent signals. J Cell Biol 1995;128:537-547.

8. Page RM, Baumann K, Tomioka M, Perez-Revuelta BI, et al. Generation of Abeta38 and Abeta42 is independently and differentially affected by familial Alzheimer disease-associated presenilin mutations and gamma-secretase modulation. J Biol Chem 2008;283:677-683.

9. Poorkaj P, Muma NA, Zhukareva V, Cochran EJ, et al. An R5L tau mutation in a subject with a progressive supranuclear palsy phenotype. Ann Neurol 2002;52:511-516.

10. Hayashi S, Toyoshima Y, Hasegawa M, Umeda Y, et al. Late-onset frontotemporal dementia with a novel exon 1 (Arg5His) tau gene mutation. Ann Neurol 2002;51:525-530.

11. Cortini F, Fenoglio C, Guidi I, Venturelli E, et al. Novel exon 1 progranulin gene variant in Alzheimer`s disease. Eur J Neurol 2008;15:1111-1117.

12. Guerreiro RJ, Baquero M, Blesa R, Boada M, et al. Genetic screening of Alzheimer`s disease genes in Iberian and African samples yields novel mutations in presenilins and APP. Neurobiol

Aging 2010;31:725-731.

13. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/) [accessed December 1, 2012]

14. Walker ES, Martinez M, Brunkan AL, Goate A. Presenilin 2 familial Alzheimer`s disease mutations result in partical loss of function and dramatic changes in Abeta 42/40 ratios. J Neurochem 2005;92:294-301.

15. To MD, Gokgoz N, Doyle TG, Donoviel DB, et al. Functional characterization of novel presenilin-2 variants identified in human breast cancers. Oncogene 2006;25:3557-3564.

16. Meeus B, Verstraeten A, Crosiers D, Engelborghs S, et al. DLB and PDD: a role for mutations in dementia and Parkinson disease genes ? Neurobiol Aging 2012;33:629.e5-e18.

17. Tedde A, Nacmias B, Ciantelli M, Forleo P, et al. Identification of New Presenilin Gene Mutations in Early-Onset Familial Alzheimer Disease. Arch Neurol 2003;60:1541-1544.

18. Tomaino C, Bernardi L, Anfossi M, Costanzo A, et al. Presenilin 2 Ser130Leu mutation in a case of late-onset „sporadic“ Alzheimer`s disease. J Neurol 2007;254:391-393.

19. Li D, Parks SB, Kushner JD, Nauman D, et al. Mutations in Gens in Dilated Cardiomyopathy and Heart Failure. Am J Hum Genet 2006;79:1030-1039.

SUPPLEMENTARY TABLES

Supplementary Table 1: Phenotype of carriers of rare variants.

gene	diagnosis	variant	AoO	DD	family history		IS	B	R	RT	PI	D	L-Dopa test	TCS	PSG	hyposmia	MRI	DiffD
APP	PD+D	A201V	70	8	n/a		B, RT	+	+	+	+	+	n/a	n/d	RBD, RLS	n/d	n/a	PSP
APP	PD+D	c.722_724delCTT	56	15	+	grandmother with PD	RT	+	+	++	+	+	+	+	n/d	+	temp & hippocampal atrophy, punctate signal alterations BG	
APP	PD	c.1091-3, E599K	60	1	(+)	father with D	n/a	+	+	-	+	-	-	n/d	n/d	+	chalky deposits in GP bilaterally	
APP	PD+D	E599K	76	3	-		n/a	+	+	+	+	+	+	n/d	n/d	n/d	NPH, slight gen atrophy	
APP	PD	E599K	69	7	-		B	+	+	(+)	+	-	+	n/d	n/d	n/d	microangiopathy	
APP	PD+D	c.1992_1994delGGA	59	8	n/a		RT	+	+	(+)	+	+	+	n/d	n/d	n/d	slight gen atrophy, microangiopathy	
APP	PD+D	T663M	64	5	+	3 aunts with PD	B,D	+	+	-	+	++	n/d	+	n/d	+	microangiopathy	LBD
APP	PD+D	G709S	64	8	+	uncle with PD	RT	+	+	(+)	+	+	+	n/d	n/d	n/d	microangiopathy	
APP	PD+D	c.2212-10_-11delTT	45	24	n/a		n/a	+	+	(+)	+	+	n/d	n/d	RLS	n/d	n/a	
GRN	PD+D	G35R	56	18	n/a		B	++	+	-	+	++	(+)	n/d	n/d	n/d	GE signal diminished dorsal of putamen, some frontotemp atrophy, slight thining of brain stem	LBD, MSA
GRN	PD+D	R433W	68	5	n/a		RT	+	++	(+)	+	+	(+)	n/d	RBD, RLS	n/d	BG calcifications, microangiopathy	
GRN	PD+D	R433W	67	10	n/a		n/a	+	++	-	++	++	n/d	n/d	n/d	n/d	NPH, slight microangiopathy	
GRN	PD+D	R433W	50	21	-		B, RT	+	+	(+)	+	+	+	n/d	n/d	n/d	unremarkable	
GRN	PD	R433W	64	2	-		B	+	+	+	(+)	-	+	+	n/d	n/d	n/a	
MAPT	PD+D	R5C	63	12	(+)	dizygotic twin with ET	B, RT	+	+	+	++	+	n/d	n/d	RBD	n/d	slight frontal atrophy, microangiopathy	
MAPT	PD	A41T	76	5	n/a		RT	+	+	++	+	(+)	+	n/d	n/d	n/d	iron deposits in dorsolateral putamen	
MAPT	PD+D	S427F	55	14	n/a		B	+	+	-	+	+	+	n/d	n/d	n/d	microangiopathy	
MAPT	PD	R448X	67	2	-		RT	+	+	+	+	++	+	+	RBD	n/d	n/a	
MAPT	PD+D	A469T	64	10	n/a		RT	+	+	++	+	+	+	n/d	n/d	n/d	slight mesencephalic and temp atrophy	PSP
MAPT	PD	A469T	63	4	-		RT	++	+	++	+	-	-	+	n/d	+	mesencephalic atrophy, loss of GE signal bilat putamen and pallidum	PSP
MAPT	PD	I695V	64	6	-		n/a	+	+	+	+	-	+	+	n/d	n/d	iron deposition in BG	AD
MAPT	PD	T762M	66	2	n/a		RT	+	+	+	+	-	+	n/d	n/d	n/d	microangiopathy	
PSEN1	PD+D	I148W	62	16	-		n/a	++	+	(+)	+	+	n/d	n/d	n/d	n/d	temp & hippocampal atrophy	
PENS2	PD+D	R62H	81	3	-		B	+	+	-	+	+	+	n/d	RBD	++	microangiopathy	
PSEN2	PD+D	R71W	65	10	n/a		RT	+	+	+	+	+	(+)	n/d	RBD, RLS	n/d	microangiopathy	
PSEN2	PD+D	S130L	73	12	n/a		B	+	+	+	+	++	n/d	n/d	n/d	n/d	temp atrophy, bilat. calcifications BG	

AoO = age of onset, DD = disease duration, n/a = not available, D = dementia, ET = essential tremor, IS = initial symptome, B = bradykinesia, R = rigor, RT = resting tremor, PI = postural instability, n/d = not done, TCS = transcranial sonography of the substantia nigra, PSG = polysomnography, RBD = REM sleep behavior disorder, RLS = restless legs syndrome, MRI = magnetic resonance imaging, BG = basal ganglia, GE = gradient echo, NPH = normal pressure hydrocephalus, GP = globus pallidus, DiffD = differential diagnosis considered, LBD = Lewy body dementia, MSA = multisystem atrophy, PSP = progressive supranuclear palsy, AD = Alzheimer`s disease

2.8 Zimprich *et al.*, A Mutation in *VPS35*, Encoding a Subunit of the Retromer Complex, Causes Late-Onset Parkinson Disease, *American Journal of Human Genetics*, 2011[VIII]

Personal contributions: I performed half of the LightScanner® high-resolution melting curve analysis and follow-up Sanger sequencing of the coding regions of *VPS35* in the Parkinson's disease case/control sample and analyzed the LightScanner® data. I contributed to the design of Tables 1 and 3.

REPORT

A Mutation in *VPS35*, Encoding a Subunit of the Retromer Complex, Causes Late-Onset Parkinson Disease

Alexander Zimprich,[1,14,*] Anna Benet-Pagès,[2,14] Walter Struhal,[3,14] Elisabeth Graf,[2,14] Sebastian H. Eck,[2] Marc N. Offman,[4] Dietrich Haubenberger,[1] Sabine Spielberger,[5] Eva C. Schulte,[2,6] Peter Lichtner,[2] Shaila C. Rossle,[4] Norman Klopp,[7] Elisabeth Wolf,[5] Klaus Seppi,[5] Walter Pirker,[1] Stefan Presslauer,[8] Brit Mollenhauer,[9] Regina Katzenschlager,[10] Thomas Foki,[1] Christoph Hotzy,[1] Eva Reinthaler,[1] Ashot Harutyunyan,[11] Robert Kralovics,[11] Annette Peters,[7] Fritz Zimprich,[1] Thomas Brücke,[8] Werner Poewe,[5] Eduard Auff,[1] Claudia Trenkwalder,[9,12] Burkhard Rost,[4] Gerhard Ransmayr,[3] Juliane Winkelmann,[2,6,13] Thomas Meitinger,[2,13] and Tim M. Strom[2,13,*]

To identify rare causal variants in late-onset Parkinson disease (PD), we investigated an Austrian family with 16 affected individuals by exome sequencing. We found a missense mutation, c.1858G>A (p.Asp620Asn), in the *VPS35* gene in all seven affected family members who are alive. By screening additional PD cases, we saw the same variant cosegregating with the disease in an autosomal-dominant mode with high but incomplete penetrance in two further families with five and ten affected members, respectively. The mean age of onset in the affected individuals was 53 years. Genotyping showed that the shared haplotype extends across 65 kilobases around *VPS35*. Screening the entire *VPS35* coding sequence in an additional 860 cases and 1014 controls revealed six further nonsynonymous missense variants. Three were only present in cases, two were only present in controls, and one was present in cases and controls. The familial mutation p.Asp620Asn and a further variant, c.1570C>T (p.Arg524Trp), detected in a sporadic PD case were predicted to be damaging by sequence-based and molecular-dynamics analyses. VPS35 is a component of the retromer complex and mediates retrograde transport between endosomes and the trans-Golgi network, and it has recently been found to be involved in Alzheimer disease.

Parkinson's disease (PD [MIM 168600]) is the second-most common neurodegenerative disorder; it affects 1%–2% of the population above the age of 60.[1] It is characterized by degeneration of dopaminergic neurons in the nigrostriatal pathway and other monoaminergic cell groups in the brainstem. This degeneration leads to bradykinesia, resting tremor, muscular rigidity, and postural instability as well as nonmotor symptoms. Up to 20% of cases with PD are reported to be familial,[2,3] but extended pedigrees with clear Mendelian inheritance are rare. Genetic studies have so far revealed mutations in five genes causing autosomal-recessive (*PARK2* [MIM 602544], *PINK1* [MIM 608309], *PARK7* [MIM 602533]) or autosomal-dominant (*SNCA* [MIM 163890], *LRRK2* [MIM 609007]) forms of PD.[4–9] Whereas the autosomal-recessive forms with early onset and *SNCA* missense mutations or duplications[10] are rare, a single *LRRK2* mutation (RefSeq number NM_198578.3: c.6055G>A [p.Gly2019Ser]) accounts for approximately 1% of sporadic cases of European origin.[11–13] A recent study revealed a strong association of PD with glucocerebrosidase (*GBA*) mutations in carriers for Gaucher [MIM 230800] disease, thus implicating a lysosomal enzyme in the pathogenesis of PD.[14,15] Genome-wide association studies revealed several low-risk susceptibility loci, among them *LAMP3* [MIM 605883] and *HIP1R* [MIM 605613], which have been reported to be implicated in the lysosomal pathway.[16–18]

We identified an Austrian family in which 16 members were affected by PD (family A, Figure 1). PD seemed to be inherited in an autosomal-dominant mode with high penetrance. Seven affected members were available for clinical and DNA investigations. Six of them exhibited at least three of the four cardinal signs of PD (akinesia, resting tremor, rigidity, and postural instability) and showed improvement after dopaminergic treatment. A single affected individual had displayed action tremors since childhood but developed L-Dopa-responsive resting tremors and akinesia only at the age of 62 years. The mean age of onset was 53 years (range 40–68 years) (Table 1). The clinical diagnosis of idiopathic PD was made by movement-disorder specialists who used UK brain bank criteria for PD.[19] All participants gave written informed

[1]Department of Neurology, Medizinische Universität Wien, 1090 Vienna, Austria; [2]Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany; [3]Department of Neurology and Psychiatry, Allgemeines Krankenhaus, 4021 Linz, Austria; [4]Institute of Bioinformatics, Technische Universität München, 85748 Garching, Germany; [5]Department of Neurology, Medizinische Universität Innsbruck, 6020 Innsbruck, Austria; [6]Department of Neurology, Technische Universität München, 81675 Munich, Germany; [7]Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany; [8]Department of Neurology, Wilhelminenspital, 1160 Vienna, Austria; [9]Paracelsus-Elena Klinik, 34128 Kassel, Germany; [10]Department of Neurology, Sozialmedizinisches Zentrum Ost-Donauspital, 1220Vienna, Austria; [11]Center for Molecular Medicine of the Austrian Academy of Sciences, 1090 Vienna, Austria; [12]Department of Clinical Neurophysiology, Georg-August-Universität Göttingen, 37079 Göttingen, Germany; [13]Institute of Human Genetics, Technische Universität München, 81675 Munich, Germany
[14]These authors contributed equally to this work
*Correspondence: alexander.zimprich@meduniwien.ac.at (A.Z.), timstrom@helmholtz-muenchen.de (T.M.S.)
DOI 10.1016/j.ajhg.2011.06.008.

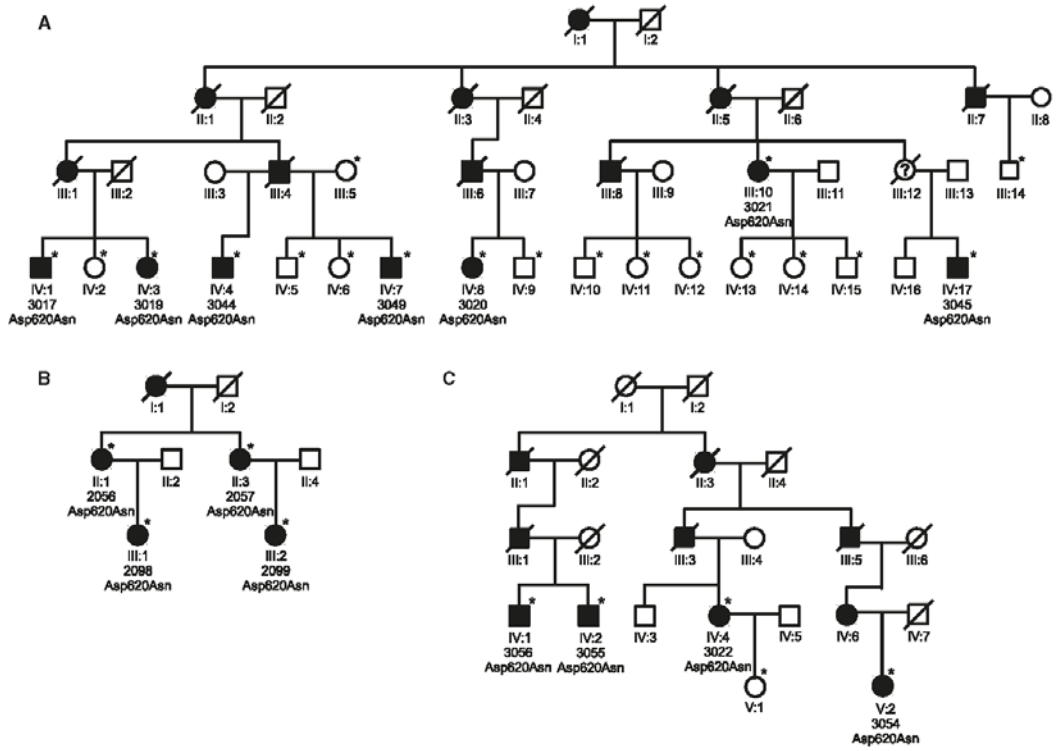


Figure 1. Pedigrees of Families A, B, and C
Unaffected family members are indicated by open symbols, affected members by closed symbols. Asterisks denote individuals genotyped for p.Asp620Asn. To maintain confidentiality, we have not shown genotypes of unaffected individuals. A question mark within a symbol denotes an unknown phenotype. Diagonal bars through symbols denote deceased individuals.

consent. The study was approved by the institutional review board of the Medizinische Universität Wien and the Hessische Landesärztekammer Wiesbaden.

To identify the disease-causing variant, we selected two second cousins (#3017 and #3020) for exome sequencing. We assumed that any rare variants common in both individuals would be disease-causing candidates. Selecting distantly related members of the pedigree should minimize the proportion of alleles shared by descent. Exome sequencing was performed on a Genome Analyzer IIx system (Illumina) after in-solution enrichment of exonic sequences (SureSelect Human All Exon 38 Mb kit, Agilent). We sequenced two lanes of a flowcell for both samples, each as 54 bp paired-end runs. Read alignment was performed with BWA (version 0.5.8) to the human genome assembly hg19 (Table S1, available online). Single-nucleotide variants and small insertions and deletions (indels) were detected with SAMtools (v 0.1.7). We filtered called variants to exclude those present in 72 control exomes from patients with other unrelated diseases. We further excluded all variants that were present in dbSNP 131 and had an average heterozygosity of more than 0.02. Variant annotation was performed with custom scripts. This approach left ten heterozygous nonsynonymous variants shared by both affected individuals (Table 2; see also Table S2).

Only a single heterozygous variant in the *VPS35* gene (RefSeq number NM_018206.4: c.1858G>A [p.Asp620Asn]) fulfilled two further criteria of being possibly causative: (1) it was found in all seven affected members investigated and (2) was absent in approximately 680 KORA S4 general-population samples (Tables 2 and 3).[20] We next screened 486 unrelated PD patients from Austria for the p.Asp620Asn variant by MALDI-TOF mass spectroscopy (Sequenom MassArray system). We detected two additional index patients carrying this mutation (families B and C; Figure 1 and Table 1). The variant was detected in all eight affected individuals investigated in both families. It was not present in a second set of 554 Austrian controls or in an additional 1014 KORA-AGE controls (Table 3). The variant was further detected in three clinically unaffected family members in families A, B, and C. Because the unaffected individuals are all younger than 60 years of age, either they are all presymptomatic or the mutation is nonpenetrant in these subjects.

Table 1. Clinical Findings for PD Patients Carrying Variants in VPS35

Family	Patient	Variation	AaO	DD	IS	B	R	RT	PI	L-Dopa/DA	Other Features
A	3017	p.Asp620Asn	48	7	B	+	+	−	+	+	
A	3019	p.Asp620Asn	40	5	B	+	+	+	+	+	
A	3020	p.Asp620Asn	46	7	PI	+	+	−	+	+	
A	3021	p.Asp620Asn	68	16	PI	+	+	+	+	+	
A	3049	p.Asp620Asn	49	4	RT	+	+	+	−	+	
A	3044	p.Asp620Asn	64	3	PI	+	+	+	+	+	
A	3045	p.Asp620Asn	63	1	RT	+	−	+	−	+	action tremor since childhood
B	2056	p.Asp620Asn	61	15	RT	+	+	+	+	+	fluctuations, dyskinesias
B	2057	p.Asp620Asn	56	8	RT	+	+	+	+	+	fluctuations, dyskinesias
B	2098	p.Asp620Asn	46	0.5	RT	−	−	+	−	untreated	depression, action tremor, pathologic DAT SPECT
B	2099	p.Asp620Asn	51	5	B	+	+	+	−	+	fluctuations, pathologic DAT SPECT
C	3022	p.Asp620Asn	61	5	RT	+	+	+	−	+	dyskinesias
C	3055	p.Asp620Asn	46	12	RT	+	+	+	−	+	
C	3054	p.Asp620Asn	53	9	B	+	+	−	−	+	dyskinesias
C	3056	p.Asp620Asn	43	10	B	+	+	+	+	+	dyskinesias
	211	p.Arg524Trp	37	9	MG	+	+	+	−	+	mild action tremor since youth; 75% motor improvement on levodopa-test; DBS for fluctuations and dyskinesias; pathologic DAT SPECT
	524	p.Leu774Met	51	7	RT	+	+	+	−	+	marked postural tremor
	243	p.Leu774Met	73	9	RT	+	+	+	+	+	dyskinesias, pathologic DAT SPECT
	806	p.Ile241Met	72	2	Postural tremor	+	−	+	+	+	hyposmia (6/12 sniffing sticks), DAT SPECT pathologic, pathologic crying
	90/05	p.Met57Ile	62	13	RT	+	+	+	+	+	dementia (MMSE 23), dysphagia and dysarthria, hyposmia by history, depression

Abbreviations are as follows: AaO, age at onset; DD, disease duration in years; IS, initial symptoms; B, bradykiesia; R, rigidity; RT, resting tremor; PI, postural instability; L-Dopa/DA, response to L-Dopa and/or dopamine agonist; MG, micrographia; DBS, deep brain stimulation.

Cross-species alignment of VPS35 from plants, fungi, invertebrates, and vertebrates showed complete conservation of amino acid Asp620 (Figure S1). The likely consequence of the p.Asp620Asn variant was predicted to be damaging by PolyPhen2,[21] SNAP,[22] and SIFT.[23] We therefore concluded that the variant p.Asp620Asn is indeed very likely to be causative for PD in families A, B, and C.

To determine whether the variant p.Asp620Asn occurred on the same haplotype, we genotyped 20 individuals from families A–C with oligonucleotide SNP arrays (HumanOmni2.5-Quad, Illumina). Haplotyping and linkage analysis were performed with the Merlin software.[24] The haplotypes carrying the variant p.Asp620Asn in families A–C are depicted in Table S3. Family A and B shared a common haplotype across 21 Mb between markers rs1072594 and rs4444336. Family C, however, showed only a common region of 65 kb across *VPS35*. Different alleles were located at markers rs56168099 and rs74459547, 25 kb upstream and 11 kb downstream of *VPS35*, respectively (Table S3). Because the two intragenic markers did not differ, we could not determine whether the three families shared an old common haplotype or whether the mutation has recently arisen on two different haplotypes.

To assess the prevalence of other *VPS35* mutations among PD cases and the general population, we screened all 17 coding exons for variations by dye-binding/high-resolution DNA melting curve analysis (LightScanner HR I 384, Idaho Technology) in 860 cases (484 Austrian and

Table 2. Exome Sequencing: Rare, Heterozygous, Nonsynonymous Variations Shared by Two Individuals of Pedigree A

				Variations		Control Genotypes			
Gene	**Position (hg19)**	**dbSNP**	**Transcript**	**Nucleotide**	**Amino Acid**	**1/1**	**1/2**	**2/2**	**Segregation**
PLK3	chr1:45270359		NM_004073.2	c.1543T>A	p.Ser515Thr	669	0	0	4 of 7
C8A	chr1:57383357	rs41285938	NM_000562.2	c.1723C>T	p.Pro575Ser				5 of 7
ADCY10	chr1:167787479	rs41270737	NM_018417.4	c.4313A>G	p.Asn1438Ser				2 of 7
LAMB2	chr3:49166460		NM_002292.3	c.1724G>A	p.Arg575Gln	647	28	0	5 of 7
NOM1	chr7:156762317		NM_138400.1	c.2503G>A	p.Ala835Thr	670	0	0	3 of 7
KIF22	chr16:29816237		NM_007317.1	c.1780G>A	p.Asp594Asn	665	6	0	6 of 7
SEZ6L2	chr16:29899021		NM_012410.2	c.947G>A	p.Arg316His	660	4	0	7 of 7
VPS35	chr16:46696364		NM_018206.4	c. 1858G>A	p.Asp620Asn	1069[a]	0	0	7 of 7
NLRP1	chr17:5421150		NM_001033053.2	c.3985G>A	p.Val1329Ile	666	4	0	3 of 7
NEURL4	chr17:7221197		NM_001005408.1	c.4109G>A	p.Arg1370Gln				3 of 7

Rare variations revealed by exome sequencing were checked in 670 controls (KORA S4) by MALDI-TOF analysis. The variant allele was denoted as "2," the reference allele as "1."
[a] This number includes additional 554 Austrian control individuals investigated by a TaqMan genotyping assay. Segregation shows the number of affected pedigree A individuals who carry the variant allele.

376 German cases) and 1014 controls. For controls, we used a population-based cohort (KORA AGE) with a mean age of 76 years but excluded eight individuals known to be on medications for PD (Table 3). Exons 2 to 12 are located within a region that is duplicated 12 Mb upstream. Primers were designed to specifically amplify these exons (Table S4). The screening revealed six further rare coding SNVs in addition to p.Asp620Asn (Table 4). Including p.Asp620Asn, we identified four different nonsynonymous missense variants only present in cases, two only present in controls, and one present in cases and controls. Two of the variants unique to PD cases were predicted to be damaging by all three methods (c.1858G>A [p.Asp620Asn]; c.1570C>T [p.Arg524Trp]), and one was predicted by PolyPhen2 to be possibly damaging (c.723T>G, p.Ile241Met). The other variants were predicted to be benign by all methods. Family information was only available for the patient carrying the p.Arg524Trp variant. The only available family member was her mother, aged 74 years. She was found to also carry the variant and showed mild extrapyramidal signs, including intermittent resting tremor of the left fingers and mild postural tremor of both upper limbs, but no bradykinesia. However, a DAT SPECT examination showed normal striatal binding, excluding the possibility of an early stage of PD in this subject. Of note, the screening did not reveal any common nonsynonymous coding SNVs. Furthermore, common nonsynonymous coding SNVs were not found in the 72 control exomes from patients with other unrelated diseases, nor were any recorded in the dbSNP database (version 131).

Table 3. Summary of the Samples Used in This Study

Cohort	**Sample Size**	**Mean Age (SD)**	**Females/Males**
Austrian PD cases[a]	486	58.7 (11.3)	172/314
German PD cases[b]	376	71.1 (9.4)	119/257
KORA S4 controls[c]	680	54.7 (11.9)	280/400
KORA-AGE controls[d]	1014	76.0 (6.6)	508/505
Austrian controls[e]	554	46 (15.2)	254/300

Patients presenting with atypical or secondary (e.g., vascular) parkinsonian disorders as well as patients with known mutations were excluded.
[a] The Austrian cases were recruited at the Department of Neurology, Medizinische Universität Wien, Vienna, as well as in affiliated departments on a consecutive basis. A positive family history for PD was reported from 131 patients. A positive family history was defined by at least one other affected first- or second-degree related family member.
[b] The German PD population originated from the Paracelsus-Elena Klinik, Kassel, a hospital specializing in movement disorders.
[c] This control population was recruited from the KORA S4 survey, comprising individuals who were aged 25–74 years and were examined during 1999–2001.
[d] The KORA-AGE samples were collected in 2009 as a gender- and age-stratified subsample of the KORA S1–S4 studies comprising participants born before 1944. KORA S1–S4 surveys comprise four independent cross-sectional population-based studies in the region of Augsburg, Southern Germany, and were conducted in 5 year intervals. Patients for whom PD was suspected on the basis of questionnaire data were excluded.
[e] These control samples were recruited through the Department of Neurology, Medical University of Vienna, as subjects without known history of a neurological disorder and included, for example, blood donors or unrelated companions or spouses of patients.

VPS35 is a component of the retromer complex and is involved in retrograde transport from the endosomes back to the trans-Golgi network.[25] This multi-protein complex consists of the cargo-recognition VPS26-VPS29-VPS35 heterotrimer and a membrane-targeting heterodimer or homodimer of SNX1 and/or SNX2 (vps5).[25,26] All proteins involved are evolutionarily conserved and have been previously described in *Saccharomyces cerevisiae*. The best characterized cargo proteins of the retromer complex are the cation-independent mannose 6-phosphate receptor

Table 4. Rare *VPS35* Variants in Cases and Controls

ID Cases	KORA AGE Controls	Heterozygous Nucleotide Change	Amino Acid Change	Predicted Impact on Protein			Exon/ Intron	Genomic Position (hg19, chr16)	KORA S4 Controls		
				(i)	(ii)	(iii)			1/1	1/2	2/2
Nonsynonymous											
-	1	c.151G>A	p.Gly51Ser	+	+	+	3	46,716,,039			
90/05	-	c.171G>A	p.Met57Ile	+	+	+	3	46,716,019	670	0	0
-	1	c.245C>G	p.Thr82Arg	+	+	+	4	46,715,367			
806	-	c.723T>G	p.Ile241Met	±	+	+	7	46,711,308	667	0	0
[211]	-	c.1570C>T	p.Arg524Trp	−	−	−	13	46,702,919	671	0	0
[Families A-C]	-	c.1858G>A	p.Asp620Asn	−	−	−	15	46,696,364	669	0	0
243, 524	2	c.2320C>A	p.Leu774Met	+	+	+	17	46,694,455			
Synonymous											
53097	-	c.492A>G	p.Glu164Glu				5	46,714,597	671	0	0
-	1	c.954A>T	p.Gly315Gly				9	46,708,542			
53496	-	c.1881C>T	p.Ala627Ala				15	46,696,341	668	5	0
45, 117, 53626	1	c.2145A>G	p.Leu715Leu				16	46,695,696	666	2	0
53667	-	c.2241C>T	p.Ile747Ile				17	46,694,534	667	2	0
53063	-	c.2346A>G	p.Glu782Glu				17	46,694,429	671	0	0
-	1	c.2361G>A	p.Glu787Glu				17	46,694,414			
Noncoding											
2212	2	c.1-35C>T					5′UTR	46,723,080	667	2	0
-	2	c.1-29C>T					5′UTR	46,723,074			
95, 2206	3	c.3+24A>G					1	46,723,019	662	6	0
159, 528	1	c.102+33G>A					2	46,717,387	668	2	0
[157, 2023]	-	c.103-77T>C					3	46,716,164	668	0	0
-	1	c.199+9T>G					3	46,715,982			
213	-	c.506+6T>C					5	46,714,577	644	0	0
53093	-	c.720+18C>T					6	46,712,773			
-	1	c.914+38T>C					8	46,710,457			
52824	-	c.1161-87A>C					10	46,706,471			
52791	-	c.1161-70G>A					10	46,706,454	668	0	0
-	1	c.1368+16C>T					11	46,706,161			
[2028]	-	c.1369-11G>A					12	46,705,783	669	0	0
-	1	c.1525-17delT					12	46,702,985			
-	1	c.1647+14T>C					13	46,702,828			
320	-	c.2212-45T>C					16	46,694,608	670	0	0
[352]	-	c.2391+7A>G					3′UTR	46,694,377			
-	1	c.2391+8A>G					3′UTR	46,694,376			

Variants for 863 cases and 1014 KORA AGE controls were determined by dye-binding/high-resolution DNA melting curve analysis and confirmed by Sanger sequencing. The table lists the case ID and the number of detected variant alleles of the cases and KORA AGE samples, respectively. Genotypes of identified variants were further investigated by MALDI-TOF analysis in approximately 680 KORA S4 controls. For the KORA S4 samples, the variant allele was denoted as "2," the reference allele as "1." cDNA numbering is based on reference gene NM_018206.4 for *VPS35*, where +1 corresponds to the A of ATG start translation codon. Familial cases are given in square brackets. Three methods were used for predicting the impact of SNPs on the protein. (1) PolyPhen2, (2) SNAP, and (3) SIFT; "+" indicates a benign impact, "±" indicates a possibly damaging impact, and "−" indicates a damaging impact. We detected a further nonsynonymous variant (c.1093C>T [p.Arg365Cys], genomic position 46,708,293) in a patient carrying two *PARKIN* variants (c.exon3_4del and p.Arg275Trp). This variant was not present in 670 KORA S4 and 1014 KORA AGE controls. It is predicted to be possibly damaging by all three methods. This patient's brother is also affected by PD. He carries the 2 PARKIN variants but not the *VPS35* variant.

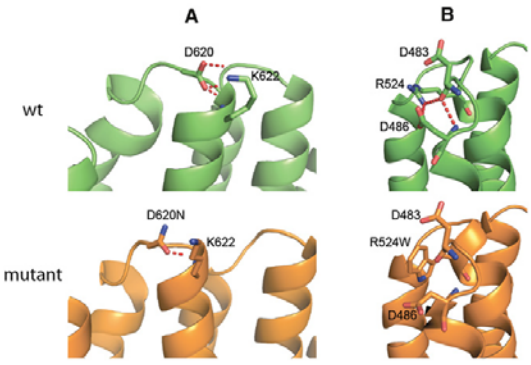


Figure 2. Hydrogen-Bonding Capacities for Wild-Type Asp620 and Arg524 and the Variants p.Asp620Asn and p.Arg524Trp
Hydrogen bonds (HB) are shown as red dashed lines. Asp60 and Arg524 are in green; p.Asp620Asn and p.Arg524Trp are in orange. (A) Asp620 forms a HB to Lys622 and shows an additional salt-bridge interaction. p.Asp620Asn forms fewer HBs, and no electrostatic interaction is possible.
(B) Arg524 forms a HB network with Asp483 and Asp486. This network is broken by the p.Arg524Trp substitution.

(CI-MPR)[27] and Vps10p in mammals and *Saccharomyces cerevisiae*, respectively; these proteins transport hydroxylases to the lysosomes or lysosomal vacuoles. Recently, additional cargo proteins and functions of VPS35 have been described.[28,29] Most interesting in our context is the involvement of the retromer into the retrograde transport of SORL1, a VPS10P-domain receptor protein that has been implicated in Alzheimer disease.[30,31] The crystal structure of the C-terminal part of VPS35 has been resolved.[32] The three variants p.Asp620Asn, p.Arg524Trp, and p.Leu774Met are located in this part of the protein, and we have investigated their impact on protein stability by using molecular dynamics (MD) simulations. We manually introduced the mutations to the crystal structure and modeled the side chains by using scwrl 4.0.[33] All MD simulations were performed via GROMACS 4.5,[34] with the all-atom force field AMBER03[35] and the water model TIP3P[36] as parameters. All three proteins are found on the edge of helices interacting with VPS29. Wild-type residue Asp620 forms frequent hydrogen bonds (HBs) to Lys622, but these bonds are less frequent in the p.Asp620Asn variant (Figure 2A). Similarly, Arg524 is involved in a triple HB network together with residues Asp483 and Asp486, but this network is broken by the introduction of p.Arg524Trp (Figure 2B). Both changes result in the loss of salt bridges and cause the protein to be locally more flexible, as shown by root-mean-square fluctuation (RMSF) profiles (Figure S2). In contrast to the effect predicted for p.Arg524Trp and p.Asp620Asn, the p.Leu774Met variant was not predicted to have a strong impact on protein stability.

In summary, we identified rare *VPS35* missense variants that are potentially pathogenic. One of these variants, p.Asp620Asn, cosegregates with late-onset PD in three unrelated families. The observation that the three families share only a small common haplotype across *VPS35*, the high conservation of VPS35, the predicted structural changes, and the protein's known involvement in lysosomal trafficking together provide strong support for the p.Asp620Asn variant's being causative for late-onset PD, although we identified only a single familial mutation. The penetrance of p.Asp620Asn is high but not complete and might be lower for the other variants. The proportion of PD caused by *VPS35* variants is expected to be low. Although exome sequencing provides perfect access to rare-variant detection, both large families and large collections of cases and controls remain a crucial resource for the identification of disease genes.

Supplemental Data

Supplemental Data include two figures and four tables and can be found with this article online at http://www.cell.com/AJHG/.

Acknowledgments

We thank all patients and their families for participating in this study. We also thank C. Fischer and B. Schmick for technical assistance and S. Schmidegg, S. Hoedl, and M. Guger for clinical examination of family A. This work was supported by a grant from the German Ministry for Education and Research (01GR0804-4). The KORA study was financed by the Helmholtz Zentrum München, the German Federal Ministry of Education and Research, the State of Bavaria, the German National Genome Research Network (NGFNplus: 01GS0823), and the Munich Center of Health Sciences (MCHealth) as part of LMUinnovativ. M.N.O., S.C.R., and B.R. were supported by the Alexander von Humboldt Foundation.

Received: May 8, 2011
Revised: June 15, 2011
Accepted: June 21, 2011
Published online: July 14, 2011

Web Resources

The URLs for data presented herein are as follows:

ExonPrimer, http://ihg.helmholtz-muenchen.de/exonprimer.html
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org
UCSC Genome Browser, http://genome.ucsc.edu

References

1. Lang, A.E., and Lozano, A.M. (1998). Parkinson's disease. First of two parts. N. Engl. J. Med. *339*, 1044–1053.
2. Bonifati, V., Fabrizio, E., Vanacore, N., De Mari, M., and Meco, G. (1995). Familial Parkinson's disease: A clinical genetic analysis. Can. J. Neurol. Sci. *22*, 272–279.
3. Payami, H., Larsen, K., Bernard, S., and Nutt, J. (1994). Increased risk of Parkinson's disease in parents and siblings of patients. Ann. Neurol. *36*, 659–661.
4. Zimprich, A., Biskup, S., Leitner, P., Lichtner, P., Farrer, M., Lincoln, S., Kachergus, J., Hulihan, M., Uitti, R.J., Calne, D.B.,

et al. (2004). Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron *44*, 601–607.

5. Bonifati, V., Rizzu, P., van Baren, M.J., Schaap, O., Breedveld, G.J., Krieger, E., Dekker, M.C., Squitieri, F., Ibanez, P., Joosse, M., et al. (2003). Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science *299*, 256–259.
6. Kitada, T., Asakawa, S., Hattori, N., Matsumine, H., Yamamura, Y., Minoshima, S., Yokochi, M., Mizuno, Y., and Shimizu, N. (1998). Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature *392*, 605–608.
7. Paisán-Ruíz, C., Jain, S., Evans, E.W., Gilks, W.P., Simón, J., van der Brug, M., López de Munain, A., Aparicio, S., Gil, A.M., Khan, N., et al. (2004). Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. Neuron *44*, 595–600.
8. Polymeropoulos, M.H., Lavedan, C., Leroy, E., Ide, S.E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R., et al. (1997). Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science *276*, 2045–2047.
9. Valente, E.M., Abou-Sleiman, P.M., Caputo, V., Muqit, M.M., Harvey, K., Gispert, S., Ali, Z., Del Turco, D., Bentivoglio, A.R., Healy, D.G., et al. (2004). Hereditary early-onset Parkinson's disease caused by mutations in PINK1. Science *304*, 1158–1160.
10. Johnson, J., Hague, S.M., Hanson, M., Gibson, A., Wilson, K.E., Evans, E.W., Singleton, A.A., McInerney-Leo, A., Nussbaum, R.L., Hernandez, D.G., et al. (2004). SNCA multiplication is not a common cause of Parkinson disease or dementia with Lewy bodies. Neurology *63*, 554–556.
11. Gilks, W.P., Abou-Sleiman, P.M., Gandhi, S., Jain, S., Singleton, A., Lees, A.J., Shaw, K., Bhatia, K.P., Bonifati, V., Quinn, N.P., et al. (2005). A common LRRK2 mutation in idiopathic Parkinson's disease. Lancet *365*, 415–416.
12. Nichols, W.C., Pankratz, N., Hernandez, D., Paisán-Ruíz, C., Jain, S., Halter, C.A., Michaels, V.E., Reed, T., Rudolph, A., Shults, C.W., et al; Parkinson Study Group-PROGENI investigators. (2005). Genetic screening for a single common LRRK2 mutation in familial Parkinson's disease. Lancet *365*, 410–412.
13. Di Fonzo, A., Rohé, C.F., Ferreira, J., Chien, H.F., Vacca, L., Stocchi, F., Guedes, L., Fabrizio, E., Manfredi, M., Vanacore, N., et al; Italian Parkinson Genetics Network. (2005). A frequent LRRK2 gene mutation associated with autosomal dominant Parkinson's disease. Lancet *365*, 412–415.
14. Sidransky, E., Nalls, M.A., Aasly, J.O., Aharon-Peretz, J., Annesi, G., Barbosa, E.R., Bar-Shira, A., Berg, D., Bras, J., Brice, A., et al. (2009). Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N. Engl. J. Med. *361*, 1651–1661.
15. Aharon-Peretz, J., Rosenbaum, H., and Gershoni-Baruch, R. (2004). Mutations in the glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. N. Engl. J. Med. *351*, 1972–1977.
16. Simón-Sánchez, J., Schulte, C., Bras, J.M., Sharma, M., Gibbs, J.R., Berg, D., Paisan-Ruiz, C., Lichtner, P., Scholz, S.W., Hernandez, D.G., et al. (2009). Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat. Genet. *41*, 1308–1312.
17. Satake, W., Nakabayashi, Y., Mizuta, I., Hirota, Y., Ito, C., Kubo, M., Kawaguchi, T., Tsunoda, T., Watanabe, M., Takeda, A., et al. (2009). Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. Nat. Genet. *41*, 1303–1307.
18. Nalls, M.A., Plagnol, V., Hernandez, D.G., Sharma, M., Sheerin, U.M., Saad, M., Simón-Sánchez, J., Schulte, C., Lesage, S., Sveinbjörnsdóttir, S., et al; International Parkinson Disease Genomics Consortium. (2011). Imputation of sequence variants for identification of genetic risks for Parkinson's disease: A meta-analysis of genome-wide association studies. Lancet *377*, 641–649.
19. Hughes, A.J., Daniel, S.E., Kilford, L., and Lees, A.J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: A clinico-pathological study of 100 cases. J. Neurol. Neurosurg. Psychiatry *55*, 181–184.
20. Wichmann, H.E., Gieger, C., and Illig, T. (2005). KORA-gen–resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen *67* (*Suppl. 1*), 26–30.
21. Ramensky, V., Bork, P., and Sunyaev, S. (2002). Human non-synonymous SNPs: server and survey. Nucleic Acids Res. *30*, 3894–3900.
22. Bromberg, Y., and Rost, B. (2007). SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Res. *35*, 3823–3835.
23. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249.
24. Abecasis, G.R., Cherny, S.S., Cookson, W.O., and Cardon, L.R. (2002). Merlin—Rapid analysis of dense genetic maps using sparse gene flow trees. Nat. Genet. *30*, 97–101.
25. Bonifacino, J.S., and Rojas, R. (2006). Retrograde transport from endosomes to the trans-Golgi network. Nat. Rev. Mol. Cell Biol. *7*, 568–579.
26. Rojas, R., Kametaka, S., Haft, C.R., and Bonifacino, J.S. (2007). Interchangeable but essential functions of SNX1 and SNX2 in the association of retromer with endosomes and the trafficking of mannose 6-phosphate receptors. Mol. Cell. Biol. *27*, 1112–1124.
27. Damen, E., Krieger, E., Nielsen, J.E., Eygensteyn, J., and van Leeuwen, J.E. (2006). The human Vps29 retromer component is a metallo-phosphoesterase for a cation-independent mannose 6-phosphate receptor substrate peptide. Biochem. J. *398*, 399–409.
28. Braschi, E., Goyon, V., Zunino, R., Mohanty, A., Xu, L., and McBride, H.M. (2010). Vps35 mediates vesicle transport between the mitochondria and peroxisomes. Curr. Biol. *20*, 1310–1315.
29. Korolchuk, V.I., Schütz, M.M., Gómez-Llorente, C., Rocha, J., Lansu, N.R., Collins, S.M., Wairkar, Y.P., Robinson, I.M., and O'Kane, C.J. (2007). Drosophila Vps35 function is necessary for normal endocytic trafficking and actin cytoskeleton organisation. J. Cell Sci. *120*, 4367–4376.
30. Willnow, T.E., Petersen, C.M., and Nykjaer, A. (2008). VPS10P-domain receptors—Regulators of neuronal viability and function. Nat. Rev. Neurosci. *9*, 899–909.
31. Rogaeva, E., Meng, Y., Lee, J.H., Gu, Y., Kawarai, T., Zou, F., Katayama, T., Baldwin, C.T., Cheng, R., Hasegawa, H., et al. (2007). The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. Nat. Genet. *39*, 168–177.
32. Hierro, A., Rojas, A.L., Rojas, R., Murthy, N., Effantin, G., Kajava, A.V., Steven, A.C., Bonifacino, J.S., and Hurley, J.H.

