TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Entwicklungsgenetik

Genome-wide and Regional Case-Control Association Studies in the Genetic Analysis of Restless Legs Syndrome

Barbara Schormair

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. H.-R. Fries

Prüfer der Dissertation:

1. Univ.-Prof. Dr. W. Wurst
2. apl. Prof. Dr. J. Winkelmann
3. Univ.-Prof. Dr. Th. A. Meitinger

Die Dissertation wurde am 07.10.2009 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 30.04.2010 angenommen.

Genome-wide and Regional Case-Control Association Studies in the Genetic Analysis of Restless Legs Syndrome

Kumulative Arbeit

Barbara Schormair

Für meine Eltern

Table of contents

Abbreviations

ARF	attributable risk fraction
AUC	area under the curve
B	Bonferroni
bp	base pair
BTBD9	BTB (POZ) domain containing 9
CDCV	common disease common variant
chr	chromosome
CNV	copy number variation
CI	confidence interval
CNS	central nervous system
CSF	cerebrospinal fluid
DNA	deoxyribonucleic acid
DZ	dizygotic
ESRD	end-stage renal disease
GRR	genotypic relative risk
GWAS	genome-wide association study
HWE	Hardy-Weinberg equilibrium
iRLS	idiopathic RLS
IRLSSG	International Restless Legs Syndrome Study Group
kb	kilobase
LBX1	ladybird homeobox 1
LBXCOR1	Lbxcor1 homolog (mouse)
LD	linkage disequilibrium
L-DOPA	levodopa
MAF	minor allele frequency
Mb	megabase
MEIS1	Meis homeobox 1
MRI	magnetic resonance imaging
MZ	monozygotic
NO	nitric oxide
NOS1, nNOS	neuronal nitric oxide synthase
OR	odds ratio
P	P value
PLM	periodic limb movements
PLMW	periodic limb movements while awake
PLMS	periodic limb movements in sleep
PTPRD	protein tyrosine phosphatase receptor type delta
RNA	ribonucleic acid
RLS	restless legs syndrome
ROC	receiver operating characteristic
SNP	single nucleotide polymorphism
sRLS	symptomatic RLS
TH	tyrosine hydroxylase
uRLS	uremic RLS
VNTR	variable number of tandem repeats

Summary

Restless legs syndrome (RLS) is a common neurological disorder with an age-dependent prevalence of 10% in Caucasian populations. It is characterized by an urge to move the legs usually accompanied by unpleasant or painful sensations. Apart from the idiopathic form of RLS there is also a symptomatic form in other medical conditions such as iron deficiency, end stage renal disease or pregnancy. It is a complex disease determined by both genetic and non-genetic factors and interactions between them. RLS shows strong familial aggregation (familial RLS) and family and twin studies have found heritability estimates of about 60% highlighting the significant genetic contribution to the phenotype. The susceptibility to common complex diseases is presumed to be influenced substantially by common genetic variants with allele frequencies > 1% (common disease/common variant hypothesis). Only recently, it has become feasible to conduct genetic association studies for these variants covering large candidate genomic regions or even the entire genome. These studies use common single nucleotide polymorphisms (SNPs) as markers and rely on the correlation of the genotypes of tightly linked variants on the population level due to linkage disequilibrium (LD) between them. SNPs in strong LD serve as surrogates for each other and thus it is sufficient to assay only a subset of so-called tagging SNPs to cover the majority of the common variation in the genome.

The aim of this work was the identification of common susceptibility variants for idiopathic RLS. We conducted large-scale association studies of different designs: hypothesis-based regional studies of previously identified linkage regions (RLS-1 on chromosome (chr.) 12 and RLS-3 on chr. 9) and a hypothesis-free genome-wide study.

These approaches identified associated variants in five genomic loci. Within the loci the association signals either map to single genes (*NOS1* on chr. 12, *MEIS1* on chr. 2, *BTBD9* on chr. 6, and *PTPRD* on chr. 9) or span parts of two adjunct genes (*MAP2K5*/*LBXCOR1* on chr. 15). Compelling statistical evidence and replication in independent populations ensure that the signals in *MEIS1*, *BTBD9*, *MAP2K5*/*LBXCOR1*, and *PTPRD* represent true-positive associations. The *NOS1* association was inconsistent in our study and thus needs evaluation in further independent samples to either confirm the association or identify the signal as a false-positive. The association signals primarily delineate the genomic regions that contain the causal variants and possibly also the disease-related genes. Further analysis of these regions is necessary for assigning a causal role to any of the associated variants or identifying the true causal variants, which will then provide clues on the molecular mechanisms underlying the

disease. *MEIS1*, *MAP2K5*, *LBXCOR1*, and *PTPRD* are known to play a role in embryonic development, which poses the question if RLS has components of a developmental disorder. However, these genes could have different functions in adult life that could be causally related to RLS.

Interestingly, the variants in *MEIS1*, *BTBD9*, *MAP2K5*/*LBXCOR1* and *PTPRD* showed equal contributions to disease risk in familial and sporadic RLS. An association study in uremic RLS, the most common form of symptomatic RLS, indicated equal contributions of *MEIS1* and *BTBD9* to idiopathic and symptomatic RLS. This suggests a common pathophysiology and overlapping predisposition mechanisms. A possible disease model could be a base-line predisposition to RLS conferred by these variants, which is influenced by proprietary additional factors in the different forms of RLS, e.g. renal failure in uremic RLS.

The associated SNPs are common and confer a moderate increase in disease risk (odds ratios (ORs) of 1.3 to 1.7). A strong risk haplotype was found in *MEIS1* with an OR of 2.7. These common variants together account for a large proportion of RLS cases in the general population, confirming to a certain extent the notion of the common disease common variant hypothesis. However, they explain only a small fraction of the heritability and the familial aggregation of RLS. This limits their potential for individual risk prediction and renders genetic testing meaningless at present. Due to limitations in power and the scope of genetic variants detectable in current designs, the studies are far from being exhaustive and many other variants remain to be identified. These will include further common variants with smaller effect sizes, structural variants, epigenetic modifications, and rare variants of larger effect sizes. Once a more complete picture of the underlying genetic variation in RLS is obtained, the causal pathomechanisms and molecular pathways will start to be unravelled and even meaningful genetic risk profiling leading to tailored prevention strategies and therapy options might come into reach.

Zusammenfassung

Das Restless Legs Syndrom (RLS) ist mit einer altersabhängigen Prävalenz von circa 10% in der kaukasischen Bevölkerung eine der häufigsten neurologischen Erkrankungen. Charakteristisch für die Erkrankung ist ein außerordentlicher Bewegungsdrang der Beine, der von unangenehmen oder schmerzhaften Empfindungen in den Beinen verursacht oder begleitet wird. Es tritt sowohl als eigenständige Krankheit auf (idiopathisches RLS) als auch als Komorbidität in anderen Erkrankungen wie Eisenmangelanämie, Urämie oder auch in der Schwangerschaft (symptomatisches RLS). Das RLS ist eine komplexe Erkrankung, die durch genetische und nicht-genetische Faktoren bestimmt wird. Es findet sich eine deutliche familiäre Aggregation (familiäres RLS). Zwillings- und Familienstudien zeigen eine Heritabilität von ca. 60% für das RLS, die die Bedeutung genetischer Faktoren in seiner Ätiologie zeigt. Man nimmt an, dass die Suszeptibilität für häufige komplexe Erkrankungen durch häufige genetische Varianten mit Allelfrequenzen > 1% bestimmt wird („common disease common variant“ Hypothese). Seit kurzem ist es möglich, genomweite oder auf Kandidatenregionen fokussierte Assoziationsstudien für diese Varianten durchzuführen. Diese Studien verwenden Einzelnukleotid-Polymorphismen (single nucleotide polymorphisms, SNPs) und basieren auf der Korrelation von Genotypen eng beieinander liegender Varianten innerhalb einer Bevölkerung aufgrund von Kopplungsungleichgewicht (linkage disequilibrium, LD) zwischen ihnen. SNPs, die in hohem LD miteinander stehen, können sich gegenseitig ersetzen. Daher genügt es, eine Auswahl sogenannter tagging SNPs zu typisieren, um den Großteil der häufigen Varianten im Genom abzudecken.

Das Ziel dieser Arbeit war die Identifizierung häufiger Suszeptibilitätsvarianten für das idiopathische RLS. Zu diesem Zweck wurden zwei Formen von Assoziationsstudien eingesetzt. Zum einen wurden in fokussierten Studien bereits bekannte Kopplungsregionen (RLS-1 auf Chromosom (Chr.) 12 und RLS-3 auf Chr. 9) untersucht, zum anderen wurde als hypothesenfreier Ansatz eine genomweite Assoziationsstudie durchgeführt.

Insgesamt wurden mit dem RLS assoziierte Varianten in fünf genomischen Loci gefunden. In diesen Loci sind die Assoziationssignale entweder in einzelnen Genen lokalisiert (*NOS1* auf Chr. 12, *MEIS1* auf Chr. 2, *BTBD9* auf Chr. 6 und *PTPRD* auf Chr. 9) oder erstrecken sich über Teilbereiche zweier benachbarter Gene (*MAP2K5*/*LBXCOR1* auf Chr. 15). Die Assoziationen in *MEIS1*, *BTBD9*, *MAP2K5*/*LBXCOR1* und *PTPRD* sind hochsignifikant und ihre Echtheit ist durch Replikation in weiteren unabhängigen Studienpopulationen bestätigt. Für die Assoziation zu *NOS1* waren die Ergebnisse widersprüchlich, daher bedarf es weiterer

Replikationsstudien, um dieses Signal entweder zu bestätigen oder als falsch-positiv zu identifizieren. Die Assoziationssignale grenzen in erster Linie nur die Kandidatenregionen ab, die die kausalen Varianten und krankheitsrelevanten Gene enthalten. Erst die weitere detaillierte Analyse dieser Regionen wird Aufschluss darüber geben können, welche Varianten und Gene tatsächlich an der Krankheitsentstehung beteiligt sind und welche molekularen Mechanismen in der Pathophysiologie eine Rolle spielen könnten. Da *MEIS1*, *MAP2K5*, *LBXCOR1* und *PTPRD* für die Embryonalentwicklung relevant sind, stellt sich die Frage, ob das RLS Komponenten einer Entwicklungsstörung hat. Andererseits könnten diese Gene in späteren Lebensphasen andere noch unbekannte Funktionen haben, deren Veränderung oder Ausfall das RLS auslösen könnte.

Interessanterweise ist die Risikoerhöhung durch die assoziierten Varianten in *MEIS1*, *BTBD9* *MAP2K5*/*LBXCOR1* und *PTPRD* bei familiärem und bei sporadischem RLS gleich groß. Eine Assoziationsstudie des urämischen RLS, der häufigsten symptomatischen Form, deutete ebenfalls einen vergleichbaren Beitrag von *MEIS1* und *BTBD9* bei idiopathischem und symptomatischem RLS an. Dies lässt gemeinsame Mechanismen in der Pathophysiologie und Prädisposition vermuten. Daraus ließe sich als mögliches Krankheitsmodell eine Prädisposition für RLS aufgrund dieser Varianten ableiten, die dann in den verschiedenen Formen des RLS jeweils durch zusätzliche spezifische Faktoren beeinflusst wird wie z. B das Nierenversagen beim urämischen RLS.

Die assoziierten SNPs sind häufig und führen zu einer moderaten Erhöhung des Erkrankungsrisikos (Odds Ratios (ORs) von 1.3 bis 1.7). Ein stärkerer Effekt ist mit einem Haplotyp in *MEIS1* verbunden, der eine OR von 2.7 aufweist. Gemeinsam sind diese häufigen Varianten für einen Großteil der RLS Fälle in der Allgemeinbevölkerung verantwortlich und unterstützen so die „common disease common variant“ Hypothese. Sie können jedoch nur einen kleinen Anteil der Heritabilität und der familiären Aggregation des RLS erklären. Dadurch ist ihr Potential für individuelle Risikoprognosen und damit auch der Einsatz in genetischen Testverfahren beschränkt. Viele weitere Varianten sind noch zu entdecken, da die Power der Studien und das Spektrum der mit ihnen detektierbaren genetischen Varianten begrenzt waren. Darunter fallen zusätzliche häufige Varianten mit geringerem Effekt, strukturelle Varianten, epigenetische Modifikationen und seltene Varianten mit stärkerem Effekt. Ein umfassenderes Bild der Suszeptibilitätsvarianten für RLS ist wichtig für die Identifizierung der zugrundeliegenden Pathomechanismen. Weiterhin können dann möglicherweise aussagekräftigere genetische Risikoprofile als Basis für maßgeschneiderte Präventionsmaßnahmen und Therapieoptionen erstellt werden.

1 Introduction

1.1 Association studies in the genetic mapping of common diseases

1.1.1 Rationale and basic principles

Genetics contribute significantly to human disease and the current main strategy to localize the causally related genes or genetic variants is mapping them in the genome based on a correlation of the disease phenotype with naturally occurring DNA variants, termed markers in such studies [1]. For common diseases association studies have become the most frequently employed approach. This choice is based primarily on the complex multifactorial etiology and the supposed genetic architecture of these diseases, which result in a limited power of the classic gene mapping approach of linkage analysis [2, 3]. Linkage studies rely on the detection of a cosegregation of marker loci and a disease in affected families. Such a cosegregation occurs because physically close genetic loci are linked, i.e. inherited together more often than expected by chance due to the reduced recombination frequency between them in meiosis. Therefore, marker loci segregating with the disease are assumed to be located near the disease-causing gene and delineate a corresponding candidate genomic region [4, 5]. To be successful, such studies require clear-cut segregation patterns in families. This is the case in Mendelian diseases where a single major gene is sufficient to determine the outcome, i.e. where the causal variants have a large effect on the risk of disease [3, 6]. Common diseases, however, are mostly determined by multiple genetic and environmental (non-genetic) factors and interactions between them, with each factor presumably having only a modest effect on the risk of disease [3, 6-8]. Due to this interplay of many different factors with partly equal contributions, they do not show clearly observable segregation patterns in families [3, 6, 9]. Moreover, the susceptibility to common complex diseases is thought to be influenced substantially by common genetic variants with allele frequencies > 1%, a notion known as the common disease/common variant (CDCV) hypothesis [2, 10, 11]. Such common variants are presumed to confer only minor increases in disease risk [7, 11]. The low power of linkage analysis for detecting these common, modest effect variants was demonstrated in an influential study by Risch and Merikangas in 1996, where the authors also advocated association analysis as the more powerful mapping strategy for common complex diseases [2]. Association studies do not rely on tracing transmission patterns in families, but compare allele or genotype frequencies of genetic variants between affected and unaffected

subjects at the population level. A variant is associated with the disease if there is a statistically significant difference in the frequencies between affected and unaffected individuals [5, 9, 12]. Due to the relative ease of collecting large numbers of samples compared to family-based study designs, most association studies are conducted as case-control studies, where unrelated affected (= cases) and unaffected subjects (= controls) are compared. Cases have a definite diagnosis of the disease, whereas controls can either be negative for it or be randomly selected from the general population without being screened for the disease [12-14]. Single nucleotide polymorphisms (SNPs) are the logical choice as markers in these studies, because they account for the majority (up to 90%) of common variation in the human genome and are found in high density throughout the genome with an expected frequency of on average one SNP every 300 basepairs (bp) [15, 16]. SNPs are changes of one nucleotide and in most cases have only two alleles, enabling easy genotyping [16]. Based on population genetics theory around 11 million SNPs with allele frequencies > 1% were expected to exist in the human genome [15, 16]. The most recent build of the public SNP database dbSNP already contains over 17 million annotated SNPs, of which 6.5 million have been validated so far and can reliably serve as markers for association studies [17]. The statistical analysis of biallelic SNPs in case-control studies is relatively straightforward. The null hypothesis of no association between each single SNP and disease status can be tested in 2x3 (based on genotype counts) or 2x2 (based on allele counts) contingency tables by applying standard χ^2 tests for independence, e.g. a Pearson or Fisher exact test [12, 18]. These tests perform well in dominant, recessive or multiplicative genetic models but have less power in additive models, where each risk allele contributes the same amount to the risk for disease, i.e. two copies of the risk allele double the risk. For this situation, which is often assumed in complex diseases, the Armitage trend test is a very powerful statistical test [14, 18, 19]. Another approach frequently used is logistic regression analysis, which allows the inclusion of possible affection status modifiers such as age or gender as additional covariates in the analysis [18, 20]. SNP-based association studies can be focused on candidate genes or individual genomic regions based on prior hypotheses about their biological function or their identification in linkage studies [13]. Only very recently hypothesis-free genome-wide association studies (GWAS) have become feasible due to huge advances in high-throughput low-cost genotyping technologies for SNPs [21], the development of appropriate statistical methods for such large-scale analyses [18, 19], and the cataloguing of SNP markers across the entire genome [22-24].

1.1.2 Relevance of linkage disequilibrium to study design and analysis

The most straightforward reason for an association between a genetic variant and a disease is a causal role of the variant. This type of association is referred to as direct [9, 12]. However, functional candidate variants for direct association studies are difficult to define and testing all common variants or at least all known SNPs in the genome is not yet feasible [7, 12, 25]. Therefore, the current SNP-based association studies rely on indirect association where the associated variant is not the causal variant but merely correlated with it in the population due to linkage disequilibrium (LD) between them [7, 24]. LD is defined as the non-random association of alleles at genetic loci, i.e. the alleles at these loci are not inherited independently from each other but are found together more often than expected based on their frequencies [26, 27]. The basis of LD is the joint population ancestry shared by these variants. When a new DNA sequence variation is introduced by mutation, it occurs on a certain ancestral haplotype, i.e. on the background of a specific combination of alleles at the polymorphic loci in its vicinity. This combination of alleles is in complete LD at this point and through subsequent generations the LD between them will be reduced mainly by recombination so that only alleles at very tightly linked loci will stay in strong LD [12, 16, 24, 26]. However, several other forces such as natural selection, genetic drift, gene conversion, mating patterns, population demographics, and mutation also influence the extent of LD [26, 28]. These different influences create great variation in the patterns of LD across the genome. Although a general decrease in LD between loci with increasing distance and thus increasing probability of recombination between them is observed, physical closeness of loci does not necessarily entail strong LD between them and great distances between loci do not impede high degrees of LD. Detailed studies of LD in genomic regions found complete LD between loci several kb apart and weak to no LD between loci separated only by a few hundred bp [29-34]. These studies revealed that the genome is organized in a block-like structure. Blocks of high LD are separated by regions with little to no LD, which often correspond to hotspots of recombination [24, 32, 34-36]. Within each block a small number of common haplotypes captures most of the chromosomes in the population [24, 34, 36]. The size of the blocks was found to vary between 10 and 200kb with the mean size estimated to be 22kb in populations of European and Asian ancestry and 11kb in populations of African ancestry [32, 34, 36]. The strength of LD is usually measured between pairs of loci. Several statistical measures have been proposed, but only D' and r^2 are commonly used [37]. Both scale between 0 (no LD) and 1 (absolute LD) but differ in their properties and thus in their application area. D' is

commonly used to assess recombination patterns and recombination history [34, 36, 38]. It is equal to 1 (termed "complete LD") as long as no recombination occurs between the two loci, independent of the allele frequencies at both loci. Intermediate values of D' have no clear interpretation and its values are inflated by small sample sizes and low allele frequencies [26]. r^2 is the preferred measure in LD-based association studies, because it quantifies the statistical correlation between the alleles at the two loci [26, 27]. It equals 1 (termed "perfect LD") only if there has been no recombination and the allele frequencies at both loci are identical. In this case, every occurrence of one specific allele at each of the markers perfectly predicts the allele at the other locus and they can be used as surrogates for each other [26]. Indirect association studies utilize the LD and haplotype block structure of the genome by choosing a subset of SNPs (tagging SNPs or tagSNPs) that can serve as surrogates for variants in high LD with them [16, 24, 39]. Due to the high variability of LD within the genome and also between different human populations, genome-wide LD maps for the various populations are a prerequisite for the selection of SNP markers in such association studies. To this end, the International HapMap project was launched in 2002 [40]. By 2007 (HapMap phase II), over 3.1 million SNPs had been genotyped in four selected populations of European, African, and Asian ancestry, equal to a density of one SNP every 1000 bp [22]. The most recent release in 2009 contains genotypes of over 4 million SNPs in these populations [41]. Several methods are available to choose tagging SNPs for a focused association study of a genomic region [24, 42]. Some are based on the haplotype structure and choose tagging SNPs for all common haplotypes in this region. Others use the pairwise LD between markers given by r^2 and require an $r^2 \geq 0.8$ between the tagging SNP and the tagged SNP [43]. For GWAS there is a range of commodity arrays containing between 300,000 to 1,000,000 SNPs. These arrays have been shown to have a coverage ranging from 68% (Affymetrix 500K and 5.0 arrays) to 89% (Illumina HumanHap650Y array) of the common variation in the genome in the HapMap population of European descent, defined as the percentage of SNPs tagged at $r^2 \geq 0.8$ [24].

The LD and haplotype structure of the genome also adds power to the statistical analysis because single marker analysis can be extended to haplotype analysis [18, 19]. A haplotype can capture a genetic effect that requires a specific combination of alleles at multiple SNPs. Moreover, it can potentially tag an untyped causal variant that is not captured by any individual SNP [14].

In LD-based studies the probability is very low that an associated variant is the true causal variant. Statistical analysis can only narrow the association signal down to the level of an LD

block. Within this block, any of the probably multiple associated SNPs can be causal or in LD with the causal variant. Therefore, these studies entail the need for further analysis of the associated region, e.g. by sequencing or by further association studies using more SNPs or including potential functional SNPs like nonsynonymous coding SNPs [1, 44-46].

1.1.3 Study power and detection range

The power of any statistical test is defined as the probability of correctly rejecting the null hypothesis when it is truly false, i.e. it represents the chance of detecting a true effect in the study. With regard to genetic association studies this means detecting a genuine association between a genetic variant and the disease [13]. Determining the power of a study is important in both study design and interpretation of the results. Studies are only sensible if they are sufficiently powered to detect the type of variants expected for common diseases. Also, if a study does not yield any significant associations, post-hoc power considerations can offer an explanation. A number of open-source software tools are available for such power calculations and two of them were used in this work: the web-based genetic power calculator [47, 48] and the program CaTS [49, 50].

Several factors directly influence the power of an association study [7, 51]. One important determinant is the sample size. The larger the study sample, the higher the power [46]. Another aspect is the strength of LD between the marker locus and the disease locus. Power is greatest if marker and disease locus are identical (direct association) or in perfect LD ($r^2 = 1$) with each other. A decrease in LD leads to a loss of power, which can be easily derived from the inverse relationship between the r^2 value and the required sample size for a certain study power. If a sample size of N, e.g. 1000, was required for detecting a direct association, an r^2 of 0.5 between marker and disease locus would mean an increase of the sample size by $1/r^2$ for identical power, in this case 2000 [26, 27]. The key determinants of study power, however, are the effect size of the genetic variant, i.e. the extent to which it influences the risk for disease, and the allele frequencies at both marker and disease locus [38]. The effect size can be represented as the genotypic relative risk (GRR), which measures the relative increase in risk for individuals with a certain genotype, e.g. homozygous for the risk allele, compared to another genotype, e.g. homozygous for the non-risk allele. In case-control association studies the standard measure is the odds ratio (OR), which is the odds of exposure to the risk allele among the cases divided by the odds of exposure to the risk allele among the controls. An OR > 1 equals an increase in risk conferred by the variant, an OR < 1 indicates a protective effect, and an OR =1 means no contribution of the variant to disease risk [38]. The

observed effect size of a variant screened as a marker in an association study depends on the OR of the true causal variant for the disease, the LD between the marker and the causal locus, and the allele frequencies at both marker and causal locus [7]. In general, higher allele frequencies and larger effect sizes increase power. Differences in the allele frequencies between the marker locus and the disease locus lead to a decrease in study power [7, 51]. Common variants with minor allele frequencies (MAFs) below 10% and ORs > 1.3 require very large samples sizes of several thousands and variants with lower ORs are even more difficult to detect. Less frequent variants with MAFs below 1% are not detectable in association studies of feasible size unless they have large effect sizes (ORs of ≥ 2). However, such variants are thought to be untypical for common diseases [38, 52]. Currently, most association studies have a sample size of 1000 to 5000 cases and are only sufficiently powered to detect variants with MAFs above 5% and ORs of 1.2 and above [7, 18, 25]. Due to the fact that these studies are based on the analysis of common SNPs (MAFs > 5%), the detection scope of association analysis is limited predominantly to such variants [53]. It is possible to detect common structural variants like insertions, deletions or copy number variations (CNVs) if they are in strong LD with some of the genotyped SNPs [54, 55]. The latest generation of commercial SNP chips also offers additional probes for the detection of CNVs [56, 57]. However, rarer SNPs and structural variants are not detectable by association studies as they are implemented at present [14, 45, 46, 53].

1.1.4 Minimizing false-positive and false-negative associations

Association analysis faces two problems, that of false-negative results (type II errors), where a genuine association signal is not detected and that of false-positive results (type I errors), where spurious association arises due to confounding factors. Both types of errors decrease the overall power of the study and thus it is essential to minimize the possibility of such erroneous results [3, 13, 16]. There are three main sources of false-positive and false-negative results: phenotype and genotype misclassification errors, the simultaneous testing of many hypotheses (problem of multiple testing) and the presence of population stratification in the study sample [39, 44, 51]. For each of these issues numerous approaches have been developed to minimize their influence and/or to correct the data adequately in order to maximize the chances of the detected associations being true positives. Phenotypic misclassification can occur in both cases and controls. Diseases with a high variability in symptom manifestation and potential genetic heterogeneity are prone to misdiagnosis and moreover a case sample might consist of subjects suffering from the same disease due to a different genetic

predisposition, obscuring true-positive associations. Therefore, the diagnostic screening of patients has to be extensive and rigorous. It is a frequently followed course of action to focus on a defined subgroup of cases, e.g. those with a familial history of the disease or with an early age of onset in order to minimize the sample heterogeneity [13, 14, 44, 51]. The presence of affected subjects in the control group preferably should be excluded by screening the controls for the disease and including only truly unaffected subjects in the study (super-controls). However, it has been demonstrated that the loss of power due to the use of un-screened common population-based controls can be balanced by increasing the sample size [14, 58]. Genotypic misclassification such as incorrectly typing heterozygotes as homozygotes becomes increasingly problematic the more SNPs are genotyped. Another problem in genotype quality is differential missingness of genotype data between cases and controls. If a certain genotype of a SNP is called with different success rates in cases and controls this can change allele frequencies and LD structures and lead to false-positive or false-negative results [14, 18]. For these reasons, association studies have to apply a strict quality control to genotype data before statistical analysis. Standard quality criteria are minimal callrate requirements per SNPs and also per individual genotyped. Most studies set the threshold for inclusion in the analysis at a callrate ≥ 95%. Another measure is testing the genotype data in controls for significant deviations (commonly used threshold: $P < 10^{-4}$) from Hardy-Weinberg equilibrium (HWE), which can indicate genotyping errors, and discard such SNPs [18, 19]. Furthermore, the allele frequency is used as a filter. Genotypes of low frequency SNPs are more susceptible to errors and thus most studies exclude SNPs with MAFs < 5% from the analysis [19]. To exclude technical artifacts, SNPs can be retyped on a different genotyping platform [14].

The problem of multiple testing is a major issue inflating the type I error rate, especially in genome-wide association studies testing 500,000 or more SNPs simultaneously [16, 18]. In order to control the type I error rate in a study, a significance level α is defined, which is the accepted probability of a false positive result in an experiment. Associations are significant only if their P value is below this threshold. It is usually set at 5% for testing a single hypothesis, i.e. a P value < 0.05 is required for statistical significance. Testing more than one hypothesis at a time increases the probability of false positive results by chance. In order to maintain a type I error rate of 5% over multiple tests, the point-wise significance level for each individual test has to be lowered. A simple yet robust method to correct for multiple testing is the Bonferroni correction, which divides the standard significance level α = 5% by the number of tests performed in order to obtain a new point-wise significance level α′

corrected for multiple testing [18, 39]. This is equivalent to multiplying the nominal P value by the number of tests and keeping the 5% significance level, i.e. a result is significant if the corrected P value is < 0.05. The Bonferroni method is frequently used in association studies at present and was also used in this dissertation [44]. However, it is very conservative, because it does not consider the correlation of SNPs due to LD but treats each SNP as an independent test. An alternative approach is the approximation of the required significance level by permutation-based procedures. These methods retain the genotype data and thus the LD structure, but shuffle the affection status randomly between cases and controls, effectively destroying any biologically meaningful association. This permutation process is repeated thousands of times and each time the P values are calculated to obtain their distribution under no association. Comparing the actual distribution and the permutation-based distribution then allows adjustment of the significance level and the calculation of empirical P values corrected for multiple testing [18, 39].

Another critical issue in case-control association studies is confounding due to population stratification [16, 18, 39]. It is present when the sample contains several subpopulations or when the study population originated from a recent admixture of multiple ethnic groups. Since different human populations differ in allele frequencies and also in disease prevalence, the presence of such subgroups alone will cause statistically significant differences in allele frequencies of genetic variants between cases and controls which are not due to a biological association between the variants and the disease. It is a major concern in case-control association studies, because affected and unaffected subjects are unrelated subjects from the general population. Family-based association designs are protected against this confounding, because affected and unaffected subjects definitely have the same ethnic background [16, 18, 39]. Careful matching of cases and controls to ensure ethnic homogeneity of the sample is therefore a prerequisite for successful association studies [14]. Moreover, the extent of population stratification in the study sample should be assessed and adequate statistical corrections for it should be applied. Several methods have been developed to perform such corrections [18, 19]. The two most frequently used approaches, which were also applied in this work, are genomic control [59], and the principal components analysis approach implemented in the software EIGENSTRAT [60]. Genomic control is suitable for both region-focused and genome-wide studies. This method requires the genotyping of at least 100 neutral or null SNPs, i.e. SNPs not related to the disease, throughout the genome. The distribution of the association test statistic obtained from these SNPs is used to calculate an inflation factor λ, reflecting the effect of the stratification. Then the test statistic calculated for

each marker genotyped in the study is divided by λ to correct for this effect [59]. In contrast, EIGENSTRAT is only applicable to genome-wide studies. It infers the principal components of the genetic variation in the genotype dataset and groups individuals accordingly. This creates a set of matched cases and controls and allows the identification and removal of outlying individuals with differing ancestry [60].

The most conclusive evidence that a detected association is a true-positive is its exact replication in independent populations. This means confirming the association of the same allele or haplotype with the same direction and comparable size of the ORs as in the original study [14, 61, 62].

1.2 Restless legs syndrome (RLS): a common complex disease

1.2.1 Clinical characteristics and epidemiology

The diagnosis of RLS is based on the description of the symptoms by the patient, as there is no objective biochemical or physiological test for establishing the presence of RLS to date. Generally accepted diagnostic criteria were set up by the International Restless Legs Syndrome Study Group (IRLSSG) in 1995 [63] and further refined in 2003 [64] and are now the gold standard for diagnosing RLS.

Four diagnostic criteria (essential criteria) must be fulfilled for the definite diagnosis of RLS. These represent the key characteristics of RLS as described by the patients [64]:

(1) Presence of an urge to move the legs usually accompanied or caused by uncomfortable and unpleasant sensations in the legs.

The urge to move the legs is the key characteristic of RLS. It can also be present without any further sensory symptoms. The sensations (paresthesias and/or dysesthesias) are usually located deep inside the leg, affecting muscle or bone and are described as “some kind of movement inside the legs” by the patients (descriptive terms include for example “soda bubbling in the veins” or “ants crawling”). About 50% of the patients perceive the sensations as painful [65].

(2) Symptoms begin or worsen during rest or inactivity.

RLS is a quiescegenic disorder where symptoms are induced or at least aggravated by rest. The term “rest” here encompasses both physical inactivity and reduced mental activity. Although most patients experience symptoms while sitting or lying, no specific body position is associated with causing the symptoms.

(3) Total or partial relief of symptoms by movement.

Moving the affected limbs e.g. by walking around or stretching the legs promptly relieves symptoms at least for the duration of this activity. This relief can be complete or partial.

(4) Symptoms are worse in the evening or at night or only occur at this time.

A diurnal rhythm of both sensory and motor symptoms has been established by several studies [66-68]. Symptom intensity peaks in the hours around midnight and is lowest in the late morning hours.

In addition to these mandatory criteria, the IRLSSG also agreed on three supportive clinical features. These are observed in a large proportion or RLS patients and their presence or absence can help to establish the diagnosis in arguable cases [64].

The first supportive feature is a positive family history of the patient, which is defined as having at least one first-degree relative affected with RLS. The second supportive feature is the positive response to dopaminergic treatment. The dopamine precursor levodopa (L-DOPA) and a variety of dopamine-receptor agonists are the first line treatment option for RLS [64, 69]. The third supportive feature is the presence of periodic limb movements (PLM) in sleep (PLMS) or while awake (PLMW). PLM are involuntary stereotypic jerks of the legs characterised by a dorsiflexion of the ankle and the big toe, sometimes also with a flexion of the knee and the hip. A PLMS index (number of PLMS per hour of sleep) > 5 is generally considered as pathological [70]. More than 80% of RLS patients display pathological PLMS [71], but their occurrence is not limited to RLS. An elevated PLMS index is also observed in other diseases such as narcolepsy or attention-deficit hyperactivity syndrome, and in healthy elderly individuals [72-74].

Consequences of RLS are sleep disruption, depression, anxiety, and generally reduced mental and physical health [75, 76]. Moreover, an increased risk for cardiovascular disorders and hypertension in RLS patients has been reported [77, 78]. The age of onset of the disorder shows high variability, ranging from early childhood to old age [79-81]. The frequency and severity of the symptoms also vary widely, from only mild occasional to daily severe symptoms [64].

RLS is a very common disorder in Caucasian populations of Europe and North America with an overall age-dependent prevalence of approximately 10% [82-90]. It is less common in South East Europe with on average 3.5% prevalence [91-93] and even more so in Asian populations [94-99]. Here the prevalence is 1 to 2% with the exception of one study in Korea

where a prevalence of 3.9% was found [100]. These estimates refer to RLS of any symptom frequency and severity. Some studies assessed the severity of the symptoms and found a prevalence of roughly 2% for severe and therefore clinically relevant RLS in Caucasian populations [82, 85, 86, 88]. Ethnicity seems to play a role in RLS frequency but this is still a matter of debate. Although a reduced prevalence is observed in Asian and East European countries, the methodological differences in study design and case classification or the varying sociocultural background could cause this difference [75]. Only one study assessed the prevalence of RLS in people of African descent and found similar prevalences in African-Americans and Caucasian Americans [86].

Both female gender and age are consistently reported as risk factors for RLS. Women are affected approximately twice as often as men and prevalence was found to increase in an age-dependent manner in the majority of studies, with a 2-3 fold increased prevalence in the age group 60-69 years compared to the age group 20-29 years [75, 82-84, 88-91, 93, 100].

1.2.2 Classification and endophenotypes

1.2.2.1 Idiopathic (primary) and symptomatic (secondary) RLS

RLS can be divided into two forms according to the underlying etiology. Idiopathic or primary RLS (iRLS) is a stand-alone disease with no obvious other cause. It accounts for the majority of RLS cases with clinical studies reporting 70 – 87% of cases as idiopathic [65, 101]. Patients display no further clinical symptoms except the RLS symptoms and their physical, neurological, neurophysiological and laboratory examinations are normal. In contrast, symptomatic or secondary RLS (sRLS) develops as a consequence of other medical conditions present in the patient. Both forms show the same core features [63, 80, 102]. The most well-established causes of symptomatic RLS are iron deficiency, pregnancy, and end-stage renal disease [103]. The relationship between RLS and iron deficiency has been documented already in the first clinical studies of RLS where 24% of patients with iron-deficiency suffered from RLS [104]. Several studies have demonstrated that administration of iron can lead to complete abolishment of RLS symptoms in the patients, confirming the link between RLS and iron deficiency [94, 105, 106]. The prevalence of RLS in pregnant women is estimated at 19 to 26% [107-110]. In most cases, the symptoms develop in the third trimester and usually resolve within a few months after delivery [109, 110]. The best-characterized symptomatic form of RLS is uremic RLS (uRLS) in end-stage renal disease (ESRD). The prevalence has been studied in numerous ESRD patient populations in different

countries using the IRLSSG criteria and although a wide range of estimates was found (11 - 70%), the prevalence was consistently higher than in the general population [103, 111-117]. The patient populations under study included both patients on hemodialysis and peritoneal dialysis and did not find any differences in RLS prevalence with regard to the type of dialysis [112, 115]. Moreover, a large number of demographic, clinical and biochemical parameters have been compared between RLS-positive and RLS-negative ESRD patients but to date no consistent association has been found. Increasing duration of dependence on dialysis, frequency of dialysis sessions, phosphate metabolism, and anemia were reported as associated to RLS in ESRD patients in some studies but not in others [111, 112, 114, 117-120]. RLS is a severe complaint in ESRD and is associated with an increased mortality [121, 122]. Studies of the clinical presentation found subtle differences between uRLS and iRLS which suggest an accentuation of the motor component and a higher severity of RLS in dialysis patients [123, 124].