(2007). Functional architecture of the retromer cargo-recognition complex. Nature *449*, 1063–1067.

33. Krivov, G.G., Shapovalov, M.V., and Dunbrack, R.L., Jr. (2009). Improved prediction of protein side-chain conformations with SCWRL4. Proteins *77*, 778–795.
34. Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theory Comput. *4*, 435–447.
35. Duan, Y., Wu, C., Chowdhury, S., Lee, M.C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., et al. (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J. Comput. Chem. *24*, 1999–2012.
36. Mahoney, M.W. (2000). A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. J. Chem. Phys. *112*, 8910–8922.

2.9 Schulte *et al.*, Rare variants in *LRRK1* and Parkinson`s disease, *Neurogenetics*, 2013[IX]

Personal contributions: I recruited part of the family and participated in designing the study. I also analyzed the exome sequencing data and participated in the CNV analysis. I designed the multiplex PCRs for Sequenom®-based genotyping and analyzed genotyping data for both the segregation analysis in the family and the frequency assessment in the Parkinson's disease case/control sample. I performed *in silico* predictions of variant pathogenicity. Further, I performed the LightScanner® high-resolution melting curve analysis and follow-up Sanger sequencing of the coding regions of both *LRRK1* and *EEF1D* in the Parkinson's disease case/control sample and analyzed the LightScanner® data and calculated the burden tests. I wrote the manuscript and designed all tables and figures except for Figure 3 and 4.

Neurogenetics
DOI 10.1007/s10048-013-0383-8

ORIGINAL ARTICLE

Rare variants in *LRRK1* and Parkinson's disease

Eva C. Schulte · Daniel C. Ellwanger · Sybille Dihanich · Claudia Manzoni · Katrin Stangl · Barbara Schormair · Elisabeth Graf · Sebastian Eck · Brit Mollenhauer · Dietrich Haubenberger · Walter Pirker · Alexander Zimprich · Thomas Brücke · Peter Lichtner · Annette Peters · Christian Gieger · Claudia Trenkwalder · Hans-Werner Mewes · Thomas Meitinger · Patrick A. Lewis · Hans H. Klünemann · Juliane Winkelmann

Received: 28 July 2013 / Accepted: 15 October 2013

Abstract Approximately 20 % of individuals with Parkinson's disease (PD) report a positive family history. Yet, a large portion of causal and disease-modifying variants is still unknown. We used exome sequencing in two affected individuals from a family with late-onset PD to identify 15 potentially causal variants. Segregation analysis and frequency assessment in 862 PD cases and 1,014 ethnically matched controls highlighted variants in *EEF1D* and *LRRK1* as the best candidates. Mutation screening of the coding regions of these genes in 862 cases and 1,014 controls revealed several novel

Electronic supplementary material The online version of this article (doi:10.1007/s10048-013-0383-8) contains supplementary material, which is available to authorized users.

E. C. Schulte · J. Winkelmann
Klinik und Poliklinik für Neurologie, Klinikum rechts der Isar, Technische Universität München, Ismaningerstr. 22, 81675 Munich, Germany

E. C. Schulte · B. Schormair · E. Graf · S. Eck · P. Lichtner · T. Meitinger · J. Winkelmann
Institut für Humangenetik, Helmholtz Zentrum München, Munich, Germany

D. C. Ellwanger · H.-W. Mewes
Department of Genome-Oriented Bioinformatics, Technische Universität München, Freising, Germany

S. Dihanich · C. Manzoni · P. A. Lewis
Department of Molecular Neuroscience, UCL Institute of Neurology, University College London, London, UK

K. Stangl · H. H. Klünemann
Klinik und Poliklinik für Psychiatrie, Psychosomatik und Psychotherapie, Universität Regensburg, Regensburg, Germany

B. Schormair · P. Lichtner · T. Meitinger · J. Winkelmann
Institut für Humangenetik, Technische Universität München, Munich, Germany

B. Mollenhauer · C. Trenkwalder
Paracelsus Elena Klinik, Kassel, Germany

B. Mollenhauer
Neurochirurgische Klinik, Georg August Universität Göttingen, Göttingen, Germany

D. Haubenberger · W. Pirker · A. Zimprich
Department of Neurology, Medical University of Vienna, Vienna, Austria

T. Brücke
Neurologische Klinik, Wilhelminenspital, Vienna, Austria

A. Peters
Institut für Epidemiologie II, Helmholtz Zentrum München, Munich, Germany

C. Gieger
Institut für Genetische Epidemiologie, Helmholtz Zentrum München, Munich, Germany

C. Trenkwalder
Neurologische Klinik, Georg August Universität Göttingen, Göttingen, Germany

T. Meitinger · J. Winkelmann
Munich Cluster for Systems Neurology (SyNergy), Munich, Germany

P. A. Lewis
School of Pharmacy, University of Reading, Whiteknights, Reading, UK

J. Winkelmann (✉)
Department of Neurology and Neurosciences, Stanford University, Palo Alto, USA
e-mail: winkelmann@lrz.tum.de

Published online: 16 November 2013

non-synonymous variants in both genes in cases and controls. An in silico multi-model bioinformatics analysis was used to prioritize identified variants in *LRRK1* for functional follow-up. However, protein expression, subcellular localization, and cell viability were not affected by the identified variants. Although it has yet to be proven conclusively that variants in *LRRK1* are indeed causative of PD, our data strengthen a possible role for *LRRK1* in addition to *LRRK2* in the genetic underpinnings of PD but, at the same time, highlight the difficulties encountered in the study of rare variants identified by next-generation sequencing in diseases with autosomal dominant or complex patterns of inheritance.

Keywords Parkinson's disease · *LRRK1* · *EEF1D* · Exome sequencing

Introduction

Characterized by resting tremor, bradykinesia, rigidity, and postural instability, Parkinson's disease (PD) is a prominent neurodegenerative disorder. Genetic factors contribute to the risk of PD—both sporadic and familial. Although up to 20 % of PD cases are believed to be familial [1, 2], thus far, rare genetic variants in only a few genes have been unequivocally shown to underlie these familial forms. They include *PARK2*/*PARKIN*, *PINK1*, *PARK7*/*DJ-1*, *SNCA*, and *LRRK2* [3–8]. While all of these were identified by classical linkage analysis in large, multi-generation families, recently, next-generation sequencing has enabled the identification of disease-causing variants in smaller families and—what is especially important with regard to the investigation of neurodegenerative conditions with an onset late in life—without the need of genotypic information from more than one generation of affected individuals. Recently, exome sequencing was used to identify *VPS35* as an additional gene involved in late-onset familial PD [9, 10]. Still, to date, the identified genes only explain a small portion of the genetic burden in familial PD. It is likely that genetic factors involved in bringing about the PD phenotype comprise both genetic variants of strong effect as well as variants of weaker effect which contribute to disease risk or phenotypic modification. A thorough understanding of the entire spectrum of genetic alterations implicated in the disease is necessary to better understand disease pathogenesis and to provide more specific treatment options in the future.

Here, we describe whole exome sequencing in a German family with autosomal dominant late-onset PD in whom known PD-linked mutations has previously been excluded in an attempt to pinpoint the disease-causing genetic variant. Two variants in leucine-rich repeat kinase 1 (*LRRK1*) and eukaryotic translation elongation factor 1 delta (*EEF1D*) emerged as the best candidate variants.

Materials and methods

Participants

The family was evaluated by neurologists specializing in movement disorders. All family members received a detailed neurologic exam. Information on deceased family members was gathered from medical and family records. Cases and controls used in genotyping and variant screening have been reported previously [9, 11] and are described in more detail in the supplement. Ethics review board approval and participants' written informed consent were obtained prior to the initiation of the study.

Analysis of copy number variation

Genome-wide copy number variant (CNV) analysis was carried out using Affymetrix Whole-Genome 2.7 M Array in conjunction with the Chromosome Analysis Suite with a confidence index of 85, a minimum homozygous region size of 10 kb and a minimum probe count of 5.

Exome sequencing

Exome sequencing was performed on a Genome Analyzer IIx (Illumina) after in-solution enrichment of exonic sequences (SureSelect Human All Exon 38 Mb kit, Agilent). For both samples, two lanes of a flow cell were sequenced, each as 54-bp paired-end runs. Read alignment was carried out with BWA (version 0.5.8) to the human genome assembly hg19. Single nucleotide variants (SNVs) and small insertions and deletions (indels) were detected with SAMtools (version 0.1.7). Prior to exome sequencing, presumably causal mutations in known Parkinson's disease genes (*SNCA*, *PARK2*, *DJ1*, *PINK1*, and *LRRK2* (p.G2019S only)) had been excluded. Moreover, no known PD-linked variants were identified in either V:8 or V:17 by exome sequencing.

Genotyping

All 15 candidate variants were genotyped in 862 cases (376 of German (age 71.1±9.4 years, 31.6 % female) and 486 of Austrian (age 58.7±11.3 years, 35.4 % female) origin) and 1,014 population-based controls pertaining to the KORA-AGE cohort (age 76±6.6 years, 50.1 % female) using MALDI-TOF mass spectrometry on the Sequenom® platform. Association was tested using the allelic test in PLINK.

Variant screening

We used Idaho®'s LightScanner high-resolution melting curve analysis to screen the eight coding exons of *EEF1D* for

variants in the same set of 862 cases and 1,014 controls. For technical reasons, a part of exon 3 of *EEF1D* could not be evaluated. For *LRRK1*, the ras of complex proteins (ROC, p.631 to 826), the C-terminal of ROC (COR, p.827 to 1241), and the kinase (p.1242 to 1525) domains as determined by an InterproScan sequence search or extracted from the literature [12] were screened. In the case of altered melting patterns suggestive of variants, Sanger sequencing ensued. Significance was judged using the χ^2 test.

Bioinformatic prioritization of variants

We collected a set of reference SNVs known to impair LRRK1 function. After computing a multiple sequence alignment using ClustalW based on *LRRK1*/*LRRK2* pairs in 18 organisms, we introduced mutations into *LRRK1* which mimic non-synonymous *LRRK2* mutations related to PD (rs33939927 (p.Arg1441Gly), rs35801418 (p.Tyr1699Cys), rs34637584 (p.Gly2019Ser), rs35870237 (p.Ile2020Thr)), and added *LRRK1* variants with a reported functional impact [12] to the set of reference SNVs. An in silico approach was applied to determine the disease potential of reference SNVs and the novel, non-synonymous *LRRK1* variants. To reduce the error rates of single models in predicting the functional effect of a given variant on the protein, we implemented a multi-model ensemble combining prediction results of six publically available prediction algorithms into a combined Pscore (Fig. 3a, online methods). Additionally, a Dscore was computed, scoring the severity of structural changes between the wild type and the variant peptide based on the mean square deviation (online methods). By combining the Pscore and the Dscore, we computed a single overall mutation score (Mscore), rating the disease potential of an SNV between 0 (harmless polymorphism) and 1 (disease mutation) (online methods). SNVs were then ranked by their Mscore, and hierarchical clustering was conducted by Ward's minimum variance agglomeration method and Euclidean distance matrix and analyzed in *R*; *p* values were calculated by multi-scale bootstrap resampling [13]. Also see supplement.

Cellular analyses

Cellular analyses were carried out as previously described [14]. For a detailed description, see supplement.

Results

Pedigree and clinical phenotype

We describe a five-generation family from Southern Germany in which six members were affected by PD and the pattern of inheritance seems to be autosomal dominant with reduced penetrance (Fig. 1). Clinical assessment revealed a tremor-dominant, levodopa-responsive Parkinson's syndrome with an age of onset at 56.7±1.15 years in all living affected individuals (V:8, V:9, and V:17, Online Resources Tabl 1). Further, all three affected individuals also showed positive Babinski signs and suffered episodes of depression. Mild to moderate cognitive impairment especially with regard to visuoconstruction, memory, and attention was present in all individuals. Dopamine transporter SPECT (DAT-SPECT) performed in two affected individuals (V:8 and V:17) revealed reduced tracer uptake in the putamen and asymmetrically in the caudate nucleus, in line with a diagnosis of PD.

The affected parent and aunt (IV:5 and IV:7) of the proband died before initiation of the study. An additional cousin, V:1, had Parkinson's syndrome but also suffered from multiple sclerosis. She also died before initiation of the study. Lastly, a second cousin removed by four generations is also known to suffer from late-onset PD. The prevalence of PD in the general population is approximately 1 % [15]. Accordingly, we expected to find at least one phenocopy in this extended pedigree of 114 individuals. Since no additional family members on her side of the family showed signs of PD and since she shared none of the candidate variants common to the other three affected individuals examined, we concluded that it is unlikely that PD in her case is due to the same genetic variant as in the other affected individuals.

Identification of candidate variants by exome sequencing and segregation analysis

A genome-wide CNV scan revealed no structural variation≥ 10 kb common to two affected members of the family (V:8 and V:17, Fig. 1). Exome sequencing was performed for the same individuals. This generated 6.57 gigabases (Gb) of alignable sequence for V:8 (average coverage=70.93, >8× coverage=90.65 %) and 6.67 Gb for V:17 (average coverage=76.29, >8× coverage=92.23 %). All detected variants shared by the two affected individuals (16,283 variants) were filtered against variants annotated in dbSNP132 as well as in-house exomes (n=1076) of individuals with unrelated diseases and variants with a minor allele frequency (MAF)≥0.01 were excluded from the follow-up, leaving 71 coding variants. Of these, 36 variants were predicted to alter the amino acid sequence (i.e., missense, nonsense, stop-loss, splice site, or frameshift variants and indels) and were genotyped in a third affected individual (V:9) (Online Resources Fig 1). Fifteen variants in 15 genes were present in all three affected individuals and were pursued further by Sanger sequencing-based testing for segregation in 32 members of the family belonging to generation V. Under the assumption that a given variant would be causal for PD, penetrances ranged between 30 and 50 %, with variants in

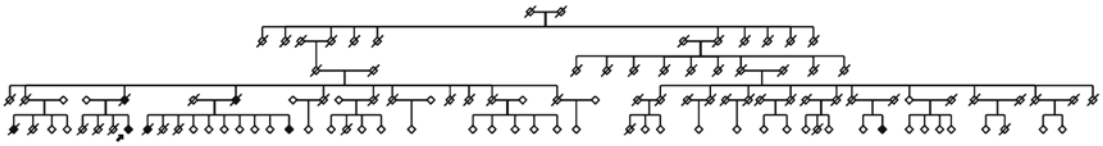


Fig. 1 Pedigree of family used for exome sequencing. *Open symbols* indicate unaffected family members; affected individuals are denoted by *closed symbols*. An *arrow* denotes the proband. Sex was obscured and birth order was altered to protect privacy. A *diagonal line* indicates a deceased individual

LRRK1, *EEF1D*, and *ARHGAP39* reaching the highest predicted penetrances (Table 1).

Frequency assessment of candidate variants in a case/control cohort

We genotyped the remaining 15 variants in a case/control sample, consisting of 862 individuals with PD and 1,014 KORA-AGE general population controls (Table 1). Two assays (*UGT1A9* p.Val167Ala, *TUBB6* p.Thr275Ala) did not meet quality control thresholds and were excluded from the analysis. The remaining 13 variants were, overall, very rare. Six (*EEF1D* p.Ala549Val, *MUC17* p.Gln4310X, *CCDC60* p.Arg155His, *NAAA* p.Arg211Trp, *PTPRN2* p.Glu317Lys, and *GLP2R* p.Ile61Met) were validated in the proband but were otherwise not found again in the 1,876 individuals tested. *FCGBP* p.Glu4657fs was present in one additional PD patient but not in controls, and *ZNF438* p.Thr454Ile was found in the proband and one control. Four additional variants annotated in dbSNP132 were identified at similar frequencies in cases and controls (*ARHGAP39* p.Arg667Gln and *MFSD3* p.Met311Thr) or were more common in controls than in cases (*AQP4* p.Met202Thr and *BRCA2* p.The1524Val). *LRRK1* p.Arg1261Gln was found in eight controls and in four cases of our case/control sample (MAF 0.23/0.40 %). However, in four other control samples (680 additional KORA general population controls (0.07 %), 1,076 in-house exomes (0.05 %), 1,000 genomes (0.00 %), and NHLBI-ESP exomes (0.09 %)), MAFs were significantly lower, and the variant was, therefore, also analyzed further.

Mutational screening of *EEF1D* and *LRRK1* in case/control cohort

While no single clear candidate for a causal variant emerged, two genes—*EEF1D* and *LRRK1*—were interesting with regard to functional considerations and predicted penetrance for PD in the family. The translation machinery has recently been implicated in PD pathogenesis [16, 17]. Also, the *EEF1D* p.Ala549Val variant was not found again in 3,064 individuals (genotyping cohort plus in-house exomes) and was also not annotated in the 1000 Genomes database. *LRRK1*, the paralog of the well-established PD gene *LRRK2*, has been shown to regulate endosomal protein transport, thus linking it to the lysosomal pathway [18] which may be compromised in PD [19, 20]. Formation of heterodimers between LRRK1 and LRRK2 has also been reported [21, 22]. We screened the coding regions of these genes in 862 Austrian and German PD cases and 1,014 controls searching for additional variants. This cohort comprised the same individuals used for the above frequency assessment of exome variants. We identified seven (six non-synonymous, one del) novel variants predicted to change the amino acid sequence of *EEF1D*. These were rare and occurred with similar frequencies in cases (five individuals with a variant) and controls (four individuals with a variant) ($p>0.5$, χ^2 test, Online Resources Tab 2). Variants did not cluster in a specific part of the gene (Fig. 2). The ROC, COR, and kinase domains of *LRRK1* harbored a total of 20 novel amino acid sequence-changing variants (19 non-synonymous, 1 del) and 2 previously reported non-synonymous variants (rs56003881, rs41531245). Variants were found at similar frequencies in both groups (30 in cases, 31 in controls) ($p>0.5$, χ^2 test, Online Resources Tab 2). While small numbers preclude quantitative analyses, it is noteworthy that within the first 20 bp of the kinase domain, variants were present in both cases and controls, while beyond p.1262, all non-synonymous variants identified in the kinase domain occurred in cases only (Fig. 2). None of the individuals harboring *LRRK1* variants were also positive for known *LRRK2* variants p.Arg1441Cys, p.Tyr1699Cys, p.Gly2019Ser, or p.Ile2020Thr.

Prioritization of *LRRK1* variants using a novel bioinformatics algorithm

Since heterodimer formation between *LRRK1* and *LRRK2* has been described [21, 22], we decided to further assess the identified variants in *LRRK1*. To this end, we used a novel bioinformatics algorithm based on a multi-model ensemble of prediction algorithms and structural analysis to select variants in *LRRK1* for functional follow-up. Mutation scores were calculated for the 19 novel, non-synonymous *LRRK1* variants identified in both cases and controls, the *LRRK1* variants (p.Lys746Glu, p.Phe1022Cys, p.Gly1411Arg and p.Ile1412Thr) corresponding to four known pathogenic *LRRK2* mutations (p.Arg1441Gly, p.Tyr1699Cys,

Table 1 Fifteen rare, non-synonymous variants shared by individuals V:8, V:9, and V:17 of family PARK_0001

Genomic position (hg19)	Gene	Alleles	In-house exomes	Genotyping		dbSNP132	1,000 Genomes	NHLBI-ESP	Transcript	Variation		Penetrance for PD (%)
				Cases	Controls					Nucleotide	Amino acid	
			$n=1076$	$n=862$	$n=1,014$		AF/DP	(EA only, hetero/ total individuals)				$n=32$
chr15:101593219	*LRRK1*	1	1	4	8		Not found	8/4,250	NM_024652.3	c.3782G>A	p.Arg1261Gln	50.00
chr8:144662740	*EEF1D*	1	0	1	0		Not found	Not found	NM_032378.4	c.1646G>A	p.Ala549Val	50.00
chr8:145771154	*ARHGAP39*	1	3	2	2	rs11994207	0.004:1,747	20/4,279	NM_025251.1	c.2000C>T	p.Arg667Gln	50.00
chr7:100694947	*MUC17*	1	0	1	0		Not found	Not found	NM_001040105.1	c.12928C>T	p.Gln4310X	42.85
chr8:145736082	*MFSD3*	1	3	4	2	rs35905340	0.003:1,950	19/4,280	NM_138431.1	c.932 T>C	p.Met311Thr	42.85
chr12:119926578	*CCDC60*	1	0	1	0		Not found	Not found	NM_178499.3	c.464G>A	p.Arg155His	42.85
chr19:40366263	*FCGBP*	1	1	2	0		Not found	Not found	NM_003890.2	c.13971_13971delC	p.Glu4657fs	42.85
chr4:76846923	*NAAA*	1	0	1	0		Not found	Not found	NM_001042402.1	c.631G>A	p.Arg211Trp	37.50
chr7:157931118	*PTPRN2*	1	1	1	0		Not found	3/4,297	NM_002847.3	c.949C>T	p.Glu317Lys	37.50
chr10:31137973	*ZNF438*	1	0	1	1		Not found	Not found	NM_001143769.1	c.1361G>A	p.Thr454Ile	37.50
chr2:234581080	*UGT1A9*	1	1	n/a	n/a		Not found	13/4,287	NM_021027.2	c.500 T>C	p.Val167Ala	33.33
chr18:24440758	*AQP4*	1	2	1	7	rs72557975	0.003:2,587	9/4,291	NM_004028.3	c.605A>G	p.Met202Thr	33.33
chr13:32913062	*BRCA2*	1	1	2	6	rs56386506	Not found	1/4,299	NM_000059.3	c.4570 T>G	p.Phe1524Val	30.00
chr17:9729563	*GLP2R*	1	0	1	0		Not found	Not found	NM_004246.1	c.183C>G	p.Ile61Met	30.00
chr18:12325722	*TUBB6*	1	0	n/a	n/a		Not found	16/4,284	NM_032525.1	c.823A>G	p.Thr275Ala	30.00

The rare variants common to all three affected individuals were genotyped in 862 cases and 1,014 controls. Penetrance with regard to the PD phenotype was assessed in 32 family members belonging to the same generation as the affected individuals

AF allele frequency, *DP* sequencing depth (number of reads), *EA* European American, *hetero* heterozygotes

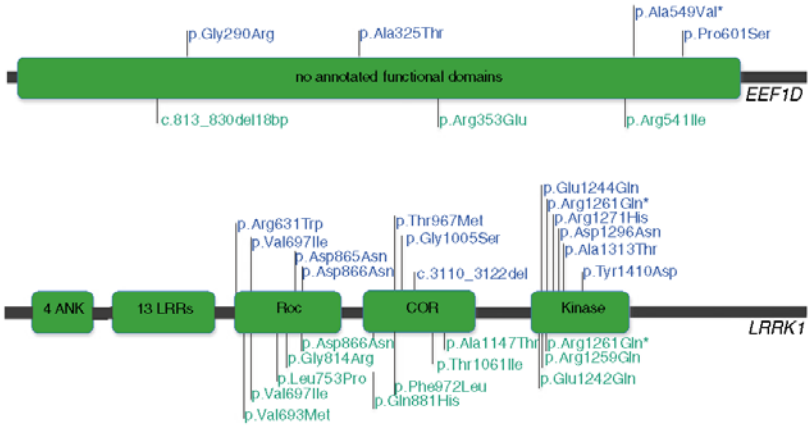


Fig. 2 Location of *EEF1D* and *LRRK1* variants identified in variant screening in relation to known functional domains. An *asterisk* denotes the variant identified by exome sequencing Variants printed in blue and annotated above the gene were found in cases, variants in green and below the gene were found in controls

p.Gly2019Ser, and p.Ile2020Thr) and three artificial variants known to abolish *LRRK1* GTP-binding (p.Lys651Ala) and kinase activity (p.Lys746Gly and p.Lys1270Trp) (Fig. 3a) [12]. Hierarchical clustering showed that three of the novel variants (p.Arg631Trp, p.Arg1271His and p.Tyr1410Asp)—present only in PD cases—clustered with the *LRRK1* equivalents of *LRRK2* p.Arg1441Gly, p.Tyr1699Cys, p.Gly2019Ser, and p.Ile2020Thr as well as the kinase- and GTP-binding dead amino acid substitutions (Fig. 3b). Accordingly, these three variants in addition to the initial variant identified by exome sequencing (p.Arg1261Gln) were selected for functional follow-up.

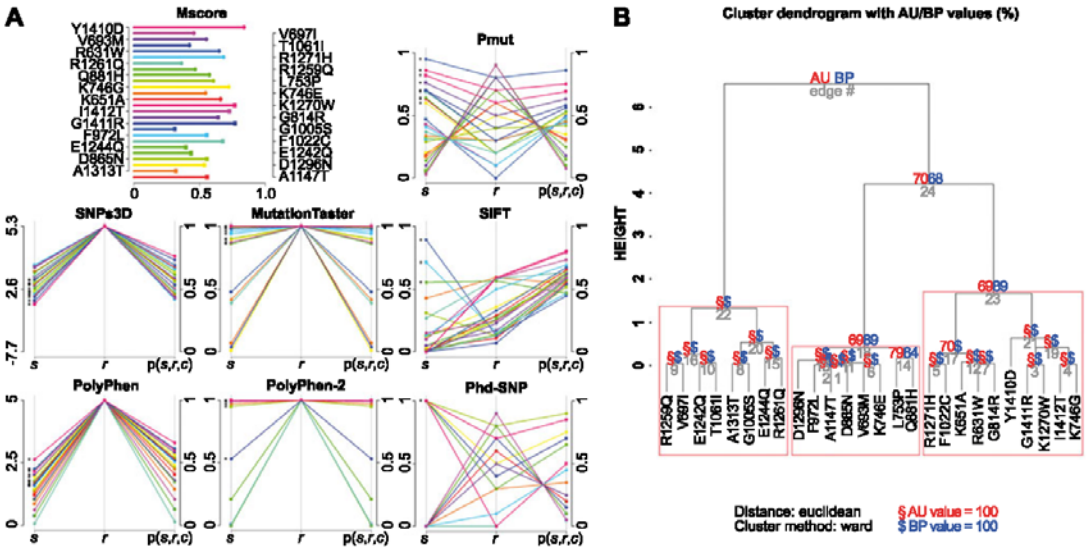


Fig. 3 Prediction of pathogenic potential of newly identified variants. **a** For each variant (*colored lines*) the predicted score s of an individual algorithm, its reliability r, and the transformed score $p(s, r, c)$ are shown. Variants holding a predicted disease-causing potential (class=1) were respectively marked with an *asterisk*. The diverse results among each single algorithm motivated the calculation of one combined score (*Pscore*), which was adjusted by additional structural analyses (*Dscore*) resulting in a mutation score (*Mscore*). The highest scoring variant is p.Tyr1410Asp (Mscore=0.839), a variant only present in PD cases, followed by the *LRRK2* equivalent of Gly2019Ser (Mscore=0.771), the loss of autophosphorylation mutation Lys1270Trp (Mscore=0.768), and two variants abolishing kinase activity: Ile1412Thr (Mscore=0.728) and Lys746Gly (Mscore=0.723). **b** Hierarchical clustering with Ward's minimum variance agglomeration method and Euclidean distance matrix shows that three of the novel variants which were only found in individuals with PD (p.Arg631Trp, p.Arg1271His, and p.Tyr1410Asp) cluster with the *LRRK1* equivalents of *LRRK2* p.Arg1441Gly, p.Tyr1699Cys, p.Gly2019Ser, and p.Ile2020Thr as well as the *LRRK1* kinase- and GTP-binding dead amino acid substitutions

Functional assessment of *LRRK1* variants

In SHSY5Y neuroblastoma cells, levels of protein expression as assessed by Western blot were not changed by any of the four newly identified variants or the artificial variants ablating GTP-binding (p.Lys651Ala) or kinase activity (p.Lys1270Trp) (Fig. 4a). Likewise, the presence of these variants was not associated with significant toxicity as measured by MTT assay (Fig. 4b) and did not alter cytoplasmic localization of myc-tagged LRRK1 (Fig. 4c). Like others [23, 24], we could not detect LRRK1 kinase activity above the background and could not determine whether activity was altered by the variants.

Discussion

In an unbiased, whole exome approach, we identified a variant in *LRRK1* (p.Arg1261Gln) as a candidate for a potentially causal variant in familial PD. Although this finding is intriguing and functionally plausible, we are unable to conclude that this is indeed the cause of PD in our family. For one, the variant was found in both cases and controls in our larger case/control sample. Yet, the actual variant frequency in controls appears to be lower than that found in the KORA-AGE cohort (8 in 1,014 KORA-AGE vs. 0 in 1,000 genomes, 1 in 1,076 in-house exomes, and 1 in additional 680 KORA controls), and it could be possible that KORA-

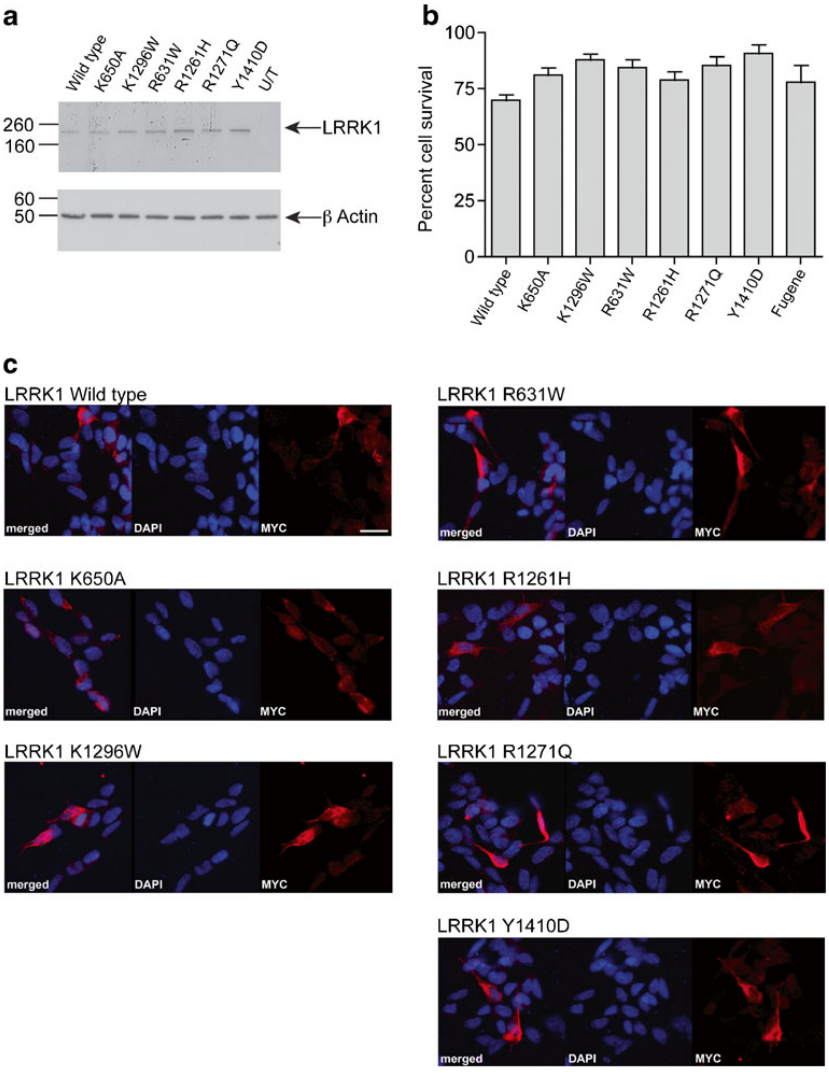


Fig. 4 Cellular expression of LRRK1 and mutant variants. **a** Western blot analysis of myc-tagged LRRK1 expression in SHSY5Y cells with beta actin-loading control. **b** Analysis of LRRK1 toxicity as measured by MTT assay in SHSY5Y cells. No significant toxicity was associated with wild-type LRRK1, artificial mutations in LRRK1, or disease-associated coding changes. Data is expressed as percentage of untransfected control cells, mean, and standard error measurement displayed. **c** Immunocytochemistry analysis of myc-tagged LRRK1 constructs. Staining for myc is shown separately and merged. All tagged constructs displayed a diffuse cytoplasmic staining pattern. *Scale bar* =20 μm

AGE is enriched for the *LRRK1* p.Arg1261Gln variant due to a founder effect or that is also present in controls because its PD-related nature depends on a specific genetic context. Secondly, the other 14 identified rare variants shared by all three affected individuals also represent potential candidates. Especially, *EEF1D* p.Ala549Val, which was not found again in any individual genotyped (n=3064, case/control sample and in-house exomes) or the 1,000 genomes or the NHLBI-ESP exomes, represents another good contender. In general, these findings draw attention to the fact that in many cases, very large populations will need to be evaluated to conclusively judge the disease-related nature of a rare variant such as those identified by exome sequencing. Most recent studies show that while the power to detect associations for genes harboring rare variants varies widely across genes, only <5 % of genes achieved 80 % power even assuming high odds ratios (OR) of 5 when tested in 400 cases and 400 controls [25]. Ultimately, it is also possible that the truly causal variant was not picked up in this study because it lies outside the targeted regions of the exome.

The fact that both the *LRRK1* and *EEF1D* variants were also found in three unaffected members of the family each per se does not contradict potential causality as it is known from other autosomal dominant forms of PD that even among members of a single family, penetrance of known PD mutations can vary widely. Of individuals who harbor the *LRRK2* p.Gly2019Ser mutation, for example, only 28 % will develop PD by the age of 59 years [26]. Thus, predicted penetrances of the variants identified in our family are in line with what is reported in the literature for other forms of autosomal dominant PD.

Both LRRK1 and LRRK2 belong to the Roco family of proteins. These proteins are likely to perform a number of different functions as they are not only characterized by a conserved Ras-like GTPase domain called ROC and a characteristic COR domain of unknown function but also harbor kinase and protein–protein interaction domains [27]. While a contribution of mutations in *LRRK2* to disease development in PD seems firmly established, the role of *LRRK2* paralog *LRRK1* is unclear. It is known that LRRK1 and LRRK2 form heterodimers in HEK293T cells [21, 22] and that both proteins are expressed in similar tissues. Accordingly, a hypothetical role for LRRK1 in addition to LRRK2 is plausible.

The precise role of LRRK2 in PD pathogenesis, however, has not been fully established. Accordingly, even if one were to assume a similar role of LRRK1 in disease development, exactly which function of the protein would be involved in the disease is uncertain. Therefore, the lack of a functional effect on protein expression levels, subcellular localization, and cell viability of the *LRRK1* variants we identified does not equate to a definitely missing role of *LRRK1* in PD. Interestingly, *Lrrk1* has also been implicated in a quantitative trait locus for dopaminergic amacrine cell number in the murine retina [28]. Further, the recent link between LRRK1 and endosomal protein trafficking [18, 29] is also very intriguing in light of the fact that one of the postulated pathomechanisms for LRRK2 in PD involves aberrant lysosomal function or localization [20, 30, 31].

However, studies addressing the role of both common and rare genetic variants in *LRRK1* with regard to PD do not seem to substantiate the conception of *LRRK1* as a "PD gene" [21, 32–34]. While none of these studies found a common or rare variant clearly linked to PD, nonetheless, across three studies ([32, 34] and our study), the p.Thr967Met variant has only been identified in seven out of 1,552 cases but not in any of 1,535 controls ($p_{\text{nominal}} \leq 0.01$, χ^2 test; not significant after correction for multiple testing, OR=14.90 (95 % confidence interval=0.85 to 261.18)). Yet, in a family with multiple individuals with PD, the variant did not segregate with the phenotype [34]. Evidence also suggests that variants in *LRRK1* are able to modify the PD phenotype. Tunisian individuals with *LRRK2* p.Gly2019Ser showed a trend towards a 6-year earlier age of onset when they also carried *LRRK1* p.Leu416Met [20]. In line with this, it has been demonstrated for other genetic disorders that genetic variants at related loci can both drive and modify a given phenotype depending on the variant and the genetic context [35]. At the moment, both functional and genetic data addressing a role of *LRRK1* as a PD gene are inconclusive. Nonetheless, it is interesting that in our unbiased whole exome approach, one of the top candidate variants for a genetic factor underlying or contributing to the PD phenotype in our family is a non-synonymous variant in the kinase domain of *LRRK1* and that other individuals suffering from PD harbor *LRRK1* variants (p.Tyr1410Asp) only with one amino acid away from the location which is equivalent to the prominent p.Gly2019Ser mutation of *LRRK2*.

In summary, all variants shared by the three affected individuals in our family represent potential causal or modifying alleles in PD. As is the case for all rare and very rare variants, establishing definitive causality is difficult, and only the identification of additional PD families harboring these variants or their analysis in sufficiently powered case/control studies will tell whether these variants do indeed hold a role in bringing about PD.

Acknowledgments We are gratefully indebted to Katja Junghans, Susanne Lindhof, Jelena Golic, Sybille Frischholz, and Regina Feldmann at the Institut für Humangenetik, Helmholtz Zentrum München, Munich, Germany, for their expert technical assistance in performing Sequenom genotyping and Light Scanner analyses. This study was funded by in-house institutional funding from Technische Universität München and Helmholtz Zentrum München, Munich, Germany. Recruitment of case and control cohorts was supported by institutional (HelmholtzZentrumMünchen, Munich, Germany) and government funding from the German Bundesministerium für Bildung und Forschung (03.2007-02.2011 FKZ 01ET0713). PAL is funded by a Parkinson's UK research fellowship

(F1002). The work performed by PAL, SD, and CM was also funded in part by grants from the Michael J. Fox Foundation LRRK2 consortium and by the Wellcome Trust/MRC Joint Call in Neurodegeneration award (WT089698) to the UK Parkinson's Disease Consortium (UKPDC) whose members are from the UCL Institute of Neurology, the University of Sheffield, and the MRC Protein Phosphorylation Unit at the University of Dundee.