In addition to these established symptomatic forms, RLS has also been described in Parkinson disease, type 2 diabetes mellitus, multiple sclerosis, attention-deficit hyperactivity syndrome, neuropathies, and incidentally also in other diseases. However, the evidence for these being symptomatic forms of RLS is still inconclusive, since some studies report increased prevalences compared to the general population whereas others do not [103, 125, 126].

1.2.2.2 Familial (early-onset) and sporadic (late-onset) endophenotypes

Currently, two complementary classification schemes are used for defining endophenotypes in RLS. One approach is to classify patients according to the presence or absence of a positive family history. These subtypes of RLS are referred to as **familial** (**hereditary) RLS** and **sporadic RLS**. The other classification system uses age-of-onset of symptoms to group patients into **early-onset** and **late-onset RLS** [127].

A large proportion of idiopathic cases have a positive family history, with most studies reporting estimates between 40 and 65% [71, 79, 80, 84, 88, 101]. Familial RLS is also found in symptomatic cases but to a much lower extent than in idiopathic RLS (e.g. 42.3% in iRLS vs. 11.7% in sRLS in [80].

Patients with familial RLS consistently have been shown to have an earlier age of onset of the disease and a chronic-progressive disease course compared to sporadic RLS [71, 80, 81, 101, 102, 127, 128]. However, the exact cut-off value between early-onset and late-onset is still a matter of debate, since it varied between 30 and 45 years in studies [80, 101]. Apart from the

course of the disease and the family history, there seem to be no further differences in clinical characteristics among the two subtypes [65, 71, 80].

1.2.3 Pathophysiology

At present, the exact pathophysiological mechanisms of RLS and the involved neuronal systems or anatomical structures are not known. However, there is accumulating evidence for dysfunctions in subcortical areas of the central nervous system (CNS), possibly due to impairments of dopaminergic systems and brain iron metabolism. So far, there is no evidence for any neurodegenerative processes [129-131]. Neuroimaging studies have reported an increased cortical excitability in RLS patients [132-136]. This could reflect a dysfunction in subcortical structures leading to reduced supraspinal inhibition affecting both the cortex and the spinal cord [127]. Corresponding results were obtained in reflex studies which found spinal cord hyperexcitability in RLS patients [137-139]. Altered temperature perception and increased sensitivity to certain pain stimuli in RLS patients indicate also alterations in somatosensory pathways [140, 141].

The powerful therapeutic effect of dopamine and dopamine agonists in RLS suggests dysfunctions in dopaminergic systems of the CNS as a cause of RLS symptoms. The involvement of subcortical dopaminergic systems in the regulation of motor control and sensory perception further supports this idea since a dysfunction in these pathways could explain both the motor and sensory components of RLS [142]. Moreover, the activity of the dopaminergic system follows a circadian pattern, with its nadir at night [143, 144]. This correlates well with the diurnal variation of RLS symptoms. Two of the various dopaminergic pathways in the CNS are especially interesting for RLS: the nigrostriatal (A9) system, which controls generation of voluntary movement and the diencephalospinal (A11) system, the main source of dopamine in the spinal cord [142, 145]. A11 neurons act inhibitory on both afferent sensory neurons and preganglionic sympathetic neurons and are thought to be involved in pain modulation and the control of autonomic and motor functions [146, 147]. In addition, the A11 cell bodies are located in close vicinity to the suprachiasmatic nucleus, the main control center for circadian rhythms [148, 149]. The inhibitory effect of the A11 neurons in the control of motor and sensory pathways in the spinal cord and their proximity to the suprachiasmatic nucleus make this dopaminergic system an interesting candidate for RLS, since dysfunction of the A11 neurons could account for the spinal hyperexcitability found in RLS patients [145]. However, biological evidence for structural or functional changes of the dopaminergic systems in RLS is still limited. Neuroimaging studies of the pre- and

postsynaptic dopaminergic status in the nigrostriatal system are inconsistent but suggest an impairment and possible hypoactivity of this system in RLS [127, 142, 150-155]. The A11 neurons have been studied mainly in animal models with lesions of the A11 neurons. These animals showed increased activity and locomotion which could be reduced using dopamine agonists and increased by iron deficiency. They also showed a decrease of dopamine in the spinal cord and of iron stores in the brain [149, 156, 157]. Although it reproduces some aspects of RLS, this model has some limitations. It is argued that the lesioning was not selective for A11 and could have included other diencephalic dopaminergic neurons and also the nigrostriatal system [148]. Moreover, the increased activity was measured during the active phase of the animals, at a time which is not compatible with the circadian pattern of RLS [158]. A recent study of the A11 system in post-mortem brain of RLS patients also failed to detect any signs of cell loss or neurodegenerative processes [159].

Altered brain iron metabolism is also proposed as a cause of RLS. Iron deficiency, pregnancy, and end-stage renal disease are the most common causes of symptomatic RLS and all involve a deranged iron status. Treatment with oral or intravenous iron has been shown to ameliorate symptoms [94]. These clinical observations of reduced availability of iron in RLS are supported by findings from neuroimaging studies and analyses of iron and proteins involved in storage and transport of iron in blood and cerebrospinal fluid (CSF). MRI studies showed reduced iron content in the substantia nigra [160, 161]. Furthermore, CSF studies revealed a decrease in ferritin, which is the main iron storage molecule, and an increase in transferrin, the main iron transport molecule, in RLS patients [162, 163]. The mechanisms underlying the reduced iron content are not known. One hypothesis is reduced accessibility of iron within the substantia nigra due to abnormal storage of iron in astrocytes instead of oligodendrocytes. Another possibility is an impaired acquisition of iron by the dopaminergic cells of the substantia nigra [94]. These cells showed reduced expression of the transferrin receptor, which is necessary for ferritin uptake into the cell, in post-mortem RLS brain [130]. It is difficult to reconcile the abnormalities found in the iron and dopamine systems to form a common pathway for RLS. Iron status and dopamine synthesis are linked through the enzyme tyrosine hydroxylase (TH), which catalyzes the essential step in the synthesis of dopamine and requires iron as a cofactor. Iron deficiency might therefore reduce the activity and lead to the observed hypofunction of the dopaminergic system [142]. Animal models of iron deficiency partly reproduce the dopaminergic abnormalities found in RLS patients but also show an increase in extracellular dopamine and intracellular TH [94]. A recent study has

found increased TH in the substantia nigra of RLS patients and postulated an overly activated dopaminergic system in RLS caused by a brain iron deficiency [164].

1.2.4 Genetic epidemiology

A substantial body of evidence for a genetic contribution to RLS has been accumulated. Foremost, the familial aggregation of RLS with 40 to 60% of RLS patients reporting further affected first-degree relatives is a strong indication of a genetic predisposition to disease (see chapter 1.2.2.2). This is confirmed by heritability estimates of 54 to 69.4% from twin studies and familial aggregation analysis [165-167]. Furthermore, the risk of RLS in relatives of affected individuals is significantly higher than the risk for relatives of healthy controls. One study found an almost sixfold increase in risk for first-degree relatives of RLS patients [168]. Another study analysed specific relative pairs and found a tenfold increase for parent-offspring pairs and a 16-fold increase for siblings [167]. Two studies have examined monozygotic (MZ) and dizygotic (DZ) twins for RLS and consistently found higher concordance rates in MZ twins (61% and 53,7%, respectively) compared to DZ twins (45% and 15.4%, respectively), which also supports a substantial genetic influence in RLS [165, 166].

Regarding the mode of inheritance of RLS, early pedigree studies of single families in RLS suggested an autosomal-dominant mode of inheritance with high penetrance and broad variation in the expressivity. Age of onset, disease course, frequency and severity of symptoms, and the presence of either both or predominantly sensory or motor symptoms were found to vary considerably within and between families [169-172]. Complex segregation analyses in families confirmed the autosomal-dominant mode of inheritance but indicated a multifactorial component and genetic heterogeneity in RLS with additional genetic and non-genetic influences on the phenotype [173, 174].

Linkage studies in RLS-affected families have identified six linkage regions to date (Table 1). All but the RLS-1 locus were found assuming an autosomal-dominant mode of inheritance. For RLS-1 to RLS-3 several independent studies already confirmed the linkage signal (Table 1). In addition, suggestive evidence for linkage was found in three further genomic regions, chromosome 4q25-26, 17p11-13 [175] and 19p13 [176], also based on an autosomal-dominant mode of inheritance. However, these linkages can only explain a minority of all RLS cases in the population and might represent rare Mendelian forms of RLS. There are several RLS families with non-Mendelian ratios, i.e. more than 50% of the offspring are affected, and for numerous families linkage to the known loci was excluded and no other

linkage loci were identified [177-180]. The high frequency of RLS in the population, the heritability estimates, the observed inter- and intrafamilial variation in the clinical presentation of symptoms, and the existence of sporadic and symptomatic forms all support a complex genetic basis of RLS.

Table 1: Published RLS linkage regions.

Locus (OMIM)	Chromosomal location	Inheritance mode	Original study	Confirmatory studies
RLS-1	12q22-23.3	AR (pseudodominant)	[181]	[177], [180], [182]
RLS-2	14q13-22	AD	[183]	[179], [184]
RLS-3	9p24-22	AD	[167]	[185], [184], [186]
RLS-4	2q33	AD	[187]	-
RLS-5	20p13	AD	[188]	-
*	16p12.1	AD	[189]	-

The linkage regions for RLS are given with their OMIM identifiers [190], the original study and any confirmatory studies, the chromosomal location by chromosome band and the proposed inheritance mode; AR, autosomal – recessive; AD, autosomal – dominant. *, no identifier assigned.

Candidate gene studies for RLS have been largely unsuccessful so far. Sequencing of exons and splice sites of several genes located in linkage regions (RLS-1, RLS-3, 16p12.1, and 19p13) did not identify any causative mutations [167, 176, 186, 189, 191-193]. Only a few case-control association studies have been performed to date and these have been limited to candidate genes and individual candidate polymorphisms. One study investigated known functional polymorphisms in eight genes which are involved in dopaminergic transmission and metabolism, *DRD1-5*, *DAT*, *TH* and *DBH*, but did not detect any association [194]. Another study investigating SNPs in the *DMT1* gene, which encodes an important brain iron transporter, also did not show any association [192]. The only study to link a gene to RLS has been an association study of the monoamine oxidase isoenzymes MAOA and MAOB which are involved in the degradation of dopamine and other neuroactive amines [195]. The authors found an association of the high activity allele of a functional variable number of tandem repeat (VNTR) polymorphism in the promoter region of *MAOA* to RLS only in females. However, this association has to be considered preliminary because it still lacks replication in further independent populations, one of the prerequisites for discerning true and spurious associations [61].

1.3 Aims of this thesis

Genetic analysis of RLS has confirmed the significant genetic contribution to the phenotype and identified linked regions on several chromosomes. The hitherto used methods for disease gene mapping in RLS were genome-wide linkage studies and association studies or sequencing of candidate genes based on their biological function or their position within a linkage region. However, none of these approaches has led to the identification of causally related genes or variants yet. Recent progress in cataloguing common SNP variation in the human genome, deciphering the haplotype block structure of the genome, and in high-throughput genotyping of SNPs have enabled the conduction of large-scale association studies of candidate regions or even the whole genome.

The primary aim of this thesis was to conduct such large-scale case-control association studies in idiopathic RLS in order to find common susceptibility loci with moderate effect sizes, which could not be detected by the mapping methods used so far. Based on the CDCV hypothesis, there should be several of this kind of loci for RLS since it is a common disease with a prevalence of 10% in the general population. We conducted two types of association studies:

a) Regional studies of previously identified linkage regions:

Here we focused on the two most robust linkage regions identified to date, RLS-1 on chromosome 12, and RLS-3 on chromosome 9 in separate multi-stage studies using different mapping strategies. For RLS-1 a gene-centered design was chosen, whereas for RLS-3 the whole linkage region was analyzed based on the coverage given by the Affymetrix 500K array.

b) Genome-wide association study:

In a completely hypothesis-free approach we conducted a genome-wide association study of idiopathic RLS compared to controls drawn from the general population. A two-stage design was used with a genome-wide scan based on the Affymetrix 500K array in stage one, followed by a replication stage including only the most significantly associated SNPs from stage one.

The secondary aim of the thesis was to investigate the possibility of a common genetic basis in idiopathic and symptomatic RLS. To this end we conducted a case-control association study of the iRLS-associated variants identified in our studies in end-stage renal disease patients with and without RLS symptoms.

2 Results and Discussion

2.1 Idiopathic RLS

2.1.1 Association of variants in *NOS1 (nNOS)* on chromosome 12 (RLS-1)

Published manuscript 1 (Appendix 1)

2.1.1.1 Study rationale and design

The RLS-1 locus on chromosome 12 has been confirmed in several independent families since its initial description but the analysis of candidate genes in this region has been limited and unsuccessful so far [177, 180-182, 191-193].

We screened this region in an association study using a three-stage design. In the explorative phase (stage 1) we genotyped tagging SNPs as well as nonsynonymous and synonymous coding and splice-site SNPs in all known transcripts of the RLS1 region in a Caucasian RLS case-control sample using the GoldenGate Genotyping Assay technology (Illumina). In the replication phase (stage 2) the most significant SNPs from stage 1 were then genotyped in an independent Caucasian RLS case-control sample using the iPLEX genotyping technology (Sequenom). In the third stage we conducted a finemapping of *NOS1* (*nNOS*) which was the only gene with significant association identified in stage 2.

All RLS cases were diagnosed according to the IRLSSG criteria [64]. Symptomatic cases were excluded. Size and demographic data of both case samples are given in Table 2. Age and sex matched controls of stage 1 were recruited for the absence of psychiatric phenotypes (depression and anxiety) but also screened negatively for RLS symptoms. Controls of stage 2 were age and sex matched population-based controls from the KORA (S4) study and had not been screened for RLS symptoms [196].

Table 2: Demographic data and sample size for RLS cases of stage 1 and 2

	Stage 1 (exploratory sample)	Stage 2 (replication sample)
N individuals	367	551
N females	263	390
Mean age (SD)	57.44 (9.66)	61.03 (10.59)
Mean age at onset (SD)	35.58 (15.73)	41.51 (18.53)
Positive family history (%)	80.38	57.35

Mean age and age at onset are given in years. N, number; SD, standard deviation.

2.1.1.2 Association analysis results and finemapping of *NOS1*

Stage 1: Explorative Case-Control Study

Statistical analysis by both Armitage trend and genotypic test revealed 79 SNPs reaching a significance level of $P_{nominal} < 0.05$ (Supplementary Table 2a of Appendix 1). These showed a prominent clustering in three chromosomal regions: 17 SNP within positions 99,792,852 to 101,016,651bp, 10 SNPs within 107,084,765 to 108,133,377bp, and 16 SNPs within 115,621,107 to 116,597,588bp of chromosome 12 (Figure 1 in Appendix 1). The most significant signals were obtained in *NOS1* with the synonymous coding SNP rs2293054 ($P_{nominal(Armitage)} = 0.0005$) and the intronic SNP rs6490121 ($P_{nominal\ (Armitage)} = 0.0021$).

Stage 2: Replication Case-Control Study

A total of 24 SNPs with a $P_{nominal} < 0.01505$ of stage 1 were genotyped in the replication case-control sample. Only SNP rs7977109 in intron 3 of the *NOS1* gene reached the level of significance before and after correction for multiple testing using the Westfall-Young method with $P_{nominal\ (Armitage)} = 0.00175$ and $P_{Westfall\text{-}Young} = 0.04895$. The other SNPs did not reach significance in the replication stage (Supplementary Table 2b of Appendix 1).

Stage 3: High-Density Mapping and mutation screening of *NOS1*

For the finemapping of the association signal, the *NOS1* gene and 10kb of flanking sequence were covered with tagging SNPs. Five SNPs in *NOS1* with a $P_{nominal} < 0.05$ in stage 1 and known nonsynonymous coding and splice site SNPs (dbSNP build 125) were also included. Two SNPs had nominally significant P-values in both samples (rs7977109, rs693534), whereas eight and four SNPs showed significance only in the explorative or replication sample, respectively (Figure 1). After correction for multiple testing three SNPs were significant in the explorative (rs4766836, rs2293054, rs6490121) and the replication (rs7977109, rs530393, rs816292) sample, respectively (Supplementary Table 3 of Appendix 1). Sequencing of all coding exons and splice sites of *NOS1* in 23 RLS subjects from the replication sample revealed no causative mutations.

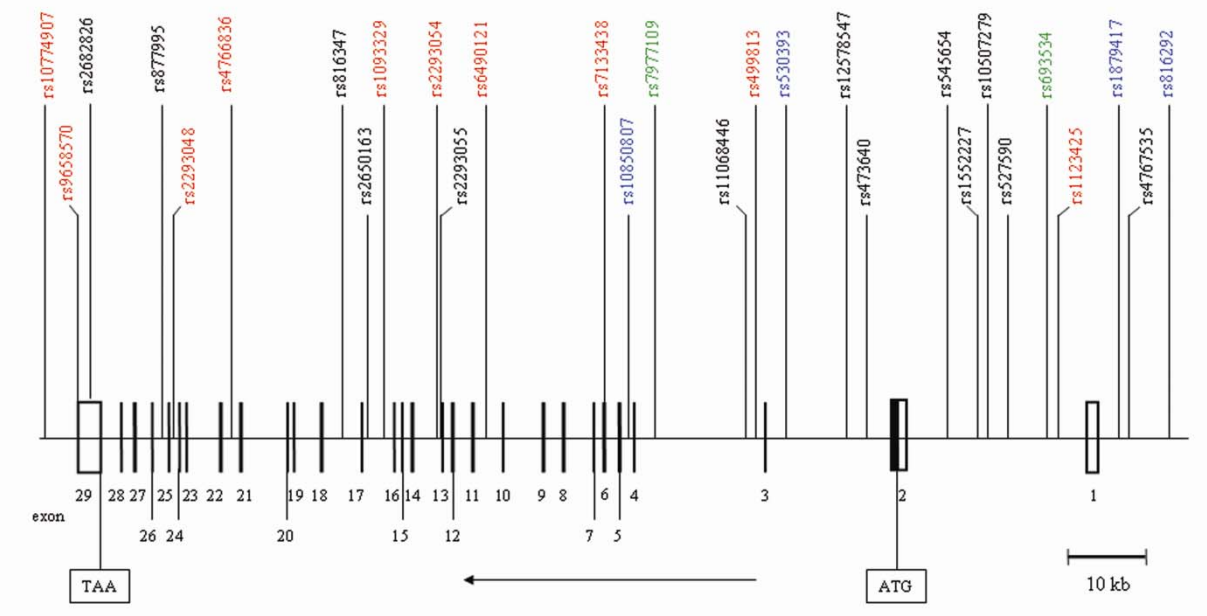


Figure 1: Genomic organization of the *NOS1* gene and positions of the SNPs in the high-density mapping of *NOS1*. White boxes denote untranslated exons or 5-UTR and 3-UTR. Black boxes denote coding sequences. The start codon (ATG) and the stop codon (TAA) are shown. The arrow specifies the direction of transcription on the reverse strand of chromosome 12. SNPs are marked with the following color code: black, no association result; red, association result in the explorative study; blue, association result in the replication study; green, association result in the explorative and replication study. An association result is defined as $P_{nominal} < 0.05$ in the genotypic test or/and Armitage test.

2.1.1.3 Discussion

This was the first systematic investigation of common polymorphisms in a known RLS linkage locus (RLS-1). We used a gene-based approach, which means that other putative functional elements located upstream or downstream of genes could have been missed unless they were in strong LD with SNP markers within genes. Of the initial 37 genes with significant association in stage 1, only *NOS1* was replicated in stage 2. Subsequent high-density analysis of this gene revealed several associated SNPs in both the explorative and the replication sample. However, these SNPs were distributed over the entire gene. The exact location of the associated SNPs varied between the two study samples (Figure 1). Moreover two SNPs (rs7977109 and rs693534) were significantly associated in the explorative and replication sample but had different allele frequencies and opposite directions of the association. This implies that the same allele is a risk allele in one but a protective allele in the other sample and is also referred to as flip-flop phenomenon [197]. Similar observations have been described in schizophrenia, another complex phenotype. Here discrepancies of the location of the associated SNPs within *DBNDD1* were found between the original and subsequent studies [198]. Furthermore, opposite alleles of variants in the *COMT* gene were

found to be associated in different studies [197]. Differences in the LD structure or the allele frequencies between individual study populations or population stratification were suggested as possible sources of such a complex pattern of association signals. Although these factors can lead to false-positive associations, the observed associations can also be genuine. The causal variant may be associated with different marker alleles in the different populations, or there could be allelic heterogeneity with different risk alleles in the different populations [198]. Moreover, the risk to develop a complex disease is not influenced by only one risk locus but by several and possibly additional environmental influences. Theoretical modeling has shown that such multilocus effects can obscure the association signal in single-marker association studies [197]. In our study samples there is potential for the abovementioned confounding factors. We conducted the explorative study with "extreme phenotypes" in both cases and controls. The cases of the explorative sample were specifically enriched for the familial endophenotype of RLS (early-onset of symptoms, positive family history), whereas the replication sample was a heterogeneous mix of familial and sporadic RLS cases. The controls of the explorative sample were recruited for the absence of any symptoms of RLS. In contrast, the controls of the replication sample were population-based controls and not scrutinized for possible symptoms of RLS [196]. Despite these differences between the study populations, the LD structure in *NOS1* is similar for cases and controls (Table 1 in Appendix 1 and Figure 2). We checked for population stratification by the method of genomic control [59, 199] and found inflation factors below or close to 1, indicating no stratification in our populations (Table 2 in Appendix 1). Finally, we cannot exclude the possibility that the association signals are spurious and arose due to chance. Therefore, the findings need to be confirmed in further studies in independent populations. Our own genome-wide association study (see the following chapter) did not reveal any significant associations in the *NOS1* region with the lowest nominal P value being 0.021. However, none of the associated SNPs from our study were genotyped in the GWAS and the LD between the associated SNPs of the *NOS1* study and the SNPs in the GWAS is not strong enough to use any of these as a surrogate marker. In conclusion, the GWAS data does not give any further information regarding the genuineness of the *NOS1* association.

Nevertheless the examples from schizophrenia clearly demonstrate that complex association patterns should not be discarded from further analysis as they may very well point to a true association [197, 198]. Looking at the underlying LD structure of the genomic region around *NOS1* using HapMap data (Rel. 24/phase II, Nov08) showed a relatively well defined LD block containing the *NOS1* gene.

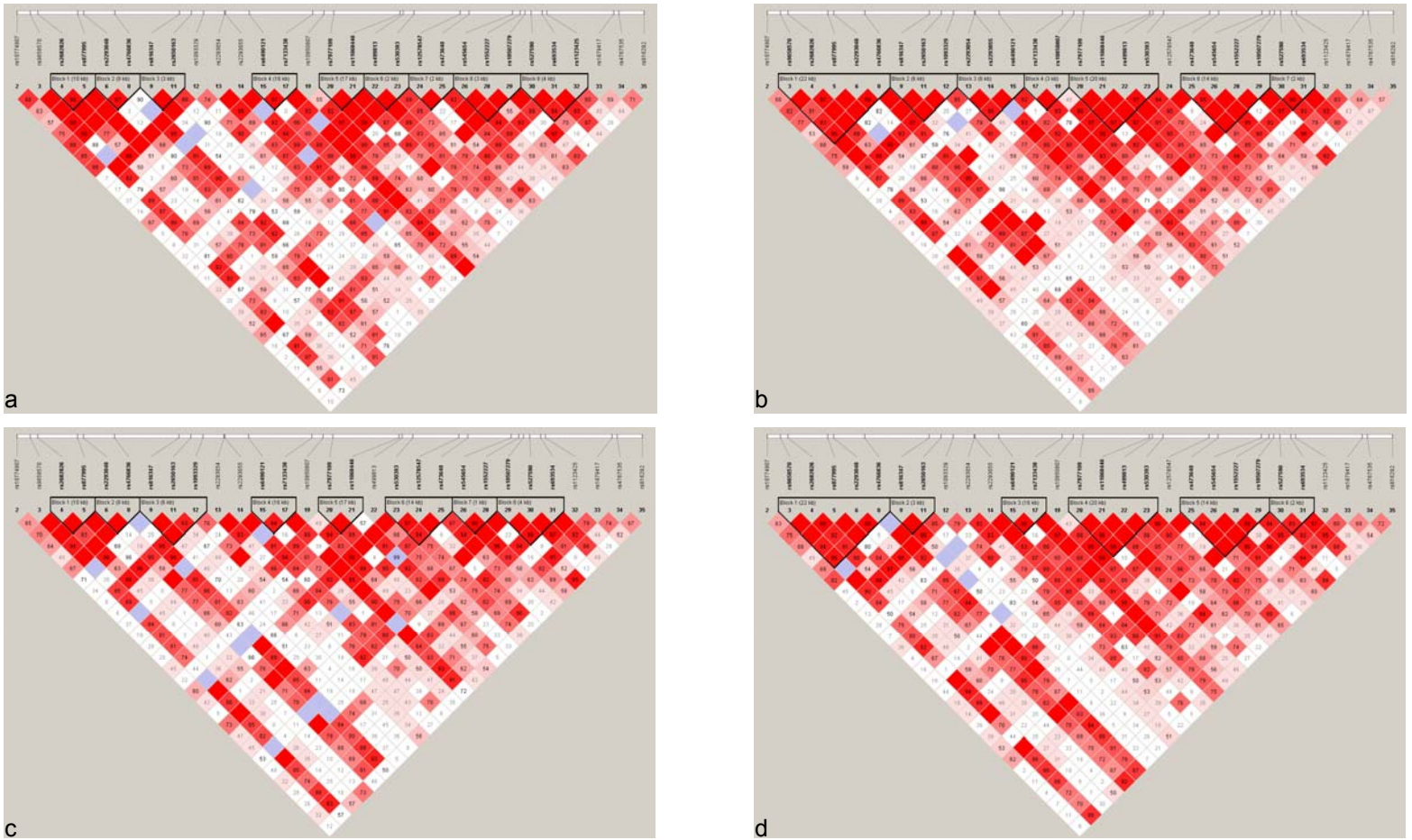


Figure 2: LD structure of the different study populations. Shown as D' values of 29 SNPs. Pairwise LD measured as D' was calculated for 29 SNPs from the stage 3 high-density mapping in both populations using the method of Gabriel et al. [34] as implemented in Haploview [200]. Shading represents the magnitude of LD, with a white-to-red gradient reflecting lower to higher values.

a) LD structure in cases of stage 1
b) LD structure in cases of stage 2
c) LD structure in controls of stage 1
d) LD structure in controls of stage 2

All associated SNPs mapped to this LD block and the closest neighboring genes *FBXO21* and *KSR2* were not part of this block. It is therefore very likely that the true disease-associated variant resides within the *NOS1* gene or possibly also regulatory elements contained in the same LD block. However, the true causal variant or variants still remain to be identified.

NOS1 is an interesting candidate for RLS since it is implicated in several pathways which are relevant for the etiology of RLS. This enzyme catalyzes the synthesis of nitric oxide (NO), an intercellular messenger or an "atypical neurotransmitter" in the central nervous system [201-203]. NOS1 action in the CNS has been associated with pain perception as well as the control of sleep wake regulation and the NO-arginine pathway participates in the modulation of the dopaminergic transmission [201, 203-209]. An interplay of the opioidergic system and NO could also be relevant in the pathophysiology of RLS. The inhibition of NOS1 enhanced the morphine-induced antinociception at the spinal cord level [210] and a positive effect of opioidergic substances on RLS symptoms is well known [211]. The neuroanatomical level of a possible interplay of the nitridergic and dopaminergic or opioidergic neurotransmission, in connection to RLS, however, is not known.

2.1.2 Association of variants in *MEIS1*, *BTBD9*, and *MAP2K5/LBXCOR1* – results from a genome-wide association study

Published manuscript 2 (Appendix 2)

2.1.2.1 Study rationale and design

We conducted a GWAS in RLS cases and a large control cohort from the general population in order to identify RLS susceptibility variants in a completely hypothesis-free and unbiased fashion. The study design involved an exploratory stage (stage 1), followed by replication in two further case/control samples (stages 2a and 2b, Figure 1 in Appendix 2). Stage 1 was the GWAS experiment where we genotyped 401 cases and 1,644 controls on the Affymetrix 500K Array Set. SNPs for replication in stages 2a (903 cases / 891 controls) and 2b (255 / 287) were chosen based on a nominal P value $\leq 10^{-5}$ in stage 1 and supplemented with neighboring SNPs based on the LD structure to increase the coverage of the regions (Supplementary Table 3 in Appendix 2). Calculations showed a power of > 90% to detect variants with an odds ratio > 1.5 in the combined stage 1 and 2a samples. All regions, which were significantly associated to RLS in the replication and in the joint analysis of both stages, were scrutinized for additional association signals and possible causal variants in a finemapping step in the stage 2a sample. Here we genotyped tagging SNPs for all genes and 10kb of flanking sequence in these regions and also all coding-sequence and splice-site SNPs (according to dbSNP build 127 and Ensemble release 44) using the iPLEX genotyping technology (Sequenom).

All cases were diagnosed in face-to-face interviews according to the IRLSSG criteria [64] and symptomatic cases were excluded from the study. Cases and controls of stage 1 and 2a were of European descent and recruited in Germany, whereas the stage 2b samples were of French-Canadian ancestry. Demographic data of the study populations can be found in Supplementary Table 8 of Appendix 2.

2.1.2.2 Association results of the genome-wide exploratory stage

The application of a stringent quality control protocol resulted in only 236,758 SNPs of the initial 500,568 SNPs for statistical analysis. Armitage trend test with age and sex as covariates revealed 13 SNPs with nominal P values $\leq 10^{-5}$ mapping to six genomic regions (Figure 3). After correction for multiple testing (Bonferroni correction for 236,758 tests) a single SNP located in the *MEIS1* gene reached genome-wide significance (rs2300478, $P_{corrected} < 0.0002$).

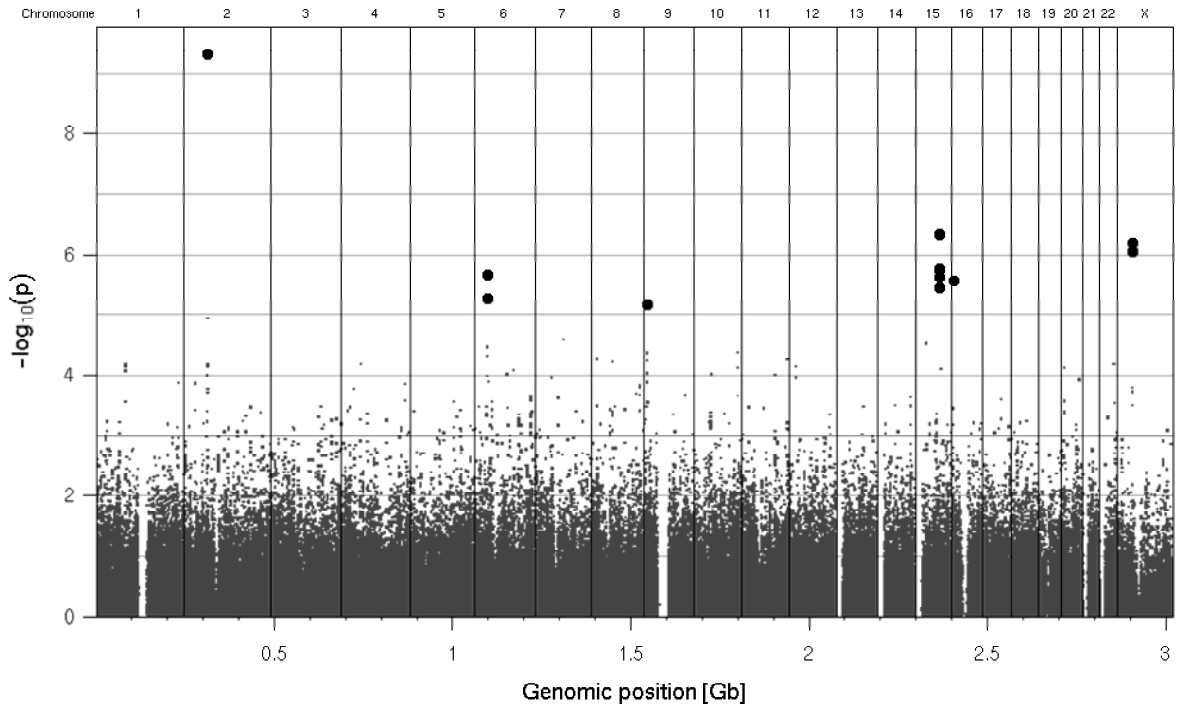


Figure 3. Genome-wide association study for RLS susceptibility loci. The analysis compared 393 sucessfully genotyped RLS cases with 1,602 population-based KORA controls. The x-axis is genomic position and the y-axis is log10(P). Thirteen SNPs that passed inclusion criteria for the replication study of stage 2 are highlighted in bold. Note that the P values of three SNPs on chromosome 15 are very similar and these SNPs appear as one single dot.

2.1.2.3 Replication, finemapping, and haplotype analysis

Three of the six candidate regions from the genome-wide stage were confirmed in both replication samples 2a and 2b. They were nominally significant in the single samples and remained significant after correction for multiple testing according to Westfall and Young ($P_{corrected} < 0.05$). All SNPs reached genome-wide significance in the joint analysis of all stages after correction of multiple testing according to Bonferroni (Table 3). The first region was on chromosome 2p located in a 32 kb LD block containing exon 9 of *MEIS1*. The second region with significant association was on chromsome 6p within a 113 kb LD block in intron 5 of the *BTBD9* gene. The third associated region on chromsome 15q contained a 48 kb LD block overlapping the 3' end of the *MAP2K5* gene and the adjacent *LBXCOR1* gene. The regions on chromosome 9 and 16 were only nominally significant in the stage 2a sample but did not reach significance after correction for multiple testing. The region on the X chromosome did not reach nominal significance in both replication samples.

Table 3: SNPs with significant association which were successfully genotyped in all three case/control samples, located in three different genomic regions.

dbSNP ID	Chr	Genome position	Gene	OR (95 % CI)	Stage 2a P_{nom}	Stage 2b P_{nom}	Stage 1&2a&2b $P_{corrected}$ (B)
rs2300478	2p	66,634,957	MEIS1	1.74 (1.57-1.92)	$5.93x10^{-12}$	$1.77x10^{-4}$	$8.08x10^{-23}$
rs9296249	6p	38,473,819	BTBD9	1.67 (1.49-1.89)	$1.61x10^{-6}$	$2.891x0^{-3}$	$9.44x10^{-13}$
rs9357271	6p	38,473,851	BTBD9	1.66 (1.48-1.87)	$1.85x10^{-6}$	$1.45x10^{-3}$	$1.50x10^{-12}$
rs12593813	15q	65,823,906	MAP2K5	1.50 (1.36-1.66)	$4.95x10^{-5}$	$7.91x10^{-3}$	$2.51x10^{-10}$
rs11635424	15q	65,824,632	MAP2K5	1.51 (1.37-1.67)	$2.54x10^{-5}$	$3.78x10^{-3}$	$8.64x10^{-11}$
rs4489954	15q	65,859,129	MAP2K5	1.51 (1.36-1.67)	$2.60x10^{-5}$	$1.63x10^{-2}$	$6.35x10^{-10}$
rs3784709	15q	65,859,329	MAP2K5	1.52 (1.37-1.68)	$7.46x10^{-5}$	$5.78x10^{-4}$	$9.61x10^{-11}$
rs1026732	15q	65,882,139	MAP2K5	1.53 (1.39-1.70)	$2.78x10^{-5}$	$2.94x10^{-3}$	$1.44x10^{-11}$
rs6494696	15q	65,890,260	[MAP2K5/ LBXCOR1]	1.52 (1.38-1.69)	$5.20x10^{-5}$	$2.94x10^{-3}$	$4.74x10^{-11}$

Genome positions refer to the Human March 2006 (hg18) assembly. [MAP2K5/LBXCOR1] denotes intergenic position of SNP. OR, Odds ratio; CI, confidence interval; P_{nom}, nominal P value. $P_{corrected}$, P value corrected for multiple testing using Bonferroni's method correcting for 236,758 SNPs (B).

The subsequent finemapping in the stage 2a samples was aimed at pinning down the localization of the association signals and assessing known functional and therefore possibly causal SNPs for an association. We included the known non-synonymous and synonymous coding SNPs, splice site and frameshift SNPs, but these were either monomorphic or had a MAF < 1% in our sample and thus could not be analyzed for association. The finemapping confirmed the position and size of the original candidate regions as defined in stage 1 of the study. Moreover, the association signal was found to be clearly restricted to the respective LD blocks of the regions (Figure 4).