References

1. Payami H, Larsen K, Bernard S, Nutt J (1994) Increased risk of Parkinson's disease in parents and siblings of patients. Ann Neurol 36:659–661
2. Bonifati V, Fabrizio E, Vanacore N, De Mari M, Meco G (1995) Familial Parkinson's disease: a clinical genetic analysis. Can J Neurol Sci 22:272–279
3. Zimprich A, Biskup S, Leitner P et al (2004) Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron 44:601–607
4. Bonifati V, Rizzu P, van Baren MJ et al (2003) Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science 299:256–259
5. Kitada T, Asakawa S, Hattori N et al (1998) Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature 392: 605–608
6. Paisan-Ruiz C, Jain S, Evans EW et al (2004) Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. Neuron 44:595–600
7. Polymeropoulos MH, Lavedan C, Leroy E et al (1997) Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 276:2045–2047
8. Valente EM, Abou-Sleiman PM, Caputo V et al (2004) Hereditary early-onset Parkinson's disease caused by mutations in PINK1. Science 304:1158–1160
9. Zimprich A, Benet-Pages A, Struhal W et al (2011) A Mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. Am J Hum Genet 89:168–175
10. Vilarino-Güell C, Wider C, Ross OA et al (2011) VPS35 mutations in Parkinson disease. Am J Hum Genet 89:162–167
11. Wichmann HE, Gieger C, Illig T (2004) KORAgen—resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen 67(Suppl 1):26–30
12. Korr D, Toschi L, Donner P, Pohlenz HD, Kreft B, Weiss B (2006) LRRK1 protein kinase activity is stimulated upon binding of GTP to its Roc domain. Cell Signal 18:910–920
13. Shimodaira H (2004) Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. Ann Statistics 32:2616–2641
14. Greggio E, Lewis PA, van der Brug MP et al (2007) Mutations in LRRK2/dardarin associated with Parkinson disease are more toxic than equivalent mutations in the homologous kinase LRRK1. J Neurochem 102:93–102
15. Elbaz A, Bower JH, Maraganore DM et al (2002) Risk tables for parkinsonism and Parkinson's disease. J Clin Epidemiol 55:25–31
16. Chartier-Harlin MC, Dachsel JC, Vilarino-Guell C et al (2011) Translation initiator EIF4G1 mutations in familial Parkinson disease. Am J Hum Genet 89:398–406
17. Gehrke S, Imai Y, Sokol N, Lu B (2010) Pathogenic LRRK2 negatively regulates microRNA-mediated translational repression. Nature 466:637–641
18. Hanafusa H, Ishikawa K, Kedashiro S et al (2011) Leucine-rich repeat kinase LRRK1 regulates endosomal trafficking of the EGF receptor. Nat Comm 2:158
19. Sidransky E, Nalls MA, Aasly JO et al (2009) Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N Engl J Med 361:1651–1661
20. Tong Y, Yamaguchi H, Giaime E et al (2010) Loss of leucine-rich repeat kinase 2 causes impairment of protein degradation pathways, accumulation of alpha-synuclein, and apoptotic cell death in aged mice. Proc Natl Acad Sci U S A 107:9879–9884
21. Dachsel JC, Nishioka K, Vilarino-Güell C et al (2010) Heterodimerization of Lrrk1-Lrrk2: implications for *LRRK2*-associated Parkinson disease. Mech Age Dev 131:210–214
22. Klein CL, Rovelli G, Springer W, Schall C, Gasser T, Kahle PJ (2009) Homo- and heterodimerization of ROCO kinases: LRRK2 kinase inhibition by the LRRK2 ROCO fragment. J Neurochem 111: 703–715
23. Deng X, Dzamko N, Prescott A et al (2011) Characterization of a selective inhibitor of the Parkinson's disease kinase LRRK2. Nat Chem Biol 7:203–205
24. Civiero L, Vancraenenbroeck R, Belluzzi E et al (2012) Biochemical characterization of highly purified leucine-rich repeat kinases 1 and 2 demonstrates formation of homodimers. PLoS One 7:e43472
25. Tennessen JA, Bigham AW, O'Connor T et al (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337:64–69
26. Healy DG, Falchi M, O'Sullivan SS et al (2008) Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case–control study. Lancet Neurol 7:583–590
27. Marin I, van Egmond WN, van Haastert PJM (2008) The Roco protein family: a functional perspective. FASEB 22:3103–3110
28. Whitney IE, Raven MA, Ciobanu DC, Williams RW, Reese BE (2009) Multiple genes on chromosome 7 regulate dopaminergic amacrine cell number in the mouse retina. Invest Ophthal Vis Sci 50:1996–2003
29. Ishikawa K, Nara A, Matsumoto K, Hanafusa H (2012) EGFR-dependent phosphorylation of leucine-rich repeat kinase LRRK1 is important for proper endosomal trafficking of EGFR. Mol Biol Cell 23:1294–1306
30. Gómez-Suaga P, Luzón-Toro B, Churamani D et al (2012) Leucine-rich repeat kinase 2 regulates autophagy through a calcium-dependent pathway involving NAADP. Hum Mol Genet 21:511–525
31. Dodson MW, Zhang T, Jiang C, Chen S, Guo M (2012) Roles of the *Drosophila LRRK2* homolog in Rab7-dependent lysosomal positioning. Hum Mol Genet 21:1350–1363
32. Haugarvoll K, Toft M, Ross OA, White LR, Aasley JO, Farrer MJ (2007) Variants in the LRRK1 gene and susceptibility to Parkinson's disease in Norway. Neurosci Lett 416:299–301
33. Chung SJ, Armasu SM, Biernacka JM et al (2011) Common variants in PARK loci and related genes and Parkinson's disease. Mov Disord 26:280–288
34. Taylor JP, Hulihan MM, Kachergus JM et al (2007) Leucine-rich repeat kinase 1: a paralog of LRRK2 and a candidate gene for Parkinson's disease. Neurogenetics 8:95–102
35. Davis EE, Zhang Q, Liu Q et al (2011) TTC21B contributes both causal and modifying alleles across the ciliopathy spectrum. Nat Genet 43:189–196

2.10 Schulte *et al.*, Rare variants in *PLXNA4* and Parkinson`s disease, *PLoS One*, 2013[X]

Personal contributions: I recruited part of the family and participated in designing the study. I also analyzed the exome sequencing data. I designed the multiplex PCRs for Sequenom®-based genotyping and analyzed genotyping data for the frequency assessment in the Parkinson's disease case/control sample. I carried out the segregation analysis in the family by Sanger sequencing. I participated in genome-wide genotyping and linkage analysis. Furthemore, I supervised and helped in the LightScanner® high-resolution melting curve analysis and follow-up Sanger sequencing of the coding regions of *PLXNA4* in the Parkinson's disease case/control sample and the analysis of the LightScanner® data and calculated the burden tests. I generated the patient-specific fibroblast cell line, performed the cell viability assay and the immunocytochemistry. Lastly, I wrote the manuscript and designed all tables and figures except for Supplementary Table 2 and Figures 2 and 4.

OPEN ACCESS Freely available online

PLOS ONE

Rare Variants in *PLXNA4* and Parkinson's Disease

Eva C. Schulte[1,2], Immanuel Stahl[1,2], Darina Czamara[3,4], Daniel C. Ellwanger[5], Sebastian Eck[2], Elisabeth Graf[2], Brit Mollenhauer[6,7], Alexander Zimprich[8], Peter Lichtner[2,9], Dietrich Haubenberger[8], Walter Pirker[8], Thomas Brücke[10], Benjamin Bereznai[11], Maria J. Molnar[11], Annette Peters[12], Christian Gieger[13], Bertram Müller-Myhsok[3,4], Claudia Trenkwalder[6,7], Juliane Winkelmann[1,2,4,9,14]*

1 Neurologische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität, München, Munich, Germany, **2** Institut für Humangenetik, Helmholtz Zentrum München, Munich, Germany, **3** Max-Planck Institut für Psychiatrie, Munich, Germany, **4** Munich Cluster for Systems Neurology (SyNergy), Munich, Germany, **5** Chair for Genome-Oriented Bioinformatics, Technische Universität München, Life and Food Science Center Weihenstephan, Freising-Weihenstephan, Germany, **6** Paracelsus Elena Klinik, Kassel, Germany, **7** Neurochirurgische Klinik, Georg August Universität, Göttingen, Germany, **8** Department of Neurology, Medical University of Vienna, Vienna, Austria, **9** Institut für Humangenetik, Technische Universität München, Munich, Germany, **10** Department of Neurology, Wilhelminenspital, Vienna, Austria, **11** Center for Molecular Neurology, Department of Neurology, Semmelweis University, Budapest, Hungary, **12** Institute for Epidemiology II, Helmholtz Zentrum München, Munich, Germany, **13** Institute for Genetic Epidemiology, Helmholtz Zentrum München, Munich, Germany, **14** Department of Neurology and Neurosciences, Stanford University, Palo Alto, California, United States of America

Abstract

Approximately 20% of individuals with Parkinson's disease (PD) report a positive family history. Yet, a large portion of causal and disease-modifying variants is still unknown. We used exome sequencing in two affected individuals from a family with late-onset familial PD followed by frequency assessment in 975 PD cases and 1014 ethnically-matched controls and linkage analysis to identify potentially causal variants. Based on the predicted penetrance and the frequencies, a variant in *PLXNA4* proved to be the best candidate and *PLXNA4* was screened for additional variants in 862 PD cases and 940 controls, revealing an excess of rare non-synonymous coding variants in *PLXNA4* in individuals with PD. Although we cannot conclude that the variant in *PLXNA4* is indeed the causative variant, these findings are interesting in the light of a surfacing role of axonal guidance mechanisms in neurodegenerative disorders but, at the same time, highlight the difficulties encountered in the study of rare variants identified by next-generation sequencing in diseases with autosomal dominant or complex patterns of inheritance.

Citation: Schulte EC, Stahl I, Czamara D, Ellwanger DC, Eck S, et al. (2013) Rare Variants in *PLXNA4* and Parkinson's Disease. PLoS ONE 8(11): e79145. doi:10.1371/journal.pone.0079145

Editor: Christian Wider, Centre Hospitalier Universitaire Vaudois (CHUV), Switzerland

Received July 31, 2013; **Accepted** September 18, 2013; **Published** November 11, 2013

Funding: The study was funded in its majority through institutional funding from Technische Universität München and Helmholtz Zentrum München, Munich, Germany. Dr. DH received an NINDS Intramural Competitive Fellowship and research support from the Austrian Science Fund (Erwin Schroedinger Fellowship, project#J2783-B09) and the NINDS Intramural Research Program. Dr. BB and Dr. MJM received support from the Hungarian National Innovation Office (TAMOP-4-2-1/B-03/11 KMR-2010-001). Recruitment and management of the KORA cohort is funded by both institutional (Helmholtz Zentrum München, Munich, Germany) and government funding from the German Bundesministerium für Bildung und Forschung (03.2007-02.2011 FKZ 01ET0713). DC was supported by the Deutsche Forschungsgemeinschaft (German Research Foundation) within the framework of the Munich Cluster for Systems Neurology (EXC 1010 SyNergy). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have read the journal's policy and have the following conflicts of interest to declare: Dr. ECS, Mr. IM, Dr. DC, Mr. DCE, Mr. SE and Mrs. EG report no disclosures. Dr. BM received travel compensation from Novartis and Boehringer-Ingelheim, lecturing fees from Orion and Glaxo-Smith-Kline, grant support from GE Healthcare, Boehringer-Ingelheim, Desitin, TEVA-Pharma, and serves as a consultant to Bayer-Schering Pharma and the Michael J. Fox Foundation for Parkinson's Disease Research. Dr. DH received speaker honoraria from Ipsen Pharma. Dr. WP received speaker honoraria from AOP Orphan Pharma, Medtronic Inc., Novartis, Boehringer-Ingelheim, Abbott Pharm and UCB as well as travel compensation from Ipsen Pharma, Boehringer-Ingelheim, and Medtronic Inc. Dr. BB declares no conflicts of interest. Dr. MJM serves/has served on scientific advisory boards for Genzyme Europe B.V., received speaker honoria from Roche and serves as the Editor-in-Chief of the Hungarian edition of Neurology. Dr. AZ, Dr. TB, Dr. PL, Dr. AP, Dr. CG and Dr. BMM report no disclosures. Dr. CT serves on scientific advisory boards for Boehringer Ingelheim and UCB, has received speaker honoraria from Boehringer Ingelheim, UCB, and Mundipharma as well as travel compensation from UCB, Boehringer-Ingelheim, and Mundipharma. Dr. JW serves on a scientific advisory board for UCB, has received speaker honoraria from UCB and Vifor Pharma. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: winkelmann@lrz.tum.de

Introduction

Characterized by resting-tremor, bradykinesia, rigidity, and postural instability, Parkinson's disease (PD) is one of the most prominent neurodegenerative disorders. Genetic factors contribute significantly to the risk of developing PD–both sporadic and familial. Although up to 20% of PD cases are believed to be familial [1,2], thus far, variants in only a few genes have been unequivocally shown to underlie familial PD. These include *PARK2*, *PINK1*, *PARK7*, *SNCA*, and *LRRK2* [3–8]. While all of these genes were identified by classical linkage analysis in large, multi-generation families, recently, next-generation sequencing has enabled the identification of disease-causing variants in smaller families and with an onset later in life without the need of genotypic information from more than one generation of affected individuals. By exome sequencing, *VPS35* was identified as a gene involved in late-onset familial PD [9,10]. Still, to date, the identified genes only explain a small portion of the genetic "burden" in PD. However, a thorough understanding of the genetic alterations implicated in disease development is necessary to better comprehend disease pathogenesis and to provide more specific and, thus, more effective treatment options in the future.

Here, we describe exome sequencing of a German family with autosomal dominant late-onset PD in an attempt to pinpoint the disease-causing genetic variant.

Methods

Ethics Statement

Ethics review board approval was obtained from the ethics review board at Klinikum rechts der Isar, Technische Universität München, and Bayerische Landesärztekammer, both Munich, Germany, Hessische Landesärztekammer, Frankfurt, Germany, the ethics review board at Medical University Vienna, Vienna, Austria, and the ethics review board at Semmelweis University, Budapest, Hungary. Participants' written informed consent was obtained.

Participants

All living family members received a detailed neurologic exam by neurologists specializing in movement disorders. Cases and controls used in genotyping and variant screening have been reported previously [10,11] and are described in more detail in the supplement.

Exome Sequencing

Exome sequencing was performed with DNA isolated from lymphozytes of IV:11 and IV:18 on a Genome Analyzer IIx system (Illumina) after in-solution enrichment of exonic sequences (SureSelect Human All Exon 38 Mb kit for IV:11 and 50 Mb kit for IV:18, Agilent) as 76 bp paired-end runs. Read alignment was carried out with BWA (version 0.5.8). Single-nucleotide variants and small insertions and deletions (indels) were detected with SAMtools (version 0.1.7). Raw sequencing data are available upon request.

Genotyping

All ten candidate variants tested for segregation by Sanger sequencing were genotyped in 975 cases and 1014 population-based controls pertaining to the KORA-AGE cohort using MALDI-TOF masspectrometry on the Sequenom® platform. Demographic data are given in the supplement. Association was tested by allelic statistics as implemented in PLINK.

Linkage Analysis

We genotyped six family members (IV:11, IV:14, IV:16, IV:18, IV:20 and IV:21) with oligonucleotide SNP arrays (500 K, Illumina). Parametric linkage analysis was performed using a subset of 12,875 SNPs using MERLIN and an autosomal dominant model with incomplete penetrance of 70%.

Variant Screening

We used Idaho®'s LightScanner high-resolution melting curve analysis to screen the coding regions and exon/intron boundaries of *PLXNA4* for variants. 862 cases and 940 population-based controls pertaining to the KORA-AGE cohort were included in the screening. Demographic data are given in the supplement. In the case of an altered melting pattern, Sanger sequencing ensued to identify the underlying variant. Group comparisons between cases and controls were performed for each gene and each variant separately using Fisher's Exact and χ^2 tests as appropriate.

Cell Viability and Immunocytochemistry

Cultured primary fibroblasts from IV:11 and an offspring were stained using a live/dead staining (Invitrogen) and analyzed by FACS and stained with anti-PLXNA4 (1:100, Sigma) and analyzed by fluorescence microscopy. Details are given in the supplement.

Construction of a Qualitative Systems Biological Model

To investigate the role of *PLXNA4* in the PD biological system, we applied an integrative modeling approach to construct a qualitative multifactorial interaction network linking *PLXNA4* and genetic factors associated with PD. An interactome with known and predicted interactions of *PLXNA4* and its direct neighbors was prepared based on four commonly used databases and integrated to known PD pathways from KEGG and CIDeR as well as a manual literature search. For a detailed description see supplement.

Results

Pedigree and Clinical Phenotype

We describe a five-generation family from Central Germany in which four members were affected by PD and the pattern of inheritance seems to be autosomal dominant with reduced penetrance (Figure 1). Clinical assessment revealed tremor-dominant, levodopa-responsive parkinsonism with an age of onset at 60 and 67 years of age in the two affected individuals examined (Table S1 in File S1). Both individuals also reported subjective cognitive impairment. Restless legs syndrome was present in IV:11 as well as one of her children. Transcranial ultrasound showed bilateral hyperechogenicity of the substantia nigra in IV:18 but was not performed in IV:11. MRI was in line with a diagnosis of PD in both. The affected parent (III:7) and aunt (III:5) of IV:11 were deceased before initiation of the study, so that no detailed phenotype information is available. Moreover, another aunt (III:2) on the same side of the family was reported to have suffered from an unclassified form of dementia.

Identification of Candidate Variants by Exome Sequencing and Frequency Assessment of Candidate Variants in a Case/Control Cohort

Exome sequencing was performed using DNA from two second cousins (IV:11 and IV:18, Figure 1A). This generated 11.68 gigabases (Gb) of alignable sequence for IV:11 (average coverage = 108.46, base pairs with >8 reads = 93.67%) and 15.02 Gb for patient IV:18 (average coverage = 154.13, base pairs with >8 reads = 94.74%). All 28,803 detected variants shared by the two affected individuals were filtered against in-house exomes (n = 1739) of individuals with unrelated diseases. Here, variants were allowed to be present in ≤1% of exomes. Moreover, synonymous and non-coding variants as well as all variants annotated in dbSNP135 with a minor allele frequency (MAF) ≥0.01 were excluded from the follow-up (Figure S1). No known variants believed to play a causative role in PD were found in either IV:11 or IV:18.

All ten remaining missense, nonsense, stoploss, splice site or frameshift variants and indels were genotyped in 975 cases and 1014 population-based controls (Table 1). The variants were, overall, very rare. Two (*PLXNA4* p.Ser657Asn and *OGN* p.Leu124fs) were validated in the individual in whom they were first identified but were otherwise not found again in the 1989 individuals tested. *CPNE1* p.Ser1831Thr was present in the index case as well as one additional control individual and *GOLGA4* p.Gln425Arg was identified in one additional PD patient. The other six variants (*RBM28* p.Asp300Gly, *IMPDH1* p.His296Arg, *ARPP21* p.Ala576Thr, *PHF2* p.Ser840Asn, *SLC22A13* p.Arg16His and *SPANXE* p.Leu42Ile) were not as rare (MAF≥0.03%) and

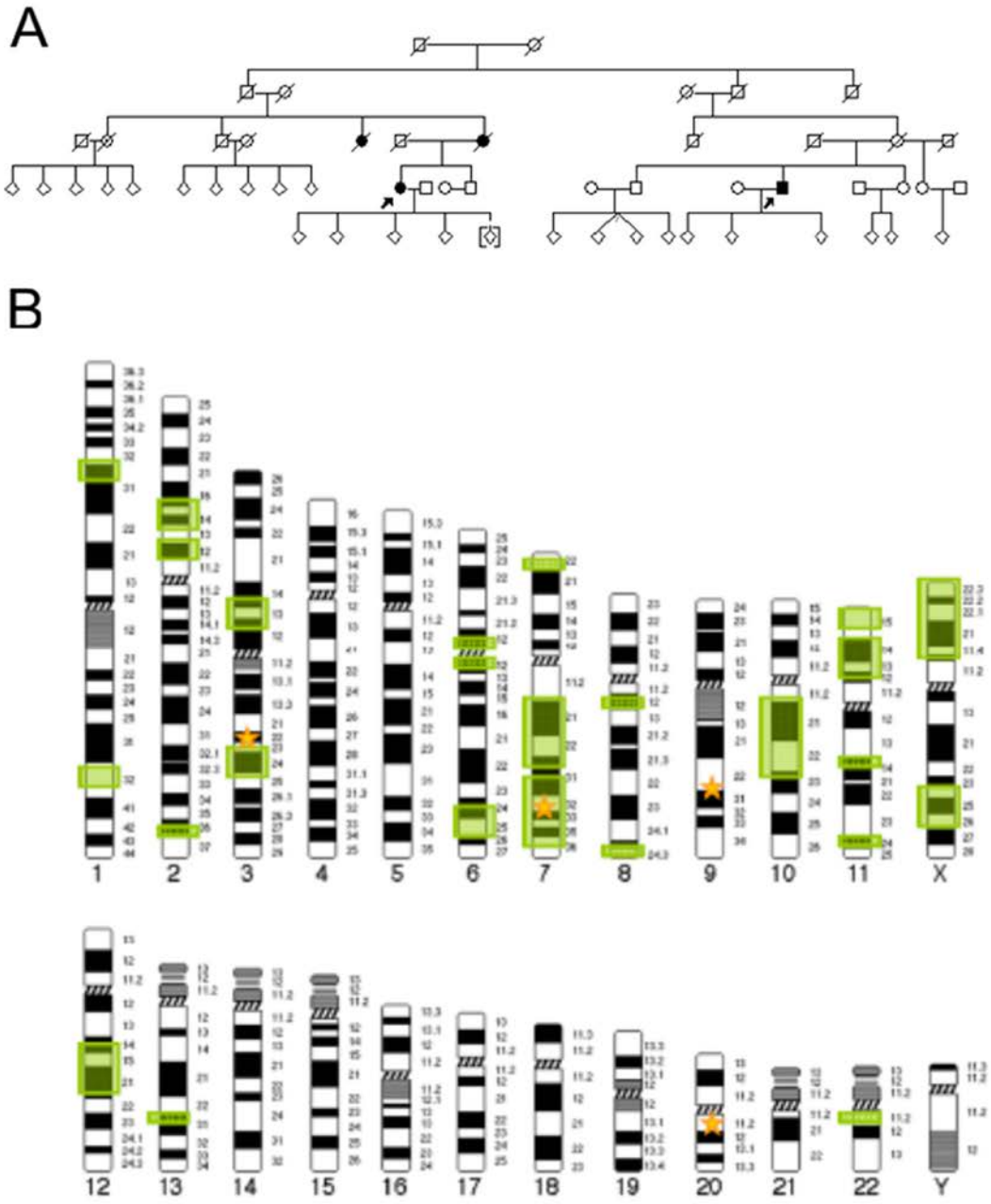


Figure 1. Pedigree and Linkage Analysis. (A) Pedigree of family used for exome sequencing. Open symbols indicate unaffected family members, affected individuals are denoted by closed symbols. An arrow denotes the individuals whose exomes were sequenced. Sex was obscured and birth order was altered to protect privacy. A diagonal line indicates a deceased individual. (B) 25 genomic regions on 12 chromosomes with logarithm of the odds (LOD) score≥0.5 were identified by linkage analysis. Green boxes represent genomic regions with LOD≥0.5, yellow stars represent the location of the four candidate genes remaining after frequency assessment (*GOLGA4*-chr3, *PLXNA4*-chr7, *OGN*-chr9, *CPNE1*-chr20). *PLXNA4* on chromosome 7 represents the only of the four genes overlapping a genomic region with LOD≥0.5.
doi:10.1371/journal.pone.0079145.g001

Table 1. Ten Rare, Non-synonymous Variants Shared by Individuals IV:11 and IV:18 of Family PARK_0005.

Genomicposition (hg19)	gene	number of alleles	*in house* exomes (n = 1739)	genotyping cases (n = 975)	genotyping controls (n = 1014)	dbSNP135	1000genomes	NHLBI-ESP [39] (EA only)	transcript	variation		penetrance for PD in % (n = 6)	PolyPhen2
										nucleotide	amino acid		
chr3:35780947	*ARPP21*	1	12	9	11	rs151173813	0.0031:2218	A = 37/G = 8563	NM_001267617.1	c.1726G>A	p.Ala576Thr	N/A	benign
chr3:37365968	*GOLGA4*	1	4	2	0	rs139536585	not found	G = 8/A = 8592	NM_001172713.1	c.1274A>G	p.Gln425Arg	66.67%	benign
chr7:127950857	*RBM28*	1	1	6	5	rs148028531	0.0007:14795	C = 20/T = 8580	NM_018077.2	c.2273T>C	p.Asp758Gly	40.00%	poss. damaging
chr7:128037009	*AMPDH1*	1	7	6	5	rs61751223	0.0052:2280	C = 23/T = 8577	NM_000883.3	c.887T>C	p.His296Arg	40.00%	benign
chr7:131910932	*PLXNA4*	1	0	1	0	novel	not found	not found	NM_020911.1	c.1970C>T	p.Ser657Asn	40.00%	prob. damaging
chr9:95155422–95155423	*OGN*	1	0	1	0	novel	not found	not found	NM_0140573	c.372_373 delAA	p. Leu124fs	50.00%	frameshift
chr20:34219872	*CPNE1*	1	0	1	1	novel	not found	not found	NM_003915.5	c.547A>T	p.Ser183Thr	66.67%	benign
chr9:96436037	*PHF2*	1		15	26	rs41276200	0.002:2389	A = 120/G = 8480	NM_005392.3	c.2519G>A	p.Ser840Asn	N/A	benign
chr3:38307398	*SLC22A13*	1		20	23	rs72542450	0.01:1459	A = 84/G = 8516	NM_004256.3	c.47G>A	p.Arg16His	N/A	benign
chrX:140785792	*SPANXE*	1,2		14/7	9/15	novel	not found	not in database	NM_145665.1	c.124G>T	p.Leu42Ile	N/A	not scored

The rare variants common to the two affected individuals were genotyped in 975 cases and 1014 controls. Penetrance with regard to the PD phenotype was assessed in 6 family members belonging to the same generation as the affected individuals. EA = European American.

doi:10.1371/journal.pone.0079145.t001

found at similar frequencies in both cases and controls and were, therefore, regarded to be unlikely candidates (Table 1).

Segregation Analysis and Genotyping of Additional *PLXNA4* Variants

The remaining four variants shared by the two affected individuals (Table 1) were pursued further by Sanger-sequencing-based testing for segregation in 6 family members belonging to generation IV. Under the assumption that a given variant would be causal for PD, penetrance ranged between 40.0 and 66.6% in 6 individuals belonging to generation IV. Moreover, on careful scrutiny of the exome data, both index patients were found to harbor one additional, variant of *PLXNA4* (p.Phe40Leu (rs145024048, 111/8489 in NHLBI-ESP exomes) for IV:11 and p.Arg302His (rs143813209, 3/8597 in NHLBI-ESP exomes) for IV:18). These two variants were also genotyped in 15 additional members of the family. *PLXNA4* p.Phe40Leu was found in 5 additional individuals and p.Arg302His was found in 7 additional family members. Importantly and contrary to the exome sequencing data, by Sanger sequencing, IV:11 was also found to harbor the *PLXNA4* p.Arg302His variant. The combination of the *PLXNA4* index variant and p.Phe40Leu was present only in IV:11, while the index variant and *PLXNA4* p.Arg302His were found in a total of 7 individuals belonging to the pedigree. None of the three additional candidate genes harbored additional non-synonymous coding variants in either IV:11 or IV:18.

Linkage Analysis

In order to further prioritize genes for follow-up, we performed parametric linkage analysis. In doing so, we identified 25 genomic regions with a suggestive linkage signal (LOD≥0.5) (Figure 1B). Only one of these regions, located on chromosome 7 (chr7:106,254,234 to 134,663,671; maximum two-point LOD score = 0.76), contained one of the four candidate genes identified during exome sequencing, lending further support to the potential causality of variants in *PLXNA4*.

Mutational Screening of *PLXNA4* in Case/Control Cohort

Linkage analysis highlighted the variant in *PLXNA4* as a potentially causal or modifying variant for the PD phenotype in our family. Also, the affected amino acid in *PLXNA4* is highly conserved in all vertebrates and two of three commonly used prediction algorithms [12–14] predicted it to be "damaging". Accordingly, we screened the 32 coding exons as well as the exon/intron boundaries of *PLXNA4* in 862 Austrian and German cases and 940 controls in order to assess a fuller spectrum of rare genetic variation found. For the most part, this cohort comprised the same individuals used for the above frequency assessment. In *PLXNA4*, a total of 38 novel (37 non-synonymous, 1 deletion) and 6 known variants (rs143813209, rs113830939, rs112682233, rs62622406, rs117458710 and rs73155258, all non-synonymous) resulting in a change in the amino acid sequence were identified (Table S2 in File S1). The large majority (86.21%) of variants were very rare, with MAF≤0.2% in controls. Overall, a similar number of cases (n = 107) and controls (n = 117) harbored at least one variant predicted to result in a changed amino acid sequence ($p>0.05$, χ^2 test). The same held true when only variants with MAF≤1.0% (46 cases vs. 52 controls, $p>0.05$, χ^2 test) were evaluated. Very rare variants with MAF≤0.2%, however, were more common in cases (n = 33) than controls (n = 18) ($p<0.02$, χ^2 test). Three cases but no controls were compound heterozygous for a non-synonymous variant in *PLXNA4*. Variants were located throughout the entire gene (Figure 2A).

Of the individuals harboring a rare non-synonymous variant in *PLXNA4*, information regarding family history was available for 17 individuals: 3 reported a first or second degree relative with PD and a positive history of essential tremor was present in the mother and a maternal uncle in one additional individual. The only brother of the individual harboring the *PLXNA4* p.Arg302Cys amino acid change was also found to have PD and to harbor this variant. However, the family was too small for formal segregation analysis.

When analyzed by means of three commonly used prediction algorithms (PolyPhen2, MutationTaster, SIFT) [12–14], the number of non-synonymous single nucleoide variants (SNVs) classified as functionally "damaging" (SNVs classified as "probably damaging" by PolyPhen2, "disease causing" by MutationTaster and "damaging" by SIFT) was greater in cases than in controls. This was especially prominent and statistically significant for PolyPhen2 when only very rare variants with MAF≤0.2% in controls were analyzed (PolyPhen2:19 variants in cases vs. 9 variants in controls, $p = 0.033$, χ^2 test; MutationTaster: 26 in cases vs. 14 in controls, $p = 0.028$, χ^2 test; SIFT/PROVEAN: 10 in cases vs. 2 in controls, $p = 0.018$, Fisher's Exact test) (Figure 2B). Deletions, which were only found in cases, cannot be assessed by PolyPhen2 and were, therefore, omitted from the analysis using this algorithm.

Functional Assessment of *PLXNA4* p.Ser657Asn in Fibroblasts

In fibroblast cell lines generated from both the index patient and an offspring who does not harbor the *PLXNA4* p.Ser657Asn variant (other variants not given to protect privacy) cell viability was similar (Figure 3A). Based on the results from the above mutation screening as well as the fact that *PLXNA4* is known to be expressed in the brain [15] and a role for axonal guidance factors similar to *PLXNA4* already postulated in PD [16], we further analyzed subcellular localization of the protein in the two cell lines but could not detect a difference (Figure 3B).

Modeling a Potential Role of *PLXNA4* in the PD Network

Beyond a proposed general role of axonal guidance pathways in the development of neurodegeneration [16,17], it is interesting to note that *PLXNA4* can be place into a network containing several firmly established PD genes (*SNCA, PARK2, DJ-1, LRRK2*), although both known and less reliable projected interactions have to be utilized (Figure 4).

Discussion

In an unbiased, whole-exome approach, we identified a variant in *PLXNA4* (p.Ser657Asn) as a candidate for a potentially causal variant in familial PD. Although this finding is intriguing and functionally plausible, we cannot conclude that this variant in *PLXNA4* is indeed the cause of PD in our family. Also, it is interesting that both affected individuals were found to harbor two or three non-synonymous variants in *PLXNA4*, thus, highlighting the possibility that a "multi-hit" model within the same gene or pathway could play a role with regard to phenotype expressivity.

Three of the final four variants (*PLXNA4* p.Ser657Asn, *OGN* p.L124fs and *CPNE1* p.Ser183Thr) are extremely rare and were only found in other family members but not in approximately 8,978 other individuals of European descent (genotyping sample (n = 1989), in-house exomes (n = 1739), 1000genomes (n = 1000) and NHLBI-ESP exomes (n = 4250)). This is interesting in light of the fact that–with regard to drug target genes–it was recently shown that the rarer a given variant the more likely it is

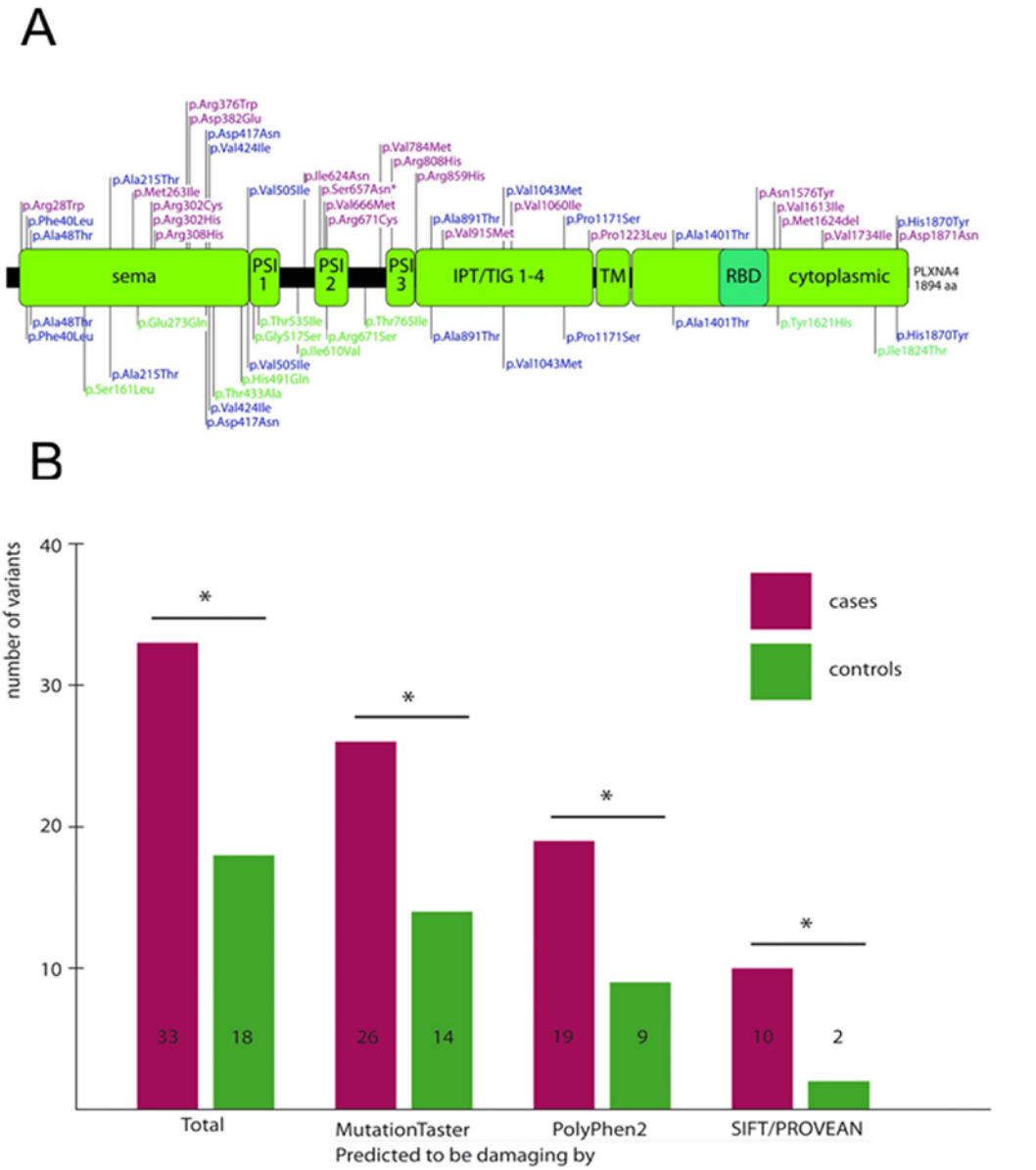


Figure 2. Mutation Screening of *PLXNA4* in PD case/control cohort. (A) Location of *PLXNA4* variants identified in variant screening in relation to known functional domains. An asterisk denotes the variant identified by exome sequencing. blue = variants found in both cases and controls, green = variants found only in cases, purple = variants found only in controls. (B) Analysis of *PLXNA4* variants using SIFT/PROVEAN, PolyPhen2 and MutationTaster reveals an excess of rare non-synonymous variants predicted to be damaging. Insertions and deletions cannot be assessed by PolyPhen2 all and were, therefore, omitted from the analysis using this algorithm.
doi:10.1371/journal.pone.0079145.g002

functionally relevant [18]. Yet, on the other hand, this rarity also means that from a genetic standpoint, at the moment, one can neither confirm nor exclude the possibility of a causal or modifying role in the PD phenotype. Further, even taken together additional evidence highlighting *PLXNA4* p.Ser657Asn (suggestive linkage signal, high conservation and predicted pathogenicity, excess of

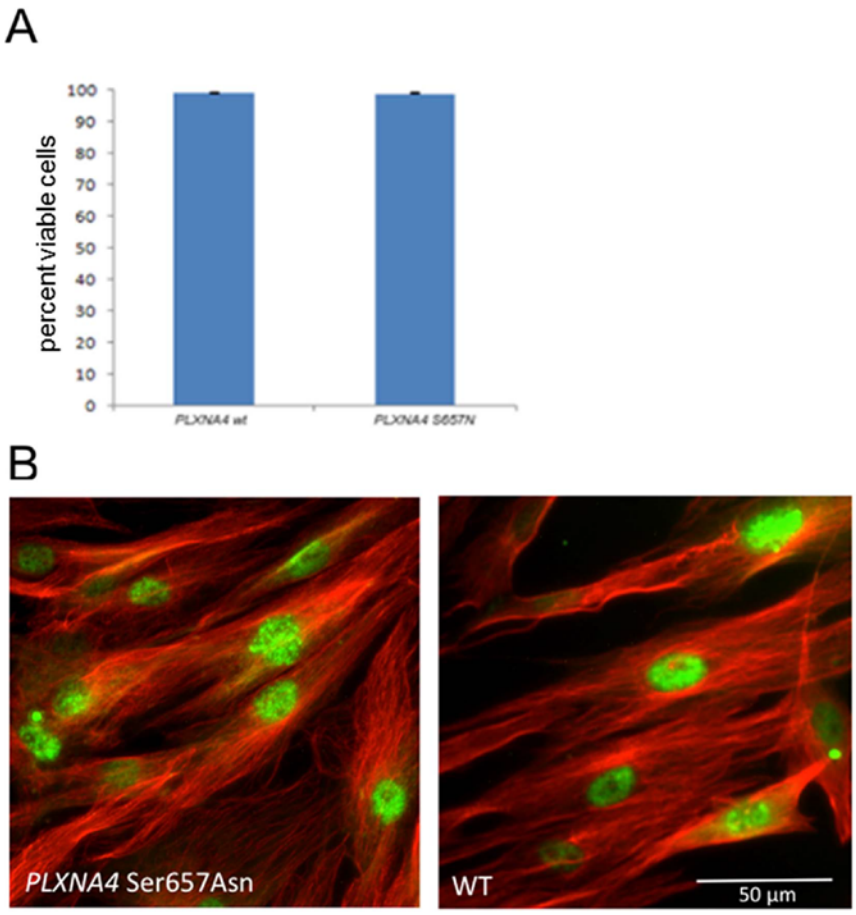


Figure 3. Assessment of cell viability and subcellular protein localization in fibroblasts. (A) The presence of *PLXNA4* p.Ser657Asn do not affect cell viability as assay by live-dead staining and FACS. (B) Immunohistochemistry shows similar subcellular localization of *PLXNA4* (anti-PLXNA4, Sigma, 1:500) in fibroblasts with and without the p.Ser657Asn amino acid substitution (scale bar = 50 µm).
doi:10.1371/journal.pone.0079145.g003

very rare coding variants in cases and functional considerations) can be viewed as suggestive at best and by no means exclude the possibility of other causative or modifying genetic factors that play a role in the PD phenotype in our family.

In general, these findings highlight the fact that in many cases very large populations will be needed to conclusively judge the disease-related nature of a rare variant. Recent studies show that while the power to detect associations for genes harboring rare variants varies widely across genes, only <5% of genes achieved 80% power even assuming high odds ratios (OR) of 5 and when tested in 400 cases and 400 controls. In the same scenario, no gene out of 12,000 genes tested achieved 80% power when assuming an OR of 1.5 [19]. Statistical evaluation is further complicated by the fact that it is not unreasonable to assume that many genes will habor both variants that are protective and predisposing with regard to a given phenotype, as was recently shown for the *APP* locus in Alzheimer's disease [20], which with the statistical analysis tools available today will always lead to an underestimation of the genetic contribution of rare variants at a given locus to a phenotype's heritability [21].

Ultimately, it is also possible that the truly causal variant was not picked up in this study because it lies outside the targeted regions of the exome. Here, the use of two enrichment kits of different sizes and different exome target definitions represents a specific weakness of the study. Also, we cannot exclude that IV:18 represents a phenocopy and that the underlying cause of PD in his case is different from that of the other affected individuals in the family. If this were the case, a much larger number of candidate variants than those assessed here could contribute to bringing about the PD phenotype in the examined family.

Moreover, copy number variants, another important player in the full spectrum of genetic variation, could, at the time of study, not yet confidently be assessed in exome sequencing data and were, therefore, not evaluated in our study. Lastly, while suggestive non-significant LOD scores have been used to prioritize variants identified in exome [22] or whole genome [23] sequencing they

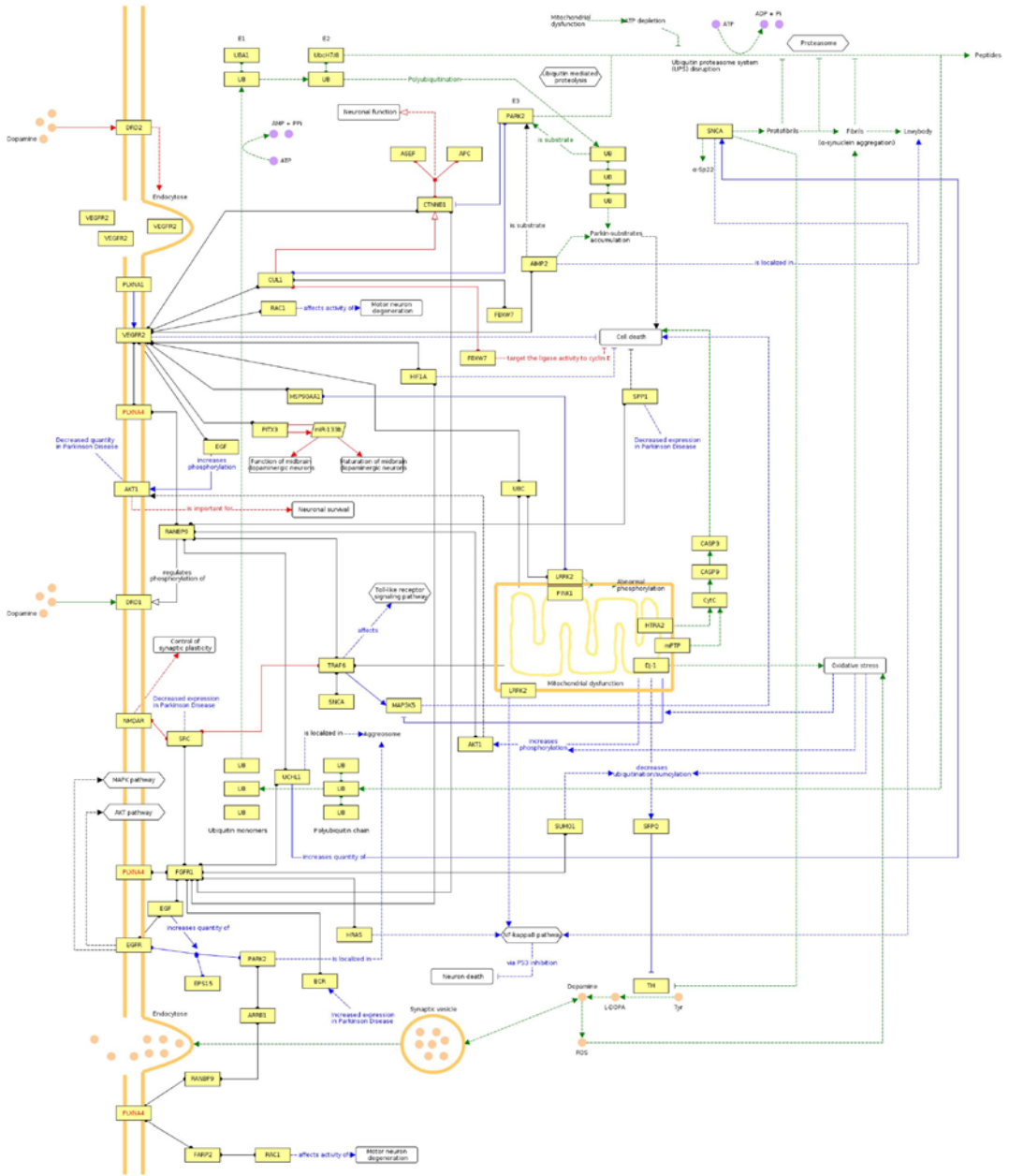


Figure 4. Qualitative multifactorial interaction network of *PLXNA4* and genetic factors with known and hypothetical relevance to PD. Edges obtained from CIDeR are highlighted in blue, PD-specific pathways from KEGG are given in green, red edges denote annotations from OMIM and edges extracted from literature, protein-protein interaction databases or high-confidence predictions are colored black. Undirected protein-protein interactions hold circular ends, directed molecular relations are marked by arcs, whereas general regulations have arrows with no filling, activations have filled arrows and inhibitions have blunted end. Dashed lines indicate indirect effects.
doi:10.1371/journal.pone.0079145.g004

also harbor the potential for the erroneous exclusion of true positives.

The fact that all four candidate variants were also found in unaffected family members, per se does not contradict potential causality as it is known from other autosomal dominant forms of PD that even among members of a single family, penetrance of known PD mutations can vary widely. Of individuals who harbor the *LRRK2* p.Gly2019Ser mutation, for example, only 28% will develop PD by the age of 59 [24]. Thus, predicted penetrance of the variants identified in our family are in line with what is reported in the literature for other forms of autosomal dominant PD.

Plexin A4, *PLXNA4*, which functions as a receptor for class 3 semaphorins, holds a firmly established role in axon guidance in the development of the central and peripheral nervous systems. For example, PlxnA4 has been shown to restrict inappropriate spreading of mossy fibers within the CA3 region of the murine hippocampus [25], to direct basal dendritic arborization in layer V cortical neurons [26] and sympathetic axons [15,27] as well as lamination and synapse formation in the outer retina [28] in the mouse.

PLXNA4 has also been implicated in neurodegenerative conditions. In the discovery stage of a large family-based GWAS assessing low-frequency (MAF≤5%) variants in late-onset Alzheimer's disease an intronic SNP in *PLXNA4* (rs277484, MAF = 2.0% in 1000genomes) yielded the most significant association signal ($p = 9.0 \times 10^{-10}$). Replication, however, is still ongoing [17]. Similarly, preliminary results have suggested decreased *PLXNA4* expression in the motor cortex of individuals with amyotrophic lateral sclerosis when compared to controls, although the sample size of the study was very limited (n = 5) [29].

PLXNA4 itself has not previously been implicated in PD. Yet, a number of studies have suggested an involvement of axonal guidance pathways in PD. An early GWAS identified a SNP in semaphorin 5A (*SEMA5A*) as the best association signal [16] and systems biology-based follow-up studies reported an overrepresentation of axonal guidance factors in subthreshold association signals [30] which were shown to predict susceptibility to PD [31]. However, both the association signal and the pathway analysis proved difficult to replicate in other cohorts [32–34] which may be due to the fact that as one of the very first GWAS it was not conducted to the current quality standards. Expression studies of different brain regions, on the other hand, have repeatedly found an overrepresentation of differentially expressed axonal guidance pathways in individuals with PD when compared to controls [30,35–37]. Axonal guidance pathways have also been implicated in the proper targeting of dopaminergic neurons from the murine mesencephalon to the ipsilateral striatum [38].

At the moment, both functional and genetic data addressing a role of *PLXNA4* as a PD gene are inconclusive. The identification of additional larger families with PD in which *PLXNA4* p.Ser657-Asn or p.Arg302His segregate with the phenotype or the replication of the finding of an excess of very rare variants (MAF≤0.02%) in an independent case/control sample would lend further support to a possible role of modifying or causal variants in *PLXNA4* in PD and to the interesting hypothesis of axonal guidance dysfunction in neurodegenerative conditions.

Supporting Information

Figure S1 Filtering scheme for variants identified by exome sequencing in the two affected family members examined.
(TIF)

File S1 Supporting Methods and Tables. Table S1 in File S1, Clinical Phenotype of Affected Individuals in PARK_0005. **Table S2 in File S1,** Non-Synonymous and Indel Variants Identified in Variant Screening of *PLXNA4*.
(DOC)

Acknowledgments

We are gratefully indebted to Katja Junghans, Susanne Lindhof, Jelena Golic, Sybille Frischholz und Regina Feldmann at the Insitut für Humangenetik, Helmholtz Zentrum München, Munich, Germany, for their expert technical assistance. We would also like to thank Lucas Schirmer, Neurologische Klinik und Poliklinik, Technische Universität München, Munich, Germany, for expert help in performing the FACS analyses.

Author Contributions

Conceived and designed the experiments: ECS DC DCE BMM CT JW. Performed the experiments: ECS IS DC DCE EG. Analyzed the data: ECS IS DC DCE SE PL BMM. Contributed reagents/materials/analysis tools: BM AZ DH WP TB BB MJM AP CG CT. Wrote the paper: ECS IS JW.

References

1. Bonifati V, Fabrizio E, Vanacore N, De Mari M, Meco G (1995) Familial Parkinson's disease: A clinical genetic analysis. Can J Neurol Sci 22: 272 279.
2. Payami H, Larsen K, Bernard S, Nutt J (1994) Increased risk of Parkinson's disease in parents and siblings of patients. Ann Neurol 36: 659 661.
3. Bonifati V, Rizzu P, van Baren MJ, Schaap O, Breedveld GJ, et al. (2003) Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science 299: 256 259.
4. Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, et al. (1998) Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature 392: 605 608.
5. Paisan-Ruiz C, Jain S, Evans EW, Gilks WP, Simon J, et al. (2004) Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. Neuron 44: 595 600.
6. Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, et al. (1997) Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 276: 2045 2047.
7. Valente EM, Abou-Sleiman PM, Caputo V, Muqit MM, Harvey K, et al. (2004) Hereditary early-onset Parkinson's disease caused by mutations in PINK1. Science 304: 1158 1160.
8. Zimprich A, Biskup S, Leitner P, Lichtner P, Farrer M, et al. (2004) Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron 44: 601 607.
9. Vilariño-Güell C, Wider C, Ross OA, Dachsel JC, Kachergus JM, et al. (2011) VPS35 Mutations in Parkinson Disease. Am J Hum Genet 89: 162 167.
10. Zimprich A, Benet-Pagès A, Struhal W, Graf E, Eck SH, et al. (2011) A Mutation in VPS35, Encoding a Subunit of the Retromer Complex, Causes Late-Onset Parkinson Disease. Am J Hum Genet 89: 168 175.
11. Schulte EC, Mollenhauer B, Zimprich A, Bereznai B, Lichtner P, et al. (2012) Variants in eukaryotic translation initiation factor 4G1 in sporadic Parkinson's disease. Neurogenetics 13: 281 285.
12. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D (2010) MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 7: 575 576.
13. Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. Nucleic Acid Res 30: 3894 3900.
14. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, et al. (2010) A method and server for predicting damaging missense mutations. Nat Methods 7: 248 249.
15. Suto F, Ito K, Uemura M, Shimizu M, Shinkawa Y, et al. (2005) Plexin-A4 Mediates Axon-Repulsive Activities of Both Secreted and Transmembrane Semaphorins and Plays Roles in Nerve Fiber Guidance. J Neurosci 25: 3628 3637.
16. Maraganore DM, de Andrade M, Lesnick TG, Strain KJ, Farrer MJ, et al. (2005) High-Resolution Whole-Genome Association Study of Parkinson Disease. Am J Hum Genet 77: 685 693.
17. Jung G, Buros J, Lunetta KL, Foroud TM, Mayeux R, et al. (2011) Identification of Rare Variants for Late-Onset Alzheimer Disease in a Family-Based Genome-Wide Association Study. Abstract No. 221. Presented at the

Annual Meeting of the American Society of Human Genetics 2011, Montreal, Canada.
18. Nelson MR, Wegmann D, Ehm MG, Kessner D, St Jean P, et al. (2012) An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science 337: 100 104.
19. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, et al. (2012) Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. Science 337: 64 69.
20. Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, et al. (2012) A mutation in *APP* protects against Alzheimer's disease and age-related cognitive decline. Nature. 488: 96 99.
21. Liu DJ, Leal SM (2012) Estimating Genetic Effects and Quantifying Missing Heritability Explained by Identified Rare-Variant Associations. Am J Hum Genet 91: 585 596.
22. Johnson JO, Mandrioli J, Benatar M, Abramzon Y, Van Deerlin VM, et al. (2010) Exome sequencing reveals VCP mutations as a cause of familial ALS. Neuron 68: 857 864.
23. Sobreira NL, Cirulli ET, Avramopoulos D, Wohler E, Oswald GL, et al. (2010) Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene. PLoS Genet 6: e1000991.
24. Healy DG, Falchi M, O'Sullivan SS, Bonifati V, Durr A, et al. (2008) Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. Lancet Neurol 7: 583 590.
25. Suto F, Tsuboi M, Kamiya H, Mizuno H, Kiyama Y, et al. (2007) Interactions between Plexin-A2, Plexin-A4, and Semaphorin 6A Control Lamina-Restricted Projection of Hippocampal Mossy Fibers. Neuron 53: 535 547.
26. Tran TS, Rubio ME, Clem RL, Johnson D, Case L, et al. (2009) Secreted semaphorins control spine distribution and morphogenesis in the postnatal CNS. Nature 462: 1065 1069.
27. Waimey KE, Huang PH, Chen M, Cheng HJ (2008) Plexin-A3 and Plexin-A4 Restrict the Migration of Sympathetic Neurons but not Their Neural Crest Precursors. Dev Biol 315: 448 458.
28. Matsuoka RL, Jiang Z, Samuels IS, Nguyen-Ba-Charvet KT, Sun LO, et al. (2012) Guidance-cue control of horizontal cell morphology, lamination, and synapse formation in the mammalian outer retina. J Neurosci 32: 6859 6868.
29. Körner S, Wilzek N, Thau S, Knippenberg S, Dengler R, et al. (2012) mRNA Expression von Axon-Leitproteinen (Axon guidance proteins) in humanem post-mortem Motorkortex bei Amyothropher Lateralsklerose. Abstract V24. Presented at the Annual Meeting of the German Society for Neurology 2012, Hamburg, Germany.
30. Edwards YJ, Beecham GW, Scott WK, Khuri S, Bademci G, et al. (2011) Identifying Consensus Disease Pathways in Parkinson's Disease Using an Integrative Systems Biology Approach. PLoS one 6: e16917.
31. Lesnick TG, Papapetropoulos S, Mash DC, Ffrench-Mullen J, Shehadeh L, et al. (2007) A Genomic Pathway Approach to a Complex Disease: Axon Guidance and Parkinson Disease. PLoS Genet 3: e98.
32. Bialecka M, Kurzawski M, Klodowska-Duda G, Opala G, Tan EK, et al. (2006) Polymorphism in semaphoring 5A (Sema5A) gene is not a marker of Parkinson's disease risk. Neurosci Lett 399: 121 123.
33. Clarimon J, Scholz S, Fung HC, Hardy J, Eerola J et al. (2006) Conflicting results regarding the semaphorin gene (SEMA5A) and the risk for Parkinson disease. Am J Hum Genet 78: 1082 1084.
34. Li Y, Rowland C, Xiromerisiou G, Lagier RJ, Schrodi SJ, et al. (2008) Neither replication nor simulation supports a role for the axon guidance pathway in the genetics of Parkinson's disease. PLoS One 3: e2707.
35. Bossers K, Heerhoff G, Balesar R, van Dongen JW, Kruse CG, et al. (2009) Analysis of gene expression in Parkinson's disease: Possible involvement of neurotrophic support and axon guidance in dopaminergic cell death. Brain Pathol 19: 91 107.
36. Sutherland GT, Matigian NA, Chalk AM, Anderson MJ, Silburn PA, et al. (2009) A cross-study transcriptional analysis of Parkinson's disease. PLoS one 4: e4955.
37. Wang K, Li M, Bucan M (2007) Pathway-based approaches for analysis of genomewide association studies. Am J Hum Genet 81: 1278 1283.
38. Kawano H, Horie M, Honma S, Kawamura K, Takeuchi K, et al. (2003) Aberrant trajectory of ascending dopaminergic pathway in mice lacking Nkx2.1. Exp Neurol 182: 103 112.
39. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA. Available: http://evs.gs.washington.edu/EVS/. Accessed October 23, 2012.

3 Discussion

3.1 Roles of Common, Low Frequency and Rare Variants in Disease Development (cf. ref. I-X)

One way in which genetic variation that affects the nucleotide make-up of genomic DNA can be characterized is by its frequency in the (general) population. As outlined in the introduction, the MAF represents a means to group variants into different frequency categories. This, however, only represents a tool for stratification and simplification whereas in reality frequencies range on a full continuous spectrum between a MAF of 50% and approximately 1.43 $x10^{-8}$ % (i.e. a variant found in only one out of seven billion people).
As highlighted by Teri Manolio and colleagues in an influential review article in 2009[26], generally, rarer variants confer larger effects on a given phenotype than more common ones. An exception to this rule arises when rare alleles come about independently again and again in a mutational hotspot and, thus, statically mimic a common allele of large effect. One example for such a situation is the *CONNEXIN26* c.35delG variant in Europeans with inherited non-syndromic hearing loss which has a relatively high carrier frequency of 2 to 4% in the general population[451,452] but instead of being a common ancestral allele it occurred multiple times independently in a mutational hotspot.[142,453] Conversely, the number of rare variants of relatively small effect sizes may actually be larger than expected. With regard to rare variants in the breast cancer genes *BRCA1* and *BRCA2*, for example, it has been postulated that the high proportion of rare variants currently classed as variants of unknown significance (VUS) because they do not show clear familial segregation and are not fully penetrant could actually harbor a significant and clinically relevant breast cancer risk both at the individual and at population level.[77]
In recent years, the realization has hit that human disease, in general, is characterized by far greater genetic heterogeneity than previously assumed.[142] And the more biologically complex the phenotype, the more heterogeneous its genetic framework.[142]
In line with these observations, in the work portrayed herein, it also becomes clear, that it is likely that a full spectrum of genetic variants of differing frequencies contribute to the genetic make-up of both RLS and PD[I-X], as exemplified by the identified allelic series in *MEIS1*[I,III,IV] that could be shown to contribute to the genetic framework of RLS.

3.1.1 Susceptibility and Causality (cf. ref. I-X)

In the most simplistic of conceptions, susceptibility to a disease is conferred by a common variant of small effect while causal alleles are rare but effect sizes so large that they themselves are sufficient to cause disease (cf. Figure 1.1). While surely too simplistic, the general trend also exists in our data: Common variants such as intronic rs9920066 located in de-etiolated homolog 1 (*Arabidopsis*), *DET1*, is a common variant with a MAF 30.3% in the 1000 genomes[124]. If it truly were an RLS-related genetic factors, which we are unable to conclude from the data obtained in our study, the OR would be low (1.11)[II] and even if this SNP were the "causal" variant underlying the observed association signal, it would only marginally increase the risk of a given individual to actually develop RLS. Susceptibility as such, is a very statistical measure. And the GWAS used to identify susceptibility alleles have often times been criticized for loosing touch with biology and producing statistically meaningful but clinically meaningless results. For example, in a 12-year follow-up study for cardiovascular disease in more than 19,000 women, 101 SNPs identified as susceptibility alleles by GWAS did not predict cardiovascular outcome[454]. Similar studies have not been performed with regard to RLS or PD.

On the other end of the spectrum, rare non-synonymous variants with MAF < 5% in the "RLS gene" *MEIS1* carry combined projected ORs of up to 30[IV], suggesting that these variants could fall into a category with effect sizes large enough to cause "Mendelian" forms of RLS[26]. Yet, whether singular rare variants in *MEIS1* can indeed be the single cause of familial RLS remains to be investigated. In both RLS and PD, the investigation of families is hampered by several factors. For one, over the past decades, research efforts with regard to both diseases have shown, that genetic heterogeneity underlying the phenotypes is likely to be large (reviewed in ref. [270,280,345]). This also seems apparent from our work. Rare variants identified in both large-scale candidate gene screens as well as whole-exome sequencing have yielded a larger number of singletons than most people in the field initially expected. For example, the candidate variant (*PLXNA4* p.S657N) identified by whole exome sequencing in a German family with suspected autosomal dominant PD, was only identified in this family and not seen again in more than 9,000 individuals examined[X]. Accordingly, although this variant is very rare with a MAF < 0.0055 % and could hold a large OR, with the datasets currently available to the statistical analysis of rare variants (NHLBI-ESP exomes, 1000 genomes, *in house* exomes), it will never be possible to demonstrate causality of such variants. In fact, some argue, that even in the context of common diseases direct causality is very difficult to resolve by large-scale association or case-control studies[142].

A second conundrum facing the establishment of variant causality in familial PD—and likely also RLS—is that of incomplete penetrance. In most of the PD families analyzed as part of this work[378,455-457], the candidate variants do not show complete penetrance. However, we cannot determine whether this is due to the fact that individuals carrying the candidate variant have just not yet developed full PD (maybe because they are not old enough yet), do not possess the additional genetic factors necessary to "unmask" the "causal" variant, have not been exposed to additional external factors that might influence variant penetrance or because the candidate variant is not the causal variant at all. Also, although several of the so-called "Mendelian" forms of PD were initially identified via family-based linkage analyses or, more recently, whole exome sequencing in families in whom the variants were highly penetrant and variants were established as "causal", later studies showed that in some individuals and families harboring the "causal" variant, penetrance is far from complete. For example, of individuals who harbor the *LRRK2* p.G2019S variant, the single most common "Mendelian" genetic factor known for PD, only 28% will develop PD by the age of 59[330].
In the vast majority of human diseases, especially those with an onset relatively late in life such as those considered in this work, it is difficult to find support for an all-or-nothing model of only one truly causal variant. Rather, it is likely that variants of varying effect sizes (and frequencies) exist and the question is whether there is a single variant that holds more than an equal share of the genetic contribution to a phenotype.

3.1.2 Allelic Series (cf. ref. I,III,IV,V)

In broad terms, an "allelic series" designates a set of allelic variants within a given genomic locus (i.e. most commonly a gene). It can be used to describe alleles differing in their position within a gene, their class (e.g. SNV, indel, CNV or synonymous, non-synonymous, intronic, etc), frequency and phenotypic expressivity. In the context of the genetics of common complex disease phenotypes, the possible contribution of allelic series consisting of genetic variants of various frequencies is particularly interesting and represents a central aspect of the work depicted herein. In connection with RLS, we evaluated seven candidate genes at five genomic loci believed to harbor common risk variants for the disease[I,III,IV,249-252] for the existence of allelic series comprising variants of different MAF (and, consequently, different effect sizes). Especially with regard to *MEIS1*, which also entails the most significant genome-wide association signals[249,250], variants of the full frequency spectrum ranging from MAF = 24 % (for rs2300478)[I] to singletons identified only in one out of 14,383 individuals (case/control sample used in ref. IV plus 1,739 in house exomes, 1,092 genomes belonging to

the 1000 genomes project[124] and exomes from approximately 4,250 individuals sequenced as part of the NHLBI-ESP[125]), corresponding to a MAF of approximately 0.0035 %, were identified in individuals with RLS[III,IV]. In aggregate, both low-frequency and rare non-synonymous variants in the coding regions and the 5`UTR of *MEIS1* were significantly more common in the individuals with RLS than in the general population[IV]. When analyzed individually, a low frequency variant (rs11693221) in the 3` UTR region of the canonical transcript of *MEIS1* also showed statistically significant association with the RLS phenotype (MAF_{cases}=13.55%, $MAF_{controls}$=3.58%; p=8.79x10^{-99}, χ^2 test; OR=4.16 (95% CI: 3.61-4.80)). Accordingly, at least with regard to the *MEIS1* locus, a complete allelic series of common to low frequency to (very) rare variants appears to contribute to the genetic framework of RLS (Figure 3.1). Moreover, by *in vivo* complementation in zebrafish, an excess of rare loss-of-function alleles seems to exist among the non-synonymous alleles found in cases when compared to controls[IV], suggesting that this finding is not merely statistically but also functionally relevant. However, as the exact function of *MEIS1* in the pathogenesis of RLS remains to be elucidated, it needs to be established whether these rare variants also disrupt RLS-relevant functions of *MEIS1*.

With regard to PD, the existence of allelic series in the form of both common and rare alleles co-existing at the same locus is already firmly established. Rare SNVs and structural variants in *SNCA*[371,372] and *LRRK2*[346] were among the first genetic factors identified in family studies of PD. Subsequently, GWAS have also highlighted independent common variants contributing to PD risk in sporadic cases in both European and Asian populations in both *SNCA*[438,439,441] and *LRRK2*[438,439,441]. At least at the *LRRK2* locus, low-frequency variants also appear to harbor both predisposing and protective effects[437] (Figure 3.1). Still, not all GWAS loci for PD harbor full allelic frequency series. At least in our case/control sample, rare variants in *MAPT*, which holds common variants strongly associated with sporadic PD[438] and other PD-like phenotypes[446], were not overrepresented among PD cases in comparison to controls[VII].

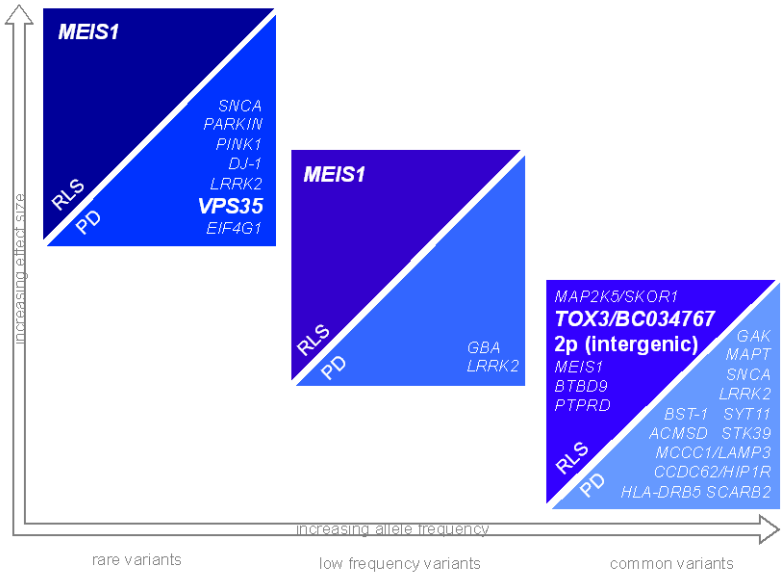


Figure 3.1: Rare, low-frequency and common genetic variants known to be involved in RLS (dark blue) and PD (light blue). In addition to common variants, low-frequency and rare variants in *MEIS1* also contribute to the genetic framework of RLS. For PD, on the other hand, strong genetic contributors are known in all frequency categories. The relative dearth of established genetic variants in the low-frequency category (1 % < MAF > 5 %) is likely due to the fact that these are difficult to identify by either family-based studies or GWAS and currently their evaluation is still largely dependent upon candidate gene approaches, although this will change as genotyping and imputation techniques improve and NGS becomes more affordable. As part of the work depicted in this dissertation, several new genes (bold) could be added to the pictogram above[I,III,IV,VIII], thus further elucidating the genetic architecture of both RLS and PD.

Yet, allelic series can also transverse diagnostic boundaries. To examine this possibility with regard to PD, we screened several genes known to harbor strong genetic factors involved in dementias (*APP*, *PSEN1* and *PSEN2* known to cause familial AD as well as *TARDBP*, *FUS*, *GRN* and *MAPT* known to bring about FTD) in a case/control sample comprising, next to controls, both PD patients with and without dementia in search for rare variants in established "dementia genes" involved in the genetics of PD[VII]. Our data suggest, that, at least with regard to *APP*, an allelic series crossing diagnostic boundaries, as has been previously described for, for example, psychiatric disorders[458,459], might exist. If so, *APP* variants contributing to the genetic burden in PD or the modification of the PD phenotype are different from those involved in AD. The fact that *APP* could be a common "neurodegeneration" rather than an AD gene is further supported by the fact that the variants identified in individuals with PD do not perturb APP function in the same way as known AD-linked variants[VII].

In the larger context of the genetics of complex diseases, these results are also of interest. They lend support to the assumption that in some complex diseases rare, low-frequency and

common variants within the same gene contribute to the genetic architecture—thus supporting both the "common disease, common variant" and the "common disease, rare variant" hypothesis. This is important for three reasons. For one, it underscores the notion that at least with respect to certain phenotypes such as, for example, RLS, some of the missing heritability[33] *will* lie in a collection of rare and very rare variants. Secondly, the identification of rare variants at a known GWAS locus argues for the value of GWAS in general, in that this approach can be used to identify genes of interest, which should then be scrutinized for the exact underlying genetic factors. Lastly, while some have criticized that candidate gene screens, whole exome sequencing and the use of SNP arrays on which coding variants are relatively overrepresented have introduced a bias towards the discovery of coding disease-linked variation[26,36], the benefit of discovering such—for the most part rare—coding variants is that they lend themselves to follow-up studies assessing biological function. This represents a large benefit especially with regard to the study of complex genetic diseases, where, in many instances such as in RLS, no strong genetic factors had been identified previously and the follow-up of GWAS association signals has proven notoriously difficult[31]. To date, at least 12 studies have been published which evaluate the role of rare and low-frequency variants at genomic loci known to harbor common variant disease associations identified in GWAS (Table 3.1). What becomes apparent is that in most traits and diseases evaluated so far, rare or low-frequency variants contribute to the genetic spectrum at at least some of the loci which also hold common variants. Yet, these loci represent the minority. This observation probably reflects both the differences in the variant framework that each individual locus contributes to the genetic architecture of a disease but also the fact that very large samples will be needed to adequately address this question especially with regard to very rare variation. Our data also illustrate these differences among the known RLS-associated GWAS loci: At some loci, such as *MEIS1*, rare variants appear to play a relatively large role, while at others, such as *TOX3* or *PTPRD*, this role may be present but does not reach statistical significance with the analyzed replication sample size (n = 3,265 cases/2,944 controls) or does not seem to exist at all[IV].

Several recent large-scale population genetics studies have described an excess of very rare functional alleles in the human genome[16,27,54]. Accordingly, one would assume that variants of lower frequency, which impact disease development, should also be very rare. In line with this assumption, very rare non-synonymous variants with MAF < 0.1 % within *MTNR1B*, the gene encoding the melatonin receptor 1B, but not with 0.1 % < MAF < 5% contribute to type 2 diabetes[70]. Our data regarding low-frequency and rare variation at the RLS-GWAS loci

illustrate that this conception holds true even when all candidate genes at the published GWAS loci are jointly analyzed. Only if solely variants with MAF < 1 % or < 0.1 % are considered, is there a significant excess of coding alleles in individuals with RLS.[IV] This would be interesting to keep in mind for future studies examining the contribution of rare variants at genomic loci known to be home to common susceptibility alleles.

complex disease/ trait	**GWAS loci evaluated**	**method**	**discovery** (cases/controls)	**replication** (cases/controls)	**indep. rare/low-freq variant association** (MAF/OR)	**aggregate rare variant association**	**ref**
type 2 diabetes	six loci	targeted reseq/ genotyping	480/480	8,379/10,575	4 variants in *IFIH1* (0.5% to 2.2%)	not evaluated	[460]
hypertriglyceridemia	*APOA5, GCKR, LPL, APOB*	GWAS/Sanger	463/1,197	438/327	none	all variants and missense/indel only across all four genes	[69]
fetal hemoglobin levels	*BCL11A, HBS1L/MYB, β-globin*	targeted reseq/genotyping	190	1,032	none	3 missense variants in *MYB* together	[461]
sick sinus syndrome	chr 14q11	GWAS/whole genome	792/37,592	7/80	*MYH6* p.Arg721Trp (0.4%/12.53)	not evaluated	[72]
LDL cholesterol level	*APOE, APOC1/2, SORT1, LDLR, APOB, PCSK9*	targeted reseq/ metabochip& 1000genomes	previously published	256/5,524	*PCSK9* p.Arg46Leu (3.7%/na) *LDLR* p.Val578Asp (0.5%/na) *APOE* p.Arg176Cys (3.7%/na)	not evaluated	[462]
age-related macular degeneration	*CFH/CFHR1/CFHR3*	haplotype analysis & targeted reseq/ genotyping	711/1,041 & 33/27	2,424/1,120	*CFH* p.Arg1210Cys (0.09%/na)	not evaluated	[73]
IBD	56 loci	targeted reseq/ Sequenom genotyping & immunochip	350/350 (pooled)	28,207/17,575	9 splice site or missense variants in 5 genes (*NOD2, IL18RAP, CARD9, IL23R, CUL2*) (around 0.2%/between 0.29 and 4.02)	not evaluated	[74]
type 2 diabetes	*MTNR1B*	targeted reseq/ genotyping	2,186/5,446	8,153/10,100	none	40 missense variants with MAF<0.1% together (OR=3.31),13 loss-of-function variants together (OR=5.67)	[70]
asthma	*ADRB2, AGT, DPP10, CFTR, CHIA, IKBKAP, IL12RB1, PLA2G7, TGFB1*	Sanger	previously published	510/515	none	non-synonymous variants in *DPP1* or *IL12RB1* together, non-coding variants ±100bp around *AGT, DPP10, IKBKAP* and *IL12RB1* and overall	[75]
celiac disease	183 non-*HLA* immune disease loci on immunochip	Immunochip	12,041/12,228	none	none but independent rare variants at 4 loci with $p<5x10^{-4}$	not evaluated	[463]
fasting proinsulin concentration	whole exome genotyping (59,029 markers)	exome chip	8229	none	*SGSM2* p.Val996Ile (1.4%, na) *MADD* p.Arg766X (3.7%, na)	not evaluated	[53]
rheumatoid arthritis	25 loci	targeted reseq/GWAS & immunochip	500/650 (pooled)	10,609/35,605	none	all coding variants across the 25 loci and marginally for non-syn variants in *IL2RA* and *IL2RB*	[71]
RLS	*MEIS1, PTPRD, TOX3, BTBD9, SKOR1, MAP2K5*	high resolution melting curve analysis	188/188 (all) 3,760/3,542 (*MEIS1* only)	3,265/2,944 none	*MEIS1* post-3`UTR (rs11693221) (3.6%/4.16)	all non-synonymous coding *MEIS1* variants of functional effect in zebrafish in vivo complementation, all 5`UTR variants in *MEIS1*	[IV]

Table 3.1: Summary of studies published to date and our present study[IV], which assess low-frequency and rare genetic variants at known GWAS loci in the context of frequency-based allelic series. LDL = low density lipoprotein, IBD = inflammatory bowel disease, reseq = resequencing, ref = reference, na = not available. Due to space considerations, only the gene symbols are given. Long versions can be found in the original publications or obtained from the internet (e.g. www.genenames.org[464]).

When putting these studies into perspective, it is also important to realize that in most cases, no single low-frequency or rare variants surpassed genome-wide thresholds for significant association. Rare variants that were significantly associated were either found in population isolates[72], in genes directly related to the trait of interest (low-density lipoprotein (LDL) receptor variants and LDL cholesterol levels)[462] or in complex diseases of a unique genetic architecture with few alleles of large effect sizes such as age-related macular degeneration[73]. If none of these situations were present, very large numbers of samples (i.e. more than 45,000) were needed to demonstrate significant association with the phenotype at genome-wide levels[74]. Hence, it is not surprising that we did not identify a single rare variant of genome-wide significance associated with RLS or PD at the examined loci[III-X].

Studies addressing the role of rare variants as part of allelic series at loci identified by GWAS have largely focused on rare coding variants. Yet, it seems not unreasonable to hypothesize that rare functionally relevant variants might also exist within promoter or enhancer regions, microRNA (miRNA) or transcription factor binding sites or other regulatory elements located within the non-coding regions within or around a gene. In view of this, we included both the 5` and the 3` UTR in our analysis of rare variation at the *MEIS1* locus and found that rare variants located in the 3`UTR, on the whole, showed a slightly protective effect ($p < 0.05$; OR = 0.83) whereas rare variants within the 5`UTR constituted the strongest risk factor, overall ($p < 1x10^{-4}$; OR = 7.62). Moreover, we identified a low frequency variant (rs11693221) located in the 3`UTR of the canonical *MEIS1* transcript, which represents the largest genetic risk factor for RLS identified to date. The excess of rare non-coding variants in the 100 bp surrounding the exons of nine genes associated with asthma[75] and the fact that fine-mapping studies located about 22 % of 36 GWAS association signals for celiac disease to either the 5` or the 3` non-coding regions (UTRs and several kb up- or downstream)[463] indicates that these regions could indeed be important to the study of complex genetic diseases. Generally, this is an interesting but little explored concept that merits further attention in the future, especially in light of the fact that rare variants in the UTRs do *not* seem to contribute much to the genetic structure of rare Mendelian diseases.

In summary, to our knowledge, *MEIS1* represents one of very few genetic loci associated with a complex genetic disease for which such an extensive allelic series comprising common, low-frequency and rare variants, located in both coding and non-coding regions in and around the gene, which all seem to contribute to the genetic architecture of the disease, has been shown[I,III,IV].

3.1.3 Mutational Load (cf. ref. IV,VII)

While allelic series traditionally consider a collection of variants on a population level, the concept of mutational load seeks to characterize the mutational spectrum of one individual not restricted to only one locus but rather across the entire genome. Especially with regard to complex genetic diseases, it seems likely that the sum of genetic alterations in a given individual will be instrumental in determining the phenotype.

Along these lines, we sought to assess the contribution of rare and low-frequency coding variants in seven genes known to play a role in familial dementias to sporadic PD under the hypothesis that variants in other neurodegeneration genes might add to the mutational load or the "neurodegenerative burden" eliciting the PD phenotype. We observed an excess of low-frequency variants in 188 individuals who had been diagnosed with PD plus dementia (PD+D) compared to 188 individuals with PD (10.11% vs. 4.26% with a variant with MAF < 5 % in any of the seven genes). Also, identified rare and low-frequency variants in *APP* were found more frequently in individuals with sporadic PD than in the general population (27 out of 975 individuals with PD with a variant vs. 13 out of 1014 controls, $p < 0.02$, χ^2 test)[VII].

This study is hampered by the fact that clinical information regarding the dementia phenotype in the individuals with PD+D is very limited, that—by our current knowledge of the vast spectrum of rare and very rare genetic variation in humans[16,27,54,59]—the sample sizes are too small and the fact that common and non-coding variants were excluded from the study. Still, even if viewed as solely preliminary, the results provide an impetus to consider the possibility that the bulk of neurodegenerative diseases are genetically much more similar than previously assumed. The "dementia gene" variants could contribute to the overall "neurodegenerative burden" that an individual carries and which reflects susceptibility across all neurodegenerative conditions. In this scenario, an excess of genetic alterations in a specific pathway could then tip the balance towards one neurodegenerative phenotype or the other or create phenotypes in which features of multiple neurodegenerative diseases coexist. Growing evidence for intersecting pathways in, for example, PD and AD[465-467] also lends support to the concept that neurodegenerative diseases might genetically arise on the background of diverse amalgamations of different or shared genetic variants (i.e. the mutational load) in a mutual set of "neurodegenerative genes". Clinically and neuropathologically, such an overlap between, for example, AD and PD, has been described many times[345,468].

In non-human organisms such as viruses and butterflies, the significance of the concept of mutational load regardless of a specific phenotype has been illustrated by the observation that augmented genome-wide mutational burden decreases fitness and may even lead to extinction

of a species[469,470]. In humans, many instances have been reported where genetic variants in different genes were shown to be necessary to cause a given phenotype[e.g.471-475]. For instance, for Bardet-Biedl Syndrome (BBS), which is inherited in an autosomal recessive fashion, it could be illustrated that in some pedigrees three variant alleles in both *BBS2* and *BBS6* were necessary to cause disease in a triallelic fashion[473]. The examples reported thus far, however, were usually rare diseases with—at most—oligogenic patterns of inheritance. Also, in the field of ciliopathies it is known that genes known to cause certain ciliopathies also have the capacity to contribute pathogenic or modifying alleles to other ciliopathies[475-478]. To date no studies have been performed which have jointly analyzed mutational load across the full frequency spectrum of genomic (or exonic) variation for any somatic disease. However, large-scale whole genome sequencing efforts currently under way along with the development of appropriate statistical analysis tools will certainly provide new insights in the near future. Most recently, evidence from cancer genomes has surfaced which suggests, that here, too, the number and deleteriousness of auxiliary genetic variants influences tumor progression[479] thus highlighting the relevance of the mutational load to cancer genetics.

3.1.4 A Complex Interplay (cf. ref. I,II,IV,VII,IX)

Examples from the ciliopathy spectrum also demonstrate that even under a mutational load model, it is unlikely that variants can simply be added up to yield the observed phenotype[475]. Variants can appear functionally benign in one context but pathogenic in another[475]. For instance, null alleles in *NPHP6* are found across the entire ciliopathy severity spectrum from very mild to lethal and it seems probable that either the stochastic situation (as in the mutational load model) or functionally related *trans* alleles modulate phenotypic expressivity[475,480]. The existence of many genes or genetic variants with pleiotropic effects[26,290,481-484] further underscores that genetic variation is context specific and is dependent upon gene-gene and gene-environment interactions. The genotype at a single or few loci is unlikely to predict the phenotype accurately, especially with regard to diseases with a complex genetic framework[475]. This is not surprising since it has been suggested that the larger the biological complexity underlying a given phenotype the larger its locus heterogeneity[142]. When trying to piece together interdependent effects of heterogeneous variants, the situation soon becomes intransparent.

In this day, when variant discovery no longer presents a limiting factor to the identification of disease genes, it is becoming all the more important to properly catalogue the vast amount of data generated. The generation of locus-specific databases[475,485] seems vital in this context.

For PD, an undesirably large number of at least six such databases exist, while with regard to RLS, it is a work in progress. Although the most complete catalogization of human genetic variation currently remains a vision of the future, it is still a prerequisite to our ability to dissect out any of the existing interactions.

Next to our lack of knowledge of the exact variants involved in genetic interactions in a particular disease, currently, the statistical evaluation of possible interactions poses an important problem due to the immense burden of multiple testing for which one needs to correct. This is also the reason for the lack of evaluation of epistasis—the statistical dependence of expression of one genetic variant upon another—in most GWAS[33,45]. When using a SNP array containing one million markers, one would have to correct the analysis of genome-wide epistatic interactions for one trillion multiple tests, making it extremely difficult to obtain statistically significant results.

In one of the projects depicted in this work, blood *trans*-eQTLs of RLS-associated SNPs were analyzed[II]. The transcriptome-wide second most significantly regulated RNA expression by the lead SNP at the *TOX3* locus was that of one of six other "RLS genes", *MAP2K5* ($p=5.21x10^{-5}$, n=760 KORA general population controls). Although one would like to believe that this finding is true, it was far away from transcriptome-wide significance (defined at $p < 8.5x10^{-8}$) and also did not replicate in a second sample of 976 SHIP-TREND general population controls. This example demonstrates the statistical difficulties encountered in the analysis of the complex interplay of genes and genetic variants and suggests that sample sizes of magnitudes larger will be needed to statistically substantiate gene-gene interactive effects. As a consequence, it could potentially be easier to functionally evaluate suspected individual interactions in animal[486,487] or cell models[488]. Such approaches have already been successfully employed for rare mono- or oligogenic diseases[471-473] but are, for the most part, lacking for complex genetic phenotypes. A murine model harboring a whole allelic series of rare to common variation in *MEIS1* or the entire common mutational burden encountered across all RLS susceptibility alleles in individuals of extreme phenotypic presentation would be extremely interesting in this context. On a more cautionary note, it will be extremely difficult to model whole interaction networks and the interaction will always be removed from its genomic context. Also, as demonstrated by a large CNV unmasking a low-frequency regulatory SNP to cause thrombocyotopenia with absent radius syndrome[489], beyond the non-structural genetic variation addressed here, structural and non-structural changes may also interact in bringing about a phenotype. This adds yet another level of complexity to be accounted for in interaction studies. Nonetheless, the *trans*-eQTL analysis performed as part

of this work[II], represents the first systematic evaluation of any form of genetic interaction in the context of RLS and could provide a starting point for future studies.

3.1.5 Differences in Genetic Architecture between RLS and PD (adapted from ref. XI,XII)

In all likelihood, all of the above constructs contribute to the genetic composition of both RLS and PD—some to a lesser and some to a greater extent. These effects are the result of selective forces, environmental impact, population history, migration and mutation rates which have shaped and continue to shape the genetic architecture of each phenotype.[26,36,45] Consequently, it is clear that genetic architecture differs across the group of genetically complex phenotypes. Table 3.2 outlines the characteristics of this genetic framework in PD and RLS.

While many similarities are shared between the genetic features of the two diseases, some differences become apparent. Overall, genetic factors appear to play a slightly more influential role in bringing about RLS as epitomized by the higher heritability estimates. Interestingly, it also looks as if fewer variants of larger (but still very moderate) effect sizes might construct the genetic scaffolding of RLS, although dependable conclusions cannot be drawn from the currently available studies. In PD, variants are found at both extremes of the frequency/effect size spectrum (and some in between). In the past, however, it has been debated, whether the rare familial forms of PD brought about by highly penetrant rare alleles of large effect size and sporadic PD might not actually represent distinct phenotypes[345]. So that, derived from what is currently known, it could be speculated that PD could either be founded upon a more heterogeneous spectrum of genetic variation than RLS—at least frequency-wise—or could be a collection of different forms of the same disease. Conversely, in the most extreme scenario which is supported by the lack of successful linkage studies in RLS[270], this could also mean that there is no single genetic variant that on its own is able to precipitate the RLS phenotype, rendering RLS a true complex genetic disease in all its facets. In how far these conjectures truly reflect the nature of the underlying genetic architecture of disease and not solely ascertainment differences impacted upon by sample number, general research intensity and focus in the field or methodological differences, cannot be fully established at the moment. However, although many of the specifics remain to be elucidated it does become clear that important differences do exist in the genetic frameworks of both diseases. The further the genetic and non-genetic factors underlying both diseases are revealed, the more apparent it will become that the missing pieces of the genetics

puzzle (i.e. the “missing heritability”) could also be missing for different reasons in the two phenotypes[26] and that different strategies will be needed to address this.