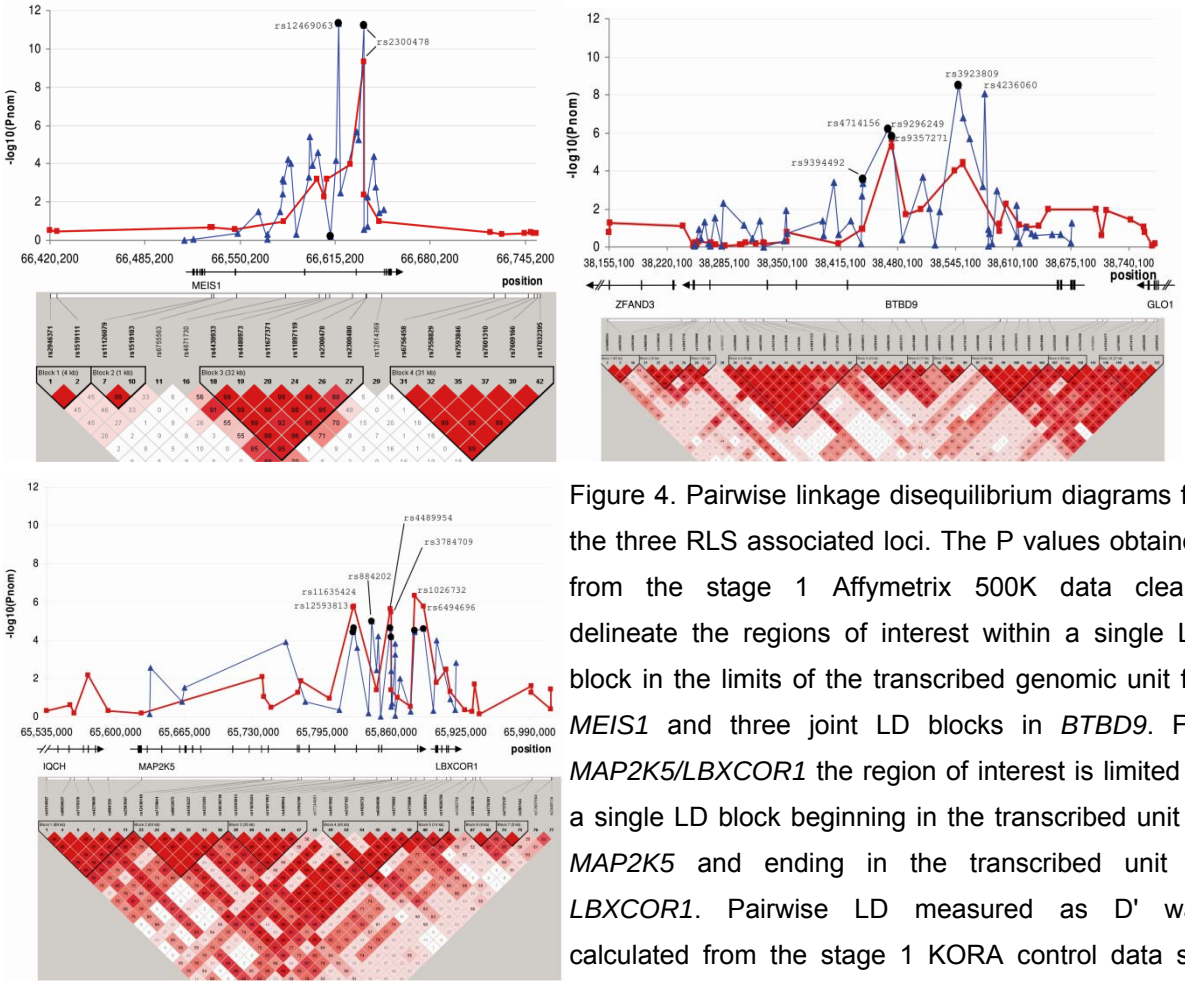


Figure 4. Pairwise linkage disequilibrium diagrams for the three RLS associated loci. The P values obtained from the stage 1 Affymetrix 500K data clearly delineate the regions of interest within a single LD block in the limits of the transcribed genomic unit for *MEIS1* and three joint LD blocks in *BTBD9*. For *MAP2K5/LBXCOR1* the region of interest is limited to a single LD block beginning in the transcribed unit of *MAP2K5* and ending in the transcribed unit of *LBXCOR1*. Pairwise LD measured as D' was calculated from the stage 1 KORA control data set using the method developed by Gabriel et al. [34] as implemented in Haploview [200]. Shading represents magnitude and significance of pairwise LD, with a white-to-red gradient reflecting lower to higher LD values. Stage 1 Affymetrix SNPs are indicated by red squares, replication SNPs by black circles and finemapping SNPs by blue triangles. X axis = $-\log_{10}(P)$. Transcriptional units are indicated by black arrows, with exons depicted as black bars. **a**, *MEIS1*; **b**, *BTBD9*; **c**, *MAP2K5*/*LBXCOR1*.

In addition to the single marker tests, we also performed a haplotype analysis of the associated regions in the stage 2 samples. For *MEIS1* this approach yielded a more significant association signal (Figure 5). A haplotype block delineated by rs3890755 to rs12469063 was more strongly associated than any single SNP (stage 2a $P_{nominal} = 5.87 \times 10^{-20}$; stage 2b $P_{nominal} = 8.51 \times 10^{-7}$). This haplotype was completely described by two SNPs: allele A of rs6710341 and allele G of rs12469063. For *BTBD9* and *MAP2K5*/*LBXCOR1* haplotype analysis confirmed the results of single SNP analysis.

The identified variants all confer a small to moderate increase in risk with ORs ranging between 1.5 and 1.7 (Table 3). The "AG" haplotype within *MEIS1* is by far the highest risk factor for developing RLS in our study with an OR of up to 2.75.

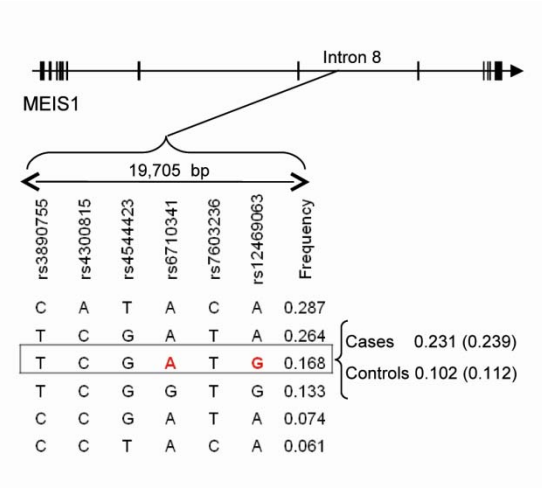


Figure 5. Haplotype structure for *MEIS1*. A haplotype consisting of six SNPs, of which rs6710341 and rs12469063 fully tagged the risk haplotype, is associated with RLS with ORs of 2.75 and 2.36 in the stage 2a and 2b samples, respectively. Frequencies for all haplotypes occurring with these six SNPs are based on cases and controls jointly, and are given for cases and controls separately for the risk haplotype. For the Canadian sample the frequencies are given in brackets and are based on the two haplotype tagging SNPs highlighted in red.

2.1.2.4 Discussion

The variants found in *MEIS1*, *BTBD9* and *MAP2K5*/*LBXCOR1* fulfill all requirements for genuine association signals. They were confirmed in two independent case-control samples and reached P values < 10^{-11} in the joint analysis, which clearly pass even the most stringent suggested threshold for genome-wide significance, a nominal P value ≤ $5x10^{-8}$ [14, 61]. Population stratification was analyzed in the stage 1 samples and the calculated inflation factor $\lambda = 1.09$ after removing of outliers as identified by an EIGENSTRAT analysis showed only minimal population substructure not responsible for the association signals in stage 1 and 2a. We were not able to analyse the substructure in stage 2b due to the lack of appropriate genotype data, since only the SNPs of the replication step were genotyped in these samples.

The moderate risk conferred by these variants is in line with our power calculations. In order to detect variants with odds ratios < 1.5 we would have to increase the sample size of our study. Nevertheless, the identified variants already have a considerable impact on the prevalence of RLS in the general population as shown by their population attributable risk fractions (ARFs). This measure represents the proportion of cases in the population that would be avoided if the risk factor, in this case the risk allele, was eliminated from the population [212]. The population ARF jointly attributable to the three identified loci was estimated in the German population at 68.6% and for the Canadian population at 74.2%. However, the ARF depends on several input factors which show large confidence intervals such as prevalence of the disease and ORs of the individual variants, and is easily overestimated. Therefore, the calculation of minimum and maximum estimates of this parameter is recommended [212]. In the combined German samples, lower limits were estimated at 9.2%, 30.32%, and 7.9% for *MEIS1*, *BTBD9*, and *MAP2K5*/*LBXCOR1* respectively. Corresponding upper limits (= ARFs) were 22.7%, 49.2%, and 20.1%. In the

Canadian sample, the lower limits were 7.5%, 31.6%, and 9.0%, and the upper limits were estimated at 22.6%, 55.0%, and 25.8%. We could not identify any statistical interaction between these loci neither in the individual samples nor in the combined German or combined German/Canadian samples.

The genome-wide stage of our study included only familial RLS cases in order to minimize the heterogeneity of our sample. The replication stage consisted of both familial and sporadic RLS cases. A comparison of these cases in the combined stage 1 and 2a data set revealed virtually indistinguishable ORs for the regions on 6p and 15q. For the region on 2p the risk was higher in familial (rs2300478: OR = 1.82 [1.55-2.14]) than in sporadic cases (OR = 1.59 [1.34-1.90]). However, confidence intervals were overlapping with no significant difference in allele distributions (P = 0.22, Supplementary Table 8 in Appendix 2). Therefore, the identified variants in these regions seem to have the same impact in both endophenotypes of RLS. Consistent with this observation, the familial relative risk figures estimated by the recurrence risk to siblings (λ_s) were all < 1.15 and clearly do not explain the familial aggregation seen in RLS. In this study we did not detect any significant association signals within the known linkage regions with the exception of nominal significance in the RLS-3 region.

The associated genes *MEIS1*, *BTBD9*, and *MAP2K5/LBXCOR1* have never been considered as candidates for RLS based on previous biological knowledge. This is a recurring theme in GWAS and can spark new concepts and hypothesis on the underlying pathophysiology of a disease. *MEIS1* is a member of a family of highly conserved TALE homeobox genes. Heterodimers of MEIS1 with PBX and HOX proteins augment the affinity and specificity of DNA binding by HOX proteins [213]. MEIS1 has been observed to be overexpressed in acute myeloid leukemia [213] and studies in xenopus showed involvement in neural crest development [214]. In addition, there are several potential links to RLS: During embryonic development MEIS1 is essential for distal limb formation [215]. MEIS1 is also part of a Hox transcriptional regulatory network that specifies spinal motor neuron pool identity and connectivity [216]. Intriguingly, spinal hyperexcitability is an established component in the genesis of periodic limb movements found in RLS subjects [137]. Specific functions of MEIS1 in postembryonic tissues still have to be established. The protein is known to be expressed in the adult mouse brain in cerebellar granule cells, the forebrain, and, interestingly, in dopaminergic neurons of the substantia nigra [217]. BTBD9 belongs to the family of BTB(POZ) proteins. These proteins are characterized by the presence of a specific domain, the BTB/POZ domain, which is an interface for protein-protein interactions [218]. Functions

of BTB(POZ) proteins include transcription repression, cytoskeleton regulation, tetramerization and gating of ion channels as well as protein ubiquitination/degradation [218, 219]. The modular nature of this protein and the universal occurrence of the particular domains of BTBD9 make an assignment of a specific function difficult at present. However, results from a concurrent GWAS in Icelandic and American subjects with RLS and/or PLMS indicate an involvement of *BTBD9* in the generation of PLMS. They found an association of the SNP rs3923809 in *BTBD9* to PLMS with or without concomitant RLS symptoms [220]. Moreover, the RLS or PLMS risk allele of rs3923809 was also associated with a 13% decrease of serum ferritin levels in this study, suggesting a role in iron metabolism [220]. This SNP was also associated in our study (Supplementary Tables 4 and 5 in Appendix 2), but since we do not have data on the occurrence of PLMS and the serum ferritin levels in our patients, we cannot validate this observation in our sample. MAPK pathways are conserved from yeast to human and are activated by a signaling cascade that mediates the transduction of extracellular signals to cytoplasmatic nuclear effectors [221]. MAP2K5 is a specific upstream activator of ERK5 and this pathway is activated by oxidative stress, hyperosmolarity and growth factors. In addition, MAP2K5 and ERK5 are abundantly expressed in heart and skeletal muscles and the MAP2K5/ERK5 MAP kinase cascade is critical at early stages of muscle cell differentiation [221]. The possible link between RLS risk alleles and known biological functions of the MAP2K5/ERK5 pathway is of particular interest since this pathway plays an important role in neuroprotection of dopaminergic neurons [222]. LBXCOR1 acts as a transcriptional corepressor of LBX1 [223]. This homeobox gene plays a critical role in the development of sensory pathways in the dorsal horn of the spinal cord that relay pain and touch [224]. The involvement of these developmental genes sheds a new light on the etiology of RLS. They might play a role in RLS pathophysiology already during embryonic development thus raising the question whether RLS has components of a developmental disorder. However, the functions of these genes in postembryonic tissues still have to be established which maintains the possibility of a dysfunction in the adult stage of life as causing RLS.

2.1.3 Association of variants in *PTPRD* on chromosome 9 (RLS-3)

Published manuscript 3 (Appendix 3)

2.1.3.1 Study rationale and design

The RLS-3 region on chromosome 9 was the only RLS linkage region where SNPs showed at least nominal significance in the GWAS (see chapter 2.1.2 of this work). Since this region is one of the best-confirmed linkage signals for RLS, we decided to enlarge our previous study sample and to focus on this region with increased power.

In order to obtain a comprehensive definition of the position and maximum extent of RLS-3 we collected data from all publications on this region [167, 184-186] and defined a target region of 31 Mb on the short arm of chromosome 9 for our association study (9p, 0.5–31.5 Mb), encompassing all published linkage peaks for RLS-3 (Figure 6).

Focused on this locus, we performed a two-stage case-control association study. The exploratory stage 1 was an extension of our GWAS by typing an additional 227 RLS cases on Affymetrix Genome-wide Human SNP 5.0 arrays and analysing these together with the existing 500K data. SNPs within our target region with a nominal P value $< 10^{-3}$ after correcting for population stratification were chosen for replication in stage 2. We genotyped these SNPs in German (1,271 cases/1,901 controls), Czech (279/368) and Canadian (285/842) samples using iPLEX Gold genotyping technology (Sequenom). The combined stage 1 and 2 samples had a power > 85% to detect variants with ORs of $\geq$ 1.3 with genome-wide significance.

All cases were diagnosed in face-to-face interviews according to the IRLSSG criteria [64] and symptomatic cases were excluded from the study (Supplementary Methods of Appendix 3). Cases of stage 1 had a higher proportion of familial RLS and the female to male ratio was approximately 2:1 in all study populations (Supplementary table 2 of Appendix 3). Controls were either population-based controls (German sample), recruited from the population visiting a hospital (Canadian sample), or selected randomly from a blood and bone marrow donor bank (Czech sample).

2.1.3.2 Association results of the exploratory phase

Following the stringent quality control protocol already used in our GWAS, 3,270 SNPs remained for statistical analysis in RLS-3. Of these, eight SNPs passed the criterion for inclusion in the replication stage (Figure 6). These SNPs were located in three genes: *PTPRD* (2 SNPs), *C9ORF52* (1 SNP), and *C9ORF93* (4 SNPs).

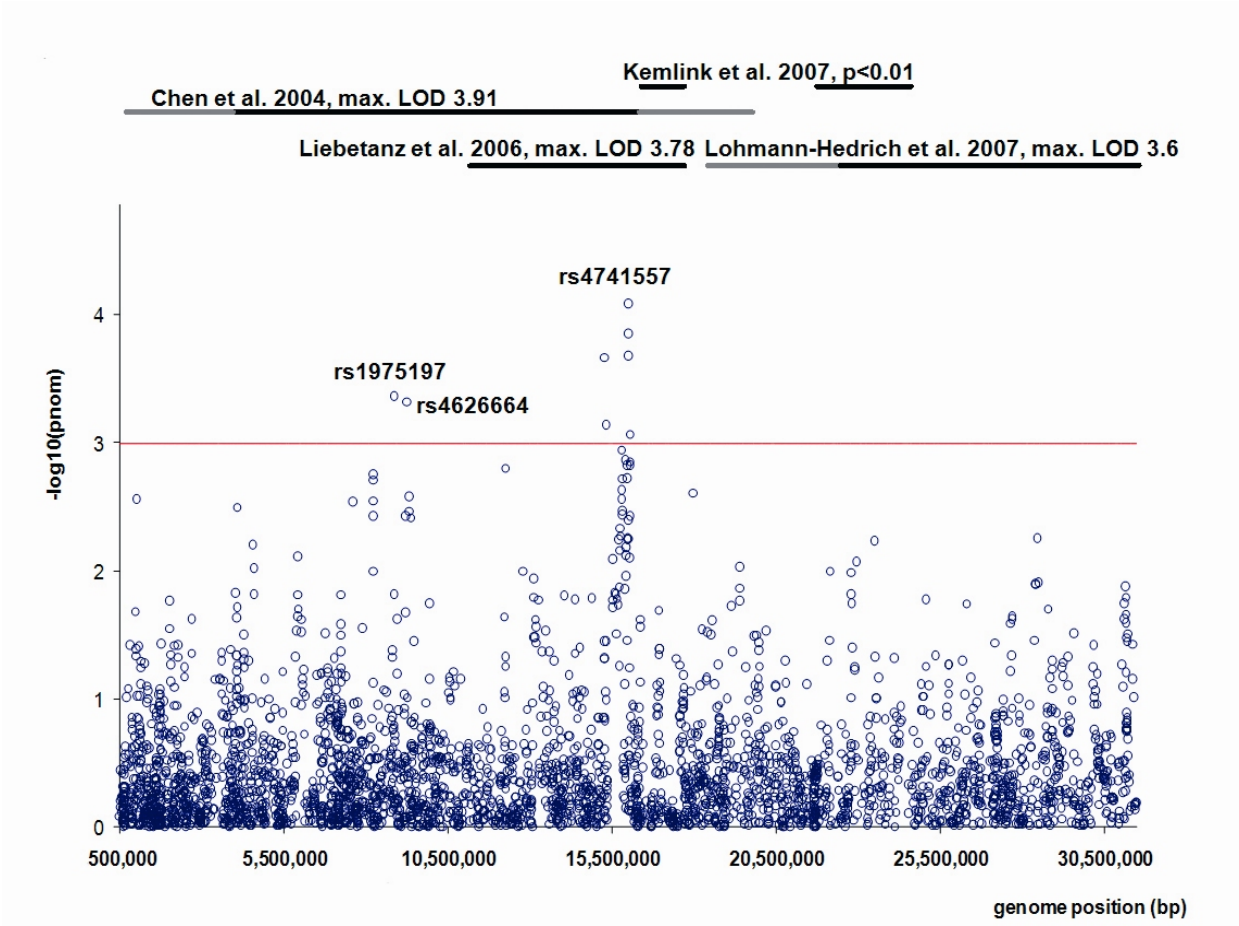


Figure 6. Association results of the exploratory phase. Results of stage 1 ($-\log_{10}$ of P nominal (P_{nom}) corrected for λ) for chromosome 9p, 0.5-31.5 Mb. The red line indicates the cut-off for selection of SNPs for replication. Position and extent of linkage signals are shown as horizontal bars. The black bars represent the narrowest suggested region as defined by intrafamilial recombination events, the grey bars extend to the maximum size. Maximum multi-point LOD scores [167, 185, 186] and the P value from non-parametric linkage analysis [184] are denoted above the bars. Genomic positions refer to the Human March 2006 assembly.

2.1.3.3 Replication of association signals within *PTPRD* and mutation screening of *PTPRD*

Of the eight stage 1 signals, only the two variants rs1975197 and rs4626664 located within the *PTPRD* gene were replicated in stage 2 (Table 1 in Appendix 3). We conducted both a separate analysis of stage 2 samples and a combined analysis of all stage 2 samples to increase the power of our analysis. Since the individual stage 2 samples showed significantly different MAFs, we included the country of origin as a covariate in the combined analysis. The association of both SNPs was significant after Bonferroni correction for multiple testing in the German subsample and the combined analysis of all stage 2 samples. In the Canadian

subsample both SNPs were nominally significant (rs4626664, $P_{nominal/\lambda\text{-corrected}}$= 0.018; rs1975197, $P_{nominal/\lambda\text{-corrected}}$= 0.024), whereas the Czech sample showed only a trend for association for the stronger signal (rs4626664, $P_{nominal/\lambda\text{-corrected}}$= 0.075), most likely explained by lack of power due to the smaller sample sizes. Comparable ORs with unidirectional allelic association were found in the individual subsamples of stage 2, indicating that the risk alleles have the same impact in the different populations despite the lack of statistical significance in the smaller subsamples (Supplementary Table 3 in Appendix 3).

For the test of genome-wide significance we combined stage 1 and 2 in a joint analysis of all samples. Both *PTPRD* SNPs reached genome-wide significance after Bonferroni (B) correction for multiple testing: rs4626664: $P_{nominal/\lambda\text{-corrected}}$= 5.91 x 10^{-10}, $P_{corrected(B)}$= 0.00012, OR= 1.44, 95% CI: 1.2-1.44; rs1975197: $P_{nominal/\lambda\text{-corrected}}$= 5.81 x 10^{-9}, $P_{corrected(B)}$= 0.0012, OR= 1.31, 95% CI: 1.31-1.59 (Table 1 in Appendix 3).

The association signals are located 0.41 Mb apart and map to intron eight and ten of *PTPRD* within two separate LD blocks. Logistic regression showed no significant interaction between these SNPs (P= 0.986) as is also evidenced by the lack of LD between them (r^2 = 0). There is also no significant interaction with the risk alleles in *MEIS1* (rs4626664, P= 0.463; rs1975197, P= 0.957), *BTBD9* (rs4626664, P= 0.487; rs1975197, P= 0.246), and *LBXCOR1/MAP2K5* (rs4626664, P= 0.510; rs1975197, P= 0.859). Haplotype analysis showed no increase in significance compared to single SNP analysis. Comparing the risk conferred by both SNPs in familial and sporadic cases in the combined stage 1 and 2 data set revealed highly similar ORs with overlapping CIs: rs1975197: OR_{fam} = 1.34 (1.15-1.57), OR_{spor} = 1.37 (1.21-1.55); rs4626664: OR_{fam} = 1.3 (1.09-1.55), OR_{spor} = 1.44 (1.26-1.65).

Mutation analysis of *PTPRD* revealed no mutations in 35 coding and ten non-coding exons of *PTPRD* comparing the sequence of nine patients from an RLS3-linked family, three index cases from RLS families in which linkage to RLS3 was not excluded and one control to the reference sequence (NM_002839). Only one not yet annotated synonymous SNP was found but did not segregate with the disease. We also found no exon deletions or duplications using quantitative real-time PCR. Among eight nonsynonymous coding SNPs genotyped in replication samples, only rs10977171 and rs35929428 were polymorphic but did not show any association (Supplementary Tables 5 and 6 of Appendix 3).

2.1.3.4 Discussion

We again used a very strict quality control approach for the exploratory stage in order to minimize false-positive signals. Moreover, population stratification was assessed in all study samples and the P values were corrected for this confounding effect. The calculated inflation factors were very low in the German stage 1 and 2 samples (λ = 1.07 and 1.10, respectively) and significantly higher in the Czech and the Canadian sample (λ = 1.23 and 1.26, respectively). This stringent approach led to the identification of two independent association signals within the *PTPRD* gene encoding the protein tyrosine phosphatase receptor type delta. PTPRD belongs to the family of type IIa receptor-like protein tyrosine phosphatases. These molecules are characterized by an extracellular part containing cell adhesion motifs and an intracellular part containing two phosphatase domains [225]. Several *PTPRD* mRNA isoforms are expressed in a developmental and tissue specific manner [226]. Both identified SNPs are located within the 5'-UTR consisting of ten non-coding exons contained in two known long splice variants expressed predominantly in fetal and adult brain tissue [226, 227]. The involvement of *PTPRD* in RLS is unknown. Studies in *PTPRD* and *PTPRS* knockout mice showed a function of these proteins in axon guidance and termination of mammalian motorneurons during embryonic development [225]. Investigations in neuroblastoma tumor tissue and cell lines identified microdeletions and aberrant splicing patterns in the 5'-UTR of *PTPRD* which may influence the mRNA stability and thereby gene expression [228].

The identified associated variants are common (MAF > 10%) with weak effects (ORs of 1.44, 95% CI = 1.3-1.6, and 1.31, 95% CI= 1.2-1.4) and cannot explain the linkage signal of RLS-3. Correspondingly, the familial relative risk figures (λ_s) were all below 1.04. There was no difference in the ORs for both SNPs between familial and sporadic cases, indicating an equal contribution to both endophenotypes similar as observed for the associated SNPs in the GWAS. The mutation screening of *PTPRD* did not detect any rare alleles with strong effects within this gene when investigating members of families with linkage to RLS-3. It is possible that an allelic series of *PTPRD* exists with yet unidentified rare high risk variants which are responsible for the linkage signal. In contrast, there could also be independent genes annotated in this region which could play a role in RLS etiology and could underlie the observed linkage of chromosome 9p. The fact that the exact position and extent of RLS-3 is not resolved at present [186] is in line with the concept of multiple susceptibility loci within one chromosomal region. This has already been described in other diseases, e.g. in prostate

cancer, where several independent susceptibility loci have been identified on chromosome 8q24 [229].

PTPRD is the fourth locus associated to RLS with genome-wide significance. The two novel association signals add another four to the previous six risk alleles from chromosomes 2p, 6p and 15q, making a total of ten possible risk alleles (referring to homozygous carriers). There is considerable interest in using susceptibility variants identified in GWAS for genetic testing and individual risk prediction [230, 231]. Since each single variant usually confers only a moderate effect, it is suggested to test multiple susceptibility loci simultaneously [230, 232, 233]. We assessed the predictive power of the RLS risk alleles by means of a receiver operating characteristic (ROC) analysis (Supplementary Methods in Appendix 3). The ROC curve depicts the sensitivity versus the specificity of the diagnostic test for each cut-off (= number of risk alleles necessary to be defined as affected) and the resulting area under the curve (AUC) is the corresponding measure for the performance of the test [233, 234]. A perfect test would produce an AUC of 1.0 whereas an AUC of 0.5 represents a test with no predictive value [234]. In our case the AUC equals 0.642, which indicates a rather poor predictive power. Accordingly, the known risk alleles for RLS have only limited usefulness for individual risk prediction. However, the identification of further associated variants might improve the prospects of meaningful genetic testing [230, 231, 235].

2.2 Symptomatic RLS

2.2.1 Association study of *MEIS1*, *BTBD9*, *MAP2K5/LBXCOR1*, and *PTPRD* in uremic RLS

Submitted manuscript 1 (Appendix 4)

2.2.1.1 Study rationale and design

The genetic basis of the symptomatic forms of RLS has never been investigated so far. It is not known if there is a genetic predisposition to symptomatic RLS and if the genetic mechanisms are overlapping in idiopathic and symptomatic RLS. We therefore conducted an association study of the variants definitely associated with iRLS in a sample of ESRD patients with uRLS.

We genotyped 10 of the most significant iRLS associated SNPs across the four genomic regions using using iPLEX Gold genotyping technology (Sequenom) in a case-control sample of European descent recruited from ESRD patients on maintenance hemodialysis. We recruited our study sample from 16 dialysis centers in Munich and the surrounding region, which had a total of 1,617 regular dialysis patients. Cases (n = 200) were diagnosed in a face-to-face interview according to the IRLSSG criteria [64], whereas control status (n = 443) was assigned by questionnaire-based self-report of absence of RLS symptoms. Power of this sample to detect the associations was 92% for *MEIS1*, 75% for *BTBD9*, 61% for *MAP2K5*/*LBXCOR1*, and 35% for *PTPRD*.

2.2.1.2 Association results

Armitage trend test revealed a significant association of variants in *MEIS1* and *BTBD9* to uRLS. Within *MEIS1*, two of three SNPs were significantly associated after correction for multiple testing: rs12469063 ($P_{corrected}$ = 0.004, OR = 1.52, 95% CI = 1.17-1.98), and rs2300478 ($P_{corrected}$ = 0.01, OR = 1.47, 95% CI = 1.13-1.91). In *BTBD9*, rs3923809 was associated ($P_{corrected}$ = 0.002, OR = 1.56, 95% CI = 1.19-2.04). For *MAP2K5*/*LBXCOR1* and *PTPRD* the nominal P values were between 0.057 and 0.3 (Table 1 in Appendix 4). Haplotype analysis confirmed the association of the known 'AG' haplotype in *MEIS1* ($P_{corrected}$ = 0.048, OR = 1.57 (95% CI = 1.10-2.23). A subanalysis with cases stratified according to their family history revealed a trend for differences in the size of the contribution of the associated loci to familial or sporadic uRLS (Supplementary Table 2 in Appendix 4). Analysing only cases with a positive family history (n = 38), revealed a significant association both to *MEIS1* and

BTBD9 (rs12469063, $P_{corrected}$ = 0.008; rs2300478, $P_{corrected}$ = 0.016, and rs3923809, $P_{corrected}$ = 0.012). Using only cases with a negative family history (n = 133), the two loci showed nominally significant P values (rs12469063, P_{nom} =0.014, $P_{corrected}$ = 0.056; rs2300478, P_{nom} =0.021, $P_{corrected}$ = 0.084, and rs3923809, P_{nom} = 0.015, $P_{corrected}$ = 0.06).

2.2.1.3 Discussion

Of the potentially available 1,617 ESRD patients, 737 agreed to participate (45.6%). The relatively low response rate was due to several reasons. A number of patients were not at the dialysis center on the day of recruitment, others had severe comorbidities such as dementia or depression and were primarily not able to participate or not interested in the study. Others in turn could not enter the study because of language and comprehension problems.

We demonstrate that sequence variants in *MEIS1* and *BTBD9* are genetic susceptibility factors for RLS in ESRD patients ($P_{corrected} \leq 0.01$). The effect size of these variants is within the same range as observed in iRLS studies of comparable sample size (ORs_{uRLS} = 1.47 to 1.56, 95% CIs = 1.19-2.04; ORs_{iRLS}= 1.43 to 1.59, 95% CIs = 1.12 – 2.2) [236, 237]. Based on the present data we can neither prove nor exclude a contribution of *MAP2K5*/*LBXCOR1* and *PTPRD* to uRLS. Although they were not significantly associated in our study, their ORs showed the same direction as in iRLS and the CIs were overlapping [236-239].

Previous genotype/phenotype analysis in iRLS patients showed *BTBD9* more associated to PLMS, the motor component of RLS, than to the sensory symptoms [220]. Therefore, its strong association to uRLS is remarkable in the context that the motor symptoms seem to be more prominent in uRLS in comparison to iRLS [123, 124].

Our result that only 19% of uRLS cases reported a positive family history is in line with previous observations showing a lower frequency of familial RLS (12%) [80] in uRLS compared to iRLS (30 - 60%) [240]. The impact of the associated variants is not statistically different between either familial or sporadic uRLS. The ORs tended to be higher in the familial subgroup but the difference was not significant ($P_{Breslow-Day} > 0.3$) and 95% CIs were overlapping which is in line with iRLS studies [236-239] (Supplementary Table 2 in Appendix 4). The reduced familial clustering of RLS in ESRD patients could be interpreted as an indicator of a stronger influence of non-genetic factors on developing the disease compared to genetic factors.

The prevalence of uRLS observed in our study (31.1%) is higher than in the only other study in German patients conducted so far (23% in [111]), which could be due to the different ascertainment strategies. It is higher than in the general population at the same age [87, 241]

suggesting that additional genetic and/or non-genetic risk factors must be present in ESRD patients. However, there is potential for ascertainment bias since ESRD patients affected by RLS are more likely to participate than RLS-negative ESRD patients. We tried to minimize this bias by informing all patients about the importance of participating even when not suffering from RLS.

Complex diseases result from genetic and non-genetic or environmental factors and their interactions. Calculating the OR for ESRD as a risk factor for RLS by comparing our ESRD patients sample to a hypothetical sample from the general population of the same size (643, of which 64 (10%) have RLS), shows an effect size of 4. Therefore reduced renal function and dependence on dialysis seem to be a strong trigger for RLS. This is supported by the observation of an abolishment of RLS symptoms after renal transplantation [118].

Finally, both iRLS and uRLS share genetic risk factors, suggesting a partial overlap in the predisposing mechanism and in the pathophysiology. Our observations should be replicated in further dialysis patient samples in order to obtain robust confirmation of the association results. It remains to be investigated if there are genetic variants specific to uRLS which are not relevant for iRLS.

3 Conclusions and future developments

This work presents the results of the first large-scale hypothesis-based regional and hypothesis-free genome-wide association studies in RLS and demonstrates the amenability of the phenotype RLS to this approach. The identification of common variants associated with idiopathic RLS in five genomic loci shows that this phenotype is influenced by such variants and to a certain extent corroborates the notion of the CDCV hypothesis for common complex diseases such as RLS.

Within each locus the association signals map to discrete LD blocks either located in a single gene (*NOS1*, *MEIS1*, *BTBD9*, and *PTPRD*) or spanning parts of two adjacent genes (*MAP2K5*/*LBXCOR1*). All but the *NOS1* signal have been replicated in independent populations. The association to *NOS1* showed a "flip-flop" phenomenon and thus needs evaluation in further independent samples to either confirm the association or identify the signal as a false-positive. Therefore, the signal in *NOS1* has to be regarded as preliminary at present and was not included in any following total risk or interaction analyses. Additional support for the associations to *MEIS1*, *BTBD9*, and *MAP2K5*/*LBXCOR1* is provided by their replication in European and US-American populations of European descent [220, 236, 237]. All replications were exact replications, i.e. the same SNPs were associated with the same direction and comparable size of the ORs, which is the most convincing evidence for the genuineness of an association signal. Furthermore, we minimized the chance for errors and biases that could lead to false-positive results by implementing a stringent genotyping quality control, a highly standardized diagnostic routine for case ascertainment, and adequate corrections for population stratification and multiple testing. Especially for the associations detected in the GWAS and the subsequent study of the RLS-3 region, the statistical evidence is compelling since all signals reach genome-wide significance when applying the most conservative approach to correct for multiple testing. Thus the signals in *MEIS1*, *BTBD9*, *MAP2K5*/*LBXCOR1*, and *PTPRD* represent true-positive associations and are the first unequivocal susceptibility variants identified for RLS.

The identified variants are not necessarily the causal variants. Due to the fact that only a small proportion of the SNPs present in the genome can be tested directly in LD-based association studies, there is a high *a priori* probability that the associated variants are not etiological variants but merely markers in LD with them. In addition, causal variants are not limited to SNPs but could also be structural or epigenetic variation not directly assayed in the association study but detected via a SNP in LD. Our association signals primarily just flag the

relevant genomic regions that contain the causal variant(s) and disease-related gene(s). With the exception of the *MAP2K5/LBXCOR1* locus, we have already narrowed down the region of interest to LD blocks in single genes. These are first-line candidates for further intensive screening in follow-up studies to detect the functional variants, although there is still the remote possibility that these LD blocks contain long-range regulatory elements belonging to more distant genes. Several approaches can be taken to identify the causal variants [46, 62]. One possibility is to test the identified variants for association in populations of different ethnicity. Divergent LD patterns in these populations might lead to divergent association patterns. These could help to exclude variants not showing consistent association or to identify variants associated in all populations, which are then more likely to be causal. This method could be especially helpful in resolving the association signal covering *MAP2K5* and *LBXCOR1*, since the HapMap sample of African ancestry shows two separate LD blocks in this region instead of the single block observed in European and Asian populations [41]. Another approach is resequencing of at least the associated LD blocks or preferably the entire genes in a large number of patients (> 500), which has become increasingly feasible with the development of the next-generation sequencing technologies. This will yield a more complete inventory of the variation in these regions including rare and structural variants which were outside the detection range of the association studies. Candidates for functional studies can then be selected from this set of variants based on the likelihood of their biologic implication. These include for example missense SNPs, variants in known regulatory elements or variants located in highly conserved non-coding regions which might possess regulatory (or structural) functions important for gene expression. The latter possibility is intriguing since the most significantly associated SNP detected in the GWAS, rs12469063 in *MEIS1*, is located in such a highly conserved region. This SNP is also part of a high risk haplotype. Only recently, a study in French-Canadian RLS patients showed that this haplotype influences the RNA and protein levels of MEIS1 [242], supporting a functional role of the SNP and/or the region tagged by the haplotype. A detailed analysis of regulatory functions in gene expression of this SNP and other associated variants in conserved regions could be performed *in-vitro* in cellular assays or *in-vivo* in model organisms such as zebrafish.

Even if the underlying causal variants and molecular mechanisms are still undetermined the known functions of the identified candidate genes can support existing concepts and generate new hypotheses on RLS pathophysiology. *NOS1* influences dopaminergic transmission in the CNS by the messenger molecule nitric oxide and could thus play a role in the dopaminergic dysfunction suspected in RLS. *BTBD9* is a gene of still unknown function but the concurrent

GWAS in RLS cases by Stefansson et al. [220] has implicated this gene in iron metabolism, the other main candidate pathway in RLS pathophysiology. A completely novel concept of RLS etiology is suggested by *MEIS1*, *MAP2K5*, *LBXCOR1*, and *PTPRD*. So far, they have been implicated in embryonic development which poses the question if RLS has components of a developmental disorder. Subtle defects in early development could result in a base-line predisposition to RLS where symptoms are subsequently triggered by environmental factors such as renal failure, pregnancy or ageing (sporadic and symptomatic RLS) or which only leads to disease in combination with an additional genetic load (familial RLS). However, it is also possible that the genes have different functions in the adult, which have not been discovered yet. These functions could be studied in the respective knockout mouse models or in transgenic mice carrying for example the *MEIS1* risk haplotype. These mice could also be analysed for RLS-related phenotypes like periodic limb movements, hyperactivity, increased pain sensitivity or disrupted sleep in order to compile more evidence for their involvement in RLS.

Another interesting aspect regarding the etiology is the observation that the effect sizes for the variants in *MEIS1*, *BTBD9*, *MAP2K5*/*LBXCOR1* and *PTPRD* are similar in familial and sporadic RLS. This indicates equal contributions of the variants to both endophenotypes and suggests a common pathophysiology, which would be in line with a base-line predisposition model as presented above, independent of the actual underlying mechanism. The same is suggested for *MEIS1* and *BTBD9* when comparing uremic RLS, the most common form of symptomatic RLS, and idiopathic RLS. If it is also the case for *MAP2K5*/*LBXCOR1* and *PTPRD* is unclear at present since the study was underpowered for the smaller effects of these variants. Symptomatic forms of RLS in particular support the notion of additional factors acting on a base-line genetic predisposition since they usually resolve once the environmental trigger, e.g. iron deficiency, renal failure or pregnancy, is eliminated.

The susceptibility variants identified in our studies are only a first glimpse at the underlying genetic architecture of RLS and more susceptibility variants remain to be discovered due to several limitations in the detection scope of the association approach. Current study designs are underpowered for small effect variants with ORs < 1.2 and provide poor coverage of rare variants and structural variation. Their design is based on common SNPs with MAFs > 5% and thus low frequency common variants (1% < MAF < 5%) are also insufficiently covered. In our specific case, additional limiting factors have to be considered. First of all, both the regional RLS-1 study and the GWAS had samples sizes below 500 cases and were only sufficiently powered to detect variants with ORs > 1.5 and MAFs > 10%. This is evident from

the increased sample size that was necessary to move the nominally significant smaller effect size signals in PTPRD found in the GWAS to genome-wide significance in the RLS-3 study. Moreover, we only carried forward a very small number of SNPs to the replication studies, thus limiting the number of false-positive results but also increasing the possibility of false-negative results. A further constraint is the incomplete genome-wide coverage of common variation on the Affymetrix SNP chips. Using all SNPs on the Affymetrix 500K and 5.0 arrays would have yielded a genome coverage of approximately 70% in our study population, but due to the strict quality control more than 50% of the SNPs present on the arrays were discarded thereby reducing the coverage further. Concurrent with the study power the effect size of our detected associations ranges between ORs of 1.3 and 1.7 for single SNPs and increases to 2.7 for the risk haplotype identified in *MEIS1*. Compared to other GWAS conducted so far, which mostly report ORs of 1.1 to 1.2, they reside in the upper tail of effect sizes for common variants [231, 243]. The estimates might be inflated due to ascertainment bias in the original study (so-called "winner's curse" [244]), but the replication studies also give ORs of 1.3 and above for these variants [236, 237]. Because of these relatively large effects for common variants and their high frequencies, they describe a large proportion of disease occurrence in the population which is indicated by their population attributable risk fractions of e.g. 9 to 23% for *MEIS1* or 30 to 49% for *BTBD9*. Because these estimates taken together do not reach 100% and because they can be inflated, more common susceptibility variants can be expected for RLS, but most likely with smaller effect sizes. In addition, the identified variants explain only a very small fraction of the inherited component of RLS as evidenced by the low familial relative risk conferred by them. This is the reason for their poor performance in predicting individual risk preventing their use in genetic testing at present. They also cannot explain the linkage signals of RLS-1 and RLS-3. The missing heritability can possibly be accounted for by the variation not detectable by current association study designs: un-tagged common variants due to incomplete coverage, common variants with small effects, structural variants, epigenetic modifications, and rare or novel variants with presumably larger effect sizes than the common variants [45, 46, 62]. The latter ones are probably responsible for the linkage signals. Moreover, gene-gene and gene-environment interactions are also likely to play an important role, but their large-scale analysis requires much larger sample sizes and detailed information on the environmental exposures of cases and controls. In the studies presented here the interaction analysis was limited to gene-gene interactions between the associated variants and did not detect any epistasis.

In order to identify the remaining genetic variants for RLS, further approaches should be tailored to assess all possible types of variation [45, 46, 62]. To detect additional common variants of similar or smaller effects and lower allele frequencies, the sample size of present GWAS should be increased substantially, preferably to several thousands or more. One way to do this is combining the data of our study with the other GWAS reported for RLS in a meta-analysis. Another possibility is to improve the coverage of our study by imputation of genotypes at untyped SNPs based on the genotypes of correlated SNPs typed in the study. The enlarged studies will be better powered to detect common and low frequency variants with moderate to small effects and also to conduct gene-gene and gene-environment interaction analysis on a larger scale. Gene-environment interactions could be especially interesting in RLS since symptomatic forms share the same genetic factors but are obviously strongly dependent on environmental triggers. However, to achieve this, data on environmental exposures has to be collected. Genome-wide studies of CNVs and epigenetic modifications are also becoming increasingly feasible and should be conducted in RLS as soon as they can be cost-efficiently realized for large sample sizes. Till then, such studies could be focused on candidate regions such as the linkage regions. The detection of rare and novel variants is facilitated by the new next-generation sequencing technologies which allow sequencing of entire genes, genomic regions or even genomes. Since whole genome sequencing of large numbers of samples is not yet affordable, these approaches are limited to individual genes or genomic regions at present.

The most important benefit obtained from the already identified variants and any additional variant revealed in future studies is highlighting of candidate genes that can provide first clues at unravelling the molecular pathways underlying RLS and might lead to new targets for the treatment in the long run. The clinical characterization together with genetic assessments will contribute to a prediction of treatment response and an optimized personalized therapy for those at increased risk for the development of the disorder.

4 Bibliography

1. Altshuler, D., Daly, M.J. and Lander, E.S. Genetic mapping in human disease. *Science*, **322** (5903): 881-888, 2008.
2. Risch, N. and Merikangas, K. The future of genetic studies of complex human diseases. *Science*, **273** (5281): 1516-1517, 1996.
3. Risch, N.J. Searching for genetic determinants in the new millennium. *Nature*, **405** (6788): 847-856, 2000.
4. Dawn Teare, M. and Barrett, J.H. Genetic linkage studies. *Lancet*, **366** (9490): 1036-1044, 2005.
5. Elston, R.C. Introduction and overview. Statistical methods in genetic epidemiology. *Stat Methods Med Res*, **9** (6): 527-541, 2000.
6. Botstein, D. and Risch, N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet*, **33** (Suppl): 228-237, 2003.
7. Zondervan, K.T. and Cardon, L.R. The complex interplay among factors that influence allelic association. *Nat Rev Genet*, **5** (2): 89-100, 2004.
8. Manolio, T.A. and Collins, F.S. Genes, environment, health, and disease: facing up to complexity. *Hum Hered*, **63** (2): 63-66, 2007.
9. Lander, E.S. and Schork, N.J. Genetic dissection of complex traits. *Science*, **265** (5181): 2037-2048, 1994.
10. Lander, E.S. The new genomics: global views of biology. *Science*, **274** (5287): 536-539, 1996.
11. Reich, D.E. and Lander, E.S. On the allelic spectrum of human disease. *Trends Genet*, **17** (9): 502-510, 2001.
12. Cordell, H.J. and Clayton, D.G. Genetic association studies. *Lancet*, **366** (9491): 1121-1131, 2005.
13. Cardon, L.R. and Bell, J.I. Association study designs for complex diseases. *Nat Rev Genet*, **2** (2): 91-99, 2001.
14. McCarthy, M.I., Abecasis, G.R., Cardon, L.R., *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*, **9** (5): 356-369, 2008.
15. Kruglyak, L. and Nickerson, D.A. Variation is the spice of life. *Nat Genet*, **27** (3): 234-236, 2001.
16. Palmer, L.J. and Cardon, L.R. Shaking the tree: mapping complex disease genes with linkage disequilibrium. *Lancet*, **366** (9492): 1223-1234, 2005.
17. National Center for Biotechnology Information. Database of Single Nucleotide Polymorphisms (dbSNP). Available from: http://www.ncbi.nlm.nih.gov/SNP. [accessed 02.10.2009]
18. Balding, D.J. A tutorial on statistical methods for population association studies. *Nat Rev Genet*, **7** (10): 781-791, 2006.
19. Ziegler, A., Konig, I.R. and Thompson, J.R. Biostatistical aspects of genome-wide association studies. *Biom J*, **50** (1): 8-28, 2008.
20. Montana, G. Statistical methods in genetics. *Brief Bioinform*, **7** (3): 297-308, 2006.
21. Fan, J.B., Chee, M.S. and Gunderson, K.L. Highly parallel genomic assays. *Nat Rev Genet*, **7** (8): 632-644, 2006.
22. Frazer, K.A., Ballinger, D.G., Cox, D.R., *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449** (7164): 851-861, 2007.

23. Sachidanandam, R., Weissman, D., Schmidt, S.C., *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, **409** (6822): 928-933, 2001.
24. The International HapMap Consortium. A haplotype map of the human genome. *Nature*, **437** (7063): 1299-1320, 2005.
25. Manolio, T.A., Brooks, L.D. and Collins, F.S. A HapMap harvest of insights into the genetics of common disease. *J Clin Invest*, **118** (5): 1590-1605, 2008.
26. Ardlie, K.G., Kruglyak, L. and Seielstad, M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet*, **3** (4): 299-309, 2002.
27. Pritchard, J.K. and Przeworski, M. Linkage disequilibrium in humans: models and data. *Am J Hum Genet*, **69** (1): 1-14, 2001.
28. Slatkin, M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet*, **9** (6): 477-485, 2008.
29. Dunning, A.M., Durocher, F., Healey, C.S., *et al.* The extent of linkage disequilibrium in four populations with distinct demographic histories. *Am J Hum Genet*, **67** (6): 1544-1554, 2000.
30. Ardlie, K., Liu-Cordero, S.N., Eberle, M.A., *et al.* Lower-than-expected linkage disequilibrium between tightly linked markers in humans suggests a role for gene conversion. *Am J Hum Genet*, **69** (3): 582-589, 2001.
31. Abecasis, G.R., Noguchi, E., Heinzmann, A., *et al.* Extent and distribution of linkage disequilibrium in three genomic regions. *Am J Hum Genet*, **68** (1): 191-197, 2001.
32. Reich, D.E., Cargill, M., Bolk, S., *et al.* Linkage disequilibrium in the human genome. *Nature*, **411** (6834): 199-204, 2001.
33. Stephens, J.C., Schneider, J.A., Tanguay, D.A., *et al.* Haplotype variation and linkage disequilibrium in 313 human genes. *Science*, **293** (5529): 489-493, 2001.
34. Gabriel, S.B., Schaffner, S.F., Nguyen, H., *et al.* The structure of haplotype blocks in the human genome. *Science*, **296** (5576): 2225-2229, 2002.
35. McVean, G.A., Myers, S.R., Hunt, S., *et al.* The fine-scale structure of recombination rate variation in the human genome. *Science*, **304** (5670): 581-584, 2004.
36. Daly, M.J., Rioux, J.D., Schaffner, S.F., *et al.* High-resolution haplotype structure in the human genome. *Nat Genet*, **29** (2): 229-232, 2001.
37. Devlin, B. and Risch, N. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics*, **29** (2): 311-322, 1995.
38. Wang, W.Y., Barratt, B.J., Clayton, D.G., *et al.* Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet*, **6** (2): 109-118, 2005.
39. Hirschhorn, J.N. and Daly, M.J. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*, **6** (2): 95-108, 2005.
40. The International HapMap Consortium. The International HapMap Project. *Nature*, **426** (6968): 789-796, 2003.
41. The International HapMap Consortium. International HapMap Project. Available from: http://hapmap.ncbi.nlm.nih.gov/index.html.en. [accessed 01.10.2009]
42. Stram, D.O. Tag SNP selection for association studies. *Genet Epidemiol*, **27** (4): 365-374, 2004.
43. Carlson, C.S., Eberle, M.A., Rieder, M.J., *et al.* Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet*, **74** (1): 106-120, 2004.
44. Pearson, T.A. and Manolio, T.A. How to interpret a genome-wide association study. *Jama*, **299** (11): 1335-1344, 2008.
45. Frazer, K.A., Murray, S.S., Schork, N.J., *et al.* Human genetic variation and its contribution to complex traits. *Nat Rev Genet*, **10** (4): 241-251, 2009.

46. McCarthy, M.I. and Hirschhorn, J.N. Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet*, **17** (R2): R156-165, 2008.
47. Purcell, S., Sham, P. Genetic Power Calculator.
Available from: http://pngu.mgh.harvard.edu/~purcell/gpc/. [accessed 07.09.2009]
48. Purcell, S., Cherny, S.S. and Sham, P.C. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*, **19** (1): 149-150, 2003.
49. Center for Statistical Genetics, University of Michigan. CaTS - Power calculator for two stage association studies.
Available from: http://www.sph.umich.edu/csg/abecasis/CaTS/index.html. [accessed 07.09.2009]
50. Skol, A.D., Scott, L.J., Abecasis, G.R., *et al.* Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet*, **38** (2): 209-213, 2006.
51. Gordon, D. and Finch, S.J. Factors affecting statistical power in the detection of genetic association. *J Clin Invest*, **115** (6): 1408-1418, 2005.
52. Iles, M.M. What can genome-wide association studies tell us about the genetics of common disease? *PLoS Genet*, **4** (2): e33, 2008.
53. Zeggini, E., Rayner, W., Morris, A.P., *et al.* An evaluation of HapMap sample size and tagging SNP performance in large-scale empirical and simulated data sets. *Nat Genet*, **37** (12): 1320-1322, 2005.
54. McCarroll, S.A., Hadnott, T.N., Perry, G.H., *et al.* Common deletion polymorphisms in the human genome. *Nat Genet*, **38** (1): 86-92, 2006.
55. Hinds, D.A., Kloek, A.P., Jen, M., *et al.* Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nat Genet*, **38** (1): 82-85, 2006.
56. McCarroll, S.A. Extending genome-wide association studies to copy-number variation. *Hum Mol Genet*, **17** (R2): R135-142, 2008.
57. McCarroll, S.A. and Altshuler, D.M. Copy-number variation and association studies of human disease. *Nat Genet*, **39** (7 Suppl): S37-42, 2007.
58. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447** (7145): 661-678, 2007.
59. Devlin, B. and Roeder, K. Genomic control for association studies. *Biometrics*, **55** (4): 997-1004, 1999.
60. Price, A.L., Patterson, N.J., Plenge, R.M., *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*, **38** (8): 904-909, 2006.
61. Chanock, S.J., Manolio, T., Boehnke, M., *et al.* Replicating genotype-phenotype associations. *Nature*, **447** (7145): 655-660, 2007.
62. Ioannidis, J.P., Thomas, G. and Daly, M.J. Validating, augmenting and refining genome-wide association signals. *Nat Rev Genet*, **10** (5): 318-329, 2009.
63. Walters, A.S. Toward a better definition of the restless legs syndrome. The International Restless Legs Syndrome Study Group. *Mov Disord*, **10** (5): 634-642, 1995.
64. Allen, R.P., Picchietti, D., Hening, W.A., *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** (2): 101-119, 2003.
65. Bassetti, C.L., Mauerhofer, D., Gugger, M., *et al.* Restless legs syndrome: a clinical study of 55 patients. *Eur Neurol*, **45** (2): 67-74, 2001.

66. Hening, W.A., Walters, A.S., Wagner, M.*, et al.* Circadian rhythm of motor restlessness and sensory symptoms in the idiopathic restless legs syndrome. *Sleep*, **22** (7): 901-912, 1999.
67. Trenkwalder, C., Hening, W.A., Walters, A.S.*, et al.* Circadian rhythm of periodic limb movements and sensory symptoms of restless legs syndrome. *Mov Disord*, **14** (1): 102-110, 1999.
68. Michaud, M., Dumont, M., Selmaoui, B.*, et al.* Circadian rhythm of restless legs syndrome: relationship with biological markers. *Ann Neurol*, **55** (3): 372-380, 2004.
69. Comella, C.L. Restless legs syndrome: treatment with dopaminergic agents. *Neurology*, **58** (4 Suppl 1): S87-92, 2002.
70. Coleman, R.M., Pollak, C.P. and Weitzman, E.D. Periodic movements in sleep (nocturnal myoclonus): relation to sleep disorders. *Ann Neurol*, **8** (4): 416-421, 1980.
71. Montplaisir, J., Boucher, S., Poirier, G.*, et al.* Clinical, polysomnographic, and genetic characteristics of restless legs syndrome: a study of 133 patients diagnosed with new standard criteria. *Mov Disord*, **12** (1): 61-65, 1997.
72. Hornyak, M., Feige, B., Riemann, D.*, et al.* Periodic leg movements in sleep and periodic limb movement disorder: prevalence, clinical significance and treatment. *Sleep Med Rev*, **10** (3): 169-177, 2006.
73. Bixler, E.O., Kales, A., Vela-Bueno, A.*, et al.* Nocturnal myoclonus and nocturnal myoclonic activity in the normal population. *Res Commun Chem Pathol Pharmacol*, **36** (1): 129-140, 1982.
74. Youngstedt, S.D., Kripke, D.F., Klauber, M.R.*, et al.* Periodic leg movements during sleep and sleep disturbances in elders. *J Gerontol A Biol Sci Med Sci*, **53** (5): M391-394, 1998.
75. Berger, K. and Kurth, T. RLS epidemiology--frequencies, risk factors and methods in population studies. *Mov Disord*, **22** (Suppl 18): S420-423, 2007.
76. Hening, W.A., Allen, R.P., Chaudhuri, K.R.*, et al.* Clinical significance of RLS. *Mov Disord*, **22** (Suppl 18): S395-400, 2007.
77. Winkelman, J.W., Finn, L. and Young, T. Prevalence and correlates of restless legs syndrome symptoms in the Wisconsin Sleep Cohort. *Sleep Med*, **7** (7): 545-552, 2006.
78. Winkelman, J.W., Shahar, E., Sharief, I.*, et al.* Association of restless legs syndrome and cardiovascular disease in the Sleep Heart Health Study. *Neurology*, **70** (1): 35-42, 2008.
79. Walters, A.S., Hickey, K., Maltzman, J.*, et al.* A questionnaire study of 138 patients with restless legs syndrome: the 'Night-Walkers' survey. *Neurology*, **46** (1): 92-95, 1996.
80. Winkelmann, J., Wetter, T.C., Collado-Seidel, V.*, et al.* Clinical characteristics and frequency of the hereditary restless legs syndrome in a population of 300 patients. *Sleep*, **23** (5): 597-602, 2000.
81. Allen, R.P. and Earley, C.J. Defining the phenotype of the restless legs syndrome (RLS) using age-of-symptom-onset. *Sleep Med*, **1** (1): 11-19, 2000.
82. Allen, R.P., Walters, A.S., Montplaisir, J.*, et al.* Restless legs syndrome prevalence and impact: REST general population study. *Arch Intern Med*, **165** (11): 1286-1292, 2005.
83. Berger, K., Luedemann, J., Trenkwalder, C.*, et al.* Sex and the risk of restless legs syndrome in the general population. *Arch Intern Med*, **164** (2): 196-202, 2004.
84. Bjorvatn, B., Leissner, L., Ulfberg, J.*, et al.* Prevalence, severity and risk factors of restless legs syndrome in the general adult population in two Scandinavian countries. *Sleep Med*, **6** (4): 307-312, 2005.

85. Hogl, B., Kiechl, S., Willeit, J.*, et al.* Restless legs syndrome: a community-based study of prevalence, severity, and risk factors. *Neurology*, **64** (11): 1920-1924, 2005.
86. Lee, H.B., Hening, W.A., Allen, R.P.*, et al.* Race and restless legs syndrome symptoms in an adult community sample in east Baltimore. *Sleep Med*, **7** (8): 642-645, 2006.
87. Rothdach, A.J., Trenkwalder, C., Haberstock, J.*, et al.* Prevalence and risk factors of RLS in an elderly population: the MEMO study. Memory and Morbidity in Augsburg Elderly. *Neurology*, **54** (5): 1064-1068, 2000.
88. Tison, F., Crochard, A., Leger, D.*, et al.* Epidemiology of restless legs syndrome in French adults: a nationwide survey: the INSTANT Study. *Neurology*, **65** (2): 239-246, 2005.
89. Ulfberg, J., Nystrom, B., Carter, N.*, et al.* Restless Legs Syndrome among working-aged women. *Eur Neurol*, **46** (1): 17-19, 2001.
90. Ulfberg, J., Nystrom, B., Carter, N.*, et al.* Prevalence of restless legs syndrome among men aged 18 to 64 years: an association with somatic disease and neuropsychiatric symptoms. *Mov Disord*, **16** (6): 1159-1163, 2001.
91. Hadjigeorgiou, G.M., Stefanidis, I., Dardiotis, E.*, et al.* Low RLS prevalence and awareness in central Greece: an epidemiological survey. *Eur J Neurol*, **14** (11): 1275-1280, 2007.
92. Sevim, S., Dogu, O., Camdeviren, H.*, et al.* Unexpectedly low prevalence and unusual characteristics of RLS in Mersin, Turkey. *Neurology*, **61** (11): 1562-1569, 2003.
93. Tasdemir, M., Erdogan, H., Boru, U.T.*, et al.* Epidemiology of restless legs syndrome in Turkish adults on the western Black Sea coast of Turkey: A door-to-door study in a rural area. *Sleep Med*, **27**: 27, 2009.
94. Allen, R.P. and Earley, C.J. The role of iron in restless legs syndrome. *Mov Disord*, **22** (Suppl 18): S440-448, 2007.
95. Mizuno, S., Miyaoka, T., Inagaki, T.*, et al.* Prevalence of restless legs syndrome in non-institutionalized Japanese elderly. *Psychiatry Clin Neurosci*, **59** (4): 461-465, 2005.
96. Nomura, T., Inoue, Y., Kusumi, M.*, et al.* Prevalence of restless legs syndrome in a rural community in Japan. *Mov Disord*, **23** (16): 2363-2369, 2008.
97. Rangarajan, S. and D'Souza, G.A. Restless legs syndrome in an Indian urban population. *Sleep Med*, **9** (1): 88-93, 2007.
98. Tan, E.K., Seah, A., See, S.J.*, et al.* Restless legs syndrome in an Asian population: A study in Singapore. *Mov Disord*, **16** (3): 577-579, 2001.
99. Tsuboi, Y., Imamura, A., Sugimura, M.*, et al.* Prevalence of restless legs syndrome in a Japanese elderly population. *Parkinsonism Relat Disord*, **15** (8): 598-601, 2009.
100. Cho, Y.W., Shin, W.C., Yun, C.H.*, et al.* Epidemiology of restless legs syndrome in Korean adults. *Sleep*, **31** (2): 219-223, 2008.
101. Whittom, S., Dauvilliers, Y., Pennestri, M.H.*, et al.* Age-at-onset in restless legs syndrome: a clinical and polysomnographic study. *Sleep Med*, **9** (1): 54-59, 2007.
102. Ondo, W. and Jankovic, J. Restless legs syndrome: clinicoetiologic correlates. *Neurology*, **47** (6): 1435-1441, 1996.
103. Garcia-Borreguero, D., Egatz, R., Winkelmann, J.*, et al.* Epidemiology of restless legs syndrome: the current status. *Sleep Med Rev*, **10** (3): 153-167, 2006.
104. Ekbom, K.A. Restless legs syndrome. *Neurology*, **10**: 868-873, 1960.
105. O'Keeffe, S.T., Gavin, K. and Lavan, J.N. Iron status and restless legs syndrome in the elderly. *Age Ageing*, **23** (3): 200-203, 1994.
106. Sun, E.R., Chen, C.A., Ho, G.*, et al.* Iron and the restless legs syndrome. *Sleep*, **21** (4): 371-377, 1998.

107. Goodman, J.D., Brodie, C. and Ayida, G.A. Restless leg syndrome in pregnancy. *BMJ*, **297** (6656): 1101-1102, 1988.
108. Suzuki, K., Ohida, T., Sone, T.*, et al.* The prevalence of restless legs syndrome among pregnant women in Japan and the relationship between restless legs syndrome and sleep problems. *Sleep*, **26** (6): 673-677, 2003.
109. Manconi, M., Govoni, V., De Vito, A.*, et al.* Restless legs syndrome and pregnancy. *Neurology*, **63** (6): 1065-1069, 2004.
110. Manconi, M., Govoni, V., De Vito, A.*, et al.* Pregnancy as a risk factor for restless legs syndrome. *Sleep Med*, **5** (3): 305-308, 2004.
111. Collado-Seidel, V., Kohnen, R., Samtleben, W.*, et al.* Clinical and biochemical findings in uremic patients with and without restless legs syndrome. *Am J Kidney Dis*, **31** (2): 324-328, 1998.
112. Gigli, G.L., Adorati, M., Dolso, P.*, et al.* Restless legs syndrome in end-stage renal disease. *Sleep Med*, **5** (3): 309-315, 2004.
113. Hui, D.S., Wong, T.Y., Li, T.S.*, et al.* Prevalence of sleep disturbances in Chinese patients with end stage renal failure on maintenance hemodialysis. *Med Sci Monit*, **8** (5): CR331-336, 2002.
114. Huiqi, Q., Shan, L. and Mingcai, Q. Restless legs syndrome (RLS) in uremic patients is related to the frequency of hemodialysis sessions. *Nephron*, **86** (4): 540, 2000.
115. Merlino, G., Piani, A., Dolso, P.*, et al.* Sleep disorders in patients with end-stage renal disease undergoing dialysis therapy. *Nephrol Dial Transplant*, **21** (1): 184-190, 2006.
116. Molnar, M.Z., Novak, M., Ambrus, C.*, et al.* Restless Legs Syndrome in patients after renal transplantation. *Am J Kidney Dis*, **45** (2): 388-396, 2005.
117. Takaki, J., Nishi, T., Nangaku, M.*, et al.* Clinical and psychological aspects of restless legs syndrome in uremic patients on hemodialysis. *Am J Kidney Dis*, **41** (4): 833-839, 2003.
118. Kavanagh, D., Siddiqui, S. and Geddes, C.C. Restless legs syndrome in patients on dialysis. *Am J Kidney Dis*, **43** (5): 763-771, 2004.
119. Roger, S.D., Harris, D.C. and Stewart, J.H. Possible relation between restless legs and anaemia in renal dialysis patients. *Lancet*, **337** (8756): 1551, 1991.
120. Siddiqui, S., Kavanagh, D., Traynor, J.*, et al.* Risk factors for restless legs syndrome in dialysis patients. *Nephron Clin Pract*, **101** (3): c155-160, 2005.
121. Unruh, M.L., Levey, A.S., D'Ambrosio, C.*, et al.* Restless legs symptoms among incident dialysis patients: association with lower quality of life and shorter survival. *Am J Kidney Dis*, **43** (5): 900-909, 2004.
122. Winkelman, J.W., Chertow, G.M. and Lazarus, J.M. Restless legs syndrome in end-stage renal disease. *Am J Kidney Dis*, **28** (3): 372-378, 1996.
123. Wetter, T.C., Stiasny, K., Kohnen, R.*, et al.* Polysomnographic sleep measures in patients with uremic and idiopathic restless legs syndrome. *Mov Disord*, **13** (5): 820-824, 1998.
124. Enomoto, M., Inoue, Y., Namba, K.*, et al.* Clinical characteristics of restless legs syndrome in end-stage renal failure and idiopathic RLS patients. *Mov Disord*, **23** (6): 811-816, 2008.
125. Merlino, G., Valente, M., Serafini, A.*, et al.* Restless legs syndrome: diagnosis, epidemiology, classification and consequences. *Neurological Sciences*, **28** (Suppl 1): S37-S46, 2007.
126. Zucconi, M. and Ferini-Strambi, L. Epidemiology and clinical findings of restless legs syndrome. *Sleep Med*, **5** (3): 293-299, 2004.
127. Allen, R.P. and Earley, C.J. Restless legs syndrome: a review of clinical and pathophysiologic features. *J Clin Neurophysiol*, **18** (2): 128-147, 2001.

128. Hanson, M., Honour, M., Singleton, A.*, et al.* Analysis of familial and sporadic restless legs syndrome in age of onset, gender, and severity features. *J Neurol*, **251** (11): 1398-1401, 2004.
129. Bucher, S.F., Trenkwalder, C. and Oertel, W.H. Reflex studies and MRI in the restless legs syndrome. *Acta Neurol Scand*, **94** (2): 145-150, 1996.
130. Connor, J.R., Boyer, P.J., Menzies, S.L.*, et al.* Neuropathological examination suggests impaired brain iron acquisition in restless legs syndrome. *Neurology*, **61** (3): 304-309, 2003.
131. Pittock, S.J., Parrett, T., Adler, C.H.*, et al.* Neuropathology of primary restless leg syndrome: absence of specific tau- and alpha-synuclein pathology. *Mov Disord*, **19** (6): 695-699, 2004.
132. Entezari-Taher, M., Singleton, J.R., Jones, C.R.*, et al.* Changes in excitability of motor cortical circuitry in primary restless legs syndrome. *Neurology*, **53** (6): 1201-1205, 1999.
133. Quatrale, R., Manconi, M., Gastaldo, E.*, et al.* Neurophysiological study of corticomotor pathways in restless legs syndrome. *Clin Neurophysiol*, **114** (9): 1638-1645, 2003.
134. Scalise, A., Cadore, I.P. and Gigli, G.L. Motor cortex excitability in restless legs syndrome. *Sleep Med*, **5** (4): 393-396, 2004.
135. Tergau, F., Wischer, S. and Paulus, W. Motor system excitability in patients with restless legs syndrome. *Neurology*, **52** (5): 1060-1063, 1999.
136. Stiasny-Kolster, K., Haeske, H., Tergau, F.*, et al.* Cortical silent period is shortened in restless legs syndrome independently from circadian rhythm. *Suppl Clin Neurophysiol*, **56**: 381-389, 2003.
137. Bara-Jimenez, W., Aksu, M., Graham, B.*, et al.* Periodic limb movements in sleep: state-dependent excitability of the spinal flexor reflex. *Neurology*, **54** (8): 1609-1616, 2000.
138. Rijsman, R.M., Stam, C.J. and de Weerd, A.W. Abnormal H-reflexes in periodic limb movement disorder; impact on understanding the pathophysiology of the disorder. *Clin Neurophysiol*, **116** (1): 204-210, 2005.
139. Scaglione, C., Vetrugno, R., Plazzi, G.*, et al.* Group I nonreciprocal inhibition in primary restless legs syndrome. *Mov Disord*, **23** (1): 96-100, 2008.
140. Schattschneider, J., Bode, A., Wasner, G.*, et al.* Idiopathic restless legs syndrome: abnormalities in central somatosensory processing. *J Neurol*, **251** (8): 977-982, 2004.
141. Stiasny-Kolster, K., Magerl, W., Oertel, W.H.*, et al.* Static mechanical hyperalgesia without dynamic tactile allodynia in patients with restless legs syndrome. *Brain*, **127** (Pt 4): 773-782, 2004.
142. Winkelman, J.W. Considering the causes of RLS. *Eur J Neurol*, **13** (Suppl 3): 8-14, 2006.
143. Carlsson, A., Svennerholm, L. and Winblad, B. Seasonal and circadian monoamine variations in human brains examined post mortem. *Acta Psychiatr Scand Suppl*, **280**: 75-85, 1980.
144. Sowers, J.R. and Vlachakis, N. Circadian variation in plasma dopamine levels in man. *J Endocrinol Invest*, **7** (4): 341-345, 1984.
145. Clemens, S., Rye, D. and Hochman, S. Restless legs syndrome: revisiting the dopamine hypothesis from the spinal cord perspective. *Neurology*, **67** (1): 125-130, 2006.
146. Skagerberg, G. and Lindvall, O. Organization of diencephalic dopamine neurones projecting to the spinal cord in the rat. *Brain Res*, **342** (2): 340-351, 1985.

147. Lindvall, O., Bjorklund, A. and Skagerberg, G. Dopamine-containing neurons in the spinal cord: anatomy and some functional aspects. *Ann Neurol*, **14** (3): 255-260, 1983.
148. Trenkwalder, C. and Paulus, W. Why do restless legs occur at rest?--pathophysiology of neuronal structures in RLS. Neurophysiology of RLS (part 2). *Clin Neurophysiol*, **115** (9): 1975-1988, 2004.
149. Zhao, H., Zhu, W., Pan, T.*, 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** (5): 1065-1076, 2007.
150. Eisensehr, I., Wetter, T.C., Linke, R.*, et al.* Normal IPT and IBZM SPECT in drug-naive and levodopa-treated idiopathic restless legs syndrome. *Neurology*, **57** (7): 1307-1309, 2001.
151. Michaud, M., Soucy, J.P., Chabli, A.*, et al.* SPECT imaging of striatal pre- and postsynaptic dopaminergic status in restless legs syndrome with periodic leg movements in sleep. *J Neurol*, **249** (2): 164-170, 2002.
152. Staedt, J., Stoppe, G., Kogler, A.*, et al.* Nocturnal myoclonus syndrome (periodic movements in sleep) related to central dopamine D2-receptor alteration. *Eur Arch Psychiatry Clin Neurosci*, **245** (1): 8-10, 1995.
153. Turjanski, N., Lees, A.J. and Brooks, D.J. Striatal dopaminergic function in restless legs syndrome: 18F-dopa and 11C-raclopride PET studies. *Neurology*, **52** (5): 932-937, 1999.
154. Ruottinen, H.M., Partinen, M., Hublin, C.*, et al.* An FDOPA PET study in patients with periodic limb movement disorder and restless legs syndrome. *Neurology*, **54** (2): 502-504, 2000.
155. Tribl, G.G., Asenbaum, S., Happe, S.*, et al.* Normal striatal D2 receptor binding in idiopathic restless legs syndrome with periodic leg movements in sleep. *Nucl Med Commun*, **25** (1): 55-60, 2004.
156. Ondo, W.G., He, Y., Rajasekaran, S.*, et al.* Clinical correlates of 6-hydroxydopamine injections into A11 dopaminergic neurons in rats: a possible model for restless legs syndrome. *Mov Disord*, **15** (1): 154-158, 2000.
157. Qu, S., Le, W., Zhang, X.*, 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** (5): 383-388, 2007.
158. Paulus, W., Dowling, P., Rijsman, R.*, et al.* Update of the pathophysiology of the restless-legs-syndrome. *Mov Disord*, **22** (Suppl 18): S431-439, 2007.
159. Earley, C.J., Allen, R.P., Connor, J.R.*, et al.* The dopaminergic neurons of the A11 system in RLS autopsy brains appear normal. *Sleep Med*, 2009.
DOI: 10.1016/j.sleep.2009.01.006
160. Allen, R.P., Barker, P.B., Wehrl, F.*, et al.* MRI measurement of brain iron in patients with restless legs syndrome. *Neurology*, **56** (2): 263-265, 2001.
161. Earley, C.J., P, B.B., Horska, A.*, et al.* MRI-determined regional brain iron concentrations in early- and late-onset restless legs syndrome. *Sleep Med*, **7** (5): 458-461, 2006.
162. Earley, C.J., Connor, J.R., Beard, J.L.*, et al.* Abnormalities in CSF concentrations of ferritin and transferrin in restless legs syndrome. *Neurology*, **54** (8): 1698-1700, 2000.
163. Mizuno, S., Mihara, T., Miyaoka, T.*, et al.* CSF iron, ferritin and transferrin levels in restless legs syndrome. *J Sleep Res*, **14** (1): 43-47, 2005.
164. Connor, J.R., Wang, X.S., Allen, R.P.*, et al.* Altered dopaminergic profile in the putamen and substantia nigra in restless leg syndrome. *Brain*, **132** (Pt 9): 2403-2412, 2009.