An intriguing but little discussed genetic phenomenon in RLS is the fact that offspring generations with a higher percentage of affected individuals than in the parent generation are frequently encountered in RLS families. In most families and for six of the seven RLS linkage loci described so far, the projected pattern of inheritance is autosomal dominant. However, often times, one encounters more than 50% of a generation of offspring showing the RLS phenotype. This is also reflected by the fact that the sibling relative risk (3.6) is much larger than the offspring relative risk (1.8)[255]. In PD, on the other hand, families with projected autosomal dominant patterns of inheritance usually show less than 50% affected individuals in the offspring generation, which is attributed to reduced penetrance. A number of possible explanations ranging from ascertainment bias in the offspring generations to selective mating and environmental contributions in a setting of genetic predisposition have been suggested but none have been investigated systematically.

	PD	RLS
features shaping genetic architecture		
selective force/fitness	none for late onset, maybe for early-onset	none selective mating has been discussed
population history	similar	similar
migration	slightly less	slightly more
mutation rates	unknown overall high in known genes	unknown overall low in known genes
environmental impact	likely, mechanism unknown	likely, mechanism unknown
epigenetics	?	?
endophenotypes	many ?	some ?
known characteristics of variant architecture		
common variants	few with average ORs	few with relatively large ORs (currently 6.8% of heritability)
low-frequency variants	*GBA* most important risk factor known *LRRK2*	collectively and individually in *MEIS1* collectively across all GWAS loci
rare variants	several known single variants in familial PD with variable penetrance	seven linkage regions collectively in *MEIS1* unclear if a single such variant exists in RLS
de novo	?	?
CNVs	?	?
translocation/inversion	?	?
heterogeneity	large	large
factors in genetic epidemiology		
predominant sex	♂	♀
ethnicity	Caucasian>Asian>African	Caucasian>Asian>African
prevalence	1%	5 to 10%
age distribution	bimodal	bimodal
inheritance pattern	AD (AR, X-linked also described)	AD (AR also described)
positive family history	approx. 15%	approx. 50%
heritability	approx. 30%	approx. 50%
penetrance	overall relatively low, age-dependent	unknown, likely incomplete and age-dependent but to a lesser extent
expressivity	variable even within a family modifiers	extremely variable even within a family polygenic/modifiers
additional features		more than 50% affected offspring in some families

Table 3.2: Central features of the genetic architecture of PD and RLS. OR = odds ratio, AD = autosomal dominant, AR = autosomal recessive.

It is also possible that this phenomenon could lie in the genetic architecture of RLS itself. RLS is a complex genetic disease and locus heterogeneity appears firmly established. Accordingly, it is possible that a given mutational burden of causal, modifying and

predisposing alleles in any number of combinations could—in some cases—be responsible for the observed deviation from expected Mendelian ratios. The possibility of epistatic interactions between any of these genetic factors only adds another layer of complexity. Furthermore, at age-dependent prevalences of up to 10% in adult populations of European descent, RLS is much more common than PD. Accordingly, bilinearity with genetic susceptibility factors contributed by both parents (including an unaffected one) could play a role in explaining the observation. Lastly, evidence also exists that acquired epigenetic footprints can be passed on from generation to generation[490], a phenomenon which could also account for the increased number of affected offspring seen in RLS[491]. Yet, at present, any of these explanations are hypothetical and are not supported by any scientific evidence.

3.2 Challenges in Analyzing Common, Low Frequency and Rare Variants and Ways to address these Challenges

The analysis of the contribution of genetic variants of different frequency—and effect sizes—to the genetic make-up of a complex genetic disease such as RLS or PD faces a number of challenges. These shall be discussed in the following. Especially with regard to the analysis of rare and very rare variants statistical considerations present the largest hurdle to overcome.

3.2.1 Statistics (cf. ref. I-IV,VII, IX,X)

As outlined in the introduction, association tests are usually employed to demonstrate statistically significant association of a common variant with a phenotype. Yet, these associations only explain a very small portion of the estimated total heritability of a phenotype in most cases[33,41,45]. Variants with very small effect sizes and the need for extremely large sample sizes[4], the large burden of multiple testing, the uncertain causality of the identified variants[26] and statistical pitfalls such as undiscovered independent associations at a given locus[463], "phantom heritability"[9] or the concept of synthetic associations[43] all hamper the statistical analysis of common variant-disease associations. The need for multiple testing is inherent in the make-up of the human genome and is, therefore, difficult to address, as is the estimation of the exact heritability of a trait as long as there is no way to account for environmental effects throughout an individual`s lifetime and across multiple generations. Finding the causal variant underlying common variant association signals is sometimes possible but the process is usually slow and tedious[4,26]. Consortia and the ongoing efforts to continuously enlarge samples sizes[41] will ameliorate the ever present lack of power to detect small-effect variants. Still, it is questionable whether large enough numbers will be reached to

detect all common variants contributing to a phenotype. This is exemplified by the study of common variation contributing to human height where it is projected that 697 associated loci of genome-wide significance would be identified in 500,000 individuals but would only explain about 19.6 % of the heritability.[40] Fine-mapping and higher-resolution SNP arrays will also be able to detect additional independent common variant associations that can be found at many identified GWAS loci and which may contribute to dissecting the "missing heritability", as has recently been illustrated with regard to celiac disease.[463]
Synthetic associations[43], i.e. associations created not by the associated common variants but by other underlying factors such as a collection of individual rare variants in the proximity, represent another statistical challenge. Both the fact that in many cases GWAS yield consistent results even across ethnic groups[492] and the fact that allele frequencies of the GWAS signals are too high to be explained by rare variants[493] argue against the fact that synthetic associations play a large role in bringing about GWAS signals in complex diseases. Albeit, in a proof of principle study, it was recently demonstrated that rare variants in the low-density lipoprotein receptor known to cause familial hypercholesterolemia are able to create synthetic associations with common polymorphisms at the locus and as far as 2.4 Mb away from the causal variant.[494] In summary, it seems likely that, at least in some cases, rare variants may actually underlie associations observed in GWAS. The graduation to performing association studies by whole genome sequencing (compare Figure 1.7) could resolve this issue.
With regard to the statistical analysis of rare variants, the situation becomes even more precarious. Here more so than in the analysis of common variants, financial constraints, workload and computational capacities limit the scale of the projects that can currently be performed even under the umbrella of large consortia. Accordingly, these conundra also reflect the largest limitation of the rare variant studies depicted in this work—sample number. It is only very recently, that the field has come to truly understand how rare most variants are in the human population. In mid-2008, it was still postulated that there would be around ten rare variants per personal exome[77]. Today, estimates are about 1000x higher[3,IX,X] and it has become clear that there is a vast excess of rare and very rare variation in the human genome[16,27,54,57,59,61]. As evident from the work presented in this thesis, sample sizes of 1000 cases and 1000 controls are much too small to draw any meaningful conclusions regarding the role of rare or very rare variants (MAF < 1%) with regard to either known or suspected disease genes[IX,X]. The addition of frequency data from the *in-house* exomes (n = 1739), the 1000 genomes (n = 1092) or the NHLBI-ESP exomes (n = approx. 4300) also did not help to

solve this dilemma as the vast majority of variants were singletons. Only in datasets of more than 3000 cases plus 3000 controls did we start to see statistically meaningful differences between the number of rare variants but only when collapsed over one or several loci[IV]. Moreover, we observed that analyzing rare variants by collapsing using fixed MAF thresholds is highly prone to confounding effects from comparatively "common" single variants around MAF = 1 %, for example, that can easily abolish any statistically significant association the other variants with an aggregate MAF around 1 % might have. Similarly, it has been argued that the existence of variants of bidirectional effects within a set of rare variants will always render rare variant statistics imperfect[141]. Use of a variable-threshold model might alleviate this problem.

These findings are in line with what has been reported for large sets of exomes (n = 2440) from which it was projected that in individuals of European ancestry only 1.67% of 12,000 genes analyzed would have 80% power to detect a rare disease-linked variant at OR = 5 if screened in a sample consisting of 400 cases and 400 controls (Figure 3.3)[27]. In no gene would a variant of OR = 1.5 be detected under the same scenario.

However, next to the effect sizes and number of disease-relevant rare variants found in a gene, the success is also very dependent upon (1) the overall variance found in the gene of interest, (2) the variants` distribution across the frequency spectrum and (3) the variants` direction of effect. Although the overall variance and thus the number of identified variants in a gene cannot be altered, increasing the number of samples naturally also increases the number of

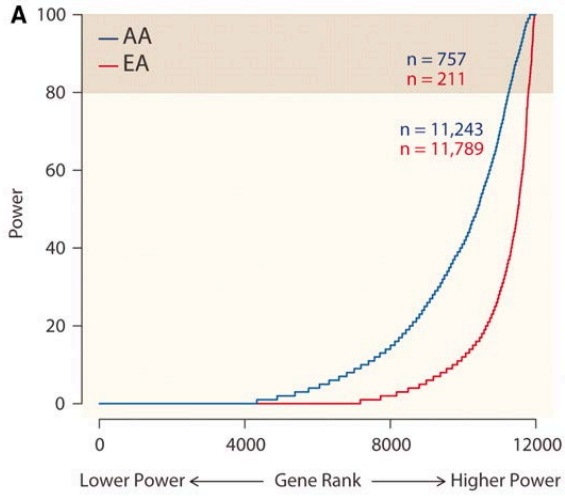


Figure 3.2: Power calculations for 12,000 genes harboring at least three SNVs and sequenced in 2440 individuals rare causal disease variants at OR = 5 in a sample of 400 cases and 400 controls. EA = European American, AA = African American. (taken from ref. [27])

identified variants and, thus, study power. To date, all studies have shown a linear increase in the number of identified variants in relation to the number of samples[16,27,54,61]. Current sample sizes are far away from reaching the point where every possible non-lethal variant has been detected although some argue that it is fair to assume that all non-lethal variation exists in at least one individual in today´s human population[142]. At present, large-scale exome sequencing studies estimate that tens of thousands of case and control samples will be needed to perform rare variant association studies[27,61]. In some extreme instances, the entire human population or a specific subpopulation (see 3.2.2 below) many not be large enough to generate statistical evidence for or against some very rare or small effect variants. It seems that only some consortia studying well-researched and well-funded phenotypes, such as, for example, human height, currently have access to the sample numbers and resources to carry out these studies. Next to the formation of cross-diagnostic consortia such as the Immunochip Consortium[495] and general large-scale sequencing endeavors such as the Beijing Genomics Institute`s "Million Human Genomes Project"[496], commercial sequencing providers might have an answer. California-based 23andMe, Inc., offers direct-to-consumer personal genome-wide genotyping and risk assessment for over 240 phenotypes[497]. Successful internet-based GWAS utilizing self-reported phenotypes have already been performed for PD[440], among others. By the end of 2013, a projected number of one million genome-wide genotyped individuals will be available for research purposes and could potentially be used in association studies of complex genetic phenotypes. Although not yet available, whole-genome and whole-exome sequencing services are sure to follow. Alternatively, *a priori* selection of individuals at the extremes of a phenotype could also lower the sample number requirements especially in the analysis of rare variants, since rare deleterious loss-of-function alleles are expected to be enriched in individuals at the extremes of a phenotypic spectrum[26,498]. This approach has already been effective in the exome-based identification of genetic variants predisposing individuals with cystic fibrosis to pseudomonal infection[6].

Another statistical dilemma is the identification of rare alleles of incomplete penetrance that either modify a phenotype[474] or are causal under an oligogenic model[77]. Such variants are near impossible to identify via family-based approaches because the variants do not show clear Mendelian segregation. However, it has been predicted that mildly deleterious low-frequency and rare variants harboring effect sizes of ORs between 2 and 3 could actually be responsible for a large portion of the genetic risk in common complex genetic diseases[65] and that a substantial portion of the "missing heritability" might be attributable to just this type of variation[26]. According to some estimates, about 20 variants with MAF = 1 % and OR = 3

could explain most of familial type II diabetes[26,56]. Others have suggested that the bulk of rare VUSs in breast cancer genes *BRCA1* and *BRCA2*, that are largely classified as such because they do not show clear familial segregation in the data currently in the public domain, may contribute 400x as much to the population attributable risk of breast cancer than the variants known to be pathogenic in these genes and would merit clinical intervention[77]. Here, too, very large numbers of samples will need to be analyzed to statistically link low-frequency variants of low penetrance to a phenotype. Myriad Genetics (Salt Lake City, UT, USA), the patent holder for the only available *BRCA* diagnostic test available in the US, already possess data on more than 1 million individuals screed and was able to use this information to reduce the percentage of VUS reported from 20% seen across European laboratories to 3%[499]. Unfortunately, the wealth of data available to Myriad Genetics is proprietary[499]. However, even when very large sample number are available, showing statistical association of rare variants of ORs < 2 with a phenotype will, even in the future, be extremely difficult (Figure 1.6).

Simpson`s paradox describes the phenomenon that trends seen in different groups of data disappear when the two groups are combined[500]. As is biologically plausible, both protective and predisposing alleles can exist at the same locus. For common susceptibility alleles, it has already been illustrated that multiple independently associated signals of different directions of effect at a given locus may exist but are only revealed after conditioning on the others[463,501]. With regard to rare variants, this was recently shown for *APP* in the context of AD[68]. Accordingly, Simpson`s paradox presents an important predicament in (rare) variant statistics where the common collapsing strategy is prone to the loss of statistical evidence due to such effects. Rare variant association tests specifically incorporating the analysis of effect direction (e.g. adaptive sums test[139] and KBAC[140]) can alleviate but not resolve this situation. Finally, the possibility that *trans* variants across the entire frequency spectrum could interact and unmask or disguise associations also exists[502]. Concealed rare variant associations between *BCL11A* and fetal hemoglobin levels in individuals with sickle cell anemia attributed to additional common variants on the same haplotype background have already been described[461]. The analysis of such interdependencies in the context of solely rare variants, however, will be very difficult to examine from a statistics perspective largely due to potentially insurmountable power impediments.

Overall, it seems likely that the “missing heritability” left in complex genetic diseases after the surge of GWAS will not likely be explained in its entirety in the near future, partly due to statistical culprits. And from a statistics perspective, the contribution of rare variants will

always be underestimated[141]. Nonetheless, it may still be possible to consider that, for all practical (and diagnostic or clinical) purposes, all heritability has been explained if the trait-associated genetic make-up of new individuals can be used to predict the actual phenotype[503]. Maybe this embodies the attainable maximum, at least in the next decade.

3.2.2 Population Specificity (cf. ref. I,II,IV,VI,VII,IX,X)

Common variation in the human genome is very old compared to rare variants[16]. As a consequence it is also much more similar across different human populations. This represents the basis for the realization that common susceptibility factors for a given complex genetic trait are often shared between human populations of European and Asian descent (only a minority of GWAS have actually been performed solely in individuals of African or Hispanic ancestry)[492]. In PD, all of the top associated loci found in Europeans have been replicated in Asians, unless associated SNPs were revealed to be monomorphic[438,439]. In some instances, different haplotype structure in different populations has also been used to finemap GWAS loci as in the case of the 16q12.2 locus associated with body mass index measurements[504]. While for five out of 18 GWAS loci for different blood lipid parameters, trans-ethic finemapping in 6,832 African Americans, 9,449 East Asians and 10,829 Europeans reduced the number of associated common variants, at two loci, distinct, ethnicity-specific signals were uncovered[505]. Accordingly, GWAS performed in different populations can be very valuable in refining GWAS association signals both by narrowing the associated LD block harboring a presumed single association signal as well as by highlighting genes or regions of genes through multiple ethnicity-specific association signals.

Although common variant associations are relatively stable across populations, for several reasons, this does not seem to be the case for rare variants. For one, rare variants are comparatively young and have neither been removed from the population by purifying selection nor have they become so frequent that they are fixed within a given population. Accordingly, any variant that developed within the last several thousand years is both likely to be rare but also to only occur in individuals who are offspring of the original founder—and, across the last millennia, these offspring are likely to belong to only one ethnic group—unless it developed independently on multiple occasions.

This represents a curse and a chance at the same time. In one of the projects depicted in this work, a rare coding variant in *APP* (p.E599K) was identified as a potential contributor to PD by candidate gene screening in a German and Austrian case/control sample[VII]. The variant was present in 14 out of 1068 individuals with PD (MAF = 0.66%) but only in 3 out of 1014

individuals belonging to the control sample (MAF = 0.15%) ($p_{nominal}$=0.01, χ^2 test)[VII]. In the NHLBI-ESP exomes, it was found in 9 out of 8591 sequenced alleles (MAF = 0.11%), a similar frequency to that seen in our controls. If one analyzes both control samples jointly (14 out of 1068 cases vs. 12 out of 5310 controls), the result becomes highly significant ($p_{nominal}=3.8x10^{-7}$, χ^2 test), once again demonstrating the power of sample numbers in rare variant statistics. Yet, any true association needs a replication. We genotyped the variant in an independent Spanish case/control sample consisting of 715 PD cases and 948 controls but did not find the *APP* p.E599K variant again in any of the 1663 individuals[VII]. Consequently, from a frequency-based perspective, the result remains inconclusive possibly due to a restriction of this rare variant to individuals of central European descent. In this context, it was illustrated employing the sequencing data of 14,002 individuals of different ancestry sequenced for 202 drug-target genes, that the number of variants observed differed not only across ethnic groups, as had been shown previously[506,507], but also within the European populations. On average, a German individual would possesses twice as many variants per kb compared as a Scandinavian individual. Also, whereas common variants appeared panmictic in the European populations, rare variants were much more dissimilar.[54]

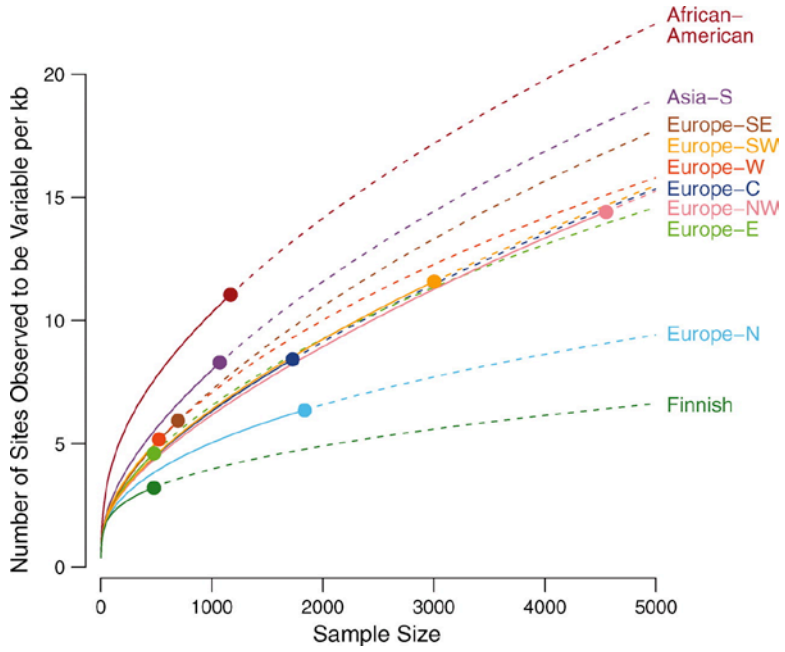


Figure 3.3: The number of variants per kb of sequence differs between individuals of different populations and subpopulations. Even within the European population, there is a stark north-to-south difference with Northern European individuals exhibiting only a third of the genetic variance of a Southern European individual. Most of this differential variance comprises rare variants. (taken from ref. [54])

This means that rare variant analysis will be most successful if performed in the most homogeneous sample possible and that it is vital to evaluate cases and controls from the same (sub-) population. On the downside, this may also mean that samples available from some subpopulations could be too few to statistically show rare variant associations in some instances.

Yet, the specificity of many rare variants to specific populations also has positive aspects. GWAS and NGS studies have largely been carried out in populations of European ancestry, although genetic variation is known to largest in populations of recent African ancestry[506,508]. The few studies performed in non-Europeans have already yielded intriguing new variants[509,510] that are just as useful in informing the biology of a disease as those identified but may be more easily ascertained than additional variants in Europeans. Similarly, population isolates may be enriched for single or a few rare variants of strong effect on a given complex genetic phenotype[511,512] and hold decreased variant and locus heterogeneity, thus facilitating variant (or association) discovery. Using the Icelandic population isolate, deCODE genetics[513] has been able to identify many low-frequency or rare variants involved in bringing about complex genetic diseases. One prominent example is the identification of a rare (MAF = 0.38%) missense variant in *MYH6* (p.Arg721Trp) predisposing to sick sinus syndrome with an OR of 12.53 that, so far, has exclusively been found in the Icelandic population[72].

In summary, the augmented population specificity of rare variants has the ability to both ease and hinder variant discovery at the same time. The analysis of the largest and most homogeneous samples possible will put this characteristic of rare variants to good use. Overall, the fact that, in theory, so many low-frequency and rare variants of moderate to high effect likely contribute to disease development across populations also holds the promise to generate better informed hypothesis regarding a disease`s underlying biology as that, in the end, is the vested interest. Analysis of low frequency and rare variants in *MEIS1* in a German/Austrian case/control sample have yielded, for the first time, variants of moderate to strong effect which can be used for functional follow-up studies inquiring into RLS pathophysiology[III,IV]. Still, whether these or other rare or low frequency variants also play a role in the genetic architecture of RLS in other European populations and especially in other Non-European populations remains to be established.

3.2.3 (High-Throughput) Functional Assessment of Identified Variants

GWAS have focused almost exclusively on statistical evidence and have de-emphasized considerations of biological relevance[26] and, with regard to rare variants, analysis—in many instances—is so underpowered that frequency assessment on its own cannot provide enough support to substantiate the involvement of single variants in disease pathogenesis (cf. Table 3.1). As a consequence, functional assessment of identified variants is needed to (1) identify that they harbor functional effects and are not benign and (2) to show that they are relevant to disease pathogenesis. With regard to (1), medium- or high-throughput methods that are, ideally, not gene-specific are needed to address the vast amount of genetic variation yielded by NGS studies. For (2), more refined disease- or gene-centered strategies commonly employed in molecular biology can be use. Accordingly, these will not be the focus of discussion below.

3.2.3.1 Cellular Assays (cf. ref. II,VII,IX,X)

Cellular phenotyping describes the quantification of the output of processes that occur at the cellular levels such as gene expression or metabolite production[488]. It can be used to evaluate intermediate phenotypes that are known to be more directly influenced by a disease-associated genetic variant than the disease phenotype itself. This is due to the fact that with increasing complexity from single cells to entire human beings, genetic effects are diluted by many layers of biological complexity. Cellular phenotypes can be quantified in primary tissues (often times as part of large-scale publicly accessible endeavors such as the Roadmap Genotype-Tissue Expression (GTEx) project[514,515]), primary or immortalized cell lines or in cells derived from patient-specific induced pluripotent stem cells (iPSCs). The different cell systems can be utilized to screen nuclear (e.g. eQTLs, methylation, chromatin and transcription factor QTLs) and cytoplasmic phenotypes (e.g. metabolite concentrations or enzyme activity) under steady-state or challenged conditions[488]. For the study of neurogenetic phenotypes, iPSCs are of particular interest as they represent the only means of obtaining living disease-relevant cells that—theoretically[516]—possess the patient`s genotype in all its complexity. In the context of PD, iPSCs have already been employed on a number of occasions to investigate the functional impact of recognized PD mutations particularly on mitochondrial biology[355,517,518]. With regard to RLS, no patient-specific iPSCs have been generated yet and the lack of a clear disease-relevant cellular phenotype further complicates the application of cellular phenotyping.

While for many decades the use of cell models to analyze genetic variants was hampered by being hypothesis-based, today, unbiased tissue studies are possible. These can be used to, for example, identify phenotype-relevant cell types if large collections of cell lines and tissues are utilized to screen for genotype-dependent functional profiles[488]. Such an approach could prove very valuable with regard to RLS, where the underlying causal cells are, at present, not known. In the work depicted herein, we sought to take advantage of blood-based eQTLs from two general population cohorts to prioritize sub-threshold association signals from a GWAS on RLS[II]. Our inability to establish new susceptibility alleles based on this approach could well be indicative of the fact that it is vital to study the cells or tissues most relevant to the phenotype[II].

Proponents of cellular phenotyping argue that genetic interactions that occur on one phenotypic level can be separated into individual linear effects in different layers and can, therefore, be examined independently[488,519]. Consequently, modern cellular phenotyping could be used to tackle the difficult task of integrating multilayered information in the functional analysis of genotype-phenotype correlations in complex genetic diseases.

3.2.3.2 Zebrafish Models (cf. ref. IV)

Another way to screen coding variants for their functional effects at medium- to high-throughput is by *in vivo* complementation assay in zebrafish embryos, as portrayed in the introduction[520-522] (cf. section 1.5.6). This assay was used to assess rare non-synonymous coding variants in *MEIS1* identified in the screening of 3760 individuals with RLS and 3542 control individuals. Although optic tectum size, the phenotypic read-out selected, is likely not directly involved in RLS pathogenesis, three interesting aspects emerged: (1) although RLS represents a comparatively mild, genetically complex phenotype, some individuals carry complete loss-of-function (i.e. null) alleles in *MEIS1* (the homozygous knock-out of which is known to be lethal in mice[293]), (2) the mechanism by which rare variants in *MEIS1* contribute to the RLS phenotype is likely a loss-of-function and (3) null but not hypomorphic alleles of *MEIS1* are enriched among individuals with RLS[IV]. Overall, this study depicts the first analysis of variants identified in the context of a complex genetic trait by *in vivo* complementation in zebrafish and one of very few functional evaluations of comprehensive sets of rare variants derived by sequencing large case/control samples of complex genetic disorders reported to date[IV,70]. Other potential applications of the complementation assay include the use in delineating causal genes from pools of candidate genes in the follow-up of

NGS experiments both in research[523] and clinical diagnostics (Prof Nicholas Katsanis & Prof Erica Davis, personal communication).

One major drawback of the *in vivo* complementation assay is that it is largely limited to coding variation. As many variants of regulatory effect are not located within the exonic regions of genes underlie the majority of GWAS association signals, the *in vivo* complementation assay in zebrafish is best suited to the study of coding variants such as those identified by targeted or whole exome sequencing. Yet, other tactics are currently in use to evaluate regulatory elements such as, for example, enhancer screens[524-529]. An additional disadvantage is the assays` limitation to the evaluation of embryonic phenotypes due to the transience of the morpholino-based knock-down. In the future, the combination of several techniques presently used to analyze different classes of variants in zebrafish could be used to construct complex genetic models of human diseases comprising both common and rare, coding and non-coding variation.

3.2.3.3 Mouse Models

Recent years have seen the emergence of more efficient techniques to generate transgenic mice than ever before. These new technologies utilize artificial restriction enzymes to induce and a cell`s endogenous machinery to repair DNA double strand breaks for *in situ* genome editing. Artificial restriction enzymes such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) or microbe-derived meganucleases can be engineered to specifically target nearly any DNA sequence and via repair mechanisms such as imperfect non-homologous end-joining and precise homology directed repair generate desired mutations. In many cases, synthetic oligonucleotides are used to further guide strand break repair. TALENs and ZFNs can be used to generate both murine knock-out[530] but also disease variant-specific knock-in models[531]. Lately, TALENs were used to generate a murine disease model of Hermansky-Pudlak syndrome harboring a specific missense mutation of *Rab38*[531], providing a foretaste of the great potential these technical advances hold for the follow-up of NGS candidate variants. ZFNs and TALENs have also been applied to create knock-out and knock-in transgenes in zebrafish[532-535] and a number of other traditional model but also non-model organisms[536].

Very recently, a novel method of nuclease-mediated genome editing using clustered regularly interspaced short palindromic repeats (CRISPRs)/CRISPR-associated (Cas) protein mRNA, usually active in the immune system of bacteria, plus single guide RNAs[487,537,538] was used to generate mice carrying multiple-allele substitutions in two genes belonging to the same family

in a single experiment[487]. This accomplishment is of vivid interest to the study of complex genetics as it signifies that the generation of complex genetic mouse models could, in the future, be a matter of weeks not years.

Although these techniques allow the generation of animal models harboring specific variants within four[487] to 18 weeks[531], this is still too labor– and time-consuming to test tens to hundreds of NGS-identified candidate variants but could prove very valuable in cases with only a handful of candidate variants and also in the creation of complex genetic animal models, for example, as models of entire allelic series such as the different frequency variants in *MEIS1* or mutational load models combining multiple GWAS-identified, potentially causal SNPs to model RLS.

At the moment, there is still an urgent need for more truly high-throughput ways to use experimentally determined functional effects in prioritizing the mass of candidate variants generated by NGS studies.

3.2.4 Regulation and Interaction (cf. ref. I,II)

Yet, being able to model single or several genetic alterations that are known to occur more frequently in individuals with a given disease hardly captures the entire picture of the disease. At the moment, it is still difficult to even fathom the enormous number of regulatory levels encoded by the human genome, let alone disentangle them. The encyclopedia of DNA elements (ENCODE) consortium[539] has spent the past decade cataloguing regulatory elements in the human genome with the result that regulation is far more complex than had been expected[539,540]. Despite the fact that the exome only makes up approximately 1.5 % of the human genome, in total, about 75 % of the genome are transcribed[539]. Accordingly the vast majority of DNA appears to serve a regulatory or yet unknown function[539] as exemplified by the 8.4 million short recognition sequences for DNA binding proteins[541] and the 3 million DNase 1 hypersensitivity sites marking regulatory DNA[542] as well as the added levels of complexity introduced by epigenetic modifications, miRNAs, feedback loops, functional redundancy and parent-of-origin specific effects[543], to name only a few. Accordingly, despite these large-scale efforts to map the human regulome in its entirety, it will take some time until we will be able to parse out how exactly the common and rare, coding and non-coding genetic variants that have been linked to a genetically complex phenotype by statistical means interact with each other and with the remainder of the genome to generate the full phenotype. Whole genome sequencing studies will provide the prerequisite for beginning to explore these levels of regulation in individuals or in a disease context.

One of the projects pursued as part of this work gives a notion of just how daunting this task may prove to be. Our study regarding *cis*-eQTLs in RLS assessed only a single level of single-order regulation[II]. However, we were unable to identify novel susceptibility factors for RLS. This may be owed to a number of factors: (1) the tissue of investigation is not the most relevant to the RLS phenotype and expression and expression regulation are likely to be tissue and cell-type specific[539,543], (2) the time point during an organism`s lifespan (development vs. adulthood) or during the circadian cycle that is most relevant to RLS is unknown and expression and expression regulation are likely to be different at different points in time, (3) we did identify many *cis*- and *trans*-eQTLs, however, due to the immense burden of multiple testing, very many of these did not reach statistical significance which is by no means synonymous with a lack of biological relevance, (4) the size of the replication sample is too small to yield statistically significant associations for variants of small effect sizes, (5) we did not attempt to analyze more than single-order interactions although it is very possible that single- or multiple-order epistatic effects exist[45], (6) array-based analyses such as GWAS and the expression studies part of this project do not render a bias-free depiction of the regulome and at best only analyze 1.5% of the transcriptionally regulated regions of the human genome[539] and (7) a last possibility could also be that it is inherent in the genetic architecture underlying RLS that there are no other genetic loci, other than those already known, of frequency and effect sizes which can be detected in the available case/control samples or that gene expression is not the most pertinent QTL.

Regulatory mechanisms involving direct DNA transcription such as those investigated by ENCODE or in the above study with regard to RLS, however, also only represent one of several layers of regulation. Protein expression, protein-protein interactions and protein phosphorylation states[543], the possibility of RNA editing[24,25] and environmental influences[45] illustrate additional levels at which regulation of pertinence to phenotypic expression may take place.

The magnitude of influence of regulation and interaction on the study of genetics was recently demonstrated in a yeast cross, where genetic interactions where found to contribute from zero to 54% to broad-sense heritability estimates in 24 of 46 traits examined[544]. The number of pairwise interactions ranged from 1 to 16 per trait[544]. In the most extreme case, a single strong interaction explained 14% of the genetic variance and 71% of the difference between broad-sense and narrow-sense heritability[544]. However, in the large majority of traits, pairwise gene-gene interactions only explained a minute fraction of the missing heritability or were not present at all[544]. Although the genetic architecture in the yeast model is infinitely less

complex than in humans, this study provides a first quantification of the possible magnitude of the role interaction effects have on complex genetic phenotypes and the “missing” heritability in humans.

3.2.5 The Great Beyond

When the concept of “missing heritability” was first described in 2008, it was already postulated that some of it could lie hidden in genetic and biological concepts that simply have not been discovered or thought of, to date[33]. In line with this, the ENCODE data revealed that approximately 75 % of the human genome is transcribed but the function of at least 60 % of these transcripts is completely unclear[539]. In light of the complexity of the human regulome, it could be possible that levels connecting genetic variation to a phenotype that have not yet been appreciated sufficiently exist. The fact that the master regulators in the non-coding RNA category—miRNAs and long intergenic non-coding RNAs (lincRNAs)—were only described in the last two decades[545,546] supports this notion. Possible examples for such levels could include transgenerational epigenetic and epistatic effects[45,547,548] where the grandparents` environmental exposure determines gene expression as has been shown with regard to murine coat color[549] or where the effect of modifier alleles increasing penetrance of *Dnd1* mutations in murine models of testicular germ cell tumors are passed on without the modifying allele itself[33,550]. One explanation for the latter could be the passing on of RNA molecules to the offspring[33] as has already been depiced in both plants[551,552] and *Caenorhabditis elegans*[553]. None of these mechanisms have been described in humans. An opposite mechanism has also been proposed which envisions that some of the “missing heritability” could lie in post-zygotic variation acquired by different non-cancerous cells throughout an organism`s lifetime[554-557] and is, therefore, not inherited. Lastly, it has also become clear that, in some cases, it is important from which parent a genetic variant was inherited[45,543]. An intronic variant in *HCCA2* on chromosome 11, for example, can either increase or reduce susceptibility to type 2 diabetes depending on the parent of origin[47]. Whichever the mechanisms may be, it is likely that our current understanding is only fragmentary but that the discovery of additional mechanisms of inheritance, interaction and regulation, which are sure to exist, will help tease apart complex genetic phenotypes.

3.3 Perspectives

The identification of common and rare variants underlying complex genetic diseases will in most cases not be able to explain the mechanism of how exactly a variant leads to a given

phenotype. The study of common and rare genetic variants in the context of common genetically complex diseases entails specific challenges depending on the method used to identify them, their frequency as well as effect sizes and penetrance. With regard to common susceptibility alleles, functional follow-up is often difficult and, if ascertained in a GWAS-approach, the identified SNP may not even be the causal one. The biggest challenge in the analysis of rare variants is the fact that very large sample numbers ranging in the hundred thousands and millions will be needed to begin to statistically judge the degree to which a given rare variant contributes to a phenotype. Still, in very many instances, this will never be possible from a statistical stand point as the involved variants are simply to rare in humans in general. Once rare variant analyses will be expanded to whole-genome sequencing as the main methodological approach and more and more rare, non-coding variants will be discovered, similar challenges in establishing the biological function of theses variants as those encountered with regard to common susceptibility alleles may emerge. A number of strategies have been devised to address the need for functional follow-up of both the common and the rare disease-linked variants and everything in between. Of these, *in vivo* complementation in zebrafish was chosen in the work depicted herein because it provides a relatively facile means to evaluate the functional effect of many coding variants within a given gene. Unfortunately, this approach is difficult to amend to non-coding variants. Novel nuclease-driven technologies using CRISPR/Cas or TALENs will hopefully provide the unique opportunity to study combinations of rare and common, coding and non-coding variants at the same or several loci in the future in the setting of complex genetics models in mice and zebrafish.

While much attention has been devoted to the study of both common and rare variants, the middle ground inhabited by the low-frequency variants has been neglected, partly due to technical challenges. Yet, as shown in the analysis of low-frequency variation in *MEIS1* depicted in this work, this may be unjust and at least with regard to the situation in RLS, truly a full spectrum of genetic variation of all different frequencies appears to contribute to disease development. Also, not only variants of different frequency but also different location with respect to the gene seem to be involved in RLS—coding variants, variants located in the 5`UTR, intronic variants and variants located in the intergenic regions very close to the annotated gene but also as far as 1.3 Mb downstream. To our knowledge, there are very few examples of loci contributing genetic variation to a phenotype in such a holistic fashion. Still, as opposed to PD, to date, no single genetic variant has been shown to be sufficient to induce the development of the RLS phenotype and the most common finding in RLS genetics has

been that of genetic heterogeneity. In the most extreme case, this could mean that there are no "causal" alleles for RLS and that the RLS phenotype is always the result of several genetic factors (possibly at a limited number of loci) acting in concert although the large pedigrees at first glance might suggest otherwise.

As in PD, many questions remain with regard to the genetic architecture of RLS (cf. Table 3.2). The answers to these questions will also tell us where—in each case—the missing heritability is likely to be found. Maybe the ultimate question is not whether a disease is monogenic or complex but rather what is the precise genetic architecture of the phenotype—or even more accurately—the individual genes with regard to a specific phenotype.

In the end, this also means that at least with regard to genetically complex neurologic diseases such as PD and RLS, where GWAS have identified a few susceptibility factors and rare variant identification by whole-exome sequencing has—in the majority of cases—proven more difficult than expected, we are still far away from an ultimate goal of individualized risk prediction and personalized medicine. And in view of the emerging multitude of regulatory layers, nobody knows if we will ever get there.

4 Summary

Genome-wide association studies (GWAS) have successfully identified common variants associated with increased susceptibility to both restless legs syndrome (RLS) and Parkinson`s diseases (PD), two common and—for the most part—genetically complex neurologic diseases. Moreover, for PD, linkage analyses have also identified rare variants of strong effect underlying familial forms, whereas for RLS, linkage analyses have not been equally successful—possibly due to the less intense research efforts in the field or a different underlying genetic architecture or variable phenotypic expressivity—and no variant of strong effect has been discovered to date. In both diseases, however, currently known genetic factors only explain a small portion of the heritability and many more factors remain to be discovered. Some of this so-called “missing heritability” could lie in a collection of additional common variants of relatively small effect such as those identified in GWAS but, in line with the “common disease, rare variant” hypothesis, rare variants are also likely to contribute to the genetic architecture of both diseases to a yet-unknown extent. In this work, a number of different strategies from exome sequencing to candidate gene screenings and GWAS were used to identify common, low-frequency and rare genetic factors, which contribute to the genetic architecture of RLS and PD.

For PD, exome sequencing studies in families identified *VPS35* as a novel causal genetic factor and *LRRK1* and *PLXNA4* as potential candidate genes in late-onset autosomal dominant PD. In *PLXNA4*, an excess of rare variants was identified in PD cases when compared to controls but awaits replication in an independent case/control sample. To explore the role of the “neurodegenerative mutational load”, we assessed low frequency and rare variants in seven genes known to be involved in familial forms of Alzheimer`s disease (AD) or frontotemoral dementia in individuals with Parkinson`s disease. Here, low-frequency and rare variants in these genes were more frequently encountered in individuals who had developed dementia during the course of PD. Also, in aggregate, variants in *APP*, which had not been previously described, were more common in individuals with PD (either with additional dementia or without) than in either controls or individuals with AD.

In the context of RLS, on-going work in the host laboratory used GWAS to identify common variants in *TOX3/BC034767* and an intergenic region on chromosome 2 as novel RLS susceptibility factors. Sub-threshold association signals from this GWAS were further analyzed by integrating expression quantitative trait loci (eQTLs). This yielded additional possible susceptibility factors that mandate further evaluation in much larger data sets. To test

whether rare and low frequency variants at the RLS-associated GWAS loci also contribute to the genetic architecture of the disease, the seven genes located at these loci were screened in a large-scale sequencing project. At the *MEIS1* locus, a low frequency variant located approximately 70 bp downstream of the 3` untranslated region (UTR) was identified as the, to date, largest genetic risk factor for RLS. Moreover, an excess of rare variants in the 5` UTR and of non-synonymous variants in the coding regions was present in individuals with RLS. *In vivo* complementation in zebrafish embryos suggested that the excess of rare non-synonymous variants is largely dependent on loss-of-function alleles.

In conclusion, the work depicted in this thesis supports the conception that rare and low frequency variants as well as common variants contribute to the genetic framework of complex diseases such as RLS and PD. It also shows that this contribution is different for different genes and that it might involve “cross-disorder” genes contributing to the total “neurodegenerative burden” across the entire genome such as in PD. As has only rarely been reported and as epitomized by the results in *MEIS1* in RLS illustrated here, in the most extensive form, a complete allelic series of variants of different frequencies and effects sizes, located within different regions in and around the gene can contribute to the genetic framework of a genetically complex phenotype.

5 References

5.1 Personal Publications Part of this Dissertation

I. Winkelmann J, Czamara D, Schormair B, Knauf F, Schulte EC, Trenkwalder C, Dauvilliers Y, Polo O, Högl B, Berger K, Fuhs A, Gross N, Stiasny-Kolster K, Oertel W, Bachmann CG, Paulus W, Xiong L, Montplaisir J, Rouleau GA, Fietze I, Vavrova J, Kemlink D, Sonka K, Nevsimalova S, Lin SC, Wszolek Z, Vilarino-Güell C, Faller MJ, Gschliesser V, Frauscher B, Falkenstetter T, Poewe W, Allen RP, Earley CJ, Ondo WG, Le WD, Spieler D, Kaffe M, Zimprich A, Kettunen J, Perola M, Silander K, Cournu-Rebeix I, Francavilla M, Fontenille C, Fontaine B, Vodicka P, Prokisch H, Lichtner P, Peppard P, Faraco J, Mignot E, Gieger C, Illig T, Wichmann HE, Müller-Myhsok B, Meitinger T. Genome-wide association study identifies novel restless legs syndrome susceptibility loci on 2p14 and 16q12.1. *PLoS Genet* **7**: e1002171 (2011).