165. Desai, A.V., Cherkas, L.F., Spector, T.D.*, et al.* Genetic influences in self-reported symptoms of obstructive sleep apnoea and restless legs: a twin study. *Twin Res*, **7** (6): 589-595, 2004.
166. Xiong, L., Jang, K., Montplaisir, J.*, et al.* Canadian restless legs syndrome twin study. *Neurology*, **68** (19): 1631-1633, 2007.
167. Chen, S., Ondo, W.G., Rao, S.*, et al.* Genomewide linkage scan identifies a novel susceptibility locus for restless legs syndrome on chromosome 9p. *Am J Hum Genet*, **74** (5): 876-885, 2004.
168. Allen, R.P., La Buda, M.C., Becker, P.*, et al.* Family history study of the restless legs syndrome. *Sleep Med*, **3** (Suppl 7): S3-7, 2002.
169. Walters, A.S., Picchietti, D., Hening, W.*, et al.* Variable expressivity in familial restless legs syndrome. *Arch Neurol*, **47** (11): 1219-1220, 1990.
170. Montplaisir, J., Godbout, R., Boghen, D.*, et al.* Familial restless legs with periodic movements in sleep: electrophysiologic, biochemical, and pharmacologic study. *Neurology*, **35** (1): 130-134, 1985.
171. Trenkwalder, C., Seidel, V.C., Gasser, T.*, et al.* Clinical symptoms and possible anticipation in a large kindred of familial restless legs syndrome. *Mov Disord*, **11** (4): 389-394, 1996.
172. Lazzarini, A., Walters, A.S., Hickey, K.*, et al.* Studies of penetrance and anticipation in five autosomal-dominant restless legs syndrome pedigrees. *Mov Disord*, **14** (1): 111-116, 1999.
173. Mathias, R.A., Hening, W., Washburn, M.*, et al.* Segregation analysis of restless legs syndrome: possible evidence for a major gene in a family study using blinded diagnoses. *Hum Hered*, **62** (3): 157-164, 2006.
174. Winkelmann, J., Muller-Myhsok, B., Wittchen, H.U.*, 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** (3): 297-302, 2002.
175. Winkelmann, J., Lichtner, P., Kemlink, D.*, et al.* New loci for restless legs syndrome map to chromosome 4q and 17p. *Mov Disord*, **21** (Suppl 15): S412, 2006.
176. Kemlink, D., Plazzi, G., Vetrugno, R.*, et al.* Suggestive evidence for linkage for restless legs syndrome on chromosome 19p13. *Neurogenetics*, **9** (2): 75-82, 2008.
177. Desautels, A., Turecki, G., Montplaisir, J.*, et al.* Restless legs syndrome: confirmation of linkage to chromosome 12q, genetic heterogeneity, and evidence of complexity. *Arch Neurol*, **62** (4): 591-596, 2005.
178. Kock, N., Culjkovic, B., Maniak, S.*, et al.* Mode of inheritance and susceptibility locus for restless legs syndrome, on chromosome 12q. *Am J Hum Genet*, **71** (1): 205-208; 2002.
179. Levchenko, A., Montplaisir, J.Y., Dube, M.P.*, et al.* The 14q restless legs syndrome locus in the French Canadian population. *Ann Neurol*, **55** (6): 887-891, 2004.
180. Winkelmann, J., Lichtner, P., Putz, B.*, et al.* Evidence for further genetic locus heterogeneity and confirmation of RLS-1 in restless legs syndrome. *Mov Disord*, **21** (1): 28-33, 2006.
181. Desautels, A., Turecki, G., Montplaisir, J.*, et al.* Identification of a major susceptibility locus for restless legs syndrome on chromosome 12q. *Am J Hum Genet*, **69** (6): 1266-1270, 2001.
182. Hicks, A., Rye, D., Kristjansson, K.*, et al.* Population-based confirmation of the 12q RLS locus in Iceland. *Mov Disord*, **20** (Suppl 10): S34, 2005.
183. Bonati, M.T., Ferini-Strambi, L., Aridon, P.*, et al.* Autosomal dominant restless legs syndrome maps on chromosome 14q. *Brain*, **126** (Pt 6): 1485-1492, 2003.

184. Kemlink, D., Polo, O., Montagna, P.*, et al.* Family-based association study of the restless legs syndrome loci 2 and 3 in a European population. *Mov Disord*, **22** (2): 207-212, 2007.
185. Liebetanz, K.M., Winkelmann, J., Trenkwalder, C.*, et al.* RLS3: fine-mapping of an autosomal dominant locus in a family with intrafamilial heterogeneity. *Neurology*, **67** (2): 320-321, 2006.
186. Lohmann-Hedrich, K., Neumann, A., Kleensang, A.*, et al.* Evidence for linkage of restless legs syndrome to chromosome 9p: are there two distinct loci? *Neurology*, **70** (9): 686-694, 2008.
187. Pichler, I., Marroni, F., Volpato, C.B.*, et al.* Linkage analysis identifies a novel locus for restless legs syndrome on chromosome 2q in a South Tyrolean population isolate. *Am J Hum Genet*, **79** (4): 716-723, 2006.
188. Levchenko, A., Provost, S., Montplaisir, J.Y.*, et al.* A novel autosomal dominant restless legs syndrome locus maps to chromosome 20p13. *Neurology*, **67** (5): 900-901, 2006.
189. Levchenko, A., Montplaisir, J.Y., Asselin, G.*, et al.* Autosomal-dominant locus for Restless Legs Syndrome in French-Canadians on chromosome 16p12.1. *Mov Disord*, **24** (1): 40-50, 2009.
190. McKusick-Nathans Institute of Genetic Medicine and National Center for Biotechnology Information. Online Mendelian Inheritance in Man, OMIM. Available from: http://www.ncbi.nlm.nih.gov/omim/. [accessed 02.10.2009]
191. Desautels, A., Turecki, G., Xiong, L.*, et al.* Mutational analysis of neurotensin in familial restless legs syndrome. *Mov Disord*, **19** (1): 90-94, 2004.
192. Xiong, L., Dion, P., Montplaisir, J.*, et al.* Molecular genetic studies of DMT1 on 12q in French-Canadian restless legs syndrome patients and families. *Am J Med Genet B Neuropsychiatr Genet*, **144B** (7): 911-917, 2007.
193. Winkelmann, J., Lichtner, P., Auer, D.*, et al.* Mutational analysis of divalent metal transporter 1 in restless legs syndrome. *Mov Disord*, **19** (Suppl 9): S358, 2004.
194. Desautels, A., Turecki, G., Montplaisir, J.*, et al.* Dopaminergic neurotransmission and restless legs syndrome: a genetic association analysis. *Neurology*, **57** (7): 1304-1306, 2001.
195. Desautels, A., Turecki, G., Montplaisir, J.*, et al.* Evidence for a genetic association between monoamine oxidase A and restless legs syndrome. *Neurology*, **59** (2): 215-219, 2002.
196. Wichmann, H.E., Gieger, C. and Illig, T. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen*, **67** (Suppl 1): S26-30, 2005.
197. Lin, P.I., Vance, J.M., Pericak-Vance, M.A.*, et al.* No gene is an island: the flip-flop phenomenon. *Am J Hum Genet*, **80** (3): 531-538, 2007.
198. Neale, B.M. and Sham, P.C. The future of association studies: gene-based analysis and replication. *Am J Hum Genet*, **75** (3): 353-362, 2004.
199. Steffens, M., Lamina, C., Illig, T.*, et al.* SNP-based analysis of genetic substructure in the German population. *Hum Hered*, **62** (1): 20-29, 2006.
200. Barrett, J.C., Fry, B., Maller, J.*, et al.* Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21** (2): 263-265, 2005.
201. Gautier-Sauvigne, S., Colas, D., Parmantier, P.*, et al.* Nitric oxide and sleep. *Sleep Med Rev*, **9** (2): 101-113, 2005.
202. Snyder, S.H. and Bredt, D.S. Nitric oxide as a neuronal messenger. *Trends Pharmacol Sci*, **12** (4): 125-128, 1991.

203. West, A.R., Galloway, M.P. and Grace, A.A. Regulation of striatal dopamine neurotransmission by nitric oxide: effector pathways and signaling mechanisms. *Synapse*, **44** (4): 227-245, 2002.
204. Kiss, J.P., Hennings, E.C., Zsilla, G.*, et al.* A possible role of nitric oxide in the regulation of dopamine transporter function in the striatum. *Neurochem Int*, **34** (4): 345-350, 1999.
205. Kiss, J.P., Zsilla, G. and Vizi, E.S. Inhibitory effect of nitric oxide on dopamine transporters: interneuronal communication without receptors. *Neurochem Int*, **45** (4): 485-489, 2004.
206. West, A.R. and Galloway, M.P. Endogenous nitric oxide facilitates striatal dopamine and glutamate efflux in vivo: role of ionotropic glutamate receptor-dependent mechanisms. *Neuropharmacology*, **36** (11-12): 1571-1581, 1997.
207. Xu, Q., Wink, D.A. and Colton, C.A. Nitric oxide production and regulation of neuronal NOS in tyrosine hydroxylase containing neurons. *Exp Neurol*, **188** (2): 341-350, 2004.
208. Zhu, X.Z. and Luo, L.G. Effect of nitroprusside (nitric oxide) on endogenous dopamine release from rat striatal slices. *J Neurochem*, **59** (3): 932-935, 1992.
209. Cragg, S.J. and Rice, M.E. DAncing past the DAT at a DA synapse. *Trends Neurosci*, **27** (5): 270-277, 2004.
210. Watanabe, C., Okuda, K., Sakurada, C.*, et al.* Evidence that nitric oxide-glutamate cascade modulates spinal antinociceptive effect of morphine: a behavioural and microdialysis study in rats. *Brain Res*, **990** (1-2): 77-86, 2003.
211. Walters, A.S., Winkelmann, J., Trenkwalder, C.*, et al.* Long-term follow-up on restless legs syndrome patients treated with opioids. *Mov Disord*, **16** (6): 1105-1109, 2001.
212. Rowe, A.K., Powell, K.E. and Flanders, W.D. Why population attributable fractions can sum to more than one. *Am J Prev Med*, **26** (3): 243-249, 2004.
213. Azcoitia, V., Aracil, M., Martinez, A.C.*, et al.* The homeodomain protein Meis1 is essential for definitive hematopoiesis and vascular patterning in the mouse embryo. *Dev Biol*, **280** (2): 307-320, 2005.
214. Maeda, R., Mood, K., Jones, T.L.*, et al.* Xmeis1, a protooncogene involved in specifying neural crest cell fate in Xenopus embryos. *Oncogene*, **20** (11): 1329-1342, 2001.
215. Mercader, N., Leonardo, E., Azpiazu, N.*, et al.* Conserved regulation of proximodistal limb axis development by Meis1/Hth. *Nature*, **402** (6760): 425-429, 1999.
216. Dasen, J.S., Tice, B.C., Brenner-Morton, S.*, et al.* A Hox regulatory network establishes motor neuron pool identity and target-muscle connectivity. *Cell*, **123** (3): 477-491, 2005.
217. Allen Institute for Brain Science. Allen Brain Atlas. Available from: http://www.brain-map.org. [accessed 02.10.2009]
218. Perez-Torrado, R., Yamada, D. and Defossez, P.A. Born to bind: the BTB protein-protein interaction domain. *Bioessays*, **28** (12): 1194-1202, 2006.
219. Stogios, P.J., Downs, G.S., Jauhal, J.J.*, et al.* Sequence and structural analysis of BTB domain proteins. *Genome Biol*, **6** (10): R82, 2005.
220. Stefansson, H., Rye, D.B., Hicks, A.*, et al.* A genetic risk factor for periodic limb movements in sleep. *N Engl J Med*, **357** (7): 639-647, 2007.
221. Dinev, D., Jordan, B.W., Neufeld, B.*, et al.* Extracellular signal regulated kinase 5 (ERK5) is required for the differentiation of muscle cells. *EMBO Rep*, **2** (9): 829-834, 2001.

222. Cavanaugh, J.E., Jaumotte, J.D., Lakoski, J.M.*, et al.* Neuroprotective role of ERK1/2 and ERK5 in a dopaminergic cell line under basal conditions and in response to oxidative stress. *J Neurosci Res*, **84** (6): 1367-1375, 2006.
223. Mizuhara, E., Nakatani, T., Minaki, Y.*, et al.* Corl1, a novel neuronal lineage-specific transcriptional corepressor for the homeodomain transcription factor Lbx1. *J Biol Chem*, **280** (5): 3645-3655, 2005.
224. Gross, M.K., Dottori, M. and Goulding, M. Lbx1 specifies somatosensory association interneurons in the dorsal spinal cord. *Neuron*, **34** (4): 535-549, 2002.
225. Uetani, N., Chagnon, M.J., Kennedy, T.E.*, et al.* Mammalian motoneuron axon targeting requires receptor protein tyrosine phosphatases sigma and delta. *J Neurosci*, **26** (22): 5872-5880, 2006.
226. Pulido, R., Krueger, N.X., Serra-Pages, C.*, et al.* Molecular characterization of the human transmembrane protein-tyrosine phosphatase delta. Evidence for tissue-specific expression of alternative human transmembrane protein-tyrosine phosphatase delta isoforms. *J Biol Chem*, **270** (12): 6722-6728, 1995.
227. Sato, M., Takahashi, K., Nagayama, K.*, et al.* Identification of chromosome arm 9p as the most frequent target of homozygous deletions in lung cancer. *Genes Chromosomes Cancer*, **44** (4): 405-414, 2005.
228. Nair, P., De Preter, K., Vandesompele, J.*, et al.* Aberrant splicing of the PTPRD gene mimics microdeletions identified at this locus in neuroblastomas. *Genes Chromosomes Cancer*, **47** (3): 197-202, 2008.
229. Haiman, C.A., Patterson, N., Freedman, M.L.*, et al.* Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet*, **39** (5): 638-644, 2007.
230. Evans, D.M., Visscher, P.M. and Wray, N.R. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum Mol Genet*, **18** (18): 3525-3531, 2009.
231. Wray, N.R., Goddard, M.E. and Visscher, P.M. Prediction of individual genetic risk of complex disease. *Curr Opin Genet Dev*, **18** (3): 257-263, 2008.
232. Yang, Q., Khoury, M.J., Botto, L.*, et al.* Improving the prediction of complex diseases by testing for multiple disease-susceptibility genes. *Am J Hum Genet*, **72** (3): 636-649, 2003.
233. Janssens, A.C., Pardo, M.C., Steyerberg, E.W.*, et al.* Revisiting the clinical validity of multiplex genetic testing in complex diseases. *Am J Hum Genet*, **74** (3): 585-588; 2004.
234. Hanley, J.A. and McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143** (1): 29-36, 1982.
235. Janssens, A.C. and van Duijn, C.M. Genome-based prediction of common diseases: advances and prospects. *Hum Mol Genet*, **17** (R2): R166-173, 2008.
236. Kemlink, D., Polo, O., Frauscher, B.*, et al.* Replication of restless legs syndrome loci in three European populations. *J Med Genet*, **46** (5): 315-318, 2009.
237. Vilarino-Guell, C., Farrer, M.J. and Lin, S.C. A genetic risk factor for periodic limb movements in sleep. *N Engl J Med*, **358** (4): 425-427, 2008.
238. Schormair, B., Kemlink, D., Roeske, D.*, et al.* PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome. *Nat Genet*, **40** (8): 946-948, 2008.
239. 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*, **39** (8): 1000-1006, 2007.
240. Winkelmann, J., Polo, O., Provini, F.*, et al.* Genetics of restless legs syndrome (RLS): State-of-the-art and future directions. *Mov Disord*, **22** (Suppl 18): S449-458, 2007.

241. Phillips, B., Young, T., Finn, L.*, et al.* Epidemiology of restless legs symptoms in adults. *Arch Intern Med*, **160** (14): 2137-2141, 2000.
242. Xiong, L., Catoire, H., Dion, P.*, et al.* MEIS1 intronic risk haplotype associated with restless legs syndrome affects its mRNA and protein expression levels. *Hum Mol Genet*, **18** (6): 1065-1074, 2009.
243. Bodmer, W. and Bonilla, C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet*, **40** (6): 695-701, 2008.
244. Zollner, S. and Pritchard, J.K. Overcoming the winner's curse: estimating penetrance parameters from case-control data. *Am J Hum Genet*, **80** (4): 605-615, 2007.

5 Appendices

Published and submitted manuscripts and personal contributions to them

Appendix 1

Research Article in *Movement Disorders*

Variants in the Neuronal Nitric Oxide Synthase (*nNOS, NOS1*) Gene are Associated with Restless Legs Syndrome

Juliane Winkelmann, Peter Lichtner, Barbara Schormair, Manfred Uhr, Stephanie Hauk, Karin Stiasny-Kolster, Claudia Trenkwalder, Walter Paulus, Ines Peglau, Ilonka Eisensehr, Thomas Illig, H.-Erich Wichmann, Hildegard Pfister, Jelena Golic, Thomas Bettecken, Benno Pütz, Florian Holsboer, Thomas Meitinger & Bertram Müller-Myhsok

Movement Disorders 23 (3), 350–358, 2008

Contributions:

In this study I performed parts of the stage 2 and 3 genotyping, and the genotyping of the genomic control SNPs for the population stratification analysis, using MALDI-TOF mass spectrometry on a Sequenom MassArray system. I participated in writing of the manuscript, designed figure 2 of the manuscript and all supplementary tables.

Movement Disorders
Vol. 23, No. 3, 2008, pp. 350–358
© 2007 Movement Disorder Society

Research Articles

Variants in the Neuronal Nitric Oxide Synthase (nNOS, NOS1) Gene are Associated with Restless Legs Syndrome

Juliane Winkelmann, MD,[1,2,3]* Peter Lichtner, PhD,[1,3] Barbara Schormair,[1,3] Manfred Uhr, MD,[2] Stephanie Hauk, MD,[1,3] Karin Stiasny-Kolster, MD,[4] Claudia Trenkwalder, MD,[5] Walter Paulus, MD,[6] Ines Peglau, MD,[7] Ilonka Eisensehr, MD,[8] Thomas Illig, PhD,[9] H.-Erich Wichmann, MD,[9] Hildegard Pfister,[2] Jelena Golic,[3] Thomas Bettecken, MD,[2] Benno Pütz, PhD,[2] Florian Holsboer, MD, PhD,[2] Thomas Meitinger, MD,[1,3] and Bertram Müller-Myhsok, MD[2]

[1]*Institute of Human Genetics, GSF-National Research Center for Environment and Health, Munich, Germany*
[2]*Max Planck Institute of Psychiatry, Munich, Germany*
[3]*Institute of Human Genetics, Technical University, Munich, Germany*
[4]*Department of Neurology, Philipps University Marburg, Marburg, Germany*
[5]*Paracelsius-Elena Klinik, Kassel, Germany*
[6]*Department of Clinical Neurophysiology, University of Göttingen, Göttingen, Germany*
[7]*Neurologische Praxis, Berlin, Germany*
[8]*Neurologische Praxis Sendlingerstrasse, Munich, Germany*
[9]*Institute of Epidemiology, GSF-National Research Center for Environment and Health, Munich, Germany*

Abstract: Sixty percent of the patients with restless legs syndrome (RLS) report a positive family history. To date five loci have been mapped on chromosome 12q, 14q, 9p, 2q, and 20p (RLS1-5) but no gene has been identified so far. To identify genes related to RLS, we performed a three-stage association study (explorative study, replication study, high-density mapping) in two Caucasian RLS case-control samples of altogether 918 independent cases and controls. In the explorative study (367 cases and controls, respectively), we screened 1536 SNPs in 366 genes in a 21 Mb region encompassing the RLS1 critical region on chromosome 12. Armitage trend test revealed three genomic regions that were significant ($P < 0.05$). In the replication study (551 cases and controls, respectively) we genotyped the most significant SNPs of Stage 1. After correction for multiple testing, association was observed with SNP rs7977109 ($P_{\text{nominal}} = 0.00175$, $P_{\text{Westfall-Young}} = 0.04895$, OR = 0.76228, 95% CI = 0.64310–0.90355), which is in the neuronal nitric oxide synthase *(NOS1)* gene. High-density mapping using altogether 34 tagging and coding SNPs of the *NOS1* gene in both case-control samples further confirmed the significant association results to *NOS1*. Ten more SNPs revealed significance with nominal *P*-values from 0.0001 to 0.0482 (genotypic test and Armitage test). Altogether, this study provides evidence for an association of variants in the *NOS1* gene and RLS, and suggests the involvement of the NO/arginine pathway in the pathogenesis of RLS. Potential usage of NO modulating agents as new treatment options for RLS have become a challenging aspect for future research of this disorder. © 2007 Movement Disorder Society

Key words: restless legs syndrome; genetics; sleep; association study; NOS.

Restless legs syndrome (RLS, *102300) is one of the most common neurological disorders with an age-dependent prevalence of up to 10% in the population older than 65 years.[1] RLS is clinically characterized by dysaesthesias usually in the lower limbs, mainly the calves, associated with an irresistible urge to move these limbs. The symptoms occur predominantly at rest, which are relieved by movement, and are worse at night, resulting in nocturnal insomnia and sleep deprivation.[2] Up to 60% of idiopathic RLS cases report a positive family history pointing to an important genetic contribution to the phenotype, although no disease-causing gene has been identified yet. Several studies demonstrated that an earlier age at onset of the disease is correlated with a higher familial contribution of RLS.[3-7]

This article contains supplementary material available via the Internet at http://www.interscience.wiley.com/jpages/0885-3185/suppmat.

*Correspondence to: Dr. Juliane Winkelmann, Institute of Human Genetics, GSF-National Research Center for Environment and Health, Ingolstädter Landstrasse 1, D-85764 Munich-Neuherberg, Germany. E-mail: julianewinkelmann@t-online.de

Received 6 March 2007; Revised 31 May 2007; Accepted 31 May 2007

Published online 30 November 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/mds.21647

Genome-wide linkage analysis in RLS families have identified so far five-loci (RLS1, chromosome 12q; RLS2 chromosome 14q; RLS3 chromosome 9p; RLS4 chromosome 2q; and RLS5 chromosome 20p) based on a recessive in the first and dominant mode of inheritance in the latter cases.[8-12] Additional studies provided evidence for further genetic heterogeneity.[13-15] The most prominent among the known loci is RLS1 at chromosome 12, which was confirmed in a number of independent families of French Canadian origin,[14] in Icelandic families,[16] as well as families of Bavarian origin using TDT statistics.[15] Analyses of candidate genes within this region have not yet led to the identification of disease-causing mutations in RLS patients.[17,18]

To identify genes related to RLS, we performed an association study in a set of unrelated RLS cases and controls screening the RLS1 locus on chromosome 12. Here, we report the findings from this scan that identified the neuronal nitric oxide synthase (nNos, NOS1, *163731) associated to RLS.

PATIENTS AND METHODS

Study Design

This is a three-stage design association study consisting of two Caucasian case-control samples. In Stage 1 (explorative study), we genotyped SNPs in all known transcripts of the RLS1 linkage region on chromosome 12 in 367 unrelated Caucasian RLS cases and 367 age and sex matched controls. In Stage 2 (replication study), the most significant SNPs from Stage 1 were genotyped in an independent Caucasian RLS sample of 551 unrelated cases and 551 age and sex matched controls. In Stage 3, a high-density mapping of the identified gene was performed in both case-control samples.

Sample-Set Characteristics

Cases of Stages 1 and 2 were diagnosed according to the diagnostic criteria of the International RLS Study Group[2] within the bounds of a personal interview conducted by RLS experts from 2000 to 2006. Age of Stage 1 cases was 57.44 ± 9.66 years mean [x ± SD (σ)], (263 women, 104 men), age at onset 35.58 ± 15.73 years (x ± σ), positive family history: n = 295, sporadic: n = 72. Age of Stage 2 cases was 61.03 ± 10.59 years (x ± σ, 390 women, 161 men), age at onset 41.51 ± 18.53 years (x ± σ), positive family history: n = 263, sporadic: n = 182, unknown family history: n = 106. Controls of Stage 1 (57.42 ± 9.63 years, x ± σ) were recruited for the absence of psychiatric phenotypes (depression and anxiety) but were also RLS negative according to the diagnostic criteria of the International RLS Study Group.[2] Controls of Stage 2 were age and sex matched population-based controls of the KORA (S4) study (age 52.02 ± 15.45 years, x ± σ).[19] Genotyping of all samples was performed with written informed consent from all participating individuals.

SNP Selection

Stage 1 SNP Selection.

We selected 1536 SNPs for Illumina Golden Gate assays on the basis of public information available in May 2005. A 21.31-Mb interval in the RLS1 linkage region on chromosome 12q23.1-12q24.31 [98,890,000–120,200,000; Human May 2004 (hg17) assembly] was targeted for SNP genotyping. This region contained 366 Ensemble transcripts that spanned 12.5 Mb of genomic sequence. We selected tagging SNPs for these genes including 10 kb of flanking sequence using the Perlegen Linkage Disequilibrium Map Data (European American sample) (http://genome.perlegen.com/browser/download.html). Perlegen grouped SNPs into bins of high LD, where at least one tagging SNP has $r^2 > 0.8$ with every other SNP in the bin, using the algorithm of Carlson et al.[20] The minor allele frequency (MAF) of the selected SNPs was >0.1. Altogether 1176 bins spanning the 366 Ensemble genes were identified. Of these, 1,029 tagging SNPs could be converted into Illumina Golden Gate assays. For the remaining 147 bins the assay design for tagging SNPs failed. In addition, we included 323 non-synonymous and synonymous coding and splice-site SNPs. In a final step, 184 SNPs from dbSNP and HapMap with a MAF > 0.1 were added to provide Ensemble genes with SNPs that could not be captured by Perlegen tagging SNPs and to enrich SNP density of large genes. The complete list of 1,536 SNPs genotyped, along with flanking sequence and expected alleles as well as the genotyping results is available in supplementary Table 1, available at http://www.interscience.wiley.com/jpages/0885-3185/suppmat.

Stage 2 SNP Selection.

A subset of 24 SNPs out of the 1,536 SNP set was genotyped in the replication sample using a Sequenom MassArray system. These SNPs were selected as the most significant in the Armitage trend test after removing SNPs with low call rates (≤90%) and with a significant deviation from Hardy-Weinberg equilibrium (HWE) ($P < 0.001$).

Stage 3 (NOS1) SNP Selection.

As we found the significantly associated SNP in the replication sample to be within the *NOS1* gene, a SNP

selection for high-density mapping of *NOS1* was performed on the basis of HapMap data (release no. 19/phase II October 2005). Genotype data from the CEPH sample (Utah residents with northern and western European ancestry) spanning the coding region and 10 kb of flanking sequence (chr12:116,052,931–116,221,534; Human May 2004 (hg17) assembly) were downloaded from the International HapMap Project. The Tagger implementation of Haploview 3.2[21] was used to select an optimal set of available tagging SNPs with MAF > 0.1 and an r^2 threshold of 0.8. Six nonsynonymous coding SNPs, one splice-site SNP and the five *NOS1* SNPs, from Stage 2 were added to the set, which brought the total to 39. Thirty-six SNPs could be converted into genotyping assays for a Sequenom MassArray system.

Genotyping

Stage 1 Genotyping (Golden Gate Assay, Illumina). Genotyping was performed using an Illumina Bead Station 500G system in accordance with the manufacturer's standard recommendations. Ninety-nine percent of samples (n = 726) were successfully genotyped. Each genotype was labeled with a quality score calculated by proprietary Illumina algorithms (GC_Score). Genotypes with a GC_Score > 0.25 were included in the analyses. Two-hundred twelve SNPs were discarded from subsequent analyses for the following reasons: low call rate (<80%) or low mean GC_Score (<0.4) (n = 88), low heterozygosity (<3 heterozygotes) (n = 112), significant deviations from HWE ($P < 0.001$ in controls) (n = 12). Thus a total of 1324 SNPs could be used for the subsequent analyses. By genotyping two samples in eight replicates we calculated a genotype error rate of 0.38%.

Stage 2 and 3 Genotyping (MassArray system, Sequenom). According to the Stages 2 and 3 SNP selection criteria (see above) 56 SNPs (Stage 2: 24 SNPs in 1,102 samples; Stage 3: 32 SNPs in 1,836 samples) were genotyped using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry on a Sequenom MassArray system. Cleaned extension products were analyzed by a mass spectrometer (Bruker Daltronik) and peaks were identified using the SpectroTYPER RT 3.3 software (Sequenom). Assays were designed by the AssayDesign software 3.0 (Sequenom) with the default parameters for the iPLEX and hME chemistry. Assay quality was scored blinded to the phenotype before the results were subjected to statistical analysis. Fifty SNPs produced genotypes, six SNPs were disregarded from further analyses (exclusion criteria: call rate $< 95\%$, HWE $P < 0.001$ in controls). From 7071 duplicate genotypes a genotype error rate of 0.08% was calculated. Five SNPs were genotyped in 730 samples on both platforms Illumina and Sequenom. The concordance rate of these genotypes was 99.59%.

DNA Sequencing

Genomic resequencing was done for the *NOS1* gene and included all 28 coding exons as well as 20–30 bp of the exon/intron junctions. Primers were designed using the ExonPrimer software (http://ihg.gsf.de/ihg/ExonPrimer.html). PCR amplified exons were sequenced directly by BigDye Terminator Cycle sequencing kit (Applied Biosystems) and were analyzed on an ABI3730 sequencer.

Statistical Analysis

Exact tests for a deviation from HWE were first performed in both case and control samples. SNPs with a significant deviation from HWE ($P < 0.001$ in controls) were removed from the analysis, as well as SNPs failing criteria for low call rates and overall quality (see above). Single locus analysis for case-control associations was performed using the Cochran-Armitage test[22,23] for linear trend. All calculations were performed using R 2.3.1. (http://www.r-project.org/). Correction for multiple testing in Stages 2 and 3 was done using the Westfall-Young minimum resampling based p correction based on a thousand permutations and correcting for both the use of Armitage and genotypic tests.[22-24] Genotype tests were performed using the Fisher exact procedure on a 2×3 table as implemented in R 2.0.1. The odds ratios given are the common odds ratios obtained in the Cochran-Armitage linear test for trend, which we formulated as a logistic regression with disease status as dependent variable and the number of copies of the minor allele as the independent variable. The odds ratio is then obtained as e^{β}, where β is the estimate of the slope of the regression line. LD structure of *NOS1* was obtained using Haploview 3.2.[21]

Population Stratification.

To exclude differences because of population stratification we estimated the amount of stratification in cases and controls performing a genomic control analysis genotyping 79 unlinked SNPs in genomic desert regions and applying the method described in Ref. 20. We estimated the factor lambda (λ) as the median of the distribution of the χ^2 values of the allelic tests using the 79 SNPs.[25] Comparison was performed between cases and controls in both the exploratory and the replication samples, but also between the two case samples and the two control samples. The list of SNPs is available on request.

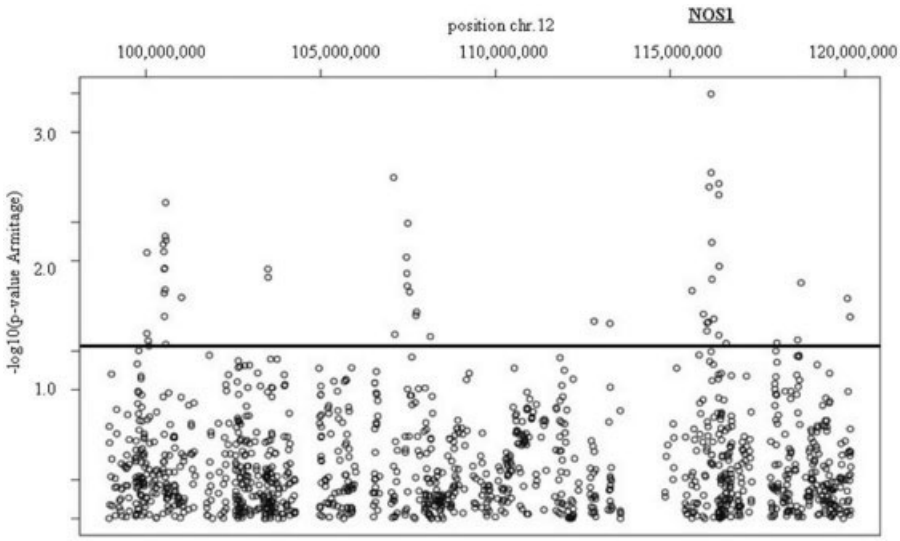


FIG. 1. *P*-values of 1324 SNPs genotyped in Stage 1 (explorative study). Only SNPs that passed all quality criteria were included. P values are specified as the –log10 of the Armitage test *P*-values. The threshold of a nominal *P*-value of 0.05 is marked with a horizontal line. Approximate position of NOS1 is shown. Genomic positions refer to the Human May 2004 (hg17) assembly. SNPs are indicated as (○).

RESULTS

Stage 1: Explorative Case-Control Study

Analysis with the Armitage and genotypic test revealed altogether 79 SNPs reaching a significant level of $P_{\text{nominal}} < 0.05$ (supplementary Table 2a). These showed a prominent clustering in the three following chromosomal regions. chr12:99,792,852–101,016,651 (17 SNPs, Armitage test; genes: *TMEM16D, SLC5A8, MYBPC1, NUP37* chr12:107,084,765–108,133,377 (10 SNPs, Armitage test; genes: *KIAA0789, SART3, ISCU, LOC338773, CORO1C, SSH1, ACACB*), and chr12: 115,621,107–116,597,588 (16 SNPs, Armitage test; genes: *FLJ21415*, *TSC*, *FBXO21*, *NOS1*, *KSR2*). The most significant signals in the Armitage test were obtained with the synonymous cSNP rs2293054 ($P_{\text{nominal}} = 0.0005$) and the intron SNP rs6490121 ($P_{\text{nominal}} = 0.0021$), both located in the *NOS1* gene (Fig. 1).

Stage 2: Replication Case-Control Study

To achieve a balance between power and cost-efficient genotyping, the 24 most significant SNPs (Armitage test: $0.00055 < P_{\text{nominal}} < 0.01505$) of Stage 1 were genotyped in the replication case-control sample (551 cases and 551 controls). A single SNP mapping to intron 3 of the *NOS1* gene reached the level of significance before and after correction for multiple testing [rs7977109: $P_{\text{nominal}} = 0.00153$ (genotypic test), $P_{\text{Westfall-Young}} = 0.04595$ (genotypic test), $P_{\text{nominal}} = 0.00175$ (Armitage test), $P_{\text{Westfall-Young}} = 0.04895$ (Armitage test)]. The other 23 SNPs that were significant in the explorative study sample were not confirmed in the replication sample (supplementary Table 2b).

Stage 3: High-Density Mapping of NOS1

In addition to five SNPs typed in both Stages 1 and 2, another 29 SNPs were used for fine mapping the *NOS1* gene in both the explorative and the replication samples. A total of 10 more SNPs (excluding NOS1 SNPs from Stage 1 were significantly associated with the RLS phenotype ($P_{\text{nominal}} < 0.05$). These SNPs are located in different regions of the *NOS1* gene. Two SNPs had nominally significant *P*-values in both samples (rs7977109, rs693534), whereas eight and four SNPs showed significance only in the explorative or replication sample, respectively (Armitage test and/or genotypic test) (supplementary Table 3, Fig. 2). Correction for multiple testing revealed three SNPs to be significant in the explorative (rs4766836, rs2293054, rs6490121) and the replication (rs7977109, rs530393, rs816292) sample. The LD pattern of NOS1 for population-based KORA controls is shown in Figure 3. Overall, the case samples show higher mean D′ and mean r^2 values than the control samples (Table 1).