II. Schulte EC, Heim K, Schurmann C, Homuth G, Lichtner P, Gieger C, Peters A, Trenkwalder C, Högl B, Frauscher B, Berger K, Fietze I, Gross N, Stiasny-Kolster K, Oertel W, Bachmann CG, Paulus W, Zimprich A, Müller-Myshok B, Prokisch H, Winkelmann J. Blood *cis*-eQTLs in Follow-Up Analysis of Genome-Wide Association Studies in Restless Legs Syndrome, submitted to *PLoS ONE*.

III. Schulte EC, Knauf F, Kemlink D, Gieger C, Lichtner P, Schormair B, Meitinger T, Winkelmann J. Variant Screening of the Coding Regions of MEIS1 in Patients with Restless Legs Syndrome. *Neurology* **76**: 1106-1108 (2011).

IV. Schulte EC[*], Kousi M[*], Tan P, Schormair B, Knauf F, Lichtner P, Trenkwalder C, Högl B, Frauscher B, Berger K, Fietze I, Gross N, Stiasny-Kolster K, Oertel W, Bachmann CG, Paulus W, Zimprich A, Peters A, Gieger C, Meitinger T, Müller-Myshok B, Katsanis N, Winkelmann J. An excess of rare loss-of-function alleles in individuals with restless legs syndrome substantiates *MEIS1* as the genetic factor underlying the GWAS locus on chromosome 2, *in preparation*.

V. Hopfner F, Schulte EC, Mollenhauer B, Bereznai B, Knauf F, Lichtner P, Zimprich A, Haubenberger D, Pirker W, Brücke T, Peters A, Gieger C, Kuhlenbäumer G, Trenkwalder C, Winkelmann J. The role of SCARB2 as susceptibility factor for Parkinson`s disease. *Mov Disord* **28**: 538-540 (2013).

VI. Schulte EC, Mollenhauer B, Zimprich A, Bereznai B, Lichtner P, Haubenberger D, Pirker W, Brücke T, Molnar MJ, Peters A, Gieger C, Trenkwalder C, Winkelmann J. Variants in Eukaryotic Translation Initiation Factor 4G1 in Sporadic Parkinson`s Disease. *Neurogenetics* **13**: 281-285 (2012).

VII. Schulte EC, Fukumori A, Mollenhauer B, Hor H, Arzberger T, Perneczky R, Kurz A, Hüll M, Lichtner P, Eckstein G, Zimprich A, Haubenberger D, Pirker W, Brücke T, Bereznai B, Molnar MJ, Lorenzo-Betancor O, Pastor P, Peters A, Gieger C, Estivill X, Meitinger T, Kretzschmar HA, Trenkwalder C, Haass C, Winkelmann J. Rare Variants in ß-Amyloid Precursor Protein (APP) and Parkinson`s Disease, *in preparation*.

VIII. Zimprich A, Benet-Pages A, Struhal W, Graf E, Eck SH, Offman MN, Haubenberger D, Spielberger S, Schulte EC, Lichtner P, Rossle SC, Klopp N, Wolf E,

Seppi K, Pirker W, Presslauer S, Mollenhauer B, Katzenschlager R, Foki T, Hotzy C, Reinthaler E, Harutyunyan A, Kralovics R, Peters A, Zimprich F, Brücke T, Poewe W, Auff E, Trenkwalder C, Rost B, Ransmayr G, Winkelmann J, Meitinger T, Strom TM. A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. *Am J Hum Genet* **89**: 168-175 (2011).

IX. Schulte EC, Ellwanger DC, Dihanich S, Manzoni C, Stangl K, Schormair B, Graf E, Eck S, Mollenhauer B, Haubenberger D, Pirker W, Zimprich A, Brücke T, Lichtner P, Peters A, Gieger C, Trenkwalder C, Mewes HW, Meitinger T, Lewis PA, Klünemann HA, Winkelmann J. Rare Variants in *LRRK1* and Parkinson`s disease, submitted to *Neurobiol Aging*.

X. Schulte EC, Stahl I, Czamara D, Ellwanger DC, Eck S, Graf E, Mollenhauer B, Zimprich A, Lichtner P, Haubenberger D, Pirker W, Brücke T, Bereznai B, Molnar MJ, Peters A, Gieger C, Müller-Myhsok B, Trenkwalder C, Winkelmann J. Rare Variants in *PLXNA4* and Parkinson's disease, submitted to *Neurology*.

XI. Schulte EC, Winkelmann J. Family- and Genome-Wide Association Studies of Restless Legs Syndrome in *Genetics of Sleep and Sleep Disorders*. (Ed.: Thorpy M, Shaw P, Tafti M). Cambridge University Press, Cambridge. 2013, *in press*.

XII. Schulte EC, Winkelmann J. Clinical Phenotype and Genetics of Restless Legs Syndrome in *Movement Disorders: Genetics and Models*. (Ed.: LeDoux MS). Elsevier, San Diego, CA. 2013, *in press*.

5.2 General References

1 C. Schaaf, J. Z. *Basiswissen Humangenetik.* (Springer, 2008).

2 Diseases, E. O. f. R. *Rare Diseases: Understanding this Public Health Priority*, 2005).

3 Bras, J., Guerreiro, R. & Hardy, J. Use of next-generation sequencing and other whole-genome strategies to dissect neurological disease. *Nat Rev Neurosci* **13**, 453-464, doi:10.1038/nrn3271 (2012).

4 Gibson, G. Rare and common variants: twenty arguments. *Nature reviews. Genetics* **13**, 135-145, doi:10.1038/nrg3118 (2011).

5 Li, J. L. *et al.* A genome scan for modifiers of age at onset in Huntington disease: The HD MAPS study. *American journal of human genetics* **73**, 682-687, doi:10.1086/378133 (2003).

6 Emond, M. J. *et al.* Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. *Nature genetics* **44**, 886-889, doi:10.1038/ng.2344 (2012).

7 Metzger, S. *et al.* Huntingtin-associated protein-1 is a modifier of the age-at-onset of Huntington's disease. *Hum Mol Genet* **17**, 1137-1146, doi:10.1093/hmg/ddn003 (2008).

8 Strachan, T. R., A. *Human Molecular Genetics.* (Garland Science, 2011).

9 Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* **109**, 1193-1198, doi:10.1073/pnas.1119675109 (2012).

10 Barnes, D. E. & Lindahl, T. Repair and genetic consequences of endogenous DNA base damage in mammalian cells. *Annu Rev Genet* **38**, 445-476, doi:10.1146/annurev.genet.38.072902.092448 (2004).

11 McKinnon, P. J. & Caldecott, K. W. DNA strand break repair and human genetic disease. *Annu Rev Genomics Hum Genet* **8**, 37-55, doi:10.1146/annurev.genom.7.080505.115648 (2007).

12 Roach, J. C. *et al.* Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328**, 636-639, doi:10.1126/science.1186802 (2010).

13 Nachman, M. W. & Crowell, S. L. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**, 297-304 (2000).

14 Kondrashov, A. S. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Human mutation* **21**, 12-27, doi:10.1002/humu.10147 (2003).

15 Rauch, A. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380**, 1674-1682, doi:10.1016/S0140-6736(12)61480-9 (2012).

16 Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216-220, doi:10.1038/nature11690 (2013).

17 Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304-1351, doi:10.1126/science.1058040 (2001).

18 Venter, J. C. A part of the human genome sequence. *Science* **299**, 1183-1184, doi:10.1126/science.299.5610.1183 (2003).

19 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921, doi:10.1038/35057062 (2001).

20 Nystrom-Lahti, M. *et al.* Founding mutations and Alu-mediated recombination in hereditary colon cancer. *Nature medicine* **1**, 1203-1206 (1995).

21 Kazazian, H. H., Jr. *et al.* Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **332**, 164-166, doi:10.1038/332164a0 (1988).

22 Solyom, S. & Kazazian, H. H., Jr. Mobile elements in the human genome: implications for disease. *Genome Med* **4**, 12, doi:10.1186/gm311 (2012).

23 Hancks, D. C. & Kazazian, H. H., Jr. Active human retrotransposons: variation and disease. *Curr Opin Genet Dev* **22**, 191-203, doi:10.1016/j.gde.2012.02.006 (2012).

24 Schrider, D. R., Gout, J. F. & Hahn, M. W. Very few RNA and DNA sequence differences in the human transcriptome. *PLoS One* **6**, e25842, doi:10.1371/journal.pone.0025842 (2011).

25 Li, M. *et al.* Widespread RNA and DNA sequence differences in the human transcriptome. *Science* **333**, 53-58, doi:10.1126/science.1207018 (2011).

26 Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753, doi:10.1038/nature08494 (2009).

27 Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64-69, doi:10.1126/science.1219240 (2012).

28 Palmer, L. J. & Cardon, L. R. Shaking the tree: mapping complex disease genes with linkage disequilibrium. *Lancet* **366**, 1223-1234, doi:10.1016/S0140-6736(05)67485-5 (2005).

29 Kruglyak, L. & Nickerson, D. A. Variation is the spice of life. *Nature genetics* **27**, 234-236, doi:10.1038/85776 (2001).

30 Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362-9367, doi:10.1073/pnas.0903103106 (2009).

31 Gloyn, A. L. & McCarthy, M. I. Variation across the allele frequency spectrum. *Nature genetics* **42**, 648-650, doi:10.1038/ng0810-648 (2010).

32 Klein, R. J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385-389, doi:10.1126/science.1109557 (2005).

33 Maher, B. Personal genomes: The case of the missing heritability. *Nature* **456**, 18-21, doi:10.1038/456018a (2008).

34 Lander, E. S. The new genomics: global views of biology. *Science* **274**, 536-539 (1996).

35 Reich, D. E. & Lander, E. S. On the allelic spectrum of human disease. *Trends Genet* **17**, 502-510 (2001).

36 Pritchard, J. K. & Cox, N. J. The allelic architecture of human disease genes: common disease-common variant...or not? *Hum Mol Genet* **11**, 2417-2423 (2002).

37 Botstein, D. & Risch, N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature genetics* **33 Suppl**, 228-237, doi:10.1038/ng1090 (2003).

38 Visscher, P. M. *et al.* Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS genetics* **2**, e41, doi:10.1371/journal.pgen.0020041 (2006).

39 Visscher, P. M. Sizing up human height variation. *Nature genetics* **40**, 489-490, doi:10.1038/ng0508-489 (2008).

40 Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832-838, doi:10.1038/nature09410 (2010).

41 Antonarakis, S. E., Chakravarti, A., Cohen, J. C. & Hardy, J. Mendelian disorders and multifactorial traits: the big divide or one for all? *Nature reviews. Genetics* **11**, 380-384, doi:10.1038/nrg2793 (2010).

42 Lander, E. S. Initial impact of the sequencing of the human genome. *Nature* **470**, 187-197, doi:10.1038/nature09792 (2011).

43 Dickson, S. P., Wang, K., Krantz, I., Hakonarson, H. & Goldstein, D. B. Rare variants create synthetic genome-wide associations. *PLoS Biol* **8**, e1000294, doi:10.1371/journal.pbio.1000294 (2010).

44 International Schizophrenia, C. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748-752, doi:10.1038/nature08185 (2009).

45 Eichler, E. E. *et al.* Missing heritability and strategies for finding the underlying causes of complex disease. *Nature reviews. Genetics* **11**, 446-450, doi:10.1038/nrg2809 (2010).

46 Kong, A. *et al.* Sequence variants in the RNF212 gene associate with genome-wide recombination rate. *Science* **319**, 1398-1401, doi:10.1126/science.1152422 (2008).

47 Kong, A. *et al.* Parental origin of sequence variants associated with complex diseases. *Nature* **462**, 868-874, doi:10.1038/nature08625 (2009).

48 Cirulli, E. T. & Goldstein, D. B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature reviews. Genetics* **11**, 415-425, doi:10.1038/nrg2779 (2010).

49 Risch, N. J. Searching for genetic determinants in the new millennium. *Nature* **405**, 847-856, doi:10.1038/35015718 (2000).

50 Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516-1517 (1996).

51 Altmuller, J., Palmer, L. J., Fischer, G., Scherb, H. & Wjst, M. Genomewide scans of complex human diseases: true linkage is hard to find. *American journal of human genetics* **69**, 936-950, doi:10.1086/324069 (2001).

52 Sulem, P. *et al.* Identification of low-frequency variants associated with gout and serum uric acid levels. *Nature genetics* **43**, 1127-1130, doi:10.1038/ng.972 (2011).

53 Huyghe, J. R. *et al.* Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nature genetics* **45**, 197-201, doi:10.1038/ng.2507 (2013).

54 Nelson, M. R. *et al.* An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* **337**, 100-104, doi:10.1126/science.1217876 (2012).

55 Price, A. L. *et al.* Pooled association tests for rare variants in exon-resequencing studies. *American journal of human genetics* **86**, 832-838, doi:10.1016/j.ajhg.2010.04.005 (2010).

56 McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature reviews. Genetics* **9**, 356-369, doi:10.1038/nrg2344 (2008).

57 Li, Y. *et al.* Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nature genetics* **42**, 969-972, doi:10.1038/ng.680 (2010).

58 Keinan, A. & Clark, A. G. Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* **336**, 740-743, doi:10.1126/science.1217283 (2012).

59 Coventry, A. *et al.* Deep resequencing reveals excess rare recent variants consistent with explosive population growth. *Nat Commun* **1**, 131, doi:10.1038/ncomms1130 (2010).

60 Tanner, C. M. *et al.* Parkinson disease in twins: an etiologic study. *JAMA : the journal of the American Medical Association* **281**, 341-346 (1999).

61 Kiezun, A. *et al.* Exome sequencing and the genetic basis of complex traits. *Nature genetics* **44**, 623-630, doi:10.1038/ng.2303 (2012).

62 Ng, S. B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272-276, doi:10.1038/nature08250 (2009).

63 Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nature genetics* **42**, 30-35, doi:10.1038/ng.499 (2010).

64 Ng, S. B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nature genetics* **42**, 790-793, doi:10.1038/ng.646 (2010).

65 Pritchard, J. K. Are rare variants responsible for susceptibility to complex diseases? *American journal of human genetics* **69**, 124-137, doi:10.1086/321272 (2001).

66 Kryukov, G. V., Pennacchio, L. A. & Sunyaev, S. R. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *American journal of human genetics* **80**, 727-739, doi:10.1086/513473 (2007).

67 Marth, G. T. *et al.* The functional spectrum of low-frequency coding variation. *Genome Biol* **12**, R84, doi:10.1186/gb-2011-12-9-r84 (2011).

68 Jonsson, T. *et al.* A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* **488**, 96-99, doi:10.1038/nature11283 (2012).

69 Johansen, C. T. *et al.* Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nature genetics* **42**, 684-687, doi:10.1038/ng.628 (2010).

70 Bonnefond, A. *et al.* Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. *Nature genetics* **44**, 297-301, doi:10.1038/ng.1053 (2012).

71 Diogo, D. *et al.* Rare, low-frequency, and common variants in the protein-coding sequence of biological candidate genes from GWASs contribute to risk of rheumatoid arthritis. *American journal of human genetics* **92**, 15-27, doi:10.1016/j.ajhg.2012.11.012 (2013).

72 Holm, H. *et al.* A rare variant in MYH6 is associated with high risk of sick sinus syndrome. *Nature genetics* **43**, 316-320, doi:10.1038/ng.781 (2011).

73 Raychaudhuri, S. *et al.* A rare penetrant mutation in CFH confers high risk of age-related macular degeneration. *Nature genetics* **43**, 1232-1236, doi:10.1038/ng.976 (2011).

74 Rivas, M. A. *et al.* Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nature genetics* **43**, 1066-1073, doi:10.1038/ng.952 (2011).

75 Torgerson, D. G. *et al.* Resequencing candidate genes implicates rare variants in asthma susceptibility. *American journal of human genetics* **90**, 273-281, doi:10.1016/j.ajhg.2012.01.008 (2012).

76 Tom Strachan, A. R. *Human Molecular Genetics*. (Garland Science, 2011).

77 Bodmer, W. & Bonilla, C. Common and rare variants in multifactorial susceptibility to common diseases. *Nature genetics* **40**, 695-701, doi:10.1038/ng.f.136 (2008).

78 Mardis, E. R. A decade's perspective on DNA sequencing technology. *Nature* **470**, 198-203, doi:10.1038/nature09796 (2011).

79 National Human Genome Research Institute, N. I. o. H. *A Catalog of Published Genome-Wide Association Studies*, <http://www.genome.gov/gwasstudies/> (2013).

80 Schadt, E. E., Turner, S. & Kasarskis, A. A window into third-generation sequencing. *Hum Mol Genet* **19**, R227-240, doi:10.1093/hmg/ddq416 (2010).

81 Wichmann, H. E., Gieger, C., Illig, T. & Group, M. K. S. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67 Suppl 1**, S26-30, doi:10.1055/s-2005-858226 (2005).

82 Lacruz, M. E. *et al.* Mental health in the aged: prevalence, covariates and related neuroendocrine, cardiovascular and inflammatory factors of successful aging. *BMC Med Res Methodol* **10**, 36, doi:10.1186/1471-2288-10-36 (2010).

83 Elston, R. C. Introduction and overview. Statistical methods in genetic epidemiology. *Stat Methods Med Res* **9**, 527-541 (2000).

84 Cordell, H. J. & Clayton, D. G. Genetic association studies. *Lancet* **366**, 1121-1131, doi:10.1016/S0140-6736(05)67424-7 (2005).

85 Illumina, I. *Illumina Omni Genotyping Arrays*, <http://www.illumina.com/applications/genotyping/omni_family.ilmn> (2013).

86 Reich, D. E. *et al.* Linkage disequilibrium in the human genome. *Nature* **411**, 199-204, doi:10.1038/35075590 (2001).

87 Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225-2229, doi:10.1126/science.1069424 (2002).

88 Pritchard, J. K. & Przeworski, M. Linkage disequilibrium in humans: models and data. *American journal of human genetics* **69**, 1-14, doi:10.1086/321275 (2001).

89 Ardlie, K. G., Kruglyak, L. & Seielstad, M. Patterns of linkage disequilibrium in the human genome. *Nature reviews. Genetics* **3**, 299-309, doi:10.1038/nrg777 (2002).

90 Hill, W. G., Goddard, M. E. & Visscher, P. M. Data and theory point to mainly additive genetic variance for complex traits. *PLoS genetics* **4**, e1000008, doi:10.1371/journal.pgen.1000008 (2008).

91 Balding, D. J. A tutorial on statistical methods for population association studies. *Nature reviews. Genetics* **7**, 781-791, doi:10.1038/nrg1916 (2006).

92 Ziegler, A., Konig, I. R. & Thompson, J. R. Biostatistical aspects of genome-wide association studies. *Biom J* **50**, 8-28, doi:10.1002/bimj.200710398 (2008).

93 Cardon, L. R. & Bell, J. I. Association study designs for complex diseases. *Nature reviews. Genetics* **2**, 91-99, doi:10.1038/35052543 (2001).

94 Fan, J. B., Chee, M. S. & Gunderson, K. L. Highly parallel genomic assays. *Nature reviews. Genetics* **7**, 632-644, doi:10.1038/nrg1901 (2006).

95 International HapMap, C. A haplotype map of the human genome. *Nature* **437**, 1299-1320, doi:10.1038/nature04226 (2005).

96 Sachidanandam, R. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928-933, doi:10.1038/35057149 (2001).

97 Nicolae, D. L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS genetics* **6**, e1000888, doi:10.1371/journal.pgen.1000888 (2010).

98 Richards, A. L. *et al.* Schizophrenia susceptibility alleles are enriched for alleles that affect gene expression in adult human brain. *Mol Psychiatry* **17**, 193-201, doi:10.1038/mp.2011.11 (2012).

99 Fransen, K. *et al.* Analysis of SNPs with an effect on gene expression identifies UBE2L3 and BCL3 as potential new risk genes for Crohn's disease. *Hum Mol Genet* **19**, 3482-3488, doi:10.1093/hmg/ddq264 (2010).

100 Moffatt, M. F. *et al.* Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* **448**, 470-473, doi:10.1038/nature06014 (2007).

101 Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-5467 (1977).

102 Sanger, F. & Coulson, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* **94**, 441-448 (1975).

103 Wittwer, C. T., Reed, G. H., Gundry, C. N., Vandersteen, J. G. & Pryor, R. J. High-resolution genotyping by amplicon melting analysis using LCGreen. *Clinical chemistry* **49**, 853-860 (2003).

104 Herrmann, M. G., Durtschi, J. D., Bromley, L. K., Wittwer, C. T. & Voelkerding, K. V. Amplicon DNA melting analysis for mutation scanning and genotyping: cross-platform comparison of instruments and dyes. *Clinical chemistry* **52**, 494-503, doi:10.1373/clinchem.2005.063438 (2006).

105 Chou, L. S., Lyon, E. & Wittwer, C. T. A comparison of high-resolution melting analysis with denaturing high-performance liquid chromatography for mutation scanning: cystic fibrosis transmembrane conductance regulator gene as a model. *Am J Clin Pathol* **124**, 330-338, doi:10.1309/BF3M-LJN8-J527-MWQY (2005).

106 Reed, G. H. & Wittwer, C. T. Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis. *Clinical chemistry* **50**, 1748-1754, doi:10.1373/clinchem.2003.029751 (2004).

107 Liew, M. *et al.* Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. *Clinical chemistry* **50**, 1156-1164, doi:10.1373/clinchem.2004.032136 (2004).

108 Zhou, L., Myers, A. N., Vandersteen, J. G., Wang, L. & Wittwer, C. T. Closed-tube genotyping with unlabeled oligonucleotide probes and a saturating DNA dye. *Clinical chemistry* **50**, 1328-1335, doi:10.1373/clinchem.2004.034322 (2004).

109 van der Stoep, N. *et al.* Diagnostic guidelines for high-resolution melting curve (HRM) analysis: an interlaboratory validation of BRCA1 mutation scanning using the 96-well LightScanner. *Human mutation* **30**, 899-909, doi:10.1002/humu.21004 (2009).

110 De Leeneer, K., Coene, I., Poppe, B., De Paepe, A. & Claes, K. Rapid and sensitive detection of BRCA1/2 mutations in a diagnostic setting: comparison of two high-resolution melting platforms. *Clinical chemistry* **54**, 982-989, doi:10.1373/clinchem.2007.098764 (2008).

111 Dawn Teare, M. & Barrett, J. H. Genetic linkage studies. *Lancet* **366**, 1036-1044, doi:10.1016/S0140-6736(05)67382-5 (2005).

112 Johnson, J. O. *et al.* Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron* **68**, 857-864, doi:10.1016/j.neuron.2010.11.036 (2010).

113 Sobreira, N. L. *et al.* Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene. *PLoS genetics* **6**, e1000991, doi:10.1371/journal.pgen.1000991 (2010).

114 Metzker, M. L. Sequencing technologies - the next generation. *Nature reviews. Genetics* **11**, 31-46, doi:10.1038/nrg2626 (2010).

115 Illumina, I. *DNA Sequencing*, <http://www.illumina.com/applications/sequencing.ilmn> (2013).

116 Singleton, A. B. Exome sequencing: a transformative technology. *Lancet neurology* **10**, 942-946, doi:10.1016/S1474-4422(11)70196-X (2011).

117 Lupski, J. R. *et al.* Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* **362**, 1181-1191, doi:10.1056/NEJMoa0908094 (2010).

118 Harismendy, O. *et al.* Population sequencing of two endocannabinoid metabolic genes identifies rare and common regulatory variants associated with extreme obesity and metabolite level. *Genome Biol* **11**, R118, doi:10.1186/gb-2010-11-11-r118 (2010).

119 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

120 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

121 Cooper, G. M. & Shendure, J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nature reviews. Genetics* **12**, 628-640, doi:10.1038/nrg3046 (2011).

122 Information, N. C. f. B. *dbSNP*, <http://www.ncbi.nlm.nih.gov/projects/SNP/> (2013).

123 Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65, doi:10.1038/nature11632 (2012).

124 1000 Genomes, E.-E. *1000 Genomes--A Deep Catalog of Human Genetic Variation*, 2013).

125 Project, N. G. O. E. S. *The NHLBI GO ESP Exome Variant Server*, <http://evs.gs.washington.edu/EVS/> (2013).

126 Cooper, G. M. *et al.* Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nat Methods* **7**, 250-251, doi:10.1038/nmeth0410-250 (2010).

127 Ng, P. C. & Henikoff, S. Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet* **7**, 61-80, doi:10.1146/annurev.genom.7.080505.115630 (2006).

128 Kilpinen, H. & Dermitzakis, E. T. Genetic and epigenetic contribution to complex traits. *Hum Mol Genet* **21**, R24-28, doi:10.1093/hmg/dds383 (2012).

129 Gorlov, I. P., Gorlova, O. Y., Sunyaev, S. R., Spitz, M. R. & Amos, C. I. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *American journal of human genetics* **82**, 100-112, doi:10.1016/j.ajhg.2007.09.006 (2008).

130 Altshuler, D., Daly, M. J. & Lander, E. S. Genetic mapping in human disease. *Science* **322**, 881-888, doi:10.1126/science.1156409 (2008).

131 Li, B. & Leal, S. M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *American journal of human genetics* **83**, 311-321, doi:10.1016/j.ajhg.2008.06.024 (2008).

132 Bansal, V., Libiger, O., Torkamani, A. & Schork, N. J. Statistical analysis strategies for association studies involving rare variants. *Nature reviews. Genetics* **11**, 773-785, doi:10.1038/nrg2867 (2010).

133 Tavtigian, S. V., Hashibe, M. & Thomas, A. Tests of association for rare variants: case control mutation screening. *Nature reviews. Genetics* **12**, 224, doi:10.1038/nrg2867-c1 (2011).

134 Asimit, J. & Zeggini, E. Rare variant association analysis methods for complex traits. *Annu Rev Genet* **44**, 293-308, doi:10.1146/annurev-genet-102209-163421 (2010).

135 Morgenthaler, S. & Thilly, W. G. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutat Res* **615**, 28-56, doi:10.1016/j.mrfmmm.2006.09.003 (2007).

136 Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368-372, doi:10.1038/nature09146 (2010).

137 Sebat, J., Levy, D. L. & McCarthy, S. E. Rare structural variants in schizophrenia: one disorder, multiple mutations; one mutation, multiple disorders. *Trends Genet* **25**, 528-535, doi:10.1016/j.tig.2009.10.004 (2009).

138 Madsen, B. E. & Browning, S. R. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS genetics* **5**, e1000384, doi:10.1371/journal.pgen.1000384 (2009).

139 Han, F. & Pan, W. A data-adaptive sum test for disease association with multiple common or rare variants. *Hum Hered* **70**, 42-54, doi:10.1159/000288704 (2010).

140 Liu, D. J. & Leal, S. M. A novel adaptive method for the analysis of next-generation sequencing data to detect complex trait associations with rare variants due to gene main effects and interactions. *PLoS genetics* **6**, e1001156, doi:10.1371/journal.pgen.1001156 (2010).

141 Liu, D. J. & Leal, S. M. Estimating genetic effects and quantifying missing heritability explained by identified rare-variant associations. *American journal of human genetics* **91**, 585-596, doi:10.1016/j.ajhg.2012.08.008 (2012).

142 McClellan, J. & King, M. C. Genetic heterogeneity in human disease. *Cell* **141**, 210-217, doi:10.1016/j.cell.2010.03.032 (2010).

143 Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248-249, doi:10.1038/nmeth0410-248 (2010).

144 Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073-1081, doi:10.1038/nprot.2009.86 (2009).

145 Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* **7**, e46688, doi:10.1371/journal.pone.0046688 (2012).

146 Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* **7**, 575-576, doi:10.1038/nmeth0810-575 (2010).

147 Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res* **19**, 1553-1561, doi:10.1101/gr.092619.109 (2009).

148 Draper, B. W., Morcos, P. A. & Kimmel, C. B. Inhibition of zebrafish fgf8 pre-mRNA splicing with morpholino oligos: a quantifiable method for gene knockdown. *Genesis* **30**, 154-156 (2001).

149 Summerton, J. Morpholino antisense oligomers: the case for an RNase H-independent structural type. *Biochim Biophys Acta* **1489**, 141-158 (1999).

150 Berger, K., Luedemann, J., Trenkwalder, C., John, U. & Kessler, C. Sex and the risk of restless legs syndrome in the general population. *Arch Intern Med* **164**, 196-202, doi:10.1001/archinte.164.2.196 (2004).

151 Hogl, B. *et al.* Restless legs syndrome: a community-based study of prevalence, severity, and risk factors. *Neurology* **64**, 1920-1924, doi:10.1212/01.WNL.0000163996.64461.A3 (2005).

152 Chen, N. H. *et al.* The prevalence of restless legs syndrome in Taiwanese adults. *Psychiatry Clin Neurosci* **64**, 170-178, doi:10.1111/j.1440-1819.2010.02067.x (2010).

153 Sevim, S. *et al.* Unexpectedly low prevalence and unusual characteristics of RLS in Mersin, Turkey. *Neurology* **61**, 1562-1569 (2003).

154 Lazzarini, A. *et al.* Studies of penetrance and anticipation in five autosomal-dominant restless legs syndrome pedigrees. *Movement disorders : official journal of the Movement Disorder Society* **14**, 111-116 (1999).

155 Pantaleo, N. P., Hening, W. A., Allen, R. P. & Earley, C. J. Pregnancy accounts for most of the gender difference in prevalence of familial RLS. *Sleep Med* **11**, 310-313, doi:10.1016/j.sleep.2009.04.005 (2010).

156 Chen, S. *et al.* Genomewide linkage scan identifies a novel susceptibility locus for restless legs syndrome on chromosome 9p. *American journal of human genetics* **74**, 876-885, doi:10.1086/420772 (2004).

157 Michaud, M., Paquet, J., Lavigne, G., Desautels, A. & Montplaisir, J. Sleep laboratory diagnosis of restless legs syndrome. *Eur Neurol* **48**, 108-113 (2002).

158 Coleman, R. M. *et al.* Daytime sleepiness in patients with periodic movements in sleep. *Sleep* **5 Suppl 2**, S191-202 (1982).

159 Hornyak, M., Feige, B., Riemann, D. & Voderholzer, U. Periodic leg movements in sleep and periodic limb movement disorder: prevalence, clinical significance and treatment. *Sleep Med Rev* **10**, 169-177, doi:10.1016/j.smrv.2005.12.003 (2006).

160 Allen, R. P. *et al.* Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology. A report from the restless legs syndrome diagnosis and epidemiology workshop at the National Institutes of Health. *Sleep Med* **4**, 101-119 (2003).

161 Benes, H. & Kohnen, R. Validation of an algorithm for the diagnosis of Restless Legs Syndrome: The Restless Legs Syndrome-Diagnostic Index (RLS-DI). *Sleep Med* **10**, 515-523, doi:10.1016/j.sleep.2008.06.006 (2009).

162 Trenkwalder, C. & Paulus, W. Restless legs syndrome: pathophysiology, clinical presentation and management. *Nat Rev Neurol* **6**, 337-346, doi:10.1038/nrneurol.2010.55 (2010).

163 Trenkwalder, C. *et al.* Circadian rhythm of periodic limb movements and sensory symptoms of restless legs syndrome. *Movement disorders : official journal of the Movement Disorder Society* **14**, 102-110 (1999).

164 Stiasny-Kolster, K., Kohnen, R., Moller, J. C., Trenkwalder, C. & Oertel, W. H. Validation of the "L-DOPA test" for diagnosis of restless legs syndrome. *Movement disorders : official journal of the Movement Disorder Society* **21**, 1333-1339, doi:10.1002/mds.20969 (2006).

165 Montplaisir, J. *et al.* Clinical, polysomnographic, and genetic characteristics of restless legs syndrome: a study of 133 patients diagnosed with new standard criteria. *Movement disorders : official journal of the Movement Disorder Society* **12**, 61-65, doi:10.1002/mds.870120111 (1997).

166 Trenkwalder, C. B., H.; Hornyak, M.; Stiasny-Kolster, K.; Winkelmann, J.; Högl, B.; Mathis, J. in *Leitlinien für Diagnostik und Therapie in der Neurologie* (ed Kommission Leitlinien der Deutschen Gesellschaft für Neurologie) 56-67 (Thieme, 2008).

167 Scholz, H. *et al.* Psychological distress of patients suffering from restless legs syndrome: a cross-sectional study. *Health and quality of life outcomes* **9**, 73, doi:10.1186/1477-7525-9-73 (2011).

168 Yngman-Uhlin, P., Johansson, A., Fernstrom, A., Borjeson, S. & Edell-Gustafsson, U. Fragmented sleep: an unrevealed problem in peritoneal dialysis patients. *Scandinavian journal of urology and nephrology* **45**, 206-215, doi:10.3109/00365599.2011.557025 (2011).

169 Brand, S., Beck, J., Hatzinger, M., Savic, M. & Holsboer-Trachsler, E. Unfavorable polysomnographic sleep patterns predict poor sleep and poor psychological functioning 3 years later in patients with restless legs syndrome. *Neuropsychobiology* **63**, 92-102, doi:10.1159/000317549 (2011).

170 Winkelmann, J. *et al.* "Anxietas tibiarum". Depression and anxiety disorders in patients with restless legs syndrome. *J Neurol* **252**, 67-71, doi:10.1007/s00415-005-0604-7 (2005).

171 Winkelmann, J. & Muller-Myhsok, B. Genetics of restless legs syndrome: a burning urge to move. *Neurology* **70**, 664-665, doi:10.1212/01.wnl.0000302178.53759.92 (2008).

172 Walters, A. S. & Rye, D. B. Review of the relationship of restless legs syndrome and periodic limb movements in sleep to hypertension, heart disease, and stroke. *Sleep* **32**, 589-597 (2009).

173 Schulte, E. C., Spieler, D. & Winkelmann, J. [Restless legs syndrome and cardiovascular risk]. *Nervenarzt* **82**, 1006-1011, doi:10.1007/s00115-010-3185-3 (2011).

174 Schulte, E. C. G., N.; Slawik, H.; Winkelmann, J. When Restless Legs Syndrome Turns Malignant. *Sleep Med* http://dx.doi.org/10.1016/j.sleep.2013.02.012 (2013).

175 Hornyak, M., Sohr, M., Busse, M. & Study, G. Evaluation of painful sensory symptoms in restless legs syndrome: experience from two clinical trials. *Sleep Med* **12**, 186-189, doi:10.1016/j.sleep.2010.11.007 (2011).

176 Scholz, H. *et al.* Dopamine agonists for restless legs syndrome. *Cochrane Database Syst Rev*, CD006009, doi:10.1002/14651858.CD006009.pub2 (2011).

177 Scholz, H. *et al.* Levodopa for restless legs syndrome. *Cochrane Database Syst Rev*, CD005504, doi:10.1002/14651858.CD005504.pub2 (2011).

178 Walters, A. S. *et al.* Successful treatment of the idiopathic restless legs syndrome in a randomized double-blind trial of oxycodone versus placebo. *Sleep* **16**, 327-332 (1993).

179 Kushida, C. A. *et al.* Randomized, double-blind, placebo-controlled study of XP13512/GSK1838262 in patients with RLS. *Neurology* **72**, 439-446, doi:10.1212/01.wnl.0000341770.91926.cc (2009).

180 Wang, J., O'Reilly, B., Venkataraman, R., Mysliwiec, V. & Mysliwiec, A. Efficacy of oral iron in patients with restless legs syndrome and a low-normal ferritin: A randomized, double-blind, placebo-controlled study. *Sleep Med* **10**, 973-975, doi:10.1016/j.sleep.2008.11.003 (2009).

181 Allen, R. P. *et al.* Clinical efficacy and safety of IV ferric carboxymaltose (FCM) treatment of RLS: a multi-centred, placebo-controlled preliminary clinical trial. *Sleep Med* **12**, 906-913, doi:10.1016/j.sleep.2011.06.009 (2011).

182 Bassetti, C. L., Mauerhofer, D., Gugger, M., Mathis, J. & Hess, C. W. Restless legs syndrome: a clinical study of 55 patients. *Eur Neurol* **45**, 67-74, doi:52098 (2001).

183 Winkelmann, J. *et al.* Clinical characteristics and frequency of the hereditary restless legs syndrome in a population of 300 patients. *Sleep* **23**, 597-602 (2000).

184 Walters, A. S. *et al.* A questionnaire study of 138 patients with restless legs syndrome: the 'Night-Walkers' survey. *Neurology* **46**, 92-95 (1996).

185 Winkelmann, J. *et al.* Complex segregation analysis of restless legs syndrome provides evidence for an autosomal dominant mode of inheritance in early age at onset families. *Ann Neurol* **52**, 297-302, doi:10.1002/ana.10282 (2002).

186 Uglane, M. T., Westad, S. & Backe, B. Restless legs syndrome in pregnancy is a frequent disorder with a good prognosis. *Acta Obstet Gynecol Scand* **90**, 1046-1048, doi:10.1111/j.1600-0412.2011.01157.x (2011).

187 Manconi, M. *et al.* Restless legs syndrome and pregnancy. *Neurology* **63**, 1065-1069 (2004).

188 Neau, J. P. *et al.* Restless legs syndrome and pregnancy: follow-up of pregnant women before and after delivery. *Eur Neurol* **64**, 361-366, doi:10.1159/000322124 (2010).

189 Collado-Seidel, V. *et al.* Clinical and biochemical findings in uremic patients with and without restless legs syndrome. *Am J Kidney Dis* **31**, 324-328 (1998).

190 Kryger, M. H., Shepertycky, M., Foerster, J. & Manfreda, J. Sleep disorders in repeat blood donors. *Sleep* **26**, 625-626 (2003).

191 Berger, K. *et al.* Iron metabolism and the risk of restless legs syndrome in an elderly general population--the MEMO-Study. *J Neurol* **249**, 1195-1199, doi:10.1007/s00415-002-0805-2 (2002).

192 Hattan, E., Chalk, C. & Postuma, R. B. Is there a higher risk of restless legs syndrome in peripheral neuropathy? *Neurology* **72**, 955-960, doi:10.1212/01.wnl.0000336341.72621.db (2009).

193 Gemignani, F. *et al.* Cryoglobulinaemic neuropathy manifesting with restless legs syndrome. *J Neurol Sci* **152**, 218-223 (1997).

194 Gemignani, F., Brindani, F., Vitetta, F., Marbini, A. & Calzetti, S. Restless legs syndrome in diabetic neuropathy: a frequent manifestation of small fiber neuropathy. *J Peripher Nerv Syst* **12**, 50-53, doi:10.1111/j.1529-8027.2007.00116.x (2007).

195 Gemignani, F. *et al.* Restless legs syndrome and polyneuropathy. *Movement disorders : official journal of the Movement Disorder Society* **21**, 1254-1257, doi:10.1002/mds.20928 (2006).

196 Lo Coco, D., Piccoli, F. & La Bella, V. Restless legs syndrome in patients with amyotrophic lateral sclerosis. *Movement disorders : official journal of the Movement Disorder Society* **25**, 2658-2661, doi:10.1002/mds.23261 (2010).

197 Ondo, W. G., Vuong, K. D. & Jankovic, J. Exploring the relationship between Parkinson disease and restless legs syndrome. *Arch Neurol* **59**, 421-424 (2002).

198 Schurks, M. & Bussfeld, P. Multiple sclerosis and restless legs syndrome: a systematic review and meta-analysis. *Eur J Neurol* **20**, 605-615, doi:10.1111/j.1468-1331.2012.03873.x (2013).

199 Hening, W. A. & Caivano, C. K. Restless legs syndrome: a common disorder in patients with rheumatologic conditions. *Semin Arthritis Rheum* **38**, 55-62, doi:10.1016/j.semarthrit.2007.09.001 (2008).

200 Moccia, M. *et al.* Restless legs syndrome is a common feature of adult celiac disease. *Movement disorders : official journal of the Movement Disorder Society* **25**, 877-881, doi:10.1002/mds.22903 (2010).

201 Weinstock, L. B. *et al.* Crohn's disease is associated with restless legs syndrome. *Inflamm Bowel Dis* **16**, 275-279, doi:10.1002/ibd.20992 (2010).

202 Rottach, K. G. *et al.* Restless legs syndrome as side effect of second generation antidepressants. *J Psychiatr Res* **43**, 70-75, doi:10.1016/j.jpsychires.2008.02.006 (2008).

203 Horiguchi, J. *et al.* Nocturnal eating/drinking syndrome and neuroleptic-induced restless legs syndrome. *Int Clin Psychopharmacol* **14**, 33-36 (1999).

204 O'Sullivan, R. L. & Greenberg, D. B. H2 antagonists, restless leg syndrome, and movement disorders. *Psychosomatics* **34**, 530-532, doi:10.1016/S0033-3182(93)71830-7 (1993).

205 Pittock, S. J. *et al.* Neuropathology of primary restless leg syndrome: absence of specific tau- and alpha-synuclein pathology. *Movement disorders : official journal of the Movement Disorder Society* **19**, 695-699, doi:10.1002/mds.20042 (2004).

206 Connor, J. R. *et al.* Neuropathological examination suggests impaired brain iron acquisition in restless legs syndrome. *Neurology* **61**, 304-309 (2003).

207 Earley, C. J., Allen, R. P., Connor, J. R., Ferrucci, L. & Troncoso, J. The dopaminergic neurons of the A11 system in RLS autopsy brains appear normal. *Sleep Med* **10**, 1155-1157, doi:10.1016/j.sleep.2009.01.006 (2009).

208 Tergau, F., Wischer, S. & Paulus, W. Motor system excitability in patients with restless legs syndrome. *Neurology* **52**, 1060-1063 (1999).

209 Stiasny-Kolster, K. *et al.* Cortical silent period is shortened in restless legs syndrome independently from circadian rhythm. *Suppl Clin Neurophysiol* **56**, 381-389 (2003).

210 Bara-Jimenez, W., Aksu, M., Graham, B., Sato, S. & Hallett, M. Periodic limb movements in sleep: state-dependent excitability of the spinal flexor reflex. *Neurology* **54**, 1609-1616 (2000).

211 Allen, R. P. & Earley, C. J. Restless legs syndrome: a review of clinical and pathophysiologic features. *J Clin Neurophysiol* **18**, 128-147 (2001).

212 Scalise, A., Cadore, I. P. & Gigli, G. L. Motor cortex excitability in restless legs syndrome. *Sleep Med* **5**, 393-396, doi:10.1016/j.sleep.2004.01.011 (2004).

213 Quatrale, R. *et al.* Neurophysiological study of corticomotor pathways in restless legs syndrome. *Clin Neurophysiol* **114**, 1638-1645 (2003).