NOS1 Resequencing (NM_000620)

To search for additional potential risk-associated sequence changes in the coding part of *NOS1*, we resequenced 23 RLS cases from the replication study cohort. This set of individuals was chosen from homozygotes of the risk allele in SNP rs7977109 providing the strongest signal in the replication case-control sample. In total, six distinct sequence variants were detected; four of these were already described in dbSNP (rs2293054, rs1047735, rs3741475, rs2293044). Two variants unre-

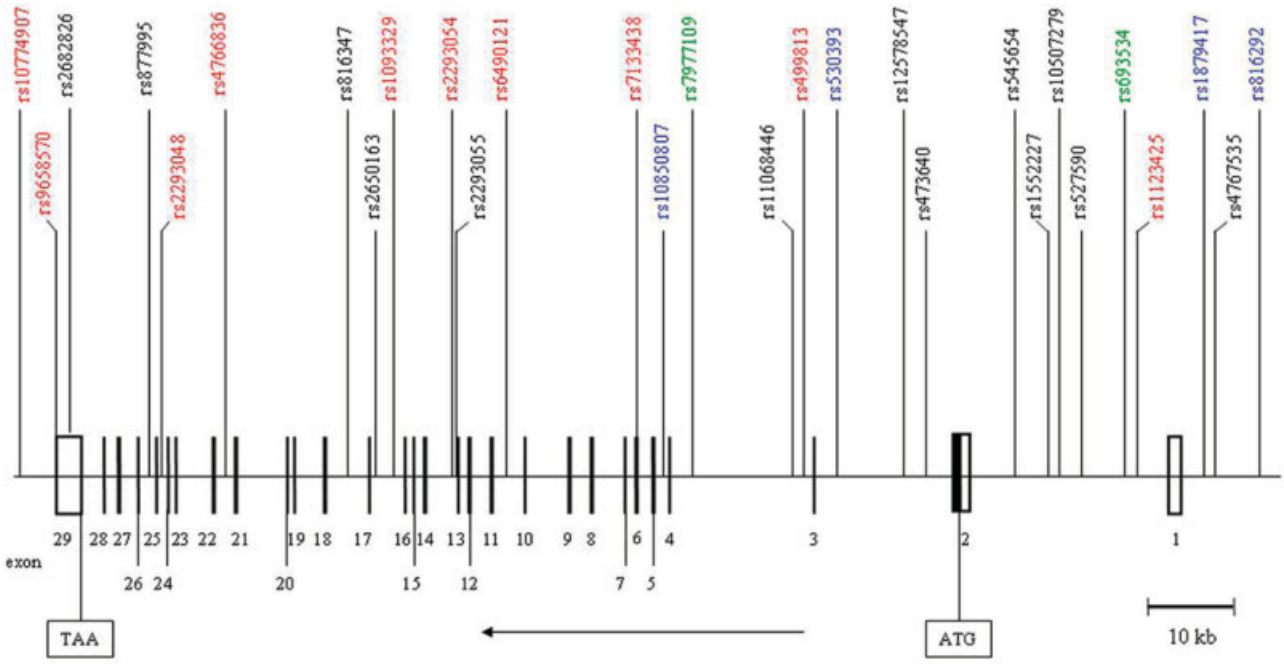


FIG. 2. Genomic organization of the *NOS1* gene and positions of the genotyped SNPs in Stage 3 (high-density maping). White boxes denote untranslated exons or 5′-UTR and 3′-UTR. Black boxes denote coding sequences. The exon-intron structure is drawn to scale. The start codon (ATG) and the stop codon (TAA) are shown. The arrow specifies the direction of transcription on the reverse strand of chromosome 12. Only SNPs that passed all quality criteria were included. SNPs are marked with the following color code: black, no association result; red, association result in the explorative study; blue, association result in the replication study; green, association result in the explorative and replication study. An association result is defined as $P_{nominal} < 0.05$ in the genotypic test or/and Armitage test. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

corded in dbSNP as of December 2005 were each identified in one heterozygote individual. One variant in exon 14 (c.2280G > A) is a synonymous coding SNP; one variant in exon 11 (c.1855A > T, p.M619L) is a nonsynonymous coding SNP. This variant was found in two of 551 RLS cases of the replication sample. Genotyping of 4,022 individuals from the German epidemiological cohort KORA S4[19] revealed a MAF of 0.35% (28 heterozygotes).

Population Stratification

Inflation factors were very close to or actually below 1 indicating that population stratification does not account for the observed *P*-values (Table 2).

DISCUSSION

This is the first systematic investigation of common polymorphisms in a known RLS linkage locus (RLS1) in two large collections of cases and controls. This hypothesis-free approach points to a possible involvement of the NO/arginine pathway in RLS disease susceptibility.

We had chosen the RLS1 linkage region given its established confirmation in independent families and populations.[14-16] Since intergenic regions were excluded from the SNP selection, the SNP set analyzed in the explorative study covered ~12.5 Mb (58%) of the entire sequence resulting in a SNP coverage of about 1 SNP per 9 kb. Following the initial screening of 1536 SNPs in an explorative case-control sample and the replication in a second independent larger sample, we obtained significant evidence for an association of RLS with sequence variations in the *NOS1* gene. Subsequent high-density analysis of the *NOS1* gene revealed several highly associated SNPs in both the explorative and the replication sample.

It was our strategy to conduct the explorative study with "extreme phenotypes" in both cases and controls. The cases of the explorative sample revealed an earlier age at onset of the disease in comparison to the replication sample (35.58 ± 15.73 years vs. 41.51 ± 18.53 years) and were enriched for familial cases (80% vs. 59%). The controls of the explorative sample can be regarded as controls in the classic sense, because they were recruited for the absence of any symptoms of RLS. In contrast, the controls of the replication study were selected from a reference population originally drawn for a population-based epidemiological survey and were not

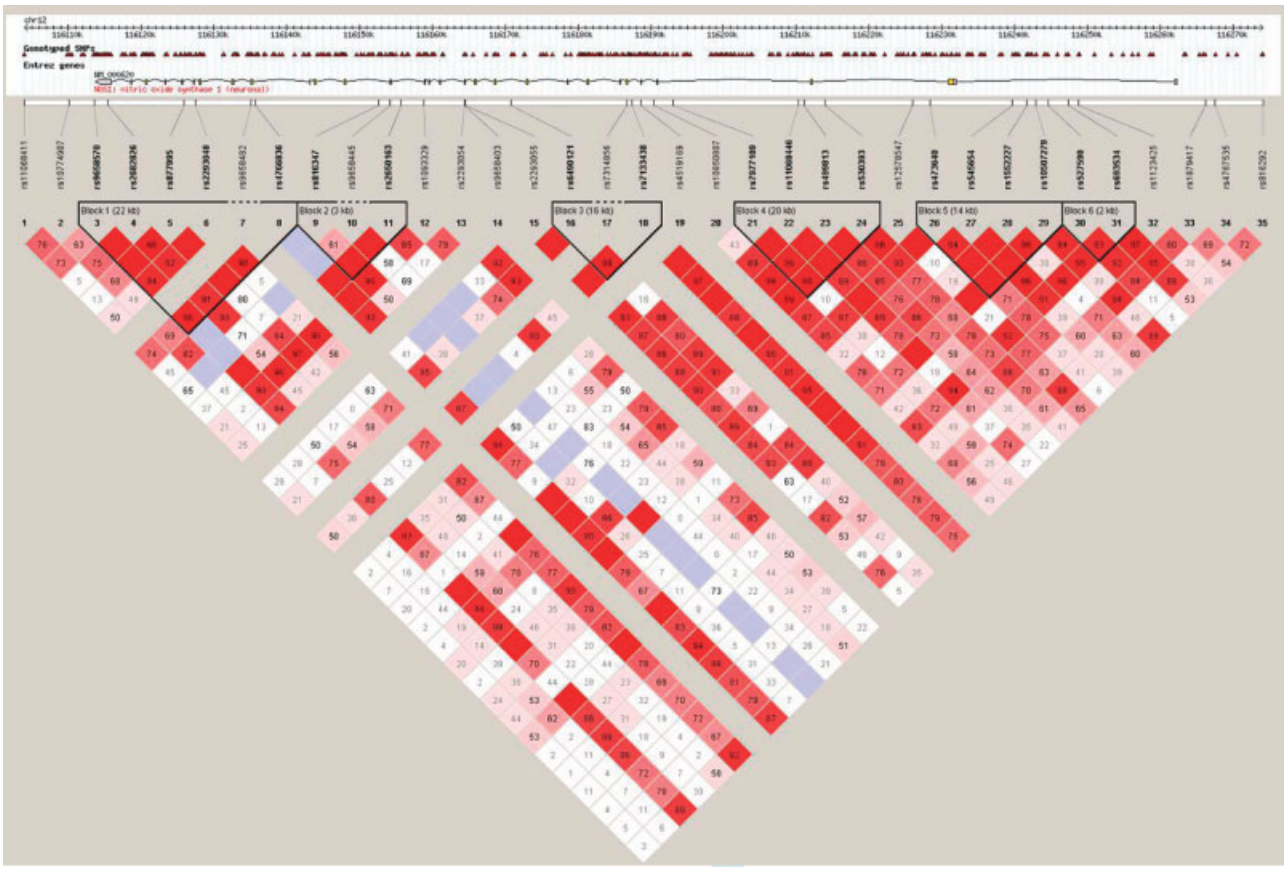


FIG. 3. LD structure of the *NOS1* gene in the population-based KORA controls of the replication sample. The lower panels show pairwise LD between SNPs. The value within each diamond represents the pairwise correlation between SNPs (measured as D′) defined by the top left and top right sides of the diamond. Diamonds without a number correspond to D′ = 1. The shading indicates the magnitude and significance of pairwise LD, with a red to white gradient reflecting higher to lower LD values. The block structure was defined by the algorithm of Gabriel et al.[26] [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

examined for possible symptoms of RLS,[19] thus they are a population reference sample rather than a classical control sample. The selection of extreme phenotypes in the explorative study should increase the chance to detect SNPs with small effects. However, also the number of not generalizable signals is expected to be increased. The second stage was then used to select SNPs that can be replicated and potentially generalized. For this second stage, we selected controls from the general population,[19] and a case group less biased for early onset and familial occurrence of the phenotype. The selection of SNPs to be

Table 1. *LD structure of the NOS1 gene in the four study samples*

Study	D′_mean	r^2_mean	tSNPs
Explorative study cases	0.639	0.099	31
Explorative study controls	0.591	0.096	34
Replication study cases	0.601	0.104	34
Replication study controls	0.593	0.097	34

LD is given as the average D′ (D′_mean) or average r^2 (r^2_mean) from all pairwise LD values obtained from Haploview 3.2. The column tSNPs indicates the number of tagging SNPs needed for covering the NOS1 gene variation using the Tagger implementation in Haploview 3.2 with pairwise tagging only and an r^2 threshold of 0.8. Analyses are based on the dataset from stage 3 (NOS1 high-density mapping).

TABLE 2. *Inflation factor* λ

Comparison	estimate of λ (unrestricted)	(95 % confidence interval for pairwise comparison of populations)
Case-control explorative	0.7360	(0.4734;1.4726)
Case-control replication	1.0057	(0.5494;1.8786)
Controls vs. controls	0.8362	(0.6259;1.1706)
Cases vs. cases	0.7125	(0.3937;0.9316)

The estimates of l are obtained as the medians of the obtained χ^2 statistics divided the median of the χ^2 distribution with 1 df. The estimates obtained are all below or very close to 1, the theroretically expected value under no stratification. This indicates that results are not due to population stratification.

genotyped was based on the results of the Armitage test only, as the genotypic and the Armitage test were found to be highly correlated.

However, obvious limitations of the study have to be considered. When analyzing single SNPs the location of the associated SNPs within the *NOS1* gene varied between the two samples (Fig. 2). This finding is similar to the situation in other complex phenotypes such as schizophrenia with discrepancies of the location of the SNPs within the dysbindin gene in the original and subsequent studies.[27] In this analysis two SNPs (rs7977109 and rs693534) showed significant association in the explorative and replication sample but showed different allele frequencies with in part opposite directions. This implies that the same allele is a risk allele in one but a protective allele in the other sample which cannot be readily explained. One obvious reason could be that the association reflects a general difference between our samples rather than an association of *NOS1* with RLS. Following we have investigated this possibility using the method of genomic controls.[25] Genotyping SNPs in genomic desert regions showed that sample stratification is most likely not the reason for this observation. A further explanation could be insufficient coverage of the *NOS1* gene, with various directions of association between truly causative variants and the SNPs studied. This is theoretically possible, but an unlikely situation in itself. Using the method of Carlson et al. ($r^2 > 0.8$), we assume a complete coverage of the *NOS* gene in the fine-mapping stage of the study. Furthermore, we exclude technical reasons as we have repeated the experiments and genotyping for the significant SNPs. An alternative but speculative explanation could be that the discrepancy in the findings between the two samples is based on an interaction of NOS with environmental factors. It appears possible, given that NOS is active in the metabolism of arginine, and levels of arginine may be variable also due to dietary habits. Finally, the differences may also be due to simply chance. A validation has only been performed in a single replication sample. Therefore, further studies in independent populations are needed to replicate and confirm our finding. Interestingly, a very recent publication discussed and investigated possible reasons for the association in varying directions coining the term flip flop phenomenon.[28] These authors conclude that association in opposite directions should not be discarded from further analysis and study as they may very well point to a true association.[28]

D' and r^2 are used to calculate the linkage disequilibrium (LD) between two makers. Overall, the case samples show higher mean D' and mean r^2 values than the control samples. While in overall the control samples appear to be fairly similar in mean D' and mean r^2 we note that the cases of the explorative sample show the highest D' and (although the mean r^2 is not the highest for all subsets) need the lowest number of tagging SNPs (31 vs. 34 for all other samples) to cover NOS1 variation. Given that in the explorative sample both cases and controls were recruited from Bavaria and not from all of Germany as in the replication case sample, this may indicate a possible founder effect in the explorative case sample. Performing genomic control analysis we saw no evidence for population admixture, which provides a baseline on which to judge our findings obtained with the *NOS1* gene.

From our results of the case-control study, we hypothesize that genetic variation in the *NOS1* gene is relevant for the etiology of RLS. The enzyme NOS catalyzes the synthesis of nitric oxide (NO) and the by-product L-citrulline in two steps from L-arginine and molecular oxygen in a Ca^{2+}/calmodulin-dependent reaction.[29] NO acts as an intercellular messenger or an "atypical neurotransmitter" in the central nervous system.[30,31] NOS1 action in the CNS has been associated with pain perception as well as the control of sleep wake regulation.[29] Furthermore, the NO-arginine pathway is intimately connected to the modulation of the dopaminergic transmission.[31-37] L-dopa/benserazide as well as dopamine agonists have a significant beneficial effect on the motor and the sensory component of RLS.[38-41] A positive response to dopaminergics supports the diagnosis of RLS[2] and it is suggested that an alteration of the dopaminergic neurotransmission is involved in the pathophysiology of the disorder. An interplay of the opioidergic system and NO could also be relevant in the pathophysiology of RLS. The inhibition of NOS enhanced the morphine-induced antinociception at the spinal cord level[42] and a positive effect of opioidergic substances on RLS symptoms is well known.[43] The neuroanatomical level of a possible interplay of the nitridergic and dopaminergic or opioidergic neurotransmission, in connection to RLS, however, is not known. Although our findings support the proposal that alterations of NO signaling contributes to the pathogenesis of RLS, Ekbom noted in a seminal essay that "The disagreeable sensation stops at once when 1/100 g of nitroglycerin is chewed, suggesting that the cause is vascular."[44] This has never been investigated systematically and today it is commonly agreed that the symptoms of RLS originate from the central nervous system. Nitroglycerin acts also in the central nervous system through the release of NO.[45] It is possible that Ekbom's observation is actually based on an effect in the

central nervous system and not in the vascular system. Different modulating effects of the NO/arginine pathway, however, might play a role in the manifestation of the symptoms of RLS.

In conclusion, this study shows an association of different variants in the NO synthesizing enzyme NOS1 and RLS. The clinical relevance and the neuroanatomical level of a possible NO/dopamine interaction in connection to RLS as well as the potential usage of NO modulating agents for new treatment options are one of the most challenging aspects in the future research of this disorder.

Acknowledgments: S.H. was partly supported by a grant from the German RLS patient organization. This work (recruiting of Stage 1 controls) has been partly funded by the German Federal Ministry of Education and Research (BMBF) in the framework of the National Genome Research Network (NGFN) Nr. 01GS0481. We are grateful to all patients who participated in this study. We thank Tim M. Strom for his help on the SNP selection and discussions throughout the project. We also thank John Favor for discussions on the manuscript and Katja Junghans, Bianka Schmick, and Viktoria Hozjan for excellent technical assistance. We gratefully acknowledge Lilo Habersack, Hans Rhese, and Jürgen Schmidt-Evers for supporting this study.

REFERENCES

1. 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 2004;164:196–202.
2. Allen RP, Picchietti D, Hening WA, Trenkwalder C, Walters AS, Montplaisir J. 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.
3. Walters AS, Hickey K, Maltzman J, et al. A questionnaire study of 138 patients with restless legs syndrome: the 'Night-Walkers' survey. Neurology 1996;46:92–95.
4. Montplaisir J, Boucher S, Poirier G, et al. Clinical, polysomnographic, and genetic characteristics of restless legs syndrome: a study of 133 patients diagnosed with new standard criteria. Mov Disord 1997;12:61–65.
5. Winkelmann J, Wetter TC, Collado-Seidel V, et al. Clinical characteristics and frequency of the hereditary restless legs syndrome in a population of 300 patients. Sleep 2000;23:597–602.
6. Allen RP, Earley CJ. Defining the phenotype of the restless legs syndrome (RLS) using age-of-symptom-onset. Sleep Med 2000;1:11–19.
7. Winkelmann J, Muller-Myhsok B, Wittchen HU, 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 2002;52:297–302.
8. Desautels A, Turecki G, Montplaisir J, Sequeira A, Verner A, Rouleau GA. Mode of inheritance and susceptibility locus for restless legs syndrome, on chromosome 12q. Am J Hum Genet 2001;71:205–208.
9. Bonati MT, Ferini-Strambi L, Aridon P, Oldani A, Zucconi M, Casari G. Autosomal dominant restless legs syndrome maps on chromosome 14q. Brain 2003;126:1485–1492.
10. Chen S, Ondo WG, Rao S, Li L, Chen Q, Wang Q. Genomewide linkage scan identifies a novel susceptibility locus for restless legs syndrome on chromosome 9p. Am J Hum Genet 2004;74:876–885.
11. Pichler I, Marroni F, Beu Volpato C, et al. Linkage analysis identifies a novel locus for restless legs syndrome on chromosome 2q in a South tyrolean population isolate Am J Hum Genet 2006;7:716–723.
12. Levchenko A, Provost S, Montplaisir JY, et al. A novel autosomal dominant restless legs syndrome locus maps to chromosome 20p13. Neurology 2006;12:900–901.
13. Kock N, Culjkovic B, Maniak S, et al. Mode of inheritance and susceptibility locus for restless legs syndrome, on chromosome 12q. Am J Hum Genet 2002;71:205–208.
14. Desautels, A., Turecki, G., Montplaisir, J, et al. Restless legs syndrome: confirmation of linkage to chromosome 12q, genetic heterogeneity, and evidence of complexity. Arch Neurol 2005;62:591–596.
15. Winkelmann J, Lichtner P, Putz B, et al. Evidence for further genetic locus heterogeneity and confirmation of RLS-1 in restless legs syndrome. Mov Disord 2006;21:28–33.
16. Hicks AA, Rye DB, Kristjansson K, et al. Population-based confirmation of the 12q RLS locus in Iceland. Mov Disord 2005;20 (Suppl 10):34.
17. Desautels A, Turecki G, Xiong L, Rochefort D, Montplaisir J, Rouleau GA. Mutational analysis of neurotensin in familial restless legs syndrome. Mov Disord 2004;19:90–94.
18. Winkelmann J, Lichtner P, Auer DP, Pastore A, Strom TM, Meitinger T. Mutational analysis of divalent metal transporter 1 in restless legs syndrome. Mov Disord 2004;19 (Suppl 9):S358.
19. 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):26–30.
20. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium Am J Hum Genet 2004;74:106–120.
21. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualisation of LD and haplotype maps. Bioinformatics 2005;21:263–265.
22. Armitage P. Tests for linear trends in proportions and frequencies. Biometrics 1955;11:375–386.
23. Cochran, WG. Some methods for strenghtening the common X^2 test. Biometrics 1954;10:417–451.
24. Westfall PH, Young SS. Resampling-based multiple resting. New York: Wiley; 1993
25. Steffens M, Lamina C, Illig T, et al. SNP-based analysis of genetic substructure in the German population. Hum Hered 2006;62:20–29.
26. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002;21:2225–2232.
27. Neale BM, Sham PC. The future of association studies: gene-based analysis and replication. Am J Hum Genet 2004;75:353–362.
28. Lin PI, Vance JM, Pericak-Vance MA, Martin ER. No gene is an island: the flip-flop phenomenon. Am J Hum Genet 2007;80:531–538.
29. Gautier-Sauvigne S, Colas D, Parmantier P, et al. Nitric oxide and sleep. Sleep Med Rev 2005;9:101–113.
30. Snyder SH, Bredt DS. Nitric oxide as a neuronal messenger. Trends Pharmacol Sci 1991;12:125–128.
31. West AR, Galloway MP, Grace AA. Regulation of striatal dopamine neurotransmission by nitric oxide: effector pathways and signaling mechanisms. Synapse 2002;44:227–245.
32. Xu Q, Wink DA, Colton CA. Nitric oxide production and regulation of neuronal NOS in tyrosine hydroxylase containing neurons. Exp Neurol 2004;188:341–350.
33. West AR, Galloway MP. Endogenous nitric oxide facilitates striatal dopamine and glutamate efflux in vivo: role of ionotropic glutamate receptor-dependent mechanisms. Neuropharmacology 1997;36:1571–1581.

34. Zhu XZ, Luo LG. Effect of nitroprusside (nitric oxide) on endogenous dopamine release from rat striatal slices. J Neurochem 1992;59:932–935.
35. Kiss JP, Hennings EC, Zsilla G, Vizi ES. A possible role of nitric oxide in the regulation of dopamine transporter function in the striatum. Neurochem Int 1999;34:345–350.
36. Kiss JP, Zsilla G, Vizi ES. Inhibitory effect of nitric oxide on dopamine transporters: interneuronal communication without receptors. Neurochem Int 2004;45:485–489.
37. Cragg SJ, Rice ME. Dancing past the DAT at a DA synapse. Trends Neurosci 2004;27:270–277.
38. Hening WA, Allen RP, Earley CJ, Picchietti DL, Silber MH. An update on the dopaminergic treatment of restless legs syndrome and periodic limb movement disorder. Sleep 2004;27:560–583.
39. Wetter TC, Stiasny K, Winkelmann J, et al. A randomized controlled study of pergolide in patients with restless legs syndrome. Neurology 1999;23:52:944–950.
40. Winkelman JW, Sethi KD, Kushida CA, et al. Efficacy and safety of pramipexole in restless legs syndrome. Neurology 2006 26;67: 1034–1039.
41. Trenkwalder C, Garcia-Borreguero D, Montagna P, et al. Ropinirole in the treatment of restless legs syndrome: results from the TREAT RLS 1 study, a 12 week, randomised, placebo controlled study in 10 European countries. J Neurol Neurosurg Psychiatry 2004;75:92–97.
42. Watanabe C, Okuda K, Sakurada C, Ando R, Sakurada T, Sakurada S. Evidence that nitric oxide-glutamate cascade modulates spinal antinociceptive effect of morphine: a behavioural and microdialysis study in rats. Brain Res 2003;990:77–86.
43. Walters AS, Winkelmann J, Trenkwalder C, et al. Long-term follow-up on restless legs syndrome patients treated with opioids. Mov Disord 2001;6:1105–1109.
44. Ekbom KA. Restless legs: clinical study of hitherto overlooked disease in legs characterized by peculiar paresthesia (Anxietas tibiarum), pain and weakness and occuring in two main forms, asthenia crurum paraesthetica and asthenia crurum dolorosa. Acta Med Scand 1945;158 (Suppl):1–123.
45. Tassorelli C, Blandini F, Costa A, Preza E, Nappi G. Glycerin-induced activation of monoaminergic transmission in the rat. Cephalalgia 2002;22:226–232.

Variants in the Neuronal Nitric Oxide Synthase (nNOS, NOS1) Gene are Associated with Restless Legs Syndrome

Supplementary Information*

* Supplementary Table 1a is available online at http://www.interscience.wiley.com/jpages/0885-3185/suppmat.

Supplementary Table 2a. Summary of association data in the explorative study for SNPs with p-values < 0.05

dbSNP ID	Gene	Position chr. 12[a]	Genotypic test p-value[b]	Armitage test p-value	OR armitage 95% CI	HWE cases	HWE controls	MAF cases[c]	MAF controls[c]
rs1514796	TMEM16D	99683454	0.013	0.341	0.894 (0.708-1.127)	0.121	0.025	0.299 (A)	0.322 (A)
rs7975520	TMEM16D	100026444	0.099	0.037	1.257 (1.014-1.559)	0.914	0.380	0.460 (C)	0.405 (C)
rs7960925	TMEM16D	100026689	0.03	0.009	1.328 (1.074-1.642)	0.833	0.510	0.461 (T)	0.394 (T)
rs164365	SLC5A8	100062796	0.071	0.042	0.768 (0.595-0.991)	0.489	0.556	0.188 (A)	0.231 (A)
rs1069476	SLC5A8	100085308	0.079	0.046	0.772 (0.598-0.996)	0.489	0.656	0.187 (G)	0.230 (G)
rs6538998	MYBPC1	100496048	0.025	0.008	0.730 (0.580-0.920)	0.573	0.903	0.247 (T)	0.310 (T)
rs7308665	MYBPC1	100508969	0.029	0.009	0.728 (0.574-0.923)	0.882	0.899	0.230 (A)	0.291 (A)
rs825067	MYBPC1	100520442	0.057	0.018	0.773 (0.625-0.957)	0.442	0.751	0.397 (G)	0.457 (G)
rs825073	MYBPC1	100522956	0.039	0.012	0.759 (0.612-0.941)	0.436	0.749	0.389 (T)	0.454 (T)
rs17031719	MYBPC1	100524732	0.079	0.028	0.735 (0.558-0.966)	0.832	0.303	0.145 (G)	0.189 (G)
rs10860758	MYBPC1	100535385	0.019	0.012	0.751 (0.601-0.938)	0.222	0.736	0.322 (T)	0.385 (T)
rs825050	MYBPC1	100546151	0.028	0.017	1.297 (1.048-1.605)	1.000	0.104	0.475 (A)	0.414 (A)
rs3817552	MYBPC1	100547631	0.024	0.007	0.675 (0.508-0.898)	1.000	0.605	0.134 (C)	0.186 (C)
rs4764808	MYBPC1	100561218	0.011	0.004	0.699 (0.549-0.890)	0.764	0.443	0.224 (G)	0.291 (G)
rs1527393	MYBPC1	100567746	0.045	0.045	1.245 (1.005-1.541)	1.000	0.053	0.485 (G)	0.433 (G)
rs1263786	MYBPC1	100571732	0.011	0.007	0.710 (0.553-0.912)	0.112	1.000	0.183 (G)	0.242 (G)
rs4764863	NUP37	101016651	0.037	0.02	0.779 (0.631-0.961)	0.336	0.666	0.472 (A)	0.465 (A)
rs950945	FLJ20641	101047368	0.034	0.802	0.973 (0.786-1.205)	0.019	0.206	0.345 (A)	0.352 (A)
rs1509674	FLJ20641	101075342	0.043	0.74	0.964 (0.776-1.198)	0.059	0.130	0.336 (G)	0.344 (G)
rs12580432	PAH	101740484	0.004	0.762	1.037 (0.818-1.315)	0.108	0.004	0.267 (T)	0.260 (T)
rs1498694	PAH	101760343	0.026	0.951	1.007 (0.817-1.240)	0.052	0.068	0.417 (T)	0.419 (T)
rs10745957	ENST00000356833	102073810	0.009	0.183	0.872 (0.713-1.067)	0.342	0.005	0.460 (G)	0.496 (G)
rs7316328	STAB2	102638407	0.021	0.06	1.249 (0.991-1.576)	0.517	0.042	0.283 (C)	0.239 (C)
rs1795848	CHST11	103413961	0.029	0.257	1.183 (0.885-1.581)	0.015	0.400	0.170 (T)	0.148 (T)
rs7314958	CHST11	103479842	0.021	0.104	1.181 (0.966-1.443)	1.000	0.002	0.453 (C)	0.409 (C)
rs11112119	CHST11	103488007	0.04	0.012	0.740 (0.584-0.935)	0.581	0.464	0.253 (T)	0.312 (T)
rs4964825	CHST11	103492633	0.048	0.014	0.770 (0.625-0.948)	1.000	0.916	0.409 (G)	0.474 (G)
rs835495	CHST11	103601476	0.032	0.872	0.982 (0.796-1.214)	0.401	0.006	0.483 (T)	0.479 (T)
rs2029736	SLC41A2	103748912	0.034	0.058	0.808 (0.647-1.008)	0.224	0.244	0.406 (G)	0.458 (G)
rs7310098	RFX4	105511514	0.015	0.168	1.174 (0.935-1.474)	0.108	0.063	0.306 (G)	0.273 (G)
rs17040284	KIAA0789	107084765	0.001	0.002	0.680 (0.530-0.872)	0.881	0.002	0.227 (A)	0.295 (A)
rs3764002	KIAA0789	107121097	0.08	0.038	1.292 (1.015-1.646)	0.296	1.000	0.283 (T)	0.235 (T)
rs4964668	KIAA0789	107137699	0.015	0.253	1.157 (0.901-1.485)	0.196	0.019	0.238 (C)	0.213 (C)
rs17040691	SART3	107450119	0.034	0.01	0.726 (0.569-0.925)	0.549	0.608	0.225 (C)	0.284 (C)
rs2075358	ISCU	107463244	0.047	0.013	0.736 (0.578-0.937)	0.657	0.608	0.227 (G)	0.284 (G)
rs10861947	ISCU	107473128	0.044	0.016	0.770 (0.623-0.952)	1.000	0.342	0.398 (C)	0.460 (C)
rs17040818	LOC338773	107491223	0.012	0.005	0.672 (0.508-0.889)	0.394	0.870	0.145 (T)	0.200 (T)
rs2111211	CORO1C	107542335	0.061	0.018	1.280 (1.044-1.571)	0.674	0.524	0.494 (A)	0.442 (A)
rs6539455	SSH1	107722369	0.043	0.027	0.701 (0.510-0.962)	1.000	0.076	0.099 (G)	0.138 (G)
rs11114068	SSH1	107740425	0.083	0.027	0.618 (0.405-0.945)	0.602	0.262	0.050 (G)	0.080 (G)
rs3742026	ACACB	108133377	0.035	0.039	1.252 (1.011-1.550)	0.450	0.176	0.417 (C)	0.365 (C)
rs7966820	GLTP	108764990	0.043	0.675	1.045 (0.851-1.283)	0.043	0.166	0.446 (C)	0.435 (C)
rs1674123	DTX1	111973950	0.033	0.185	0.864 (0.694-1.073)	0.131	0.116	0.356 (T)	0.390 (T)
rs11614425	RBM19	112756375	0.045	0.699	0.942 (0.695-1.276)	0.062	0.182	0.131 (G)	0.138 (G)
rs4767165	RBM19	112819719	0.095	0.03	1.408 (1.033-1.918)	1.000	1.000	0.152 (C)	0.113 (C)
rs10850332	TBX5	113273357	0.092	0.031	0.718 (0.531-0.971)	0.444	0.544	0.114 (C)	0.154 (C)
rs1248054	TBX5	113313871	0.018	0.507	1.103 (0.826-1.473)	0.337	0.012	0.163 (T)	0.150 (T)

rs11068062	FLJ42957	115437136	0.017	0.236	0.733 (0.439-1.224)	0.006	1.000	0.034 (G)	0.047 (G)
rs2270618	FLJ21415	115621107	0.027	0.018	1.713 (1.096-2.678)	0.711	1.000	0.08 (G)	0.049 (G)
rs2291910	TSC	115949362	0.088	0.026	1.358 (1.036-1.779)	0.620	0.841	0.199 (T)	0.154 (T)
rs1472522	FBXO21	116048956	0.05	0.035	1.252 (1.016-1.544)	0.398	0.384	0.460 (A)	0.405 (A)
rs2840100	FBXO21	116054747	0.036	0.031	1.259 (1.022-1.550)	0.342	0.383	0.461 (C)	0.405 (C)
rs2036312	FBXO21	116086100	0.031	0.03	1.259 (1.022-1.551)	0.291	0.329	0.463 (G)	0.406 (G)
rs9658570	NOS1	116113593	0.004	0.003	1.842 (1.231-2.756)	0.152	0.631	0.102 (G)	0.061 (G)
rs2293054	NOS1	116164434	0.002	0.001	1.531 (1.202-1.950)	0.143	0.769	0.312 (A)	0.233 (A)
rs6490121	NOS1	116170915	0.008	0.002	1.436 (1.139-1.810)	0.639	0.419	0.337 (G)	0.264 (G)
rs9658356	NOS1	116186738	0.012	0.014	0.209 (0.060-0.733)	1.000	1.000	0.004 (C)	0.019 (C)
rs7977109	NOS1	116193060	0.029	0.014	1.304 (1.055-1.610)	0.916	0.203	0.482 (A)	0.454 (A)
rs693534	NOS1	116247438	0.047	0.029	0.791 (0.641-0.976)	0.651	0.283	0.366 (A)	0.423 (A)
rs10507280	KSR2	116361340	0.028	0.153	0.785 (0.563-1.094)	0.148	0.139	0.100 (A)	0.124 (A)
rs1093309	KSR2	116387928	0.009	0.003	0.729 (0.591-0.899)	0.401	0.832	0.467 (G)	0.456 (G)
rs7133582	KSR2	116388113	0.003	0.003	1.448 (1.138-1.844)	0.008	0.681	0.327 (T)	0.258 (T)
rs1093307	KSR2	116388363	0.116	0.038	0.763 (0.591-0.985)	0.862	1.000	0.186 (A)	0.231 (A)
rs4766854	KSR2	116397886	0.015	0.012	0.625 (0.434-0.901)	0.422	0.597	0.073 (A)	0.112 (A)
rs11068545	KSR2	116470174	0.033	0.956	0.994 (0.800-1.235)	0.014	0.261	0.372 (A)	0.371 (A)
rs4766869	KSR2	116597588	0.046	0.045	1.425 (1.008-2.013)	0.448	0.329	0.119 (T)	0.087 (T)
rs16947978	KSR2	116719826	0.05	0.573	1.112 (0.769-1.607)	0.057	0.301	0.092 (G)	0.084 (G)
rs3884577	KIAA1853	118023890	0.004	0.051	1.237 (1.000-1.532)	0.164	0.025	0.432 (C)	0.382 (C)
rs2555276	KIAA1853	118050441	0.132	0.044	0.746 (0.560-0.992)	0.671	0.598	0.145 (A)	0.183 (A)
rs722306	HSPB8	118073205	0.023	0.618	1.064 (0.833-1.359)	0.180	0.019	0.227 (G)	0.216 (G)
rs1133026	HSPB8	118095067	0.04	0.381	1.116 (0.873-1.428)	0.188	0.060	0.234 (A)	0.215 (A)
rs278116	CIT	118642310	0.096	0.042	1.285 (1.01-1.635)	0.897	0.246	0.283 (G)	0.237 (G)
rs4766950	CIT	118740961	0.046	0.015	1.297 (1.052-1.599)	1.000	0.463	0.442 (T)	0.494 (T)
rs7961178	ENST00000315185	119868364	0.028	0.856	1.034 (0.724-1.476)	0.111	0.057	0.094 (A)	0.091 (A)
rs2254779	TCF1	119871962	0.026	0.928	0.983 (0.690-1.403)	0.110	0.057	0.094 (C)	0.092 (C)
rs503720	P2RX7	120067794	0.054	0.02	0.765 (0.611-0.959)	0.444	1.000	0.291 (A)	0.347 (A)
rs2230912	P2RX7	120084916	0.049	0.104	1.265 (0.953-1.680)	0.361	0.131	0.175 (G)	0.144 (G)
rs2393847	P2RX4	120144256	0.047	0.028	0.766 (0.605-0.971)	0.889	0.170	0.251 (G)	0.302 (G)
rs1653594	CAMKK2	120151692	0.015	0.307	0.881 (0.692-1.124)	0.020	0.195	0.218 (G)	0.241 (G)

Supplementary Table 2b. Summary of association data in the replication sample for the most significant SNPs of the explorative study

dbSNP ID	Gene	Position chr. 12[a]	Genotypic test p-value[b]	Genotypic test p-value (corrected)[c]	Armitage test p-value	Armitage test p-value (corrected)[c]	OR armitage 95% CI	HWE cases	HWE controls	MAF cases[c]	MAF controls[c]
rs7960925	TMEM16D	100026689	0.723	1.000	0.454	1.000	1.068 (0.900-1.267)	0.542	0.860	0.431 (T)	0.416 (T)
rs7308665	MYBPC1	100508969	0.972	1.000	0.811	1.000	0.977 (0.807-1.182)	0.664	0.591	0.268 (A)	0.273 (A)
rs825073	MYBPC1	100522956	0.787	1.000	0.489	1.000	1.062 (0.896-1.258)	0.796	0.931	0.445 (T)	0.431 (T)
rs3817552	MYBPC1	100547631	0.22	0.996	0.096	0.907	0.818 (0.645-1.036)	0.124	0.232	0.147 (C)	0.172 (C)
rs4764808	MYBPC1	100561218	0.097	0.908	0.050	0.713	0.824 (0.680-1.000)	0.114	0.679	0.255 (G)	0.291 (G)
rs1263786	MYBPC1	100571732	0.524	1.000	0.400	1.000	0.914 (0.741-1.127)	0.373	0.039	0.215 (G)	0.229 (G)
rs11112119	CHST11	103488007	1.000	1.000	0.925	1.000	0.991 (0.824-1.192)	0.917	0.835	0.285 (T)	0.287 (T)
rs4964825	CHST11	103492633	0.614	1.000	0.524	1.000	1.056 (0.894-1.247)	0.263	1.000	0.458 (G)	0.444 (G)
rs17040284	KIAA0789	107084765	0.602	1.000	0.340	0.999	1.098 (0.906-1.330)	1.000	0.734	0.271 (A)	0.253 (A)
rs17040691	SART3	107450119	0.736	1.000	0.961	1.000	1.005 (0.831-1.215)	0.912	0.317	0.260 (C)	0.259 (C)
rs2075358	ISCU	107463244	0.663	1.000	0.883	1.000	1.014 (0.839-1.226)	0.912	0.317	0.261 (G)	0.258 (G)
rs10861947	ISCU	107473128	0.305	0.999	0.511	1.000	1.060 (0.891-1.260)	0.038	0.931	0.452 (C)	0.438 (C)
rs9658570	NOS1	116113593	0.839	1.000	0.641	1.000	0.933 (0.697-1.248)	1.000	0.456	0.088 (G)	0.094 (G)
rs2293054	NOS1	116164434	0.293	0.999	0.230	0.996	0.894 (0.745-1.073)	0.082	0.681	0.269 (A)	0.293 (A)
rs6490121	NOS1	116170915	0.345	0.999	0.211	0.995	0.892 (0.745-1.067)	0.619	0.634	0.314 (G)	0.339 (G)
rs9658356	NOS1	116186738	0.379	1.000	0.279	0.998	0.612 (0.252-1.488)	1.000	1.000	0.007 (C)	0.012 (C)
rs7977109	NOS1	116193060	**0.002**	**0.046**	**0.002**	**0.049**	0.762 (0.643-0.903)	0.260	0.198	0.447 (A)	0.515 (A)
rs1093309	KSR2	116387928	0.698	1.000	0.579	1.000	0.954 (0.807-1.127)	0.932	0.443	0.489 (G)	0.501 (G)
rs4766950	CIT	118740961	0.098	0.909	0.112	0.934	1.143 (0.970-1.347)	1.000	0.041	0.531 (T)	0.496 (T)

Only SNPs that passed the quality criteria (call rate > 95% and HWE >0.001 controls) were included. [a]SNPs are listed by ascending position. Positions refer to the Human May 2004 (hg17) assembly. [b]Genotypic test is a Fisher exact test on a 2 x 3 table. [c]Alleles are indicated in brackets referring to dbSNP reports. [c]Correction for multiple testing was done using the Westfall-Young method (53).

Supplementary Table 3. Summary of association data of the NOS1 high-density mapping in the explorative and replication study

dbSNP ID	Position chr. 12[a]	Study	HWE cases	HWE controls	MAF HapMap[b]	MAF cases	MAF controls	Genotypic test p-value[c,d]	Genotypic test p-value[c] (corrected)[e]	Armitage test p-value[c]	Armitage test p-value[c] (corrected)[e]	OR armitage 95% CI
rs10774907	116110123	Explorative	0.708	0.091	0.293 (A)	0.291	0.249	**0.047**	0.730	0.064	0.830	1.249 (0.987 - 1.581)
		Replication	0.591	0.222		0.276	0.259	0.607	1.000	0.377	1.000	1.088 (0.903 - 1.310)
rs9658570	116113593	Explorative	0.782	0.630	0.158 (G)	0.103	0.060	**0.004**	0.117	**0.003**	0.082	1.831 (1.235 - 2.716)
		Replication	1.000	0.612		0.086	0.094	0.773	1.000	0.548	1.000	0.915 (0.684 - 1.224)
rs2682826	116115558	Explorative	0.499	0.888	0.267 (T)	0.259	0.246	0.794	1.000	0.544	1.000	1.077 (0.848 - 1.368)
		Replication	0.490	0.661		0.247	0.269	0.524	1.000	0.264	1.000	0.897 (0.741 - 1.086)
rs877995	116125970	Explorative	0.312	0.729	0.161 (A)	0.191	0.186	0.894	1.000	0.818	1.000	1.032 (0.789 - 1.349)
		Replication	0.658	0.351		0.176	0.203	0.255	1.000	0.108	0.958	0.840 (0.679 - 1.040)
rs2293048	116127545	Explorative	0.138	0.277	0.190 (T)	0.119	0.081	**0.005**	0.164	**0.017**	0.377	1.537 (1.081 - 2.185)
		Replication	0.515	0.688		0.111	0.122	0.689	1.000	0.417	1.000	0.898 (0.692 - 1.166)
rs4766836	116135736	Explorative	0.402	0.620	0.153 (A)	0.103	0.058	**0.002**	0.063	**0.001**	**0.044**	1.930 (1.289 - 2.887)
		Replication	0.579	0.428		0.086	0.091	0.896	1.000	0.689	1.000	0.942 (0.703 - 1.263)
rs816347	116152643	Explorative	0.105	0.306	0.136 (T)	0.072	0.057	0.500	1.000	0.269	1.000	1.261 (0.836 - 1.902)
		Replication	1.000	0.496		0.073	0.066	0.634	1.000	0.505	1.000	1.120 (0.803 - 1.562)
rs2650163	116155757	Explorative	0.182	0.488	0.117 (G)	0.088	0.080	0.816	1.000	0.603	1.000	1.100 (0.767 - 1.579)
		Replication	0.457	0.465		0.094	0.099	0.906	1.000	0.699	1.000	0.946 (0.714 - 1.253)
rs1093329	116158884	Explorative	0.449	0.905	0.458 (A)	0.393	0.327	**0.026**	0.534	**0.008**	0.213	1.344 (1.082 - 1.671)
		Replication	0.405	0.240		0.364	0.386	0.576	1.000	0.299	1.000	0.914 (0.771 - 1.083)
rs2293054	116164434	Explorative	0.071	0.769	0.314 (A)	0.316	0.233	**0.001**	**0.028**	**0.0003**	**0.010**	1.570 (1.233 - 2.000)
		Replication	0.079	0.605		0.268	0.294	0.284	1.000	0.198	1.000	0.887 (0.738 - 1.065)
rs2293055	116164604	Explorative	0.095	0.579	0.136 (A)	0.088	0.106	0.072	0.863	0.219	0.998	0.801 (0.563 - 1.14)
		Replication	0.427	0.615		0.090	0.095	0.928	1.000	0.722	1.000	0.95 (0.714 - 1.263)
rs6490121	116170915	Explorative	0.646	0.504	0.347 (G)	0.348	0.266	**0.003**	0.088	**0.001**	**0.023**	1.488 (1.183 - 1.872)
		Replication	0.618	0.773		0.314	0.340	0.369	1.000	0.211	1.000	0.891 (0.745 - 1.067)
rs7133438	116187456	Explorative	0.238	0.328	0.178 (T)	0.124	0.087	0.082	0.89	**0.025**	0.515	1.461 (1.049 - 2.035)
		Replication	1.000	1.000		0.117	0.135	0.434	1.000	0.200	1.000	0.845 (0.653 - 1.093)
rs10850807	116190476	Explorative	0.829	0.440	0.500(C)	0.401	0.384	0.596	1.000	0.506	1.000	1.073 (0.871 - 1.322)
		Replication	0.713	0.929		0.374	0.416	0.121	0.975	**0.044**	0.721	0.838 (0.705 - 0.995)
rs7977109	116193060	Explorative	0.600	0.398	0.500 (A)	0.480	0.463	**0.047**	0.729	**0.027**	0.541	1.263 (1.026 – 1.553)
		Replication	0.378	0.188		0.484	0.557	**0.001**	**0.037**	**0.001**	0.033	0.744 (0.625 – 0.884)
rs11068446	116210334	Explorative	0.228	1.000	0.144 (T)	0.096	0.112	0.360	1.000	0.317	1.000	0.839 (0.595 - 1.183)
		Replication	0.487	0.661		0.103	0.108	0.936	1.000	0.699	1.000	0.948 (0.722 - 1.244)

rs499813	116211120	Explorative	0.872	0.011	0.149 (A)	0.201	0.168	**0.031**	0.573	0.115	0.961	1.233 (0.950 - 1.600)
		Replication	1.000	0.563		0.165	0.183	0.536	1.000	0.282	1.000	0.886 (0.710 - 1.105)
rs530393	116213978	Explorative	0.224	0.747	0.333 (A)	0.372	0.418	0.181	0.996	0.082	0.891	0.833 (0.678 - 1.024)
		Replication	0.140	0.237		0.445	0.374	**0.001**	**0.027**	**0.001**	**0.037**	1.337 (1.127 - 1.587)
rs12578547	116226067	Explorative	0.451	0.095	0.300 (C)	0.218	0.196	0.117	0.962	0.295	1.000	1.143 (0.890 - 1.468)
		Replication	0.382	0.275		0.221	0.229	0.908	1.000	0.663	1.000	0.957 (0.786 - 1.166)
rs473640	116228434	Explorative	0.755	0.610	0.190 (C)	0.210	0.191	0.485	1.000	0.350	1.000	1.129 (0.875 - 1.458)
		Replication	0.551	0.341		0.175	0.195	0.249	1.000	0.234	1.000	0.876 (0.706 - 1.088)
rs545654	116239769	Explorative	0.350	0.296	0.492 (T)	0.492	0.503	0.314	1.000	0.678	1.000	0.958 (0.781 - 1.175)
		Replication	0.667	0.550		0.539	0.503	0.183	0.997	0.094	0.929	1.154 (0.976 - 1.365)
rs1552227	116241755	Explorative	0.697	0.532	0.333 (T)	0.273	0.300	0.534	1.000	0.261	1.000	0.880 (0.703 - 1.100)
		Replication	0.744	0.250		0.270	0.286	0.624	1.000	0.414	1.000	0.926 (0.770-1.114)
rs10507279	116242994	Explorative	0.271	0.203	0.200 (A)	0.172	0.176	0.970	1.000	0.819	1.000	0.970 (0.746 - 1.259)
		Replication	0.733	0.771		0.148	0.179	0.159	0.993	0.055	0.794	0.801 (0.637 - 1.005)
rs527590	116244638	Explorative	0.769	0.066	0.192 (T)	0.231	0.199	0.082	0.89	0.147	0.989	1.201 (0.938 - 1.539)
		Replication	0.897	0.793		0.213	0.210	0.922	1.000	0.865	1.000	1.018 (0.826 - 1.255)
rs693534	116247438	Explorative	0.653	0.236	0.366 (A)	0.362	0.422	**0.028**	0.555	**0.018**	0.413	0.778 (0.631 - 0.959)
		Replication	0.662	0.854		0.427	0.367	**0.015**	0.353	**0.004**	0.161	1.285 (1.082 - 1.526)
rs1123425	116248825	Explorative	0.092	0.126	0458 (G)	0.442	0.399	**0.017**	0.391	0.099	0.930	1.192 (0.967 - 1.468)
		Replication	0.790	0.261		0.410	0.449	0.103	0.948	0.064	0.832	0.850 (0.716 - 1.010)
rs1879417	116266235	Explorative	0.598	0.674	0.433 (C)	0.436	0.483	0.206	0.996	0.075	0.874	0.831 (0.678-1.019)
		Replication	0.345	0.222		0.484	0.444	**0.048**	0.757	0.062	0.821	1.176 (0.992 - 1.394)
rs4767535	116267595	Explorative	0.394	0.282	0.500 (C)	0.424	0.419	0.346	1.000	0.842	1.000	1.021 (0.830 - 1.256)
		Replication	1.000	0.034		0.415	0.432	0.213	1.000	0.442	1.000	0.936 (0.791 - 1.107)
rs816292	116274129	Explorative	0.693	0.221	0.267 (T)	0.272	0.268	0.832	1.000	0.870	1.000	1.019 (0.812 - 1.279)
		Replication	0.382	0.733		0.331	0.254	**0.0005**	**0.022**	**0.0001**	**0.009**	1.444 (1.199 - 1.738)

Only SNPs that passed the quality criteria (call rate > 95%, HWE> 0.001 controls and MAF> 0.05) were included. [a]SNPs are listed by ascending position. Positions refer to the Human May 2004 (hg17) assembly. [b]MAF refers to the allele indicated in brackets. [c]significant p-values ($p<0.05$) are indicated in bold. [d]Genotypic test is a Fisher exact test on a 2 x 3 table. [e]Correction for multiple testing was done using the Westfall-Young method

Appendix 2

Letter in *Nature Genetics*

Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions

Juliane Winkelmann, Barbara Schormair, Peter Lichtner, Stephan Ripke, Lan Xiong, Shapour Jalilzadeh, Stephany Fulda, Benno Pütz, Gertrud Eckstein, Stephanie Hauk, Claudia Trenkwalder, Alexander Zimprich, Karin Stiasny-Kolster, Wolfgang Oertel, Cornelius G Bachmann, Walter Paulus, Ines Peglau, Ilonka Eisensehr, Jacques Montplaisir, Gustavo Turecki, Guy Rouleau, Christian Gieger, Thomas Illig, H-Erich Wichmann, Florian Holsboer, Bertram Müller-Myhsok & Thomas Meitinger

Nature Genetics 39 (8): 1000-1006, 2007

Contributions:

In this study I selected the RLS cases for stages 1 and 2a from the database of patients recruited by collaboration partners. In addition, I performed quality control of the DNA samples prior to genotyping on the Affymetrix 500K arrays and prepared the required dilutions. I carried out the genotyping of the 500K Affymetrix arrays and the preliminary quality control of genotype calls before the clustering analysis. For the replication stage, I selected the additional SNPs for replication and the SNPs for finemapping of the associated regions, and performed the genotyping using MALDI-TOF mass spectrometry on a Sequenom MassArray system. I designed the iPLEX assays for genotyping, carried out genotype calling quality control, and prepared the genotype and phenotype data necessary for the statistical analysis. I participated in writing the manuscript and the supplementary information, designed figures 4 and 5 and table 1 of the manuscript and all tables of the supplementary information.

nature genetics

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions

Juliane Winkelmann[1–3], Barbara Schormair[1,3], Peter Lichtner[1,3], Stephan Ripke[2], Lan Xiong[4], Shapour Jalilzadeh[1,3], Stephany Fulda[2], Benno Pütz[2], Gertrud Eckstein[1,3], Stephanie Hauk[1,3], Claudia Trenkwalder[5], Alexander Zimprich[6], Karin Stiasny-Kolster[7], Wolfgang Oertel[7], Cornelius G Bachmann[8], Walter Paulus[8], Ines Peglau[9], Ilonka Eisensehr[10], Jacques Montplaisir[11,12], Gustavo Turecki[13], Guy Rouleau[4], Christian Gieger[14], Thomas Illig[14], H-Erich Wichmann[14,15], Florian Holsboer[2], Bertram Müller-Myhsok[2,16] & Thomas Meitinger[1,3,16]

Restless legs syndrome (RLS) is a frequent neurological disorder characterized by an imperative urge to move the legs during night, unpleasant sensation in the lower limbs, disturbed sleep and increased cardiovascular morbidity. In a genome-wide association study we found highly significant associations between RLS and intronic variants in the homeobox gene *MEIS1*, the *BTBD9* gene encoding a BTB(POZ) domain as well as variants in a third locus containing the genes encoding mitogen-activated protein kinase MAP2K5 and the transcription factor LBXCOR1 on chromosomes 2p, 6p and 15q, respectively. Two independent replications confirmed these association signals. Each genetic variant was associated with a more than 50% increase in risk for RLS, with the combined allelic variants conferring more than half of the risk. MEIS1 has been implicated in limb development, raising the possibility that RLS has components of a developmental disorder.

Nightwalkers, as individuals with RLS call themselves, are forced to move their legs during periods of rest especially in the evening and night to relieve uncomfortable or painful sensations in the deep calf[1]. This diurnal variation leads to impaired sleep onset, and the periodic leg movements during sleep in the majority of patients contribute to sleep disruption and a reduced quality of life as a major consequence[2]. There are recognized secondary forms of RLS such as in iron deficiency, pregnancy and end-stage renal disease and associated morbidity such as increased cardiovascular risk[2,3]. RLS is one of the most common neurological disorders, with an age-dependent prevalence of up to 10% in the elderly in North America and Europe[2]. Dopaminergic agents originally developed for Parkinson's disease have been used to treat RLS, with an unknown mode of action[2]. Neurophysiological, pharmacological and neuroimaging studies suggest that the characteristic symptoms originate in the central nervous system, yet the underlying neurobiology remains obscure[4]. A family history of RLS is present in more than 50% of affected individuals, and similar figures have been reported for heritability in twin studies[5,6]. Linkage analysis uncovered five loci based on recessive (RLS1) or dominant inheritance (RLS2–RLS5), but so far

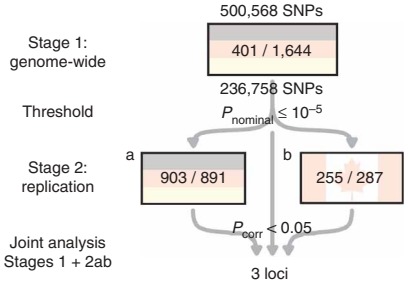


Figure 1 Study overview. Numbers refer to cases and controls and SNPs genotyped and analyzed. The 13 most significant SNPs together with neighboring SNPs were replicated in a German ('a') and a Canadian ('b') case/control sample. Three loci were confirmed in both stage 2 samples of the study.

[1]Institute of Human Genetics, GSF National Research Center of Environment and Health, D-85764 Neuherberg, Munich, Germany. [2]Max Planck Institute of Psychiatry, D-80804 Munich, Germany. [3]Technical University, Institute of Human Genetics, D-81675 Munich, Germany. [4]Laboratoire d'étude des maladies du cerveau, Centre de recherche du CHUM, Hôpital Notre-Dame, Université de Montréal, Montréal, Québec H2L 4M1, Canada. [5]Paracelsus-Elena-Hospital, 34128 Kassel, Germany. [6]Neurological Department, Medical University of Vienna, 1090 Vienna, Austria. [7]Philipps University Marburg, Department of Neurology, 35039 Marburg, Germany. [8]University of Göttingen, Department of Clinical Neurophysiology, 37070 Göttingen, Germany. [9]Neurologische Praxis, 10969 Berlin, Germany. [10]Neurologische Praxis Sendlingerstrasse, 80331 Munich, Germany. [11]Centre d'étude du sommeil, Hôpital du Sacré-Cœur de Montréal, Montréal, Québec H4J 1C5, Canada. [12]Centre de recherche en sciences neurologiques, Université de Montréal, Montréal, Québec H4J 1C5, Canada. [13]Departments of Psychiatry and Human Genetics, McGill University, Douglas Hospital, Montreal, Quebec H4H 1R3, Canada. [14]Institute of Epidemiology, GSF National Research Center for Environment and Health, 85764 Neuherberg, Munich, Germany. [15]Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, 81377 Munich, Germany, [16]These authors contributed equally to this work. Correspondence should be addressed to J.W. (janew@mpipsykl.mpg.de) or T.M. (meitinger@gsf.de).

Received 11 April; accepted 12 June; published online 18 July 2007; doi:10.1038/ng2099

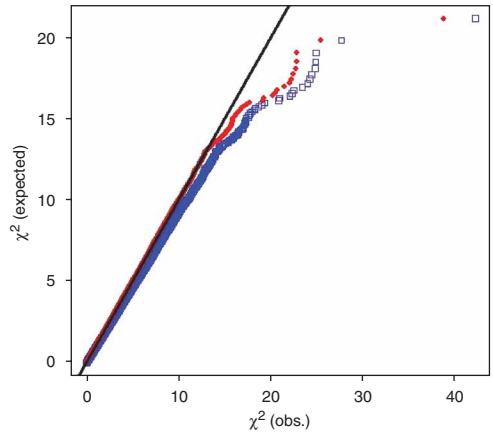


Figure 2 Extent of population stratification. The distribution of expected (under the null hypothesis) versus observed χ^2 values (all *P* values obtained in the analysis of sample 1, using Armitage trend test with age and sex as covariates) before (blue) and after (red) correction by division with λ. Adherence to the diagonal indicates lack of inflation of the statistic. As can be seen in the uncorrected plot, there is evidence for a systematic deviation toward higher-than-expected values. After the correction, there is near-perfect adherence to the diagonal for most of the values obtained, indicating that the correction performed well.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

no causally related sequence variants have been identified[5–7]. With SNP arrays becoming a mature technology, we conducted a genome-wide association study (GWAS), typing 500,568 SNPs in individuals with RLS and in a large control cohort from the general population.

Genome-wide association

The study design involved an exploratory stage (stage 1) followed by replication in two further case-control samples (stages 2a and 2b) (**Fig. 1**). In stage 1, we performed a GWAS, typing cases and controls on a single platform with the Affymetrix 500K Array Set. To enrich for risk alleles and minimize phenotypic heterogeneity, we selected subjects with familial RLS ($n = 401$). Controls were selected randomly from a population-based cohort ($n = 1{,}644$, from the KORA-S3/F3 survey, described previously)[8]. For statistical analysis, we selected SNPs by including only high-quality genotypes to reduce the number of false-positive signals (**Supplementary Table 1** online). A total of 236,758 SNPs passed all quality control filters (mean call rate = 99.48%). The effect of population stratification was negligible (inflation factor $\lambda = 1.09$ via genomic control)[9] (**Fig. 2**). Eigenvalue-based analysis showed only minimal population substructure (**Fig. 2**). An Armitage trend test uncovered four SNPs with *P* values $< 10^{-6}$ (**Fig. 3** and **Supplementary Table 2** online). After correcting for multiple testing, we identified a single SNP within *MEIS1* that reached genome-wide significance (rs2300478, $P_{corrected} < 0.0002$).

Replication of genome-wide findings

For stage 2 replication, 13 SNPs passed our inclusion criteria based on *P* value, location within a linkage peak and visual inspection of clustering data. We selected these and 15 neighboring SNPs for replication. They mapped to six discernible regions. Of these 28 SNPs, 25 were successfully genotyped in stage 2a and 24 in stage 2b (**Supplementary Table 3** online). Individuals in 2a ($n = 903$, familial or sporadic RLS) had been recruited separately using the sampling design of stage 1. Control subjects were selected from KORA-S4 ($n = 891$).

In stage 2a, we found nominally significant evidence for association in five regions, of which three withstood correction for multiple testing (**Fig. 4** and **Supplementary Table 4** online). The first region was on 2p, located in a 32-kb linkage disequilibrium (LD) block containing exon 9 of *MEIS1*. Here, two of three SNPs showed significant association ($P < 10^{-11}$). *MEIS1* is a member of a family of highly conserved *TALE* homeobox genes. Heterodimers of MEIS1 with PBX and HOX proteins augment the affinity and specificity of DNA binding by HOX proteins[10]. MEIS1 has been found to be overexpressed in acute myeloid leukemia[10], and studies in *Xenopus laevis* have shown involvement in neural crest development[11]. In addition, there are several potential links to RLS: during embryonic development, MEIS1 is essential for proximo-distal limb formation[12], and children with restless legs syndrome are often described as having growing pains[13]. MEIS1 is part of a Hox transcriptional regulatory network that specifies spinal motor neuron pool identity and connectivity[14]. Notably, spinal hyperexcitability is an established component in the genesis of periodic leg movements found in individuals with RLS[15]. Specific functions of MEIS1 in postembryonic tissues still remain to be established. The protein is known to be expressed in the adult mouse brain in cerebellar granule cells, the forebrain and, notably, in dopaminergic neurons of the substantia nigra[16].

The second region with significant association was on chromosome 6p, within a 113-kb LD block in intron 5 of the *BTBD9* gene. All five SNPs tested were significant, four of these with *P* values $< 10^{-5}$. Little is known about BTBD9 other than that it belongs to the BTB(POZ) proteins. BTB stands for *broad complex*, *tramtrack* and *bric à brac*, genes that in *Drosophila melanogaster* are required for embryonic

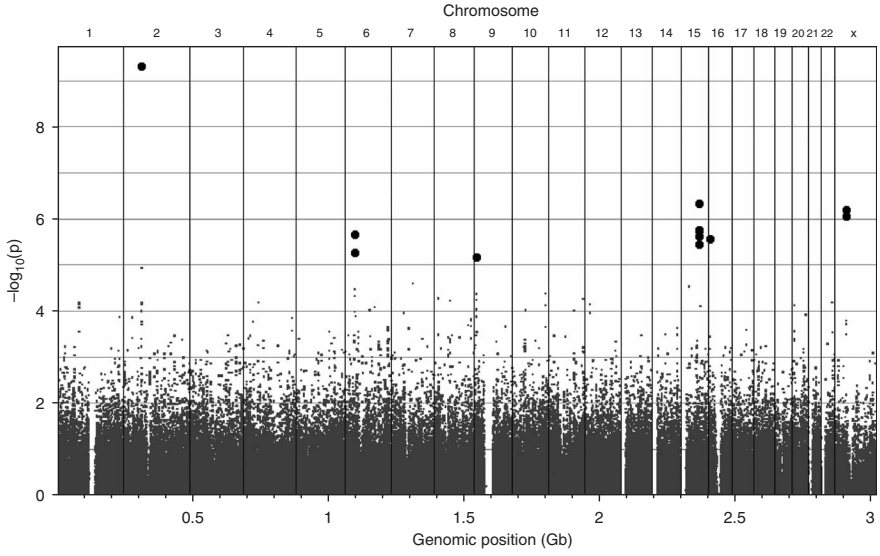


Figure 3 Genome-wide association study for RLS susceptibility loci. The analysis compared 393 successfully genotyped RLS cases with 1,602 population-based KORA controls. The *x*-axis represents genomic position, and the *y*-axis shows $-\log_{10}(P)$. Thirteen SNPs that passed inclusion criteria for the replication study of stage 2 are highlighted in bold. Note that the *P* values of three SNPs on chromosome 15 are very similar, and these SNPs appear as one single dot.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

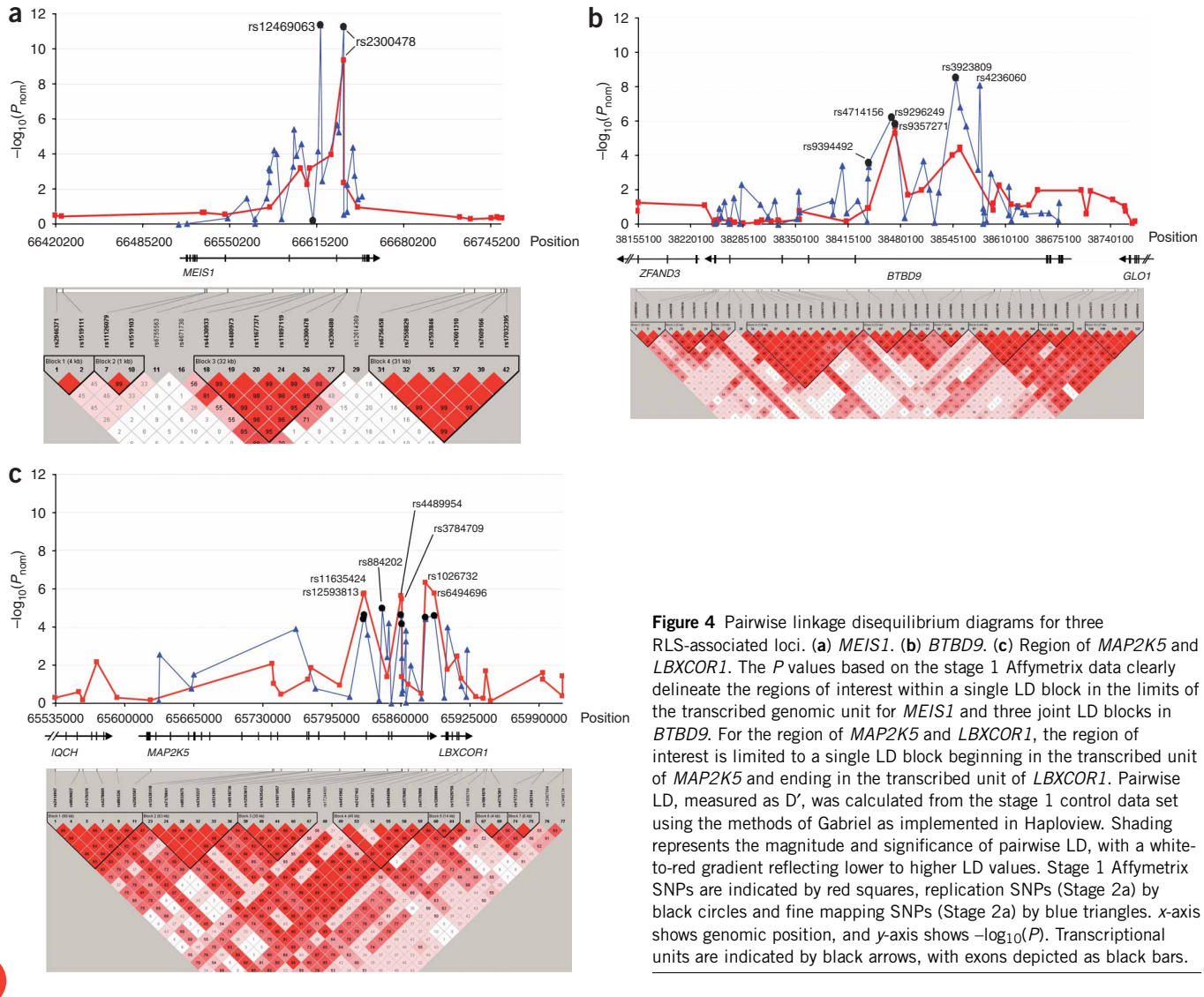


Figure 4 Pairwise linkage disequilibrium diagrams for three RLS-associated loci. (**a**) *MEIS1*. (**b**) *BTBD9*. (**c**) Region of *MAP2K5* and *LBXCOR1*. The *P* values based on the stage 1 Affymetrix data clearly delineate the regions of interest within a single LD block in the limits of the transcribed genomic unit for *MEIS1* and three joint LD blocks in *BTBD9*. For the region of *MAP2K5* and *LBXCOR1*, the region of interest is limited to a single LD block beginning in the transcribed unit of *MAP2K5* and ending in the transcribed unit of *LBXCOR1*. Pairwise LD, measured as D′, was calculated from the stage 1 control data set using the methods of Gabriel as implemented in Haploview. Shading represents the magnitude and significance of pairwise LD, with a white-to-red gradient reflecting lower to higher LD values. Stage 1 Affymetrix SNPs are indicated by red squares, replication SNPs (Stage 2a) by black circles and fine mapping SNPs (Stage 2a) by blue triangles. *x*-axis shows genomic position, and *y*-axis shows $-\log_{10}(P)$. Transcriptional units are indicated by black arrows, with exons depicted as black bars.

development, cell fate determination in the eye, metamorphosis and pattern formation in the limbs[17,18]. Functions of BTB(POZ) proteins include transcription repression, cytoskeleton regulation, tetramerization and gating of ion channels, as well as ubiquitin-dependent protein degradation[17]. The modular nature of this protein and the universal occurrence of the particular domains of BTBD9 make assignment of a specific function difficult at present.

The third region, defined by seven SNPs tested on 15q, showed significant evidence for association, with $P < 10^{-4}$. This region contains a 48-kb LD block overlapping the 3′ end of *MAP2K5*, a member of the mitogen-activated protein kinase family, and the adjacent *LBXCOR1* gene. MAPK pathways are conserved from yeast to human and are activated by a signaling cascade that mediates the transduction of extracellular signals to cytoplasmic nuclear effectors[19]. MAP2K5 is a specific upstream activator of ERK5, and this pathway is activated by oxidative stress, hyperosmolarity and growth factors. In addition, MAP2K5 and ERK5 are abundantly expressed in heart and skeletal muscles, and the MAP2K5/ERK5 MAP kinase cascade is critical at early stages of muscle cell differentiation[19]. The possible link between RLS risk alleles and known biological functions of the MAP2K5-ERK5 pathway is of particular interest, as this pathway is important in neuroprotection of dopaminergic neurons[20]. *LBXCOR1* is annotated as being downstream of *MAP2K5* and acting as a transcriptional corepressor of LBX1. This homeobox gene is critical in the development of sensory pathways in the dorsal horn of the spinal cord that relay pain and touch[21]. Three SNPs within the *PTPRD* gene in the chromosome 9 linkage region (RLS3) and one SNP on chromosome 16 in the A2BP1 gene were nominally significant.

In stage 2b, we genotyped the same SNPs in affected individuals ($n = 255$) and controls ($n = 287$) from a French-Canadian population. Here, we found nominally significant evidence for association in four regions (two SNPs on chromosome 2p, five SNPs on 6p, seven SNPs on 15q and one SNP on 16p, **Supplementary Table 5** online). The same three regions as in stages 1 and 2a remained significant after correction for multiple testing. Odds ratios (ORs) and risk alleles were very similar to those for stage 2a. **Table 1** shows those nine SNPs in the three loci confirmed in all three sample sets and in joint analysis withstanding genome-wide correction for multiple testing.

Table 1 Confirmed association results

dbSNP ID	Chr	Genome position	Gene	MAF (cases)	Risk allele	MAF (controls)
rs2300478	2p	66634957	*MEIS1*	0.367 (G)	G	0.241 (G)
rs9296249	6p	38473819	*BTBD9*	0.162 (C)	T	0.235 (C)
rs9357271	6p	38473851	*BTBD9*	0.165 (C)	T	0.238 (C)
rs12593813	15q	65823906	*MAP2K5*	0.258 (A)	G	0.330 (A)
rs11635424	15q	65824632	*MAP2K5*	0.257 (A)	G	0.330 (A)
rs4489954	15q	65859129	*MAP2K5*	0.239 (T)	G	0.311 (T)
rs3784709	15q	65859329	*MAP2K5*	0.251 (T)	C	0.321 (T)
rs1026732	15q	65882139	*MAP2K5*	0.252 (A)	G	0.327 (A)
rs6494696	15q	65890260	[*MAP2K5*/*LBXCOR1*]	0.253 (C)	G	0.326 (C)

dbSNP ID	OR (95% c.i.)	Stage 1 P_{nom}	Stage 2a P_{nom}	Stage 2b P_{nom}	Stage 1+2a+2b P_{nom}	Stage 1+2a+2b $P^{a}_{corrected}$
rs2300478	1.74 (1.57–1.92)	4.89E–10	5.93E–12	2.19E–03	3.41E–28	8.08E–23
rs9296249	1.67 (1.49–1.89)	2.19E–06	1.61E–06	4.14E–03	3.99E–18	9.44E–13
rs9357271	1.66 (1.48–1.87)	5.48E–06	1.85E–06	2.48E–03	6.31E–18	1.50E–12
rs12593813	1.50 (1.36–1.66)	1.85E–06	4.95E–05	1.57E–02	1.06E–15	2.51E–10
rs11635424	1.51 (1.37–1.67)	1.77E–06	2.54E–05	6.60E–03	3.65E–16	8.64E–11
rs4489954	1.51 (1.36–1.67)	2.44E–06	2.60E–05	1.66E–02	2.68E–15	6.35E–10
rs3784709	1.52 (1.37–1.68)	3.56E–06	7.46E–05	1.79E–03	4.06E–16	9.61E–11
rs1026732	1.53 (1.39–1.70)	4.67E–07	2.78E–05	5.22E–03	6.09E–17	1.44E–11
rs6494696	1.52 (1.38–1.69)	1.79E–06	5.20E–05	5.22E–03	2.00E–16	4.74E–11

SNPs with significant association that were successfully genotyped in all three case-control samples, located in three different genomic regions. Genome positions refer to the human March 2006 (hg18) assembly. [*MAP2K5*/*LBXCOR1*] denotes an intergenic position of the SNP. MAF, minor allele frequency; OR, odds ratio; c.i., confidence interval; P_{nom} = nominal *P* value. MAF refers to stage 2a data only; OR was calculated using combined data from all stages. *P* values for stage 1, 2a and combined analysis were calculated using logistic regression implementing an Armitage trend test and taking sex and age as covariates into account. *P* values in stage 1 and 2a resulting from this regression were further corrected for population stratification by dividing the resulting χ^2 by the inflation factor λ. $^{a}P_{corrected}$ = *P* value corrected for multiple testing using Bonferroni's method, correcting for 236,758 SNPs.