214 Entezari-Taher, M. *et al.* Changes in excitability of motor cortical circuitry in primary restless legs syndrome. *Neurology* **53**, 1201-1205 (1999).

215 Rijsman, R. M., Stam, C. J. & de Weerd, A. W. Abnormal H-reflexes in periodic limb movement disorder; impact on understanding the pathophysiology of the disorder. *Clin Neurophysiol* **116**, 204-210, doi:10.1016/j.clinph.2004.07.022 (2005).

216 Silver, N., Allen, R. P., Senerth, J. & Earley, C. J. A 10-year, longitudinal assessment of dopamine agonists and methadone in the treatment of restless legs syndrome. *Sleep Med* **12**, 440-444, doi:10.1016/j.sleep.2010.11.002 (2011).

217 Winkelman, J. W. Considering the causes of RLS. *Eur J Neurol* **13 Suppl 3**, 8-14, doi:10.1111/j.1468-1331.2006.01586.x (2006).

218 Carlsson, A., Svennerholm, L. & Winblad, B. Seasonal and circadian monoamine variations in human brains examined post mortem. *Acta Psychiatr Scand Suppl* **280**, 75-85 (1980).

219 Sowers, J. R. & Vlachakis, N. Circadian variation in plasma dopamine levels in man. *J Endocrinol Invest* **7**, 341-345 (1984).

220 Clemens, S., Rye, D. & Hochman, S. Restless legs syndrome: revisiting the dopamine hypothesis from the spinal cord perspective. *Neurology* **67**, 125-130, doi:10.1212/01.wnl.0000223316.53428.c9 (2006).

221 Lindvall, O., Bjorklund, A. & Skagerberg, G. Dopamine-containing neurons in the spinal cord: anatomy and some functional aspects. *Ann Neurol* **14**, 255-260, doi:10.1002/ana.410140302 (1983).

222 Trenkwalder, C. & Paulus, W. Why do restless legs occur at rest?--pathophysiology of neuronal structures in RLS. Neurophysiology of RLS (part 2). *Clin Neurophysiol* **115**, 1975-1988, doi:10.1016/j.clinph.2004.01.031 (2004).

223 Zhao, H. *et al.* Spinal cord dopamine receptor expression and function in mice with 6-OHDA lesion of the A11 nucleus and dietary iron deprivation. *J Neurosci Res* **85**, 1065-1076, doi:10.1002/jnr.21207 (2007).

224 Qu, S. *et al.* Locomotion is increased in a11-lesioned mice with iron deprivation: a possible animal model for restless legs syndrome. *J Neuropathol Exp Neurol* **66**, 383-388, doi:10.1097/nen.0b013e3180517b5f (2007).

225 Ondo, W. G., He, Y., Rajasekaran, S. & Le, W. D. Clinical correlates of 6-hydroxydopamine injections into A11 dopaminergic neurons in rats: a possible model for restless legs syndrome. *Movement disorders : official journal of the Movement Disorder Society* **15**, 154-158 (2000).

226 Michaud, M., Soucy, J. P., Chabli, A., Lavigne, G. & Montplaisir, J. SPECT imaging of striatal pre- and postsynaptic dopaminergic status in restless legs syndrome with periodic leg movements in sleep. *J Neurol* **249**, 164-170 (2002).

227 Eisensehr, I. *et al.* Normal IPT and IBZM SPECT in drug-naive and levodopa-treated idiopathic restless legs syndrome. *Neurology* **57**, 1307-1309 (2001).

228 Staedt, J. *et al.* Nocturnal myoclonus syndrome (periodic movements in sleep) related to central dopamine D2-receptor alteration. *Eur Arch Psychiatry Clin Neurosci* **245**, 8-10 (1995).

229 Turjanski, N., Lees, A. J. & Brooks, D. J. Striatal dopaminergic function in restless legs syndrome: 18F-dopa and 11C-raclopride PET studies. *Neurology* **52**, 932-937 (1999).

230 Tribl, G. G. *et al.* Normal striatal D2 receptor binding in idiopathic restless legs syndrome with periodic leg movements in sleep. *Nucl Med Commun* **25**, 55-60 (2004).

231 Ruottinen, H. M. *et al.* An FDOPA PET study in patients with periodic limb movement disorder and restless legs syndrome. *Neurology* **54**, 502-504 (2000).

232 Schulte, E. C. K., M.; Schormair, B.; Winkelmann, J. Iron in Restless Legs Syndrome. *International Review of Neurobiology* **in press** (2013).

233 Earley, C. J., P, B. B., Horska, A. & Allen, R. P. MRI-determined regional brain iron concentrations in early- and late-onset restless legs syndrome. *Sleep Med* **7**, 458-461, doi:10.1016/j.sleep.2005.11.009 (2006).

234 Allen, R. P., Barker, P. B., Wehrl, F., Song, H. K. & Earley, C. J. MRI measurement of brain iron in patients with restless legs syndrome. *Neurology* **56**, 263-265 (2001).

235 Haba-Rubio, J. *et al.* Restless legs syndrome and low brain iron levels in patients with haemochromatosis. *J Neurol Neurosurg Psychiatry* **76**, 1009-1010, doi:10.1136/jnnp.2003.030536 (2005).

236 Astrakas, L. G. *et al.* T2 relaxometry and fMRI of the brain in late-onset restless legs syndrome. *Neurology* **71**, 911-916, doi:10.1212/01.wnl.0000325914.50764.a2 (2008).

237 Godau, J., Schweitzer, K. J., Liepelt, I., Gerloff, C. & Berg, D. Substantia nigra hypoechogenicity: definition and findings in restless legs syndrome. *Movement disorders : official journal of the Movement Disorder Society* **22**, 187-192, doi:10.1002/mds.21230 (2007).

238 Godau, J., Klose, U., Di Santo, A., Schweitzer, K. & Berg, D. Multiregional brain iron deficiency in restless legs syndrome. *Movement disorders : official journal of the Movement Disorder Society* **23**, 1184-1187, doi:10.1002/mds.22070 (2008).

239 Margariti, P. N. *et al.* Investigation of unmedicated early onset restless legs syndrome by voxel-based morphometry, T2 relaxometry, and functional MR imaging during the night-time hours. *AJNR Am J Neuroradiol* **33**, 667-672, doi:10.3174/ajnr.A2829 (2012).

240 Schmidauer, C. *et al.* Transcranial ultrasound shows nigral hypoechogenicity in restless legs syndrome. *Ann Neurol* **58**, 630-634, doi:10.1002/ana.20572 (2005).

241 Mizuno, S., Mihara, T., Miyaoka, T., Inagaki, T. & Horiguchi, J. CSF iron, ferritin and transferrin levels in restless legs syndrome. *J Sleep Res* **14**, 43-47, doi:10.1111/j.1365-2869.2004.00403.x (2005).

242 Earley, C. J. *et al.* Abnormalities in CSF concentrations of ferritin and transferrin in restless legs syndrome. *Neurology* **54**, 1698-1700 (2000).

243 Clardy, S. L. *et al.* Is ferroportin-hepcidin signaling altered in restless legs syndrome? *J Neurol Sci* **247**, 173-179, doi:10.1016/j.jns.2006.04.008 (2006).

244 Patton, S. M., Ponnuru, P., Snyder, A. M., Podskalny, G. D. & Connor, J. R. Hypoxia-inducible factor pathway activation in restless legs syndrome patients. *Eur J Neurol* **18**, 1329-1335, doi:10.1111/j.1468-1331.2011.03397.x (2011).

245 Connor, J. R. *et al.* Decreased transferrin receptor expression by neuromelanin cells in restless legs syndrome. *Neurology* **62**, 1563-1567 (2004).

246 Connor, J. R. *et al.* Profile of altered brain iron acquisition in restless legs syndrome. *Brain* **134**, 959-968, doi:10.1093/brain/awr012 (2011).

247 Snyder, A. M. *et al.* Mitochondrial ferritin in the substantia nigra in restless legs syndrome. *J Neuropathol Exp Neurol* **68**, 1193-1199, doi:10.1097/NEN.0b013e3181bdc44f (2009).

248 Allen, R. P. & Earley, C. J. The role of iron in restless legs syndrome. *Movement disorders : official journal of the Movement Disorder Society* **22 Suppl 18**, S440-448, doi:10.1002/mds.21607 (2007).

249 Winkelmann, J. *et al.* Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. *Nature genetics* **39**, 1000-1006, doi:10.1038/ng2099 (2007).

250 Winkelmann, J. *et al.* Genome-wide association study identifies novel restless legs syndrome susceptibility loci on 2p14 and 16q12.1. *PLoS genetics* **7**, e1002171, doi:10.1371/journal.pgen.1002171 (2011).

251 Schormair, B. *et al.* PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome. *Nature genetics* **40**, 946-948, doi:10.1038/ng.190 (2008).

252 Stefansson, H. *et al.* A genetic risk factor for periodic limb movements in sleep. *N Engl J Med* **357**, 639-647, doi:10.1056/NEJMoa072743 (2007).

253 Ekbom, K. A. Restless legs. *Acta Med Scand* **158**, 1-123 (1945).

254 Oppenheim, H. *Lehrbuch der Nervenheilkunde.* (S. Karger, 1923).

255 Xiong, L. *et al.* Canadian restless legs syndrome twin study. *Neurology* **68**, 1631-1633, doi:10.1212/01.wnl.0000261016.90374.fd (2007).

256 Desai, A. V., Cherkas, L. F., Spector, T. D. & Williams, A. J. Genetic influences in self-reported symptoms of obstructive sleep apnoea and restless legs: a twin study. *Twin Res* **7**, 589-595, doi:10.1375/1369052042663841 (2004).

257 Ondo, W. G., Vuong, K. D. & Wang, Q. Restless legs syndrome in monozygotic twins: clinical correlates. *Neurology* **55**, 1404-1406 (2000).

258 Schormair, B. *et al.* MEIS1 and BTBD9: genetic association with restless leg syndrome in end stage renal disease. *J Med Genet* **48**, 462-466, doi:10.1136/jmg.2010.087858 (2011).

259 Walters, A. S., Picchietti, D., Hening, W. & Lazzarini, A. Variable expressivity in familial restless legs syndrome. *Arch Neurol* **47**, 1219-1220 (1990).

260 Trenkwalder, C., Seidel, V. C., Gasser, T. & Oertel, W. H. Clinical symptoms and possible anticipation in a large kindred of familial restless legs syndrome. *Movement disorders : official journal of the Movement Disorder Society* **11**, 389-394, doi:10.1002/mds.870110407 (1996).

261 Levchenko, A. *et al.* A novel autosomal dominant restless legs syndrome locus maps to chromosome 20p13. *Neurology* **67**, 900-901, doi:10.1212/01.wnl.0000233991.20410.b6 (2006).

262 Levchenko, A. *et al.* Autosomal-dominant locus for Restless Legs Syndrome in French-Canadians on chromosome 16p12.1. *Movement disorders : official journal of the Movement Disorder Society* **24**, 40-50, doi:10.1002/mds.22263 (2009).

263 Bonati, M. T. *et al.* Autosomal dominant restless legs syndrome maps on chromosome 14q. *Brain* **126**, 1485-1492 (2003).

264 Desautels, A. *et al.* Identification of a major susceptibility locus for restless legs syndrome on chromosome 12q. *American journal of human genetics* **69**, 1266-1270, doi:10.1086/324649 (2001).

265 Pichler, I. *et al.* Linkage analysis identifies a novel locus for restless legs syndrome on chromosome 2q in a South Tyrolean population isolate. *American journal of human genetics* **79**, 716-723, doi:10.1086/507875 (2006).

266 Lohmann-Hedrich, K. *et al.* Evidence for linkage of restless legs syndrome to chromosome 9p: are there two distinct loci? *Neurology* **70**, 686-694, doi:10.1212/01.wnl.0000282760.07650.ba (2008).

267 Skehan, E. B., Abdulrahim, M. M., Parfrey, N. A. & Hand, C. K. A novel locus for restless legs syndrome maps to chromosome 19p in an Irish pedigree. *Neurogenetics* **13**, 125-132, doi:10.1007/s10048-012-0317-x (2012).

268 Desautels, A. *et al.* Restless legs syndrome: confirmation of linkage to chromosome 12q, genetic heterogeneity, and evidence of complexity. *Arch Neurol* **62**, 591-596, doi:10.1001/archneur.62.4.591 (2005).

269 Kemlink, D. *et al.* Family-based association study of the restless legs syndrome loci 2 and 3 in a European population. *Movement disorders : official journal of the Movement Disorder Society* **22**, 207-212, doi:10.1002/mds.21254 (2007).

270 Schulte, E. C. W., J. in *Genetics of Sleep and Sleep Disorders* (ed M.; Shaw Thorpy, P.; Tafti, M.) (Cambridge University Press, 2013, in press).

271 Kock, N. *et al.* Mode of inheritance and susceptibility locus for restless legs syndrome, on chromosome 12q. *American journal of human genetics* **71**, 205-208; author reply 208, doi:10.1086/341097 (2002).

272 Liebetanz, K. M. *et al.* RLS3: fine-mapping of an autosomal dominant locus in a family with intrafamilial heterogeneity. *Neurology* **67**, 320-321, doi:10.1212/01.wnl.0000224886.65213.b5 (2006).

273 Levchenko, A. *et al.* The 14q restless legs syndrome locus in the French Canadian population. *Ann Neurol* **55**, 887-891, doi:10.1002/ana.20140 (2004).

274 Winkelmann, J. *et al.* Evidence for further genetic locus heterogeneity and confirmation of RLS-1 in restless legs syndrome. *Movement disorders : official journal of the Movement Disorder Society* **21**, 28-33, doi:10.1002/mds.20627 (2006).

275 Pichler, I. *et al.* Fine-mapping of restless legs locus 4 (RLS4) identifies a haplotype over the SPATS2L and KCTD18 genes. *J Mol Neurosci* **49**, 600-605, doi:10.1007/s12031-012-9891-5 (2013).

276 Kemlink, D. *et al.* Replication of restless legs syndrome loci in three European populations. *J Med Genet* **46**, 315-318, doi:10.1136/jmg.2008.062992 (2009).

277 Vilarino-Guell, C. *et al.* Susceptibility genes for restless legs syndrome are not associated with Parkinson disease. *Neurology* **71**, 222-223, doi:10.1212/01.wnl.0000317101.67684.e3 (2008).

278 Yang, Q. *et al.* Family-based and population-based association studies validate PTPRD as a risk factor for restless legs syndrome. *Movement disorders : official journal of the Movement Disorder Society* **26**, 516-519, doi:10.1002/mds.23459 (2011).

279 Yang, Q. *et al.* Association studies of variants in MEIS1, BTBD9, and MAP2K5/SKOR1 with restless legs syndrome in a US population. *Sleep Med* **12**, 800-804, doi:10.1016/j.sleep.2011.06.006 (2011).

280 Schulte, E. C. W., J. in *Movement Disorders: Genetics and Models* (ed M.S. LeDoux) (Elsevier, 2013, in press).

281 Azcoitia, V., Aracil, M., Martinez, A. C. & Torres, M. The homeodomain protein Meis1 is essential for definitive hematopoiesis and vascular patterning in the mouse embryo. *Dev Biol* **280**, 307-320, doi:10.1016/j.ydbio.2005.01.004 (2005).

282 Wang, Z. *et al.* GSK-3 promotes conditional association of CREB and its coactivators with MEIS1 to facilitate HOX-mediated transcription and oncogenesis. *Cancer Cell* **17**, 597-608, doi:10.1016/j.ccr.2010.04.024 (2010).

283 Maeda, R. *et al.* Xmeis1, a protooncogene involved in specifying neural crest cell fate in Xenopus embryos. *Oncogene* **20**, 1329-1342, doi:10.1038/sj.onc.1204250 (2001).

284 Mercader, N. *et al.* Conserved regulation of proximodistal limb axis development by Meis1/Hth. *Nature* **402**, 425-429, doi:10.1038/46580 (1999).

285 Penkov, D. *et al.* Analysis of the DNA-Binding Profile and Function of TALE Homeoproteins Reveals Their Specialization and Specific Interactions with Hox Genes/Proteins. *Cell Rep* **3**, 1321-1333, doi:10.1016/j.celrep.2013.03.029 (2013).

286 Dasen, J. S., Tice, B. C., Brenner-Morton, S. & Jessell, T. M. A Hox regulatory network establishes motor neuron pool identity and target-muscle connectivity. *Cell* **123**, 477-491, doi:10.1016/j.cell.2005.09.009 (2005).

287 Science, A. I. f. B. *Allen Brain Atlas*, <http://www.brain-map.org> (2004).

288 Moskow, J. J., Bullrich, F., Huebner, K., Daar, I. O. & Buchberg, A. M. Meis1, a PBX1-related homeobox gene involved in myeloid leukemia in BXH-2 mice. *Mol Cell Biol* **15**, 5434-5443 (1995).

289 Nakamura, T., Jenkins, N. A. & Copeland, N. G. Identification of a new family of Pbx-related homeobox genes. *Oncogene* **13**, 2235-2242 (1996).

290 Stankunas, K. *et al.* Pbx/Meis deficiencies demonstrate multigenetic origins of congenital heart disease. *Circ Res* **103**, 702-709, doi:10.1161/CIRCRESAHA.108.175489 (2008).

291 Smith, J. G. *et al.* Genome-wide association studies of the PR interval in African Americans. *PLoS genetics* **7**, e1001304, doi:10.1371/journal.pgen.1001304 (2011).

292 Pfeufer, A. *et al.* Genome-wide association study of PR interval. *Nature genetics* **42**, 153-159, doi:10.1038/ng.517 (2010).

293 Hisa, T. *et al.* Hematopoietic, angiogenic and eye defects in Meis1 mutant animals. *EMBO J* **23**, 450-459, doi:10.1038/sj.emboj.7600038 (2004).

294 Vilarino-Guell, C. *et al.* MEIS1 p.R272H in familial restless legs syndrome. *Neurology* **73**, 243-245, doi:10.1212/WNL.0b013e3181ae7c79 (2009).

295 Schulte, E. C. *et al.* Variant screening of the coding regions of MEIS1 in patients with restless legs syndrome. *Neurology* **76**, 1106-1108, doi:10.1212/WNL.0b013e318211c366 (2011).

296 Schulte, E. C. K., M.; Schormair, B.; Knauf, F.; Lichtner, P.; Trenkwalder, C.; Högl, B.; Frauscher, B.; Berger, K.; Fietze, I.; Gross, N.; Stiasny-Kolster, K.; Oertel, W.; Bachmann, C.G.; Paulus, W.; Zimprich, A., Peters, A.; Gieberg, C., Meitinger, T.; Müller-Myhsok, B.; Katsanis, N.; Winkelmann, J. An excess of rare loss-of-function alleles substantiates MEIS1 as the genetic factor underlying the chromosome 2 locus identified in genome-wide association studies of restless legs syndrome. *in preparation* (2013).

297 Xiong, L. *et al.* MEIS1 intronic risk haplotype associated with restless legs syndrome affects its mRNA and protein expression levels. *Hum Mol Genet* **18**, 1065-1074, doi:10.1093/hmg/ddn443 (2009).

298 Catoire, H. *et al.* Restless legs syndrome-associated MEIS1 risk variant influences iron homeostasis. *Ann Neurol* **70**, 170-175, doi:10.1002/ana.22435 (2011).

299 Godt, D., Couderc, J. L., Cramton, S. E. & Laski, F. A. Pattern formation in the limbs of Drosophila: bric a brac is expressed in both a gradient and a wave-like pattern and is required for specification and proper segmentation of the tarsus. *Development* **119**, 799-812 (1993).

300 Stogios, P. J., Downs, G. S., Jauhal, J. J., Nandra, S. K. & Prive, G. G. Sequence and structural analysis of BTB domain proteins. *Genome Biol* **6**, R82, doi:10.1186/gb-2005-6-10-r82 (2005).

301 Gross, M. K., Dottori, M. & Goulding, M. Lbx1 specifies somatosensory association interneurons in the dorsal spinal cord. *Neuron* **34**, 535-549 (2002).

302 Dinev, D. *et al.* Extracellular signal regulated kinase 5 (ERK5) is required for the differentiation of muscle cells. *EMBO Rep* **2**, 829-834, doi:10.1093/embo-reports/kve177 (2001).

303 Yuan, S. H., Qiu, Z. & Ghosh, A. TOX3 regulates calcium-dependent transcription in neurons. *Proc Natl Acad Sci U S A* **106**, 2909-2914, doi:10.1073/pnas.0805555106 (2009).

304 Uetani, N., Chagnon, M. J., Kennedy, T. E., Iwakura, Y. & Tremblay, M. L. Mammalian motoneuron axon targeting requires receptor protein tyrosine phosphatases sigma and delta. *J Neurosci* **26**, 5872-5880, doi:10.1523/JNEUROSCI.0386-06.2006 (2006).

305 Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168-176, doi:10.1038/nature05453 (2007).

306 Dittmer, S. *et al.* TOX3 is a neuronal survival factor that induces transcription depending on the presence of CITED1 or phosphorylated CREB in the transcriptionally active complex. *J Cell Sci* **124**, 252-260, doi:10.1242/jcs.068759 (2011).

307 Pulido, R., Serra-Pages, C., Tang, M. & Streuli, M. The LAR/PTP delta/PTP sigma subfamily of transmembrane protein-tyrosine-phosphatases: multiple human LAR, PTP delta, and PTP sigma isoforms are expressed in a tissue-specific manner and associate with the LAR-interacting protein LIP.1. *Proc Natl Acad Sci U S A* **92**, 11686-11690 (1995).

308 Sato, M. *et al.* Identification of chromosome arm 9p as the most frequent target of homozygous deletions in lung cancer. *Genes Chromosomes Cancer* **44**, 405-414, doi:10.1002/gcc.20253 (2005).

309 Riviere, J. B. *et al.* Association of intronic variants of the BTBD9 gene with Tourette syndrome. *Arch Neurol* **66**, 1267-1272, doi:10.1001/archneurol.2009.213 (2009).

310 Molenaar, J. J. *et al.* Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. *Nature* **483**, 589-593, doi:10.1038/nature10910 (2012).

311 Stallings, R. L. *et al.* High-resolution analysis of chromosomal breakpoints and genomic instability identifies PTPRD as a candidate tumor suppressor gene in neuroblastoma. *Cancer Res* **66**, 3673-3680, doi:10.1158/0008-5472.CAN-05-4154 (2006).

312 Elia, J. *et al.* Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. *Mol Psychiatry* **15**, 637-646, doi:10.1038/mp.2009.57 (2010).

313 Tsai, F. J. *et al.* A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. *PLoS genetics* **6**, e1000847, doi:10.1371/journal.pgen.1000847 (2010).

314 Saade, S. *et al.* Large scale association analysis identifies three susceptibility loci for coronary artery disease. *PLoS One* **6**, e29427, doi:10.1371/journal.pone.0029427 (2011).

315 Easton, D. F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087-1093, doi:10.1038/nature05887 (2007).

316 Freeman, A. *et al.* Sleep fragmentation and motor restlessness in a Drosophila model of Restless Legs Syndrome. *Curr Biol* **22**, 1142-1148, doi:10.1016/j.cub.2012.04.027 (2012).

317 DeAndrade, M. P. *et al.* Enhanced hippocampal long-term potentiation and fear memory in Btbd9 mutant mice. *PLoS One* **7**, e35518, doi:10.1371/journal.pone.0035518 (2012).

318 DeAndrade, M. P. *et al.* Motor restlessness, sleep disturbances, thermal sensory alterations and elevated serum iron levels in Btbd9 mutant mice. *Hum Mol Genet* **21**, 3984-3992, doi:10.1093/hmg/dds221 (2012).

319 de Lau, L. M. & Breteler, M. M. Epidemiology of Parkinson's disease. *Lancet neurology* **5**, 525-535, doi:10.1016/S1474-4422(06)70471-9 (2006).

320 Bower, J. H., Maraganore, D. M., McDonnell, S. K. & Rocca, W. A. Incidence and distribution of parkinsonism in Olmsted County, Minnesota, 1976-1990. *Neurology* **52**, 1214-1220 (1999).

321 de Rijk, M. C. *et al.* Prevalence of Parkinson's disease in the elderly: the Rotterdam Study. *Neurology* **45**, 2143-2146 (1995).

322 Nussbaum, R. L. & Ellis, C. E. Alzheimer's disease and Parkinson's disease. *N Engl J Med* **348**, 1356-1364, doi:10.1056/NEJM2003ra020003 (2003).

323 Elbaz, A. *et al.* Risk tables for parkinsonism and Parkinson's disease. *J Clin Epidemiol* **55**, 25-31 (2002).

324 Twelves, D., Perkins, K. S. & Counsell, C. Systematic review of incidence studies of Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society* **18**, 19-31, doi:10.1002/mds.10305 (2003).

325 Foltynie, T., Brayne, C. E., Robbins, T. W. & Barker, R. A. The cognitive ability of an incident cohort of Parkinson's patients in the UK. The CamPaIGN study. *Brain* **127**, 550-560, doi:10.1093/brain/awh067 (2004).

326 Aarsland, D., Andersen, K., Larsen, J. P., Lolk, A. & Kragh-Sorensen, P. Prevalence and characteristics of dementia in Parkinson disease: an 8-year prospective study. *Arch Neurol* **60**, 387-392 (2003).

327 Riedel, O. *et al.* Frequency of dementia, depression, and other neuropsychiatric symptoms in 1,449 outpatients with Parkinson's disease. *J Neurol* **257**, 1073-1082, doi:10.1007/s00415-010-5465-z (2010).

328 Ravina, B. *et al.* The impact of depressive symptoms in early Parkinson disease. *Neurology* **69**, 342-347, doi:10.1212/01.wnl.0000268695.63392.10 (2007).

329 Poewe, W. Treatments for Parkinson disease--past achievements and current clinical needs. *Neurology* **72**, S65-73, doi:10.1212/WNL.0b013e31819908ce (2009).

330 Healy, D. G. *et al.* Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. *Lancet neurology* **7**, 583-590, doi:10.1016/S1474-4422(08)70117-0 (2008).

331 Abbott, R. D. *et al.* Frequency of bowel movements and the future risk of Parkinson's disease. *Neurology* **57**, 456-462 (2001).

332 Tandberg, E., Larsen, J. P. & Karlsen, K. A community-based study of sleep disorders in patients with Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society* **13**, 895-899, doi:10.1002/mds.870130606 (1998).

333 Schulte, E. C. & Winkelmann, J. When Parkinson's disease patients go to sleep: specific sleep disturbances related to Parkinson's disease. *J Neurol* **258**, S328-335, doi:10.1007/s00415-011-5933-0 (2011).

334 Wasner, G. & Deuschl, G. Chapter 50 Pain in Parkinson's disease. *Handb Clin Neurol* **81**, 747-760, doi:10.1016/S0072-9752(06)80054-0 (2006).

335 Negre-Pages, L. *et al.* Chronic pain in Parkinson's disease: the cross-sectional French DoPaMiP survey. *Movement disorders : official journal of the Movement Disorder Society* **23**, 1361-1369, doi:10.1002/mds.22142 (2008).

336 Global Parkinson's Disease Survey Steering, C. Factors impacting on quality of life in Parkinson's disease: results from an international survey. *Movement disorders : official journal of the Movement Disorder Society* **17**, 60-67 (2002).

337 O'Sullivan, S. S. *et al.* Nonmotor symptoms as presenting complaints in Parkinson's disease: a clinicopathological study. *Movement disorders : official journal of the Movement Disorder Society* **23**, 101-106, doi:10.1002/mds.21813 (2008).

338 Hughes, A. J., Daniel, S. E., Kilford, L. & Lees, A. J. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry* **55**, 181-184 (1992).

339 Gibb, W. R. & Lees, A. J. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. *J Neurol Neurosurg Psychiatry* **51**, 745-752 (1988).

340 Braak, H. *et al.* Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiology of aging* **24**, 197-211 (2003).

341 Hughes, A. J., Daniel, S. E., Blankson, S. & Lees, A. J. A clinicopathologic study of 100 cases of Parkinson's disease. *Arch Neurol* **50**, 140-148 (1993).

342 Goedert, M., Spillantini, M. G., Del Tredici, K. & Braak, H. 100 years of Lewy pathology. *Nat Rev Neurol* **9**, 13-24, doi:10.1038/nrneurol.2012.242 (2013).

343 Hayashi, S. *et al.* An autopsy case of autosomal-recessive juvenile parkinsonism with a homozygous exon 4 deletion in the parkin gene. *Movement disorders : official journal of the Movement Disorder Society* **15**, 884-888 (2000).

344 Cookson, M. R., Hardy, J. & Lewis, P. A. Genetic neuropathology of Parkinson's disease. *Int J Clin Exp Pathol* **1**, 217-231 (2008).

345 Corti, O., Lesage, S. & Brice, A. What genetics tells us about the causes and mechanisms of Parkinson's disease. *Physiol Rev* **91**, 1161-1218, doi:10.1152/physrev.00022.2010 (2011).

346 Zimprich, A. *et al.* Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron* **44**, 601-607, doi:10.1016/j.neuron.2004.11.005 (2004).

347 Wszolek, Z. K. *et al.* Autosomal dominant parkinsonism associated with variable synuclein and tau pathology. *Neurology* **62**, 1619-1622 (2004).

348 Spillantini, M. G. *et al.* Alpha-synuclein in Lewy bodies. *Nature* **388**, 839-840, doi:10.1038/42166 (1997).

349 Spillantini, M. G., Crowther, R. A., Jakes, R., Hasegawa, M. & Goedert, M. alpha-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with lewy bodies. *Proc Natl Acad Sci U S A* **95**, 6469-6473 (1998).

350 Oertel, W. H., Deuschl, G., Poewe, W. *Parkinson-Syndrome und andere Bewegungsstörungen*. (Georg Thieme Verlag, 2012).

351 Helmich, R. C., Janssen, M. J., Oyen, W. J., Bloem, B. R. & Toni, I. Pallidal dysfunction drives a cerebellothalamic circuit into Parkinson tremor. *Ann Neurol* **69**, 269-281, doi:10.1002/ana.22361 (2011).

352 Raethjen, J. *et al.* Cortical correlates of the basic and first harmonic frequency of Parkinsonian tremor. *Clin Neurophysiol* **120**, 1866-1872, doi:10.1016/j.clinph.2009.06.028 (2009).

353 Ottolini, D., Cali, T., Negro, A. & Brini, M. The Parkinson disease-related protein DJ-1 counteracts mitochondrial impairment induced by the tumour suppressor protein p53

by enhancing endoplasmic reticulum-mitochondria tethering. *Hum Mol Genet*, doi:10.1093/hmg/ddt068 (2013).

354 Okatsu, K. *et al.* PINK1 autophosphorylation upon membrane potential dissipation is essential for Parkin recruitment to damaged mitochondria. *Nat Commun* **3**, 1016, doi:10.1038/ncomms2016 (2012).

355 Cooper, O. *et al.* Pharmacological rescue of mitochondrial deficits in iPSC-derived neural cells from patients with familial Parkinson's disease. *Sci Transl Med* **4**, 141ra190, doi:10.1126/scitranslmed.3003985 (2012).

356 Burman, J. L., Yu, S., Poole, A. C., Decal, R. B. & Pallanck, L. Analysis of neural subtypes reveals selective mitochondrial dysfunction in dopaminergic neurons from parkin mutants. *Proc Natl Acad Sci U S A* **109**, 10438-10443, doi:10.1073/pnas.1120688109 (2012).

357 Usenovic, M., Tresse, E., Mazzulli, J. R., Taylor, J. P. & Krainc, D. Deficiency of ATP13A2 leads to lysosomal dysfunction, alpha-synuclein accumulation, and neurotoxicity. *J Neurosci* **32**, 4240-4246, doi:10.1523/JNEUROSCI.5575-11.2012 (2012).

358 Sanchez-Danes, A. *et al.* Disease-specific phenotypes in dopamine neurons from human iPS-based models of genetic and sporadic Parkinson's disease. *EMBO Mol Med* **4**, 380-395, doi:10.1002/emmm.201200215 (2012).

359 Neudorfer, O. *et al.* Occurrence of Parkinson's syndrome in type I Gaucher disease. *QJM* **89**, 691-694 (1996).

360 Muller-Rischart, A. K. *et al.* The E3 ligase parkin maintains mitochondrial integrity by increasing linear ubiquitination of NEMO. *Mol Cell* **49**, 908-921, doi:10.1016/j.molcel.2013.01.036 (2013).

361 Rakovic, A. *et al.* Phosphatase and tensin homolog (PTEN)-induced putative kinase 1 (PINK1)-dependent ubiquitination of endogenous Parkin attenuates mitophagy: study in human primary fibroblasts and induced pluripotent stem cell-derived neurons. *J Biol Chem* **288**, 2223-2237, doi:10.1074/jbc.M112.391680 (2013).

362 Johnson, B. N., Berger, A. K., Cortese, G. P. & Lavoie, M. J. The ubiquitin E3 ligase parkin regulates the proapoptotic function of Bax. *Proc Natl Acad Sci U S A* **109**, 6283-6288, doi:10.1073/pnas.1113248109 (2012).

363 Cullen, V. *et al.* Acid beta-glucosidase mutants linked to Gaucher disease, Parkinson disease, and Lewy body dementia alter alpha-synuclein processing. *Ann Neurol* **69**, 940-953, doi:10.1002/ana.22400 (2011).

364 Yap, T. L. *et al.* Alpha-synuclein interacts with Glucocerebrosidase providing a molecular link between Parkinson and Gaucher diseases. *J Biol Chem* **286**, 28080-28088, doi:10.1074/jbc.M111.237859 (2011).

365 Lastres-Becker, I. *et al.* alpha-Synuclein expression and Nrf2 deficiency cooperate to aggravate protein aggregation, neuronal death and inflammation in early-stage Parkinson's disease. *Hum Mol Genet* **21**, 3173-3192, doi:10.1093/hmg/dds143 (2012).

366 L'Episcopo, F. *et al.* Plasticity of subventricular zone neuroprogenitors in MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) mouse model of Parkinson's disease involves cross talk between inflammatory and Wnt/beta-catenin signaling pathways: functional consequences for neuroprotection and repair. *J Neurosci* **32**, 2062-2085, doi:10.1523/JNEUROSCI.5259-11.2012 (2012).

367 Huh, S. H. *et al.* Ethyl pyruvate rescues nigrostriatal dopaminergic neurons by regulating glial activation in a mouse model of Parkinson's disease. *J Immunol* **187**, 960-969, doi:10.4049/jimmunol.1100009 (2011).

368 Lannuzel, A. *et al.* Atypical parkinsonism in Guadeloupe: a common risk factor for two closely related phenotypes? *Brain* **130**, 816-827, doi:10.1093/brain/awl347 (2007).

369 Langston, J. W., Ballard, P., Tetrud, J. W. & Irwin, I. Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis. *Science* **219**, 979-980 (1983).

370 Castello, P. R., Drechsel, D. A. & Patel, M. Mitochondria are a major source of paraquat-induced reactive oxygen species production in the brain. *J Biol Chem* **282**, 14186-14193, doi:10.1074/jbc.M700827200 (2007).

371 Polymeropoulos, M. H. *et al.* Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science* **276**, 2045-2047 (1997).

372 Singleton, A. B. *et al.* alpha-Synuclein locus triplication causes Parkinson's disease. *Science* **302**, 841, doi:10.1126/science.1090278 (2003).

373 Valente, E. M. *et al.* Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science* **304**, 1158-1160, doi:10.1126/science.1096284 (2004).

374 Kitada, T. *et al.* Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* **392**, 605-608, doi:10.1038/33416 (1998).

375 Sidransky, E. *et al.* Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N Engl J Med* **361**, 1651-1661, doi:10.1056/NEJMoa0901281 (2009).

376 Bonifati, V. *et al.* Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. *Science* **299**, 256-259, doi:10.1126/science.1077209 (2003).

377 Ramirez, A. *et al.* Hereditary parkinsonism with dementia is caused by mutations in ATP13A2, encoding a lysosomal type 5 P-type ATPase. *Nature genetics* **38**, 1184-1191, doi:10.1038/ng1884 (2006).

378 Zimprich, A. *et al.* A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. *American journal of human genetics* **89**, 168-175, doi:10.1016/j.ajhg.2011.06.008 (2011).

379 Vilarino-Guell, C. *et al.* VPS35 mutations in Parkinson disease. *American journal of human genetics* **89**, 162-167, doi:10.1016/j.ajhg.2011.06.001 (2011).

380 Wirdefeldt, K., Gatz, M., Reynolds, C. A., Prescott, C. A. & Pedersen, N. L. Heritability of Parkinson disease in Swedish twins: a longitudinal study. *Neurobiology of aging* **32**, 1923 e1921-1928, doi:10.1016/j.neurobiolaging.2011.02.017 (2011).

381 Houlden, H. & Singleton, A. B. The genetics and neuropathology of Parkinson's disease. *Acta Neuropathol* **124**, 325-338, doi:10.1007/s00401-012-1013-5 (2012).

382 Lucking, C. B. *et al.* Association between early-onset Parkinson's disease and mutations in the parkin gene. *N Engl J Med* **342**, 1560-1567, doi:10.1056/NEJM200005253422103 (2000).

383 Polymeropoulos, M. H. *et al.* Mapping of a gene for Parkinson's disease to chromosome 4q21-q23. *Science* **274**, 1197-1199 (1996).

384 Kruger, R. *et al.* Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. *Nature genetics* **18**, 106-108, doi:10.1038/ng0298-106 (1998).

385 Zarranz, J. J. *et al.* The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. *Ann Neurol* **55**, 164-173, doi:10.1002/ana.10795 (2004).

386 Nishioka, K. *et al.* Expanding the clinical phenotype of SNCA duplication carriers. *Movement disorders : official journal of the Movement Disorder Society* **24**, 1811-1819, doi:10.1002/mds.22682 (2009).

387 Farrer, M. *et al.* Comparison of kindreds with parkinsonism and alpha-synuclein genomic multiplications. *Ann Neurol* **55**, 174-179, doi:10.1002/ana.10846 (2004).

388 Ibanez, P. *et al.* Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease. *Lancet* **364**, 1169-1171, doi:10.1016/S0140-6736(04)17104-3 (2004).

389 Chartier-Harlin, M. C. *et al.* Alpha-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet* **364**, 1167-1169, doi:10.1016/S0140-6736(04)17103-1 (2004).

390 Singleton, A. & Gwinn-Hardy, K. Parkinson's disease and dementia with Lewy bodies: a difference in dose? *Lancet* **364**, 1105-1107, doi:10.1016/S0140-6736(04)17117-1 (2004).

391 Paisan-Ruiz, C. *et al.* Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron* **44**, 595-600, doi:10.1016/j.neuron.2004.10.023 (2004).

392 Zimprich, A. *et al.* The PARK8 locus in autosomal dominant parkinsonism: confirmation of linkage and further delineation of the disease-containing interval. *American journal of human genetics* **74**, 11-19, doi:10.1086/380647 (2004).

393 Funayama, M. *et al.* A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2-q13.1. *Ann Neurol* **51**, 296-301 (2002).

394 Nichols, W. C. *et al.* Genetic screening for a single common LRRK2 mutation in familial Parkinson's disease. *Lancet* **365**, 410-412, doi:10.1016/S0140-6736(05)17828-3 (2005).

395 Di Fonzo, A. *et al.* A frequent LRRK2 gene mutation associated with autosomal dominant Parkinson's disease. *Lancet* **365**, 412-415, doi:10.1016/S0140-6736(05)17829-5 (2005).

396 Gilks, W. P. *et al.* A common LRRK2 mutation in idiopathic Parkinson's disease. *Lancet* **365**, 415-416, doi:10.1016/S0140-6736(05)17830-1 (2005).

397 Paisan-Ruiz, C., Nath, P., Washecka, N., Gibbs, J. R. & Singleton, A. B. Comprehensive analysis of LRRK2 in publicly available Parkinson's disease cases and neurologically normal controls. *Human mutation* **29**, 485-490, doi:10.1002/humu.20668 (2008).

398 Ozelius, L. J. *et al.* LRRK2 G2019S as a cause of Parkinson's disease in Ashkenazi Jews. *N Engl J Med* **354**, 424-425, doi:10.1056/NEJMc055509 (2006).

399 Hulihan, M. M. *et al.* LRRK2 Gly2019Ser penetrance in Arab-Berber patients from Tunisia: a case-control genetic study. *Lancet neurology* **7**, 591-594, doi:10.1016/S1474-4422(08)70116-9 (2008).

400 Lesage, S. *et al.* LRRK2 G2019S as a cause of Parkinson's disease in North African Arabs. *N Engl J Med* **354**, 422-423, doi:10.1056/NEJMc055540 (2006).

401 Ishihara, L. *et al.* Clinical features of Parkinson disease patients with homozygous leucine-rich repeat kinase 2 G2019S mutations. *Arch Neurol* **63**, 1250-1254, doi:10.1001/archneur.63.9.1250 (2006).

402 Cookson, M. R. Cellular effects of LRRK2 mutations. *Biochem Soc Trans* **40**, 1070-1073, doi:10.1042/BST20120165 (2012).

403 Lu, Y. W. & Tan, E. K. Molecular biology changes associated with LRRK2 mutations in Parkinson's disease. *J Neurosci Res* **86**, 1895-1901, doi:10.1002/jnr.21656 (2008).

404 Guo, L. *et al.* The Parkinson's disease-associated protein, leucine-rich repeat kinase 2 (LRRK2), is an authentic GTPase that stimulates kinase activity. *Exp Cell Res* **313**, 3658-3670, doi:10.1016/j.yexcr.2007.07.007 (2007).

405 Lewis, P. A. *et al.* The R1441C mutation of LRRK2 disrupts GTP hydrolysis. *Biochem Biophys Res Commun* **357**, 668-671, doi:10.1016/j.bbrc.2007.04.006 (2007).

406 Li, X. *et al.* Leucine-rich repeat kinase 2 (LRRK2)/PARK8 possesses GTPase activity that is altered in familial Parkinson's disease R1441C/G mutants. *J Neurochem* **103**, 238-247, doi:10.1111/j.1471-4159.2007.04743.x (2007).

407 Smith, W. W. *et al.* Kinase activity of mutant LRRK2 mediates neuronal toxicity. *Nat Neurosci* **9**, 1231-1233, doi:10.1038/nn1776 (2006).

408 Greggio, E. *et al.* Kinase activity is required for the toxic effects of mutant LRRK2/dardarin. *Neurobiol Dis* **23**, 329-341, doi:10.1016/j.nbd.2006.04.001 (2006).

409 Chartier-Harlin, M. C. *et al.* Translation initiator EIF4G1 mutations in familial Parkinson disease. *American journal of human genetics* **89**, 398-406, doi:10.1016/j.ajhg.2011.08.009 (2011).

410 Leroy, E. *et al.* The ubiquitin pathway in Parkinson's disease. *Nature* **395**, 451-452, doi:10.1038/26652 (1998).

411 Lautier, C. *et al.* Mutations in the GIGYF2 (TNRC15) gene at the PARK11 locus in familial Parkinson disease. *American journal of human genetics* **82**, 822-833, doi:10.1016/j.ajhg.2008.01.015 (2008).

412 Pankratz, N. *et al.* Significant linkage of Parkinson disease to chromosome 2q36-37. *American journal of human genetics* **72**, 1053-1057, doi:10.1086/374383 (2003).

413 Maruyama, M. *et al.* Novel mutations, pseudo-dominant inheritance, and possible familial affects in patients with autosomal recessive juvenile parkinsonism. *Ann Neurol* **48**, 245-250 (2000).

414 Jones, J. M. *et al.* Loss of Omi mitochondrial protease activity causes the neuromuscular disorder of mnd2 mutant mice. *Nature* **425**, 721-727, doi:10.1038/nature02052 (2003).

415 Kawamoto, Y. *et al.* Accumulation of HtrA2/Omi in neuronal and glial inclusions in brains with alpha-synucleinopathies. *J Neuropathol Exp Neurol* **67**, 984-993, doi:10.1097/NEN.0b013e31818809f4 (2008).

416 Sharma, M. *et al.* A multi-centre clinico-genetic analysis of the VPS35 gene in Parkinson disease indicates reduced penetrance for disease-associated variants. *J Med Genet* **49**, 721-726, doi:10.1136/jmedgenet-2012-101155 (2012).

417 Periquet, M. *et al.* Parkin mutations are frequent in patients with isolated early-onset parkinsonism. *Brain* **126**, 1271-1278 (2003).

418 Shimura, H. *et al.* Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. *Nature genetics* **25**, 302-305, doi:10.1038/77060 (2000).

419 Deas, E., Plun-Favreau, H. & Wood, N. W. PINK1 function in health and disease. *EMBO Mol Med* **1**, 152-165, doi:10.1002/emmm.200900024 (2009).

420 Nuytemans, K., Theuns, J., Cruts, M. & Van Broeckhoven, C. Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7,

and LRRK2 genes: a mutation update. *Human mutation* **31**, 763-780, doi:10.1002/humu.21277 (2010).

421 Funayama, M. *et al.* Familial Parkinsonism with digenic parkin and PINK1 mutations. *Movement disorders : official journal of the Movement Disorder Society* **23**, 1461-1465, doi:10.1002/mds.22143 (2008).

422 Tang, B. *et al.* Association of PINK1 and DJ-1 confers digenic inheritance of early-onset Parkinson's disease. *Hum Mol Genet* **15**, 1816-1825, doi:10.1093/hmg/ddl104 (2006).

423 Canet-Aviles, R. M. *et al.* The Parkinson's disease protein DJ-1 is neuroprotective due to cysteine-sulfinic acid-driven mitochondrial localization. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9103-9108, doi:10.1073/pnas.0402959101 (2004).

424 Beilina, A. *et al.* Mutations in PTEN-induced putative kinase 1 associated with recessive parkinsonism have differential effects on protein stability. *Proc Natl Acad Sci U S A* **102**, 5703-5708, doi:10.1073/pnas.0500617102 (2005).

425 Khan, N. L. *et al.* Progression of nigrostriatal dysfunction in a parkin kindred: an [18F]dopa PET and clinical study. *Brain* **125**, 2248-2256 (2002).

426 Khan, N. L. *et al.* Dopaminergic dysfunction in unrelated, asymptomatic carriers of a single parkin mutation. *Neurology* **64**, 134-136, doi:10.1212/01.WNL.0000148725.48740.6D (2005).

427 Klein, C., Lohmann-Hedrich, K., Rogaeva, E., Schlossmacher, M. G. & Lang, A. E. Deciphering the role of heterozygous mutations in genes associated with parkinsonism. *Lancet neurology* **6**, 652-662, doi:10.1016/S1474-4422(07)70174-6 (2007).

428 Foroud, T. *et al.* Heterozygosity for a mutation in the parkin gene leads to later onset Parkinson disease. *Neurology* **60**, 796-801 (2003).

429 Pankratz, N. *et al.* Genes influencing Parkinson disease onset: replication of PARK3 and identification of novel loci. *Neurology* **62**, 1616-1618 (2004).

430 DeStefano, A. L. *et al.* PARK3 influences age at onset in Parkinson disease: a genome scan in the GenePD study. *American journal of human genetics* **70**, 1089-1095, doi:10.1086/339814 (2002).

431 Oliveira, S. A. *et al.* Identification of risk and age-at-onset genes on chromosome 1p in Parkinson disease. *American journal of human genetics* **77**, 252-264, doi:10.1086/432588 (2005).

432 Hicks, A. A. *et al.* A susceptibility gene for late-onset idiopathic Parkinson's disease. *Ann Neurol* **52**, 549-555, doi:10.1002/ana.10324 (2002).

433 Pankratz, N. *et al.* Genome-wide linkage analysis and evidence of gene-by-gene interactions in a sample of 362 multiplex Parkinson disease families. *Hum Mol Genet* **12**, 2599-2608, doi:10.1093/hmg/ddg270 (2003).

434 Paisan-Ruiz, C. *et al.* Characterization of PLA2G6 as a locus for dystonia-parkinsonism. *Ann Neurol* **65**, 19-23, doi:10.1002/ana.21415 (2009).

435 Di Fonzo, A. *et al.* FBXO7 mutations cause autosomal recessive, early-onset parkinsonian-pyramidal syndrome. *Neurology* **72**, 240-245, doi:10.1212/01.wnl.0000338144.10967.2b (2009).

436 Aharon-Peretz, J., Rosenbaum, H. & Gershoni-Baruch, R. Mutations in the glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. *N Engl J Med* **351**, 1972-1977, doi:10.1056/NEJMoa033277 (2004).

437 Ross, O. A. *et al.* Association of LRRK2 exonic variants with susceptibility to Parkinson's disease: a case-control study. *Lancet neurology* **10**, 898-908, doi:10.1016/S1474-4422(11)70175-2 (2011).

438 Simon-Sanchez, J. *et al.* Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nature genetics* **41**, 1308-1312, doi:10.1038/ng.487 (2009).

439 Satake, W. *et al.* Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nature genetics* **41**, 1303-1307, doi:10.1038/ng.485 (2009).

440 Do, C. B. *et al.* Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS genetics* **7**, e1002141, doi:10.1371/journal.pgen.1002141 (2011).

441 International Parkinson Disease Genomics, C. *et al.* Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* **377**, 641-649, doi:10.1016/S0140-6736(10)62345-8 (2011).

442 Hamza, T. H. *et al.* Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nature genetics* **42**, 781-785, doi:10.1038/ng.642 (2010).

443 International Parkinson's Disease Genomics, C. & Wellcome Trust Case Control, C. A two-stage meta-analysis identifies several new loci for Parkinson's disease. *PLoS genetics* **7**, e1002142, doi:10.1371/journal.pgen.1002142 (2011).

444 Pankratz, N. *et al.* Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2. *Ann Neurol* **71**, 370-384, doi:10.1002/ana.22687 (2012).

445 Lill, C. M. *et al.* Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. *PLoS genetics* **8**, e1002548, doi:10.1371/journal.pgen.1002548 (2012).

446 Hoglinger, G. U. *et al.* Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nature genetics* **43**, 699-705, doi:10.1038/ng.859 (2011).

447 Galimberti, D. & Scarpini, E. Genetics of frontotemporal lobar degeneration. *Front Neurol* **3**, 52, doi:10.3389/fneur.2012.00052 (2012).

448 Puschmann, A. *et al.* Human leukocyte antigen variation and Parkinson's disease. *Parkinsonism Relat Disord* **17**, 376-378, doi:10.1016/j.parkreldis.2011.03.008 (2011).

449 Peeraully, T. & Tan, E. K. Genetic variants in sporadic Parkinson's disease: East vs West. *Parkinsonism Relat Disord* **18 Suppl 1**, S63-65, doi:10.1016/S1353-8020(11)70021-9 (2012).

450 Pankratz, N. *et al.* Genomewide association study for susceptibility genes contributing to familial Parkinson disease. *Hum Genet* **124**, 593-605, doi:10.1007/s00439-008-0582-9 (2009).

451 Snoeckx, R. L. *et al.* GJB2 mutations and degree of hearing loss: a multicenter study. *American journal of human genetics* **77**, 945-957, doi:10.1086/497996 (2005).

452 Estivill, X. *et al.* Connexin-26 mutations in sporadic and inherited sensorineural deafness. *Lancet* **351**, 394-398, doi:10.1016/S0140-6736(97)11124-2 (1998).

453 Dror, A. A. & Avraham, K. B. Hearing loss: mechanisms revealed by genetics and cell biology. *Annu Rev Genet* **43**, 411-437, doi:10.1146/annurev-genet-102108-134135 (2009).

454 Paynter, N. P. *et al.* Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA : the journal of the American Medical Association* **303**, 631-637, doi:10.1001/jama.2010.119 (2010).

455 Schulte, E. C. S., I.; Czamara, D.; Ellwanger, D.C.; Eck, S.; Graf, E.; Mollenhauer, B.; Zimprich, A.; Lichtner, P.; Haubenberger, D.; Pirker, W.; Brücke, T.; Bereznai, B.; Molnar, M.J.; Mewes, H.W.; Peters, A.; Gieger, C.; Müller-Myhsok, B.; Trenkwalder, C.; Winkelmann, J. Rare Variants in PLXNA4 and Parkinson’s disease. *submitted* (2013).

456 Schulte, E. C. E., D.C.; Dihanich, S.; Manzoni, C.; Stangl, K.; Schormair, B.; Graf, E.; Eck, S.; Mollenhauer, B.; Haubenberger, D.; Pirker, W.; Zimprich, A.; Brücke, T.; Lichtner, P.; Peters, A.; Gieger, C.; Trenkwalder, C.; Mewes, H.W.; Meitinger, T.; Lewis, P.A.; Klünemann, H.H.; Winkelmann, J. Rare variants in LRRK1 and Parkinson’s disease. *submitted* (2013).

457 Schulte, E. C. *et al.* Variants in eukaryotic translation initiation factor 4G1 in sporadic Parkinson's disease. *Neurogenetics* **13**, 281-285, doi:10.1007/s10048-012-0334-9 (2012).

458 Talkowski, M. E. *et al.* Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* **149**, 525-537, doi:10.1016/j.cell.2012.03.028 (2012).

459 Cross-Disorder Group of the Psychiatric Genomics, C. *et al.* Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371-1379, doi:10.1016/S0140-6736(12)62129-1 (2013).

460 Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. A. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **324**, 387-389, doi:10.1126/science.1167728 (2009).

461 Galarneau, G. *et al.* Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. *Nature genetics* **42**, 1049-1051, doi:10.1038/ng.707 (2010).

462 Sanna, S. *et al.* Fine mapping of five loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability. *PLoS genetics* **7**, e1002198, doi:10.1371/journal.pgen.1002198 (2011).

463 Trynka, G. *et al.* Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nature genetics* **43**, 1193-1201, doi:10.1038/ng.998 (2011).

464 Committee, H. G. N. *HUGO Gene Nomenclature*, <http://www.genenames.org> (2013).

465 Shulman, J. M. & De Jager, P. L. Evidence for a common pathway linking neurodegenerative diseases. *Nature genetics* **41**, 1261-1262, doi:10.1038/ng1209-1261 (2009).

466 Lei, P. *et al.* Tau deficiency induces parkinsonism with dementia by impairing APP-mediated iron export. *Nat Med* **18**, 291-295, doi:10.1038/nm.2613 (2012).

467 Guo, J. P., Arai, T., Miklossy, J. & McGeer, P. L. Abeta and tau form soluble complexes that may promote self aggregation of both into the insoluble forms observed in Alzheimer's disease. *Proc Natl Acad Sci U S A* **103**, 1953-1958, doi:10.1073/pnas.0509386103 (2006).

468 Deramecourt, V. *et al.* Biochemical staging of synucleinopathy and amyloid deposition in dementia with Lewy bodies. *J Neuropathol Exp Neurol* **65**, 278-288, doi:10.1097/01.jnen.0000205145.54457.ea (2006).

469 Mattila, A. L. *et al.* High genetic load in an old isolated butterfly population. *Proc Natl Acad Sci U S A* **109**, E2496-2505, doi:10.1073/pnas.1205789109 (2012).

470 Arias, A. *et al.* Molecular dissection of a viral quasispecies under mutagenic treatment: positive correlation between fitness loss and mutational load. *J Gen Virol* **94**, 817-830, doi:10.1099/vir.0.049171-0 (2013).

471 Lemmers, R. J. *et al.* Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nature genetics* **44**, 1370-1374, doi:10.1038/ng.2454 (2012).

472 Kajiwara, K., Berson, E. L. & Dryja, T. P. Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. *Science* **264**, 1604-1608 (1994).

473 Katsanis, N. *et al.* Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. *Science* **293**, 2256-2259, doi:10.1126/science.1063525 (2001).

474 Lupski, J. R. Digenic inheritance and Mendelian disease. *Nature genetics* **44**, 1291-1292, doi:10.1038/ng.2479 (2012).

475 Zaghloul, N. A. & Katsanis, N. Functional modules, mutational load and human genetic disease. *Trends Genet* **26**, 168-176, doi:10.1016/j.tig.2010.01.006 (2010).

476 Zaghloul, N. A. *et al.* Functional analyses of variants reveal a significant role for dominant negative and common alleles in oligogenic Bardet-Biedl syndrome. *Proc Natl Acad Sci U S A* **107**, 10602-10607, doi:10.1073/pnas.1000219107 (2010).

477 Davis, E. E. *et al.* TTC21B contributes both causal and modifying alleles across the ciliopathy spectrum. *Nature genetics* **43**, 189-196, doi:10.1038/ng.756 (2011).

478 Leitch, C. C. *et al.* Hypomorphic mutations in syndromic encephalocele genes are associated with Bardet-Biedl syndrome. *Nature genetics* **40**, 443-448, doi:10.1038/ng.97 (2008).

479 McFarland, C. D., Korolev, K. S., Kryukov, G. V., Sunyaev, S. R. & Mirny, L. A. Impact of deleterious passenger mutations on cancer progression. *Proc Natl Acad Sci U S A* **110**, 2910-2915, doi:10.1073/pnas.1213968110 (2013).

480 Khanna, H. *et al.* A common allele in RPGRIP1L is a modifier of retinal degeneration in ciliopathies. *Nature genetics* **41**, 739-745, doi:10.1038/ng.366 (2009).

481 Li, D. *et al.* Mutations of presenilin genes in dilated cardiomyopathy and heart failure. *American journal of human genetics* **79**, 1030-1039, doi:10.1086/509900 (2006).

482 Wang, B. *et al.* Gamma-secretase gene mutations in familial acne inversa. *Science* **330**, 1065, doi:10.1126/science.1196284 (2010).

483 Genovese, G. *et al.* Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science* **329**, 841-845, doi:10.1126/science.1193032 (2010).

484 Patrinos, G. P. *et al.* Evidence for the molecular heterogeneity of sickle cell anemia chromosomes bearing the betaS/Benin haplotype. *Am J Hematol* **80**, 79-80, doi:10.1002/ajh.20400 (2005).

485 Horaitis, O., Talbot, C. C., Jr., Phommarinh, M., Phillips, K. M. & Cotton, R. G. A database of locus-specific databases. *Nature genetics* **39**, 425, doi:10.1038/ng0407-425 (2007).

486 Dooley, C. M. *et al.* Multi-allelic phenotyping - A systematic approach for the simultaneous analysis of multiple induced mutations. *Methods*, doi:10.1016/j.ymeth.2013.04.013 (2013).

487 Wang, H. *et al.* One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell* **153**, 910-918, doi:10.1016/j.cell.2013.04.025 (2013).

488 Dermitzakis, E. T. Cellular genomics for complex traits. *Nature reviews. Genetics* **13**, 215-220, doi:10.1038/nrg3115 (2012).

489 Albers, C. A. *et al.* Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. *Nature genetics* **44**, 435-439, S431-432, doi:10.1038/ng.1083 (2012).

490 Jablonka, E. & Raz, G. Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Q Rev Biol* **84**, 131-176 (2009).

491 Zimprich, A. Phenocopies in families with essential tremor and restless legs syndrome challenge Mendelian laws. Epigenetics might provide answers. *Parkinsonism Relat Disord* **18**, 711-716, doi:10.1016/j.parkreldis.2012.03.019 (2012).

492 Anderson, C. A., Soranzo, N., Zeggini, E. & Barrett, J. C. Synthetic associations are unlikely to account for many common disease genome-wide association signals. *PLoS Biol* **9**, e1000580, doi:10.1371/journal.pbio.1000580 (2011).

493 Wray, N. R., Purcell, S. M. & Visscher, P. M. Synthetic associations created by rare variants do not explain most GWAS results. *PLoS Biol* **9**, e1000579, doi:10.1371/journal.pbio.1000579 (2011).

494 Oosterveer, D. M. *et al.* Low-density lipoprotein receptor mutations generate synthetic genome-wide associations. *European journal of human genetics : EJHG* **21**, 563-566, doi:10.1038/ejhg.2012.207 (2013).

495 Cortes, A. & Brown, M. A. Promise and pitfalls of the Immunochip. *Arthritis Res Ther* **13**, 101, doi:10.1186/ar3204 (2011).

496 BGI. *The Million Human Genomes Project*, <http://www.genomics.cn/en/navigation/show_navigation?nid=5658> (2013).

497 23andMe. *23andMe*, <https://http://www.23andme.com> (2013).

498 Romeo, S. *et al.* Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nature genetics* **39**, 513-516, doi:10.1038/ng1984 (2007).

499 Cook-Deegan, R., Conley, J. M., Evans, J. P. & Vorhaus, D. The next controversy in genetic testing: clinical data as trade secrets? *European journal of human genetics : EJHG* **21**, 585-588, doi:10.1038/ejhg.2012.217 (2013).

500 Simpson, E. H. The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society, Series B* **13**, 238-241 (1951).

501 Spencer, C., Hechter, E., Vukcevic, D. & Donnelly, P. Quantifying the underestimation of relative risks from genome-wide association studies. *PLoS genetics* **7**, e1001337, doi:10.1371/journal.pgen.1001337 (2011).

502 Boon, C. J. *et al.* Basal laminar drusen caused by compound heterozygous variants in the CFH gene. *American journal of human genetics* **82**, 516-523, doi:10.1016/j.ajhg.2007.11.007 (2008).

503 Lee, S. H., van der Werf, J. H., Hayes, B. J., Goddard, M. E. & Visscher, P. M. Predicting unobserved phenotypes for complex traits from whole-genome SNP data. *PLoS genetics* **4**, e1000231, doi:10.1371/journal.pgen.1000231 (2008).

504 Peters, U. *et al.* A systematic mapping approach of 16q12.2/FTO and BMI in more than 20,000 African Americans narrows in on the underlying functional variation: results from the Population Architecture using Genomics and Epidemiology (PAGE) study. *PLoS genetics* **9**, e1003171, doi:10.1371/journal.pgen.1003171 (2013).

505 Wu, Y. *et al.* Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. *PLoS genetics* **9**, e1003379, doi:10.1371/journal.pgen.1003379 (2013).

506 International HapMap, C. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-861, doi:10.1038/nature06258 (2007).

507 International HapMap, C. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52-58, doi:10.1038/nature09298 (2010).

508 Campbell, M. C. & Tishkoff, S. A. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet* **9**, 403-433, doi:10.1146/annurev.genom.9.081307.164258 (2008).

509 Zheng, W. *et al.* Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nature genetics* **41**, 324-328, doi:10.1038/ng.318 (2009).

510 Yasuda, K. *et al.* Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nature genetics* **40**, 1092-1097, doi:10.1038/ng.207 (2008).

511 Kristiansson, K., Naukkarinen, J. & Peltonen, L. Isolated populations and complex disease gene identification. *Genome Biol* **9**, 109, doi:10.1186/gb-2008-9-8-109 (2008).

512 Sabatti, C. *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nature genetics* **41**, 35-46, doi:10.1038/ng.271 (2009).

513 genetics, d. *deCODE genetics*, <http://www.decode.com> (2013).

514 NCBI. *GTEx (Genotype-Tissue Expression) eQTL Browser*, <http://www.ncbi.nlm.nih.gov/gtex/GTEX2/gtex.cgi> (2013).

515 Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nature genetics* **45**, 580-585, doi:10.1038/ng.2653 (2013).

516 Gore, A. *et al.* Somatic coding mutations in human induced pluripotent stem cells. *Nature* **471**, 63-67, doi:10.1038/nature09805 (2011).

517 Soldner, F. *et al.* Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell* **146**, 318-331, doi:10.1016/j.cell.2011.06.019 (2011).

518 Seibler, P. *et al.* Mitochondrial Parkin recruitment is impaired in neurons derived from mutant PINK1 induced pluripotent stem cells. *J Neurosci* **31**, 5970-5976, doi:10.1523/JNEUROSCI.4441-10.2011 (2011).

519 Lappalainen, T., Montgomery, S. B., Nica, A. C. & Dermitzakis, E. T. Epistatic selection between coding and regulatory variation in human evolution and disease. *American journal of human genetics* **89**, 459-463, doi:10.1016/j.ajhg.2011.08.004 (2011).

520 Bandmann, O. & Burton, E. A. Genetic zebrafish models of neurodegenerative diseases. *Neurobiol Dis* **40**, 58-65, doi:10.1016/j.nbd.2010.05.017 (2010).

521 Kabashi, E. *et al.* Gain and loss of function of ALS-related mutations of TARDBP (TDP-43) cause motor deficits in vivo. *Hum Mol Genet* **19**, 671-683, doi:10.1093/hmg/ddp534 (2010).

522 Kabashi, E., Champagne, N., Brustein, E. & Drapeau, P. In the swim of things: recent insights to neurogenetic disorders from zebrafish. *Trends Genet* **26**, 373-381, doi:10.1016/j.tig.2010.05.004 (2010).

523 Golzio, C. *et al.* KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* **485**, 363-367, doi:10.1038/nature11091 (2012).

524 Levesque, M. P., Krauss, J., Koehler, C., Boden, C. & Harris, M. P. New tools for the identification of developmentally regulated enhancer regions in embryonic and adult zebrafish. *Zebrafish* **10**, 21-29, doi:10.1089/zeb.2012.0775 (2013).

525 Ishibashi, M., Mechaly, A. S., Becker, T. S. & Rinkwitz, S. Using zebrafish transgenesis to test human genomic sequences for specific enhancer activity. *Methods*, doi:10.1016/j.ymeth.2013.03.018 (2013).

526 Rinkwitz, S., Mourrain, P. & Becker, T. S. Zebrafish: an integrative system for neurogenomics and neurosciences. *Prog Neurobiol* **93**, 231-243, doi:10.1016/j.pneurobio.2010.11.003 (2011).

527 Becker, T. S. & Rinkwitz, S. Zebrafish as a genomics model for human neurological and polygenic disorders. *Dev Neurobiol* **72**, 415-428, doi:10.1002/dneu.20888 (2012).

528 Solouki, A. M. *et al.* A genome-wide association study identifies a susceptibility locus for refractive errors and myopia at 15q14. *Nature genetics* **42**, 897-901, doi:10.1038/ng.663 (2010).

529 Ragvin, A. *et al.* Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to HHEX, SOX4, and IRX3. *Proc Natl Acad Sci U S A* **107**, 775-780, doi:10.1073/pnas.0911591107 (2010).

530 Sung, Y. H. *et al.* Knockout mice created by TALEN-mediated gene targeting. *Nat Biotechnol* **31**, 23-24, doi:10.1038/nbt.2477 (2013).

531 Wefers, B. *et al.* Direct production of mouse disease models by embryo microinjection of TALENs and oligodeoxynucleotides. *Proc Natl Acad Sci U S A* **110**, 3782-3787, doi:10.1073/pnas.1218721110 (2013).

532 Bedell, V. M. *et al.* In vivo genome editing using a high-efficiency TALEN system. *Nature* **491**, 114-118, doi:10.1038/nature11537 (2012).

533 Huang, P., Zhu, Z., Lin, S. & Zhang, B. Reverse genetic approaches in zebrafish. *J Genet Genomics* **39**, 421-433, doi:10.1016/j.jgg.2012.07.004 (2012).

534 Chen, F. *et al.* High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat Methods* **8**, 753-755, doi:10.1038/nmeth.1653 (2011).

535 Zu, Y. *et al.* TALEN-mediated precise genome modification by homologous recombination in zebrafish. *Nat Methods* **10**, 329-331, doi:10.1038/nmeth.2374 (2013).

536 Wood, A. J. *et al.* Targeted genome editing across species using ZFNs and TALENs. *Science* **333**, 307, doi:10.1126/science.1207773 (2011).

537 Shen, B. *et al.* Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Res* **23**, 720-723, doi:10.1038/cr.2013.46 (2013).

538 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).

539 Consortium, E. P. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).

540 Pennisi, E. Genomics. ENCODE project writes eulogy for junk DNA. *Science* **337**, 1159, 1161, doi:10.1126/science.337.6099.1159 (2012).

541 Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83-90, doi:10.1038/nature11212 (2012).

542 Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82, doi:10.1038/nature11232 (2012).

543 Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91-100, doi:10.1038/nature11245 (2012).

544 Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T. L. & Kruglyak, L. Finding the sources of missing heritability in a yeast cross. *Nature* **494**, 234-237, doi:10.1038/nature11867 (2013).

545 Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-854 (1993).

546 Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223-227, doi:10.1038/nature07672 (2009).

547 Nadeau, J. H. Transgenerational genetic effects on phenotypic variation and disease risk. *Hum Mol Genet* **18**, R202-210, doi:10.1093/hmg/ddp366 (2009).

548 Hollick, J. B. Paramutation: a trans-homolog interaction affecting heritable gene regulation. *Curr Opin Plant Biol* **15**, 536-543, doi:10.1016/j.pbi.2012.09.003 (2012).

549 Waterland, R. A. & Jirtle, R. L. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol Cell Biol* **23**, 5293-5300 (2003).

550 Lam, M. Y., Heaney, J. D., Youngren, K. K., Kawasoe, J. H. & Nadeau, J. H. Trans-generational epistasis between Dnd1Ter and other modifier genes controls susceptibility to testicular germ cell tumors. *Hum Mol Genet* **16**, 2233-2240, doi:10.1093/hmg/ddm175 (2007).

551 Brennecke, J. *et al.* An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322**, 1387-1392, doi:10.1126/science.1165171 (2008).

552 Slotkin, R. K. *et al.* Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**, 461-472, doi:10.1016/j.cell.2008.12.038 (2009).

553 Burton, N. O., Burkhart, K. B. & Kennedy, S. Nuclear RNAi maintains heritable gene silencing in Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **108**, 19683-19688, doi:10.1073/pnas.1113310108 (2011).

554 Forsberg, L. A., Absher, D. & Dumanski, J. P. Non-heritable genetics of human disease: spotlight on post-zygotic genetic variation acquired during lifetime. *J Med Genet* **50**, 1-10, doi:10.1136/jmedgenet-2012-101322 (2013).

555 Forsberg, L. A. *et al.* Age-related somatic structural changes in the nuclear genome of human blood cells. *American journal of human genetics* **90**, 217-228, doi:10.1016/j.ajhg.2011.12.009 (2012).

556 Laurie, C. C. *et al.* Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nature genetics* **44**, 642-650, doi:10.1038/ng.2271 (2012).

557 Jacobs, K. B. *et al.* Detectable clonal mosaicism and its relationship to aging and cancer. *Nature genetics* **44**, 651-658, doi:10.1038/ng.2270 (2012).

6 Acknowledgements

Mein Dank gilt an allererster Stelle Frau Prof. Dr. Juliane Winkelmann, die meine Arbeit in den letzten Jahren durch alle ihre Höhen und Tiefen wahrhaft unermüdlich unterstützt hat und mir zu jeder Tages- und Nachtzeit mit Rat und Tat und einem offenen Ohr als Mentorin zur Seite stand. Besonders danken möchte ich ihr für all ihre Förderung und dafür, dass sie mich auch immer meine eigenen Wünsche bezüglich meiner Promotionsprojekte hat verwirklichen lassen. Und ich würde mich sehr freuen, liebe Juliane, wenn dies erst der Anfang war.

Weiterhin geht mein Dank an Herrn Prof. Dr. Thomas Meitinger und Prof. Dr. Bernhard Hemmer für ihr Interesse an meiner Arbeit und dafür, dass sie mir als Thesis Committee zur Verfügung standen und meine Projekte durch Diskussionen und Kritik vorangetrieben haben.

Ein ganz besonderes Dankeschön geht darüber hinaus an Prof. Dr. Nicholas Katsanis, der mich für ein halbes Jahr in seinem Labor am Center for Human Disease Modelling, Department of Cell Biology, Duke University, Durham, NC, USA, aufnahm, mir die große weite Welt der Zebrafischforschung eröffnete und mich durch seine intelligente und ehrliche Art begeisterte.

Mein ganz großer Dank gebührt auch allen Kollaborationspartnern in München, in Weihenstephan, in Kassel, in Greifswald, in Regensburg, in Birmingham, in Melbourne, in London, in Wien, in Innsbruck, in Durham, in Barcelona, in Prag und in Budapest. Ohne sie wäre keines der hier dargestellten Projekte möglich gewesen.

Leider reicht der Platz nicht aus, um all diejenigen namentlich zu nennen, die meinen Arbeitsalltag—sowohl in München am Helmholtz Zentrum als auch im Katsanis Lab an Duke—geprägt, durch viele interessante Diskussionen bereichert, durch Hilfestellungen und Erklärungen erleichtert und durch ihre Unterhaltung sehr abwechslungsreich gestaltet haben. Dennoch ganz vielen lieben Dank für die schönen Stunden und all Eure Hilfe !

Der allergrößte Dank an dieser Stelle gebührt natürlich meiner Familie, ohne deren liebevolle Unterstützung während der letzten 31 Jahre all dies nicht möglich gewesen wäre. Und zu guter Letzt noch ein großes Dankeschön an alle Freunde und Verwandten, die meine „Doktorarbeitsquengeleien“ auch in der zweiten Runde so geduldig ertragen haben.

7 Fellowships, Grants and Prizes

Parts of this work were supported by

- Travel grants from the Movement Disorder Society (MDS) for their annual meetings in Buenos Aires, Argentina (2010) and Sydney, Australia (2013)

- Travel grants from the Deutscher Akademischer Austausch Dienst (DAAD) for the annual meeting of the Society for Neuroscience in Washington, DC, USA (2011) and the annual meeting of the Movement Disorder Society in Sydney Australia (2013, expenses not covered by the MDS travel grant)

- Travelling fellowship from Stiftung Felgenhauer der Deutschen Gesellschaft für Neurologie (DGN) for stay at the laboratory of Prof. Dr. Yuqing Li, Center for Experimental Therapeutics, Department of Neurology, University of Alabama, Birmingham, AL, USA (2010)

- Travelling fellowship from *Disease Models and Mechanisms* and the Company of Biologists, London, UK, for stay at the laboratory of Prof. Dr. Nicholas Katsanis, Center for Human Disease Modelling, Duke University, Durham, NC, USA (2013)

- Doctoral fellowship and project funding from Kommission für Klinische Forschung (KKF) des Klinikums rechts der Isar der Technischen Universität München, Munich, Germany (2010 to 2012)

- Postdoctoral fellowship in the “Programm für Chancengleichheit für Frauen in Forschung und Lehre” der Exzellenzförderungskommission der Technischen Universität München, Munich, Germany (2011 to 2012)

- Postdoctoral fellowship from the DAAD for stay at the laboratory of Prof. Dr. Nicholas Katsanis, Center for Human Disease Modelling, Duke University, Durham, NC, USA (2012)

- Project support for “Functional Analysis of Rare Variants in Restless Legs Syndrome” from Else Kröner-Fresenius Stiftung, Bad Homburg, Germany (2013 to 2014)

- Poster Prize “Neurogenetik” at the annual meeting of the DGN in Dresden, Germany (2013)

- Charles J. Epstein/American Society for Human Genetics (ASHG) Trainee Award for Excellence in Human Genetics Research (2013, semifinalist)

8 Curriculum Vitae

Name:	Dr. Eva Christina Schulte
Geburtsdatum:	15. August 1981
Geburtsort:	Aachen
Eltern:	Prof. Dr. Dr. Horst Schulte & Marianne Schulte

Schulbildung

1996-2000	Cranbrook Kingswood School, Bloomfield Hills, MI, USA
Sommer 2000	High School Diploma und Allgemeine Hochschulreife (Durchschnittsnote 1,0) Zulassung zum Studium an den amerikanischen Universitäten Harvard, Yale und Stanford.

Studium

2000-2008	Studium der Humanmedizin an der Albert-Ludwigs Universität Freiburg im Brsg. darunter mehr-monatige Auslandsaufenthalte in Oxford, Hong Kong, Lausanne und Teneriffa
Frühjahr 2008	2. Abschnitt der Ärztlichen Prüfung (Durchschnittsnote 1,0) und USMLE Step 1 & 2
seit 2009	PhD-Promotionsstudiengang „Medical Life Science & Technology“, Technische Universität München

Wissenschaftliche Ausbildung

2004-2009	Promotionsarbeiten über trangsene Mausmodelle bei der Alzheimer Erkrankung bei Prof. Dr. Michael Hüll, Psychiatrische Universitätsklinik Freiburg (Dr. med., *summa cum laude*)
2006	6-monatiger Forschungsaufenthalt im Labor von Prof. Dr. Jie Shen, Center for Neurologic Diseases, Harvard Medical School, Boston, MA, USA
2009-2013	Promotionsarbeiten bei Prof. Dr. Juliane Winkelmann, AG Neurogenetik, Helmholtz Zentrum München und Technische Universität München
2012-2013	6-monatiger Forschungsaufenthalt im Labor von Prof. Dr. Nicholas Katsanis, Center for Human Disease Modelling, Duke University, Durham, NC, USA

Weiterbildung

seit 2009	Assistenzärztin an der Neurologischen Klinik der Klinikums rechts der Isar der Technischen Universität München (Prof. Dr. Bernhard Hemmer)

9 Publications

9.1 Publications and Submitted Manuscripts Related to the Dissertation herein

Manuscripts published or in press

1. Schulte E.C., Knauf F., Kemlink D., Gieger C., Lichtner P., Schormair B., Meitinger T., Winkelmann J. Variant Screening of the Coding Regions of MEIS1 in Patients with Restless Legs Syndrome. *Neurology* **76**: 1106-1108 (2011).

2. Schulte E.C., Mollenhauer B., Zimprich A., Bereznai B., Lichtner P., Haubenberger D., Pirker W., Brücke T., Molnar M.J., Peters A., Gieger C., Trenkwalder C., Winkelmann J. Variants in Eukaryotic Translation Initiation Factor 4G1 in Sporadic Parkinson`s Disease. *Neurogenetics* **13**: 281-285 (2012).

3. Schulte E.C., Stahl I., Czamara D., Ellwanger D.C., Eck S., Graf E., Mollenhauer B., Zimprich A., Lichtner P., Haubenberger D., Pirker W., Brücke T., Bereznai B., Molnar M.J., Peters A., Gieger C., Müller-Myhsok B., Trenkwalder C., Winkelmann J. Rare Variants in *PLXNA4* and Parkinson's disease. *PLoS One* **8**: e79145 (2013).

4. Schulte E.C., Ellwanger D.C., Dihanich S., Manzoni C., Stangl K., Schormair B., Graf E., Eck S., Mollenhauer B., Haubenberger D., Pirker W., Zimprich A., Brücke T., Lichtner P., Peters A., Gieger C., Trenkwalder C., Mewes H.W., Meitinger T., Lewis P.A., Klünemann H.A., Winkelmann J. Rare Variants in *LRRK1* and Parkinson`s disease. *Neurogenetics* (2013), epub ahead of print.

5. Zimprich A., Benet-Pages A., Struhal W., Graf E., Eck S.H., Offman M.N., Haubenberger D., Spielberger S., Schulte E.C., Lichtner P., Rossle S.C., Klopp N., Wolf E., Seppi K., Pirker W., Presslauer S., Mollenhauer B., Katzenschlager R., Foki T., Hotzy C., Reinthaler E., Harutyunyan A., Kralovics R., Peters A., Zimprich F., Brücke T., Poewe W., Auff E., Trenkwalder C., Rost B., Ransmayr G., Winkelmann J., Meitinger T., Strom T.M. A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. *Am J Hum Genet* **89**: 168-175 (2011).

6. Winkelmann J., Czamara D., Schormair B., Knauf F., Schulte E.C., Trenkwalder C., Dauvilliers Y., Polo O., Högl B., Berger K., Fuhs A., Gross N., Stiasny-Kolster K., Oertel W., Bachmann C.G., Paulus W., Xiong L., Montplaisir J., Rouleau G.A., Fietze I., Vavrova J., Kemlink D., Sonka K., Nevsimalova S., Lin S.C., Wszolek Z., Vilarino-Güell C., Faller M.J., Gschliesser V., Frauscher B., Falkenstetter T., Poewe W., Allen R.P., Earley C.J., Ondo W.G., Le W.D., Spieler D., Kaffe M., Zimprich A., Kettunen J., Perola M., Silander K., Cournu-Rebeix I., Francavilla M., Fontenille C., Fontaine B., Vodicka P., Prokisch H., Lichtner P., Peppard P., Faraco J., Mignot E., Gieger C., Illig T., Wichmann H.E., Müller-Myhsok B., Meitinger T. Genome-wide association study identifies novel restless legs syndrome susceptibility loci on 2p14 and 16q12.1. *PLoS Genet* **7**: e1002171 (2011).

7. Hopfner F., Schulte E.C., Mollenhauer B., Bereznai B., Knauf F., Lichtner P., Zimprich A., Haubenberger D., Pirker W., Brücke T., Peters A., Gieger C., Kuhlenbäumer G., Trenkwalder C., Winkelmann J. The role of SCARB2 as susceptibility factor for Parkinson`s disease. *Mov Disord* **28**: 538-540 (2013).

8. Schulte E.C., Winkelmann J. Family- and Genome-Wide Association Studies of Restless Legs Syndrome in *Genetics of Sleep and Sleep Disorders*. (Ed.: Thorpy M, Shaw P, Tafti M). Cambridge University Press, Cambridge. 2013, in press.

9. Schulte E.C., Winkelmann J. Clinical Phenotype and Genetics of Restless Legs Syndrome in *Movement Disorders: Genetics and Models*. (Ed.: LeDoux MS). Elsevier, San Diego, CA. 2013, in press.

Manuscripts submitted

1. Schulte E.C., Heim K., Schurmann C., Homuth G., Lichtner P., Gieger C., Peters A., Trenkwalder C., Högl B., Frauscher B., Berger K., Fietze I., Gross N., Stiasny-Kolster K., Oertel W., Bachmann C.G., Paulus W., Zimprich A., Müller-Myshok B., Prokisch H., Winkelmann J. Blood *cis*-eQTLs in Follow-Up Analysis of Genome-Wide Association Studies in Restless Legs Syndrome, revision process at *PLoS ONE*.

2. Schulte E.C., Fukumori A., Mollenhauer B., Hor H., Arzberger T., Perneczky R., Kurz A., Hüll M., Lichtner P., Eckstein G., Zimprich A., Haubenberger D., Pirker W., Brücke T., Bereznai B., Molnar M.J., Lorenzo-Betancor O., Pastor P., Peters A., Gieger C., Estivill X., Meitinger T., Kretzschmar H.A., Trenkwalder C., Haass C., Winkelmann J. Rare Variants in ß-Amyloid Precursor Protein (APP) and Parkinson`s Disease, *submitted*.

3. Schulte E.C.*, Kousi M.*, Tan P., Schormair B., Knauf F., Lichtner P., Trenkwalder C., Högl B., Frauscher B., Berger K., Fietze I., Gross N., Stiasny-Kolster K., Oertel W., Bachmann C.G., Paulus W., Zimprich A., Peters A., Gieger C., Meitinger T., Müller-Myshok B., Katsanis N., Winkelmann J. An excess of rare loss-of-function alleles in individuals with restless legs syndrome substantiates *MEIS1* as the genetic factor underlying the GWAS locus on chromosome 2, *submitted*.

9.2. Publications Not Related to this Dissertation

1. Schulte E.C.*, Claussen M.*, Jochim A., Haack T., Hartig M., Hempel M., Prokisch H., Haun-Jünger U., Winkelmann J., Hemmer B., Förschler A., Ilg R. Mitchondrial Membrane Protein Associated Neurodegeneration: A Novel Variant of Neurodegeneration with Brain Iron Accumulation. *Mov Disord* **28**: 224-227 (2013).

2. Schulte E.C., Gross N., Slawik H., Winkelmann J. When Restless Legs Syndrome turns malignant. *Sleep Med* **14**: 575-577 (2013).

3. Schulte E.C., Winkelmann J. When Parkinson's disease patients go to sleep: specific sleep disturbances related to Parkinson's disease. *J Neurol* 258: S328-335 (2011).

4. Schulte E.C., Spieler D., Winkelmann J. Restless-legs syndrome and cardiovascular diseases, *Der Nervenarzt* **82**: 1006-1011 (2011).

5. Schulte E.C., Sauerbrei A., Hoffmann D., Zimmer C., Hemmer B., Mühlau M. Acylcovir Resistance in Herpes Simplex Encephalitis. *Ann Neurol* **67**: 830-833 (2010).

6. Wines-Samuelson M., Schulte E.C., Smith M.J., Aoki C., Liu X., Kelleher R.J. III, Shen J. Characterization of age-dependent and progressive cortical neuronal degeneration in presenilin conditional mutant mice. *PLoS ONE* **5**: e10195 (2010).

7. Schulte E.C., Slawik H., Schüle R., Gunther T., Hüll M. Alterations in excitotoxicity and prostaglandin metabolism in a transgenic mouse model of Alzheimer's disease. *Neurochem Int* **55**: 689-696 (2009).