Fine mapping, haplotype and risk analysis

We genotyped tagging SNPs and all known coding and splice-site SNPs for fine mapping in the stage 2a samples. This confirmed the candidate regions defined by the explorative phase of the study (**Fig. 4**). Haplotype analysis for *MEIS1* delineated a haplotype block (rs3890755 to rs12469063). A haplotype completely described by allele A (rs6710341) and allele G (rs12469063) was more strongly associated than each single SNP in this block ($P = 5.87 \times 10^{-20}$, OR = 2.75 [95% confidence interval, 2.23–3.41]). This haplotype was also maximally associated in the Canadian sample ($P = 8.51 \times 10^{-7}$, OR = 2.36 [1.40–3.97], **Fig. 5**). For *BTBD9* and the *MAP2K5* and *LBXCOR1* region haplotype analysis confirmed the results of single-SNP analysis.

In exploratory analysis, we compared the ORs obtained under the allele dosage model to those obtained under the unrestricted model. For *MEIS1* and *BTBD9*, we did not find any significant difference between the models tested (MEIS1: $P = 0.714$; BTBD9: $P = 0.913$), but the allele dosage model was more parsimonious. For the *MAP2K5* and *LBXCOR1* region, the allele dosage model was significantly less likely than the unrestricted model ($P = 0.006$). Estimates pointed to a recessive model. This model was significantly better than the allele dosage model ($P = 0.009$) and not worse than the unrestricted model ($P = 0.395$). There was no difference in effect estimates between samples (**Supplementary Table 6** online).

In the combined German samples, lower limits of the sequential attributable fraction (SAFs)[22,23] were estimated at 0.092, 0.303 and 0.079 for *MEIS1*, *BTBD9* and the *MAP2K5* and *LBXCOR1* region respectively. Corresponding upper limits (equal to the population attributable risk fraction (ARF)) were 0.227, 0.492 and 0.201. In the Canadian sample, the lower limits of the SAFs were 0.075, 0.316 and 0.090, respectively, and we estimated the upper limits at 0.226, 0.550 and 0.258, respectively. We could not identify any statistical interaction between these loci, either in the individual samples or in the combined German or combined German-Canadian samples. Overall, although the single ARF and SAF estimates may be slightly overestimated, they clearly indicate that the three loci account for a large part of the phenotype in the populations studied. We estimated the ARF jointly attributable to the three loci at 68.6% in the German population and 74.2% for the Canadian population.

A comparison of familial versus sporadic cases in the combined stage 1 and 2a data set demonstrated virtually indistinguishable ORs

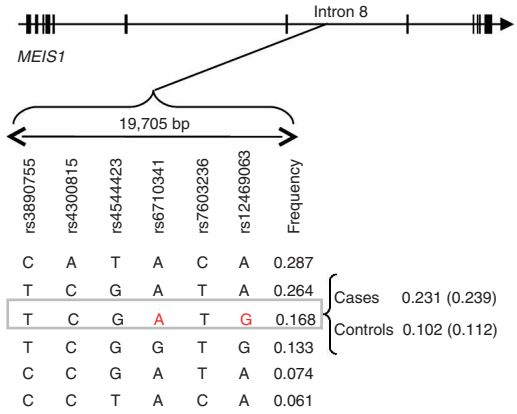


Figure 5 Haplotype structure for *MEIS1*. A haplotype consisting of six SNPs (of which rs6710341 and rs12469063 fully tagged the risk haplotype) is associated with RLS with odds ratios of 2.75 and 2.36 in the stage 2a and 2b samples, respectively. Haplotype frequencies for all haplotypes occurring with these six SNPs are based on cases and controls jointly and are given for cases and controls separately for the risk haplotype. For the Canadian sample, the frequencies are given in brackets and are based on the two tagging SNPs.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

for the regions on 6p and 15q. For the region on 2p, the risk was higher in familial (rs2300478: OR = 1.82 [1.55–2.14]) than in sporadic cases (OR = 1.59 [1.34–1.90]). However, confidence intervals were overlapping with no significant difference in allele distributions ($P = 0.22$, **Supplementary Table 7** online). The familial relative risk figures estimated by the risk to siblings λ_s were 1.13 for *MEIS1*, 1.02 for *BTBD9* and 1.03 for *MAP2K5/LBXCOR1* in the German data set, with almost identical estimates in the Canadian data.

The increasing medical attention to RLS in recent years is matched by our ignorance about its underlying molecular basis. The genetic heterogeneity of RLS has made linkage studies notoriously difficult and favors association approaches. In agreement with power calculations, an initial genome-wide screen for common variants in 400 cases and 1,600 controls enabled us to detect risk alleles with odds ratios >1.5. Sample size in the replicate was twice as high as in the initial GWAS and provided unequivocal evidence for the signals. The effects were strong enough that a second replication in a small independent sample from Canada also yielded significant signals for all three regions. A particular feature of our study design is the use of a control group from the general population. This provided us with very accurate estimates of the genotype frequencies and it avoided any bias to which a disease-negative population is prone.

The identification of significant signals in genes that have not been considered candidates from previous biological knowledge is a recurring theme in GWASs[24]. The current knowledge about *MEIS1*, *BTBD9*, *MAP2K5* and *LBXCOR1* opens new avenues of RLS research, and the involvement of developmental genes challenges us to rethink our basic concept of this widespread disease.

A major proportion of the risk for RLS is explained by variants in the loci identified. We could not derive any different contributions from any of these loci to familial versus sporadic RLS. The associated variants all convey very low familial relative risk ($\lambda_s < 1.15$ in all cases). The lack of positive results within the known linkage regions does not argue against the validity of the linkage results. The nominally significant signals detected in the RLS3 linkage region might indicate an allelic series of variants conferring weak and strong effects within the same gene.

This study is not exhaustive in identifying genetic factors contributing to RLS, and further investigations will provide a better picture of what constitutes the genetic architecture of the complex phenotype of restless legs syndrome. Future studies should investigate endophenotypes or secondary RLS cases, which might show alternative signal patterns. An interesting question is also whether the loci identified have a role in other dopaminergic disorders such as Parkinson's disease or in other associated disorders such as attention deficit hyperactivity disorder or sleep disorders. Further experimental advances might include features such as higher sample numbers in the exploratory stage, higher SNP density, modification of clustering algorithms[25], inclusion of lower frequency polymorphisms, investigation of copy number changes and use of lower statistical thresholds using *a priori* information.

METHODS

Study population and phenotype assessment. Cases of stages 1 and 2a were of European descent and were diagnosed according to standard criteria[2] in a personal interview. Familial RLS was defined by at least one affected first-degree relative. We excluded subjects with secondary RLS due to uremia, dialysis and iron deficiency.

Controls of stage 1 and 2a were of European descent and from the KORA S3/F3 and S4 surveys, representative of the general population. KORA procedures have been described[8]. For stage 1, we included 1,644 subjects from S3/F3, ages 35–84 years, and for stage 2a, 891 age- and sex-matched subjects from S4. In 2a, 102 affected individuals were outside the age range of KORA and were matched to the next age group.

Affected individuals and controls of stage 2b were of French-Canadian ancestry. Affected individuals ($n = 255$) were diagnosed according to standard criteria[2], and polysomnography was performed in 156 subjects; of those, 82.1% ($n = 128$) showed significant periodic leg movements during sleep. Controls were recruited from the general population (n = 287). Secondary cases were excluded.

Studies were performed according to the declaration of Helsinki and approved by institutional review boards in Germany, Austria, and Canada. Written informed consent was obtained from participants. For demographic data of successfully genotyped samples, see **Supplementary Table 8** online.

Genome-wide assays, SNP genotyping and quality control. Stage 1 genotyping was performed using the Affymetrix 500K Array Set. Genotypes were determined using the BRLMM algorithm with cases and controls undergoing a joint cluster analysis. From 500,568 SNPs, a total of 236,758 were selected for subsequent analyses based on stringent quality control criteria. Exclusion criteria were call rate <98% ($n = 146,297$), minor allele frequency (MAF) <10% ($n = 151,583$), deviations from HWE ($P < 0.00001$, $n = 22,536$) and low number of heterozygotes (<10, $n = 33,122$). 14,069 SNPs were monomorphic. For a detailed breakdown, see **Supplementary Table 1**.

For the 13 SNPs passing the inclusion criteria for genotyping in stages 2, visual inspection of clustering was performed using the Affymetrix SNP Signaling Tool 1.0.0.12. All clusters passed this test. To validate the stage 1 experiment, we genotyped 15 SNPs in 400 samples on another platform (Sequenom MassArray system) with a genotype discordance rate of 0.2%.

Stage 2 and fine-scale mapping were performed using MALDI-TOF mass spectrometry on a Sequenom system (Autoflex HT and SpectroTYPER RT 3.4 analysis software). Assays were designed using AssayDesign 3.1.2.2 with iPLEX Gold chemistry default parameters. **Supplementary Table 9** online lists oligonucleotide sequences of replication and fine mapping.

SNP quality control criteria leading to exclusion were call rate <97%, MAF <10% and $P < 0.001$ for deviations from HWE in controls. This resulted in an exclusion of one SNP (rs2110974) in stage 2a, two SNPs (rs2110974, rs7881785) in stage 2b and 51 SNPs in fine mapping. All coding SNPs were monomorphic. A total of 28 affected individuals and 55 controls in stage 2a and 44 affected individuals and 46 controls in stage 2b were excluded owing to low call rate (<90%) of all SNPs within a single DNA sample.

SNP selection for stage 2. We used the following inclusion criteria: (i) $P < 10^{-6}$ in stage 1 analysis (four SNPs); (ii) $P \leq 10^{-5}$ with two neighboring SNPs (± 100 kb) with $P \leq 10^{-3}$ (eight SNPs); (iii) $P \leq 10^{-4}$ for SNPs within described linkage peaks (one SNP in RLS3). For these 13 SNPs, we chose 15 additional neighboring SNPs based on LD structure for genotyping in the replication samples 2a and 2b (**Supplementary Table 3**).

SNP selection for fine mapping. SNPs in the coding regions and 10 kb of flanking sequences were selected using the Tagger algorithm ($r^2 = 0.8$) implemented in HAPLOVIEW 3.3.2 (ref. 26). In addition, all coding-region SNPs and splice-site SNPs were included. This led to 41 SNPs on chromosome 2p (38 tagging, 1 synonymous and 2 nonsynonymous), 77 SNPs on chromosome 6p (tagging only) and 46 SNPs on chromosome 15q (37 tagging, 1 synonymous, 4 nonsynonymous, 2 splice site, 2 frameshift coding). In total, 164 SNPs were selected, of which 163 were converted into genotyping assays, and 103 with a MAF >10% were analyzed.

Analysis of genetic effects. To test and correct for possible population stratification, we performed an EIGENSOFT[27,28] analysis. We used a random sample of 16,000 SNPs passing the quality criteria for the stage 1 sample and allowed for ten rounds of outlier removal. In the first six rounds, a total of 50 outliers (8 cases and 42 controls) were removed, with none removed in the remaining rounds. To assess stratification, we compared the expected distribution of P values for association versus the expected χ^2 distribution with one degree of freedom[29]. We compared the empirically observed mean of the lower 90% of the distribution of the statistics observed and divided it by its expectation[9]. This led to an inflation factor (λ) of 1.09 (**Fig. 2**).

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

We performed logistic regression analysis coding the number of minor alleles as the dependent variables, thus implementing Armitage's trend test, including age and sex as covariates and allowing for interactions between age, sex and the number of alleles. Odds ratios and confidence interval limits were obtained through logistic regression analysis. The χ^2 values resulting from these analyses (stage 1, 2a and fine mapping) were divided by λ, assuming similar conditions for both German samples.

Haplotype analysis was performed using HAPLOVIEW 3.3.2 (refs. 26,29), with the fraction of strong LD informative comparisons set at 0.9, and using UNPHASED 3.0.8, which allows the incorporation of age and sex as covariates[30]. Haplotype blocks were delineated using the method of Gabriel implemented in HAPLOVIEW[26]. ORs were obtained using logistic regression with age and sex as covariates in the stage 2 samples. χ^2 and P values in 2a were λ corrected. Differences between familial and sporadic cases were tested using Fisher's exact test. Familial attributable risks were calculated using the power calculator described in ref. 22.

Multiple testing. Using WG-PERMER, a program for rapid permutation of genome-wide data, preliminary analysis showed that P values after Westfall-Young and Bonferroni correction, with the number of tests set at the number of SNPs tested ($n = 236{,}758$), were in good agreement. This may reflect stringent criteria for SNPs to enter the analysis, resulting in low average r^2 values between SNPs. To maintain comparability across results, we show Bonferroni-corrected P values for stage 1 and the combined analysis. For stages 2a and 2b, we give Westfall-Young–corrected P values based on 10,000 permutations, as only a few candidate regions were tested with high LD between them, and thus Bonferroni would be conservative.

Power analysis. Power analysis for the combined German sample was performed using the Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/). The power of any SNP tested with MAF $\geq$ 0.2 and OR $\geq$ 1.5 (or 1/1.5 or lower) was beyond 90%. The P value used was the P value required for a significant result after Bonferroni correction ($P = 0.05/236{,}758 = 2.112 \times 10^{-7}$).

Testing the mode of inheritance. OR values and likelihoods were obtained using logistic regression analysis with age, sex and samples as covariates in the combined stage 1, 2a and 2b samples. Significance testing between models was done using the likelihood ratio test.

Attributable risk fraction. To quantify the contribution of these loci to RLS, we estimated the population attributable risk fraction (ARF)[22] and the sequential attributable fraction (SAF)[23]. We used the allele dosage model for *MEIS1* and *BTBD9* and the recessive model for *MAP2K5* and *LBXCOR1* and calculated upper and lower limits of SAFs[23]. For each locus, we used the SNP with the lowest P value, aware of the fact that this may lead to slight overestimation of the ARF. The ARF for the three loci combined was calculated by allowing for the possibility of simultaneous exposure to several of the risk genotypes.

Requests for materials: janew@mpipsykl.mpg.de or meitinger@gsf.de.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We are grateful to all patients who participated in this study. The authors also thank T.M. Strom, J. Favor, D. Vogt-Weisenhorn, W. Wurst and I. Tews for discussions and R. Feldmann, J. Golic, K. Junghans, B. Schmick, N. Trapp, M. Petzold, G. Fischer and M. Putz for technical assistance. We acknowledge L. Habersack, H. Rhese and J. Schmidt-Evers from the German RLS patient organization for supporting this study. 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 co-workers, who are responsible for the design and conduct of the KORA studies. The KORA (Cooperative Research in the Region of Augsburg) research project was initiated and financed by the National Research Centre for Environment and Health (GSF), which is funded by the German Federal Ministry of Education and Research and by the State of Bavaria. S.H. was partly supported by a grant from the German RLS patient organization. J.W. was partly supported by a grant form the Bavarian Ministry of Science, Culture and Art. The Canadian part of the study was supported by a Canadian Institutes of Health Research (CIHR) grant to G.R, J.M and G.T.

AUTHOR CONTRIBUTIONS

Study design: J.W., P.L., G.R., F.H., B.M.-M., T.M.; recruitment and biobanking of individuals with RLS: J.W., S.H., C.T., A.Z., K.S.-K., W.O., C.B., W.P., I.P., I.E., T.M.; recruitment and biobanking of KORA controls: C.G., T.I., H.-E.W.; recruitment and biobanking of Canadian affected individuals and controls: L.X., J.M., G.T., G.R.; Affymetrix genotyping: B.S., P.L., G.E.; Sequenom genotyping: B.S., P.L., S.J.; supervision of typing of all markers: J.W., P.L.; software development and data processing: S.R.,B.P.; statistical analysis: S.R., B.P., B.M.-M.; clustering of Affymetrix genotypes: S.R., B.M.-M.; manuscript writing: J.W., B.S., S.F., L.X., F.H., B.M.-M., T.M.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Walters, A.S. *et al.* A questionnaire study of 138 patients with restless legs sydrome: the 'night-walkers' survey. *Neurology* **46**, 92–95 (1996).
2. 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).
3. Winkelman, J.W., Finn, L. & Young, T. Prevalence and correlates of restless legs syndrome symptoms in the Wisconsin Sleep Cohort. *Sleep Med.* **7**, 545–552 (2006).
4. Barrière, G., Cazalets, J.R., Bioulac, B., Tison, F. & Ghorayeb, I. The restless legs syndrome. *Prog. Neurobiol.* **77**, 139–165 (2005).
5. Winkelmann, J. *et al.* Genetics of restless legs syndrome (RLS): state-of-the-art and future directions. *Mov. Disord.,* published online 7 June 2007 (doi:10.1002/mds.21587).
6. Rao, S., Winkelmann, J. & Wang, Q.K. in *Restless Legs Syndrome. Diagnosis and Treatment* (ed. Ondo,W.G.) 111–123 (Informa Healthcare, New York, 2007).
7. Kemlink, D. *et al.* Family-based association study of the restless legs syndrome loci 2 and 3 in a European population. *Mov. Disord.* **22**, 207–212 (2007).
8. Wichmann, H.E., Gieger, C., Illig, T. & MONICA/KORA Study Group KORA-gen-resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67**, S26–S30 (2005).
9. Clayton, D.G. *et al.* Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat. Genet.* **37**, 1243–1246 (2005).
10. 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 (2005).
11. Maeda, R. *et al.* Xmeis1, a protooncogene involved in specifying neural crest cell fate in *Xenopus* embryos. *Oncogene* **20**, 1329–1342 (2001).
12. Mercader, N. *et al.* Conserved regulation of proximodistal limb axis development by Meis/Hth. *Nature* **402**, 425–429 (1999).
13. Rajaram, S.S., Walters, A.S., England, S.J., Mehta, D. & Nizam, F. Some children with growing pains may actually have restless legs syndrome. *Sleep* **27**, 767–773 (2004).
14. Dasen, J.S., Tice, B.C., Brenner-Morton, S. & Jessel, T.M. A Hox regulatory network establishes motor neuron pool identity and target-muscle connectivity. *Cell* **123**, 477–491 (2005).
15. 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**, 1609–1615 (2000).
16. Allen Institute for Brain Science. Allen Brain Atlas. *http://www.brain-map.org* (2004).
17. 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 (2005).
18. Godt, D., Couderc, J.L., Cramton, S.E. & Laski, F.A. Pattern formation in the limbs of *Drosophila*: bric à 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).
19. Dinev, D. *et al.* Extracellular signal regulated kinase 5 (ERK5) is required for the differentiation of muscle cells. *EMBO Rep.* **2**, 829–834 (2001).
20. Cavanaugh, J.E., Jaumotte, J.D., Lakoski, J.M. & Zigmond, M.J. Neuroprotective role of ERK1/2 and ERK5 in a dopaminergic cell line under basal conditions and in response to oxidative stress. *J. Neurosci. Res.* **84**, 1367–1375 (2006).
21. Gross, M.K., Dottori, M. & Goulding, M. Lbx1 specifies somatosensory association interneurons in the dorsal spinal cord. *Neuron* **34**, 535–549 (2002).

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

22. Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association. *Nat. Genet.* **38**, 209–213 (2006).
23. Rowe, A.K., Powell, K.E. & Flanders, W.D. Why population attributable fractions can sum to more than one. *Am. J. Prev. Med.* **26**, 243–249 (2004).
24. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
25. Plagnol, V., Cooper, A.D., Todd, J.A. & Clayton, D.G. A method to address differential bias in genotyping in large–scale association studies. *Plos Genetics* **3**, e74 10.1371/journal.pgen.0030074 (2007).
26. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualisation of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
27. Patterson, N., Price, A.L. & Reich, D. Population structure and Eigenanalysis. *Plos Genetics* **2**, e190 (2006).
28. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
29. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
30. Dudbridge, F. UNPHASED user guide. *Technical Report 2006/5*. MRC Biostatistics Unit, Cambridge, UK, (2006).

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

Genome-wide association study in restless legs syndrome identifies common variants in three genomic regions

Juliane Winkelmann, Barbara Schormair, Peter Lichtner, Stephan Ripke, Lan Xiong, Shapour Jalilzadeh, Stephany Fulda, Benno Pütz, Gertrud Eckstein, Stephanie Hauk, Claudia Trenkwalder, Alexander Zimprich, Karin Stiasny-Kolster, Wolfgang Oertel, Cornelius G. Bachmann, Walter Paulus, Ines Peglau, Ilonka Eisensehr, Jacques Montplaisir, Gustavo Turecki, Guy Rouleau, Christian Gieger, Thomas Illig, H.-Erich Wichmann, Florian Holsboer, Bertram Müller-Myhsok, Thomas Meitinger.

Supplementary Information

Supplementary Table 1: Stage 1 SNP exclusion.

Reason for exclusion	Number of SNPs excluded
Monomorphic >10%	14,069
HWE (deviation from HWE (P < 0.00001 in controls))	854
MAF (minor allele frequency < 0.10)	72,099
HET (< 10 heterozygotes for this SNP)	0
CR (call rate < 98%)	79,851
MAF + HWE	552
MAF + HET	29,889
CR + HWE	17,453
CR + MAF	42,194
HET + MAF + HWE	50
CR + MAF + HWE	3,616
CR + HET + MAF	3,172
CR + HET + MAF + HWE	11
Total	263,810

Detailed breakdown of the SNPs that did not pass the quality control or were monomorphic and therefore not entered subsequent analysis.

Supplementary Table 2: Stage 1 association results.

SNP ID	Chr	Genome pos	Gene	HWE cases	HWE controls	MAF cases	MAF controls	MAF HapMap	OR (95% CI)	P_{nom}	$P_{corrected}$ (B)
rs2300478	2p	66,634,957	MEIS1	0.669	0.948	0.368 (G)	0.253 (G)	0.242 (G)	1.77 (1.49-2.10)	4.89E-10	<0.0002
rs9296249	6p	38,473,819	BTBD9	0.432	0.587	0.150 (C)	0.239 (C)	0.200 (C)	0.59 (0.48-0.74)	2.19E-06	0.519
rs9357271	6p	38,473,851	BTBD9	0.336	0.736	0.154 (C)	0.240 (C)	0.198 (C)	0.61 (0.49-0.75)	5.48E-06	1
rs4626664	9p	9,251,737	PTPRD	0.527	0.162	0.199 (A)	0.133 (A)	0.133 (A)	1.68 (1.36-2.08)	6.81E-06	1
rs12593813	15q	65,823,906	MAP2K5	0.016	0.296	0.249 (A)	0.340 (A)	0.317 (A)	0.64 (0.53-0.76)	1.85E-06	0.437
rs11635424	15q	65,824,632	MAP2K5	0.016	0.272	0.249 (A)	0.340 (A)	0.317 (A)	0.64 (0.53-0.76)	1.77E-06	0.419
rs4489954	15q	65,859,129	MAP2K5	0.007	0.211	0.230 (T)	0.318 (T)	0.288 (T)	0.63 (0.53-0.76)	2.44E-06	0.578
rs3784709	15q	65,859,329	MAP2K5	0.007	0.345	0.244 (T)	0.334 (T)	0.325 (T)	0.65 (0.54-0.77)	3.56E-06	0.844
rs1026732	15q	65,882,139	MAP2K5	0.019	0.245	0.241 (A)	0.337 (A)	0.317 (A)	0.62 (0.51-0.74)	4.67E-07	0.111
rs6494696	15q	65,890,260	[MAP2K5/ LBXCOR1]	0.007	0.245	0.246 (C)	0.338 (C)	0.317 (C)	0.64 (0.53-0.76)	1.79E-06	0.423
rs6500963	16p	7,407,490	A2BP1	0.288	0.408	0.253 (T)	0.343 (T)	0.246 (T)	1.4 (1.11-1.78)	2.69E-06	0.638
rs1983167	Xp	42,733,328		0.478	0.107	0.263 (A)	0.367 (A)	0.356 (A)	0.66 (0.56-0.77)	6.48E-07	0.153
rs7881785	Xp	42,739,550		0.573	0.045	0.265 (A)	0.367 (A)	0.356 (A)	0.66 (0.56-0.78)	9.01E-07	0.213

Supplementary table 2: Stage 1 association results (continued).

These 13 SNPs showed nominally significant association and were selected for replication. Chr, Chromosome; HWE, P value for the deviation from Hardy-Weinberg-Equilibrium; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval; P_{nom}, nominal P value. Genome positions (Genome pos) refer to the Human March 2006 (hg18) assembly (http://genome.ucsc.edu/index.html). [gene] denotes intergenic position of SNP. HapMap refers to HapMap rel21a_NCBI_Build35 (http://www.hapmap.org). Minor allele annotation refers to HapMap data. Nominal P values were calculated using logistic regression implementing Armitage trend test and taking sex and age as covariates into account. Odds ratios and confidence limits were calculated from logistic regression. P values resulting from this regression were further corrected for population stratification by dividing the resulting χ^2 by the inflation factor λ, i.e. by dividing with 1.09. $P_{corrected}$, P value corrected for multiple testing using Bonferroni (B).

Supplementary Table 3: Stage 2a and 2b SNP selection.

Region	Chr	Genome position	dbSNP ID	Distance (bp) to * SNP	r^2	Genotyped in stage 1	stage 2a	stage 2b
1	2p	66,670,073	rs6710341	23,031	0.576	-	+	+
		66,675,959	rs12469063	17,145	0.953	-	+	+
		66,693,104	rs2300478	*	*	+	+	+
		66,700,242	rs2110974	-7,138	0.517	-	-	-
2	6p	38,440,588	rs9394492	33,231	0.482	-	+	+
		38,469,090	rs4714156	4,729	0.948	-	+	+
		38,473,819	rs9296249	*	*	+	+	+
		38,473,851	rs9357271	-32	1	+	+	+
		38,548,948	rs3923809	-75,129	0.436	-	+	+
3	9p	9,206,980	rs10816064	44,757	0.722	-	+	+
		9,243,587	rs7872553	8,150	0.858	-	+	+
		9,251,737	rs4626664	*	*	+	+	+
		9,262,841	rs4302899	-11,104	0.769	-	+	+
4	15q	65,823,906	rs12593813	35,223	0.779	+	+	+
		65,824,632	rs11635424	34,497	0.779	+	+	+
		65,841,442	rs884202	17,687	0.809	-	+	+
		65,859,129	rs4489954	*	*	+	+	+
		65,859,329	rs3784709	-200	0.820	+	+	+
		65,882,139	rs1026732	-23,010	0.852	+	+	+
		65,890,260	rs6494696	-31,131	0.852	+	+	+
5	16p	7,406,554	rs6500961	936	0.714	-	+	+
		7,407,490	rs6500963**	*	*	+	-	-
		7,407,527	rs6500964**	-37	0.832	-	-	-
		7,408,353	rs7194617	-863	0.742	-	+	+
6	Xp	42,591,364	rs6520824	13,274	0.790	-	+	+
		42,604,638	rs1983167	*	*	+	+	+
		42,610,860	rs7881785	-6,222	1	+	+	-
		42,621,074	rs6610746	-16,436	0.335	-	+	+

28 SNPs that were chosen for genotyping in stage 2a and 2b (13 SNPs from stage 1 and 15 additional SNPs). The additional SNPs were selected based on LD structure: We choose one of the original SNPs for each of the six regions and sought one SNP with an r^2-value of ≥ 0.9 as a technical replicate, two SNPs with an $r^2 \approx 0.8$ and two SNPs with an $r^2 \approx 0.7$ upstream and downstream, respectively, of the original SNP. In regions where these criteria could not be met, neighbouring SNPs with maximum r^2 obtainable in this region were chosen.

+ indicates successfully genotyped. Genome positions (Genome pos) refer to the Human March 2006 (hg18) assembly (http://genome.ucsc.edu/index.html).
* SNP, reference SNP from which the distance within the region was computed.
r^2, r^2-values between the reference *SNP and its respective neighboring SNPs as downloaded from HapMap rel21a_NCBI_Build35 (http://www.hapmap.org).
** not genotyped in stage 2 because no PCR-primer could be designed.

Supplementary Table 4: Stage 2a association results.

dbSNP ID	Chr	Genome pos	Gene	HWE cases	HWE controls	MAF cases	MAF controls	MAF HapMap	OR (95% CI)	P_{nom}	$P_{corrected}$ (WY)
rs6710341	2p	66,611,925	MEIS1	1	1	0.141 (G)	0.142 (G)	0.158 (G)	0.96 (0.77-1.17)	6.53E-01	1
rs12469063	2p	66,617,811	MEIS1	0.661	0.923	0.363 (G)	0.236 (G)	0.241 (G)	1.78 (1.52-2.10)	6.52E-12	0.001
rs2300478	2p	66,634,957	MEIS1	0.383	1	0.367 (G)	0.241 (G)	0.242 (G)	1.78 (1.52-2.09)	5.93E-12	0.001
rs9394492	6p	38,440,588	BTBD9	0.027	0.820	0.292 (T)	0.352 (T)	0.342 (T)	0.76 (0.64-0.87)	4.07E-04	0.008
rs4714156	6p	38,469,089	BTBD9	0.262	0.924	0.163 (T)	0.240 (T)	0.214 (T)	0.61 (0.51-0.74)	6.50E-07	0.001
rs9296249	6p	38,473,819	BTBD9	0.262	0.848	0.162 (C)	0.235 (C)	0.200 (C)	0.62 (0.52-0.75)	1.61E-06	0.001
rs9357271	6p	38,473,851	BTBD9	0.178	0.849	0.165 (C)	0.238 (C)	0.198 (C)	0.63 (0.52-0.75)	1.85E-06	0.001
rs3923809	6p	38,548,947	BTBD9	0.003	0.465	0.207 (G)	0.307 (G)	0.259 (G)	0.57 (0.48-0.68)	1.75E-09	0.001
rs10816064	9p	9,206,979	PTPRD	0.895	1	0.152 (A)	0.134 (A)	0.100 (A)	1.21 (0.98-1.48)	9.04E-02	0.917
rs7872553	9p	9,243,586	PTPRD	0.306	0.547	0.156 (C)	0.133 (C)	0.117 (C)	1.27 (1.03-1.56)	3.22E-02	0.616
rs4626664	9p	9,251,737	PTPRD	0.232	0.213	0.170 (A)	0.147 (A)	0.133 (A)	1.26 (1.03-1.55)	3.01E-02	0.716
rs4302899	9p	9,262,840	PTPRD	0.407	0.107	0.172 (A)	0.151 (A)	0.167 (A)	1.24 (1.01-1.52)	4.39E-02	0.853
rs12593813	15q	65,823,906	MAP2K5	0.479	0.814	0.258 (A)	0.330 (A)	0.317 (A)	0.71 (0.60-0.83)	4.95E-05	0.001
rs11635424	15q	65,824,632	MAP2K5	0.595	0.584	0.257 (A)	0.330 (A)	0.317 (A)	0.70 (0.59-0.82)	2.54E-05	0.001
rs884202	15q	65,841,441	MAP2K5	0.653	0.481	0.250 (G)	0.328 (G)	0.319 (G)	0.69 (0.58-0.81)	1.17E-05	0.001
rs4489954	15q	65,859,129	MAP2K5	0.779	0.872	0.239 (T)	0.311 (T)	0.288 (T)	0.69 (0.59-0.82)	2.60E-05	0.001
rs3784709	15q	65,859,329	MAP2K5	0.928	0.579	0.251 (T)	0.321 (T)	0.325 (T)	0.71 (0.60-0.83)	7.46E-05	0.001
rs1026732	15q	65,882,139	MAP2K5	0.530	0.582	0.252 (A)	0.327 (A)	0.317 (A)	0.70 (0.59-0.82)	2.78E-05	0.001
rs6494696	15q	65,890,260	[MAP2K5/ LBXCOR1]	0.476	0.431	0.253 (C)	0.326 (C)	0.317 (C)	0.71 (0.60-0.83)	5.20E-05	0.001

Supplementary table 4: Stage 2a association results (continued).

dbSNP ID	Chr	Genome pos	Gene	HWE cases	HWE controls	MAF cases	MAF controls	MAF HapMap	OR (95% CI)	P_{nom}	$P_{corrected}$ (WY)
rs6500961	16p	7,406,553	A2BP1	0.489	0.600	0.326 (A)	0.363 (A)	0.283 (A)	0.84 (0.73-0.98)	3.67E-02	0.342
rs7194617	16p	7,408,352	A2BP1	0.888	0.724	0.399 (T)	0.429 (T)	0.317 (T)	0.89 (0.77-1.03)	1.28E-01	0.726
rs6520824	Xp	42,720,053		0,254	0,450	0.393 (C)	0.411 (C)	0.411 (C)	0.96 (0.83-1.09)	5.21E-01	0.596
rs1983167	Xp	42,733,328		0,491	0,215	0.339 (A)	0.358 (A)	0.356 (A)	0.95 (0.83-1.08)	4.43E-01	0.530
rs7881785	Xp	42,739,550		0,488	0,285	0.334 (A)	0.356 (A)	0.356 (A)	0.95 (0.83-1.09)	4.57E-01	0.453
rs6610746	Xp	42,749,763		0,348	0,430	0.367 (A)	0.346 (A)	0.378 (A)	1.06 (0.93-1.21)	4.17E-01	0.478

Chr, Chromosome; HWE, P value for the deviation from Hardy-Weinberg-Equilibrium; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval; P_{nom}, nominal P value. Genome positions (Genome pos) refer to the Human March 2006 (hg18) assembly (http://genome.ucsc.edu/index.html). HapMap refers to HapMap rel21a_NCBI_Build35 (http://www.hapmap.org). Minor allele annotation refers to HapMap data. Nominal P values were calculated using logistic regression implementing Armitage trend test and correcting for population stratification. $P_{corrected}$, P value corrected for multiple testing using Westfall-Young method (WY).

Supplementary Table 5: Stage 2b association results.

dbSNP ID	Chr	Genome pos	Gene	HWE cases	HWE controls	MAF cases	MAF controls	MAF HapMap	OR (95% CI)	P_{nom}	$P_{corrected}$ (WY)
rs6710341	2p	66,611,925	MEIS1	1	0.277	0.133 (G)	0.141 (G)	0.158 (G)	0.87 (0.59-1.29)	4.93E-01	1
rs12469063	2p	66,617,811	MEIS1	0.882	0.607	0.355 (G)	0.248 (G)	0.241 (G)	1.68 (1.24-2.27)	6.54E-04	0.018
rs2300478	2p	66,634,957	MEIS1	0.653	0.505	0.358 (G)	0.261 (G)	0.242 (G)	1.59 (1.18-2.15)	2.19E-03	0.059
rs9394492	6p	38,440,588	BTBD9	1	0.888	0.280 (T)	0.355 (T)	0.342 (T)	0.71 (0.53-0.95)	1.99E-02	0.429
rs4714156	6p	38,469,089	BTBD9	1	0.177	0.140 (T)	0.213 (T)	0.214 (T)	0.60 (0.41-0.87)	6.79E-03	0.173
rs9296249	6p	38,473,819	BTBD9	1	0.451	0.140 (C)	0.218 (C)	0.200 (C)	0.59 (0.41-0.85)	4.14E-03	0.109
rs9357271	6p	38,473,851	BTBD9	1	0.578	0.140 (C)	0.222 (C)	0.198 (C)	0.57 (0.40-0.83)	2.48E-03	0.067
rs3923809	6p	38,548,947	BTBD9	0.139	1	0.207 (G)	0.313 (G)	0.259 (G)	0.58 (0.42-0.8)	8.68E-04	0.024
rs10816064	9p	9,206,979	PTPRD	0.015	1	0.133 (A)	0.106 (A)	0.100 (A)	1.40 (0.93-2.09)	1.05E-01	0.954
rs7872553	9p	9,243,586	PTPRD	0.055	1	0.128 (C)	0.104 (C)	0.117 (C)	1.35 (0.90-2.04)	1.46E-01	0.988
rs4626664	9p	9,251,737	PTPRD	0.084	0.746	0.138 (A)	0.108 (A)	0.133 (A)	1.39 (0.93-2.07)	1.02E-01	0.951
rs4302899	9p	9,262,840	PTPRD	0.144	0.513	0.140 (A)	0.112 (A)	0.167 (A)	1.34 (0.91-1.98)	1.43E-01	0.986
rs12593813	15q	65,823,906	MAP2K5	0.168	0.442	0.225 (A)	0.303 (A)	0.317 (A)	0.69 (0.51-0.94)	1.57E-02	0.357
rs11635424	15q	65,824,632	MAP2K5	0.240	0.372	0.227 (A)	0.315 (A)	0.317 (A)	0.67 (0.49-0.90)	6.60E-03	0.169
rs884202	15q	65,841,441	MAP2K5	0.158	0.452	0.220 (G)	0.311 (G)	0.319 (G)	0.65 (0.48-0.88)	4.56E-03	0.120
rs4489954	15q	65,859,129	MAP2K5	0.093	0.157	0.206 (T)	0.285 (T)	0.288 (T)	0.69 (0.51-0.94)	1.66E-02	0.374
rs3784709	15q	65,859,329	MAP2K5	0.147	0.368	0.211 (T)	0.309 (T)	0.325 (T)	0.62 (0.46-0.84)	1.79E-03	0.049
rs1026732	15q	65,882,139	MAP2K5	0.158	0.548	0.220 (A)	0.309 (A)	0.317 (A)	0.66 (0.49-0.88)	5.22E-03	0.136
rs6494696	15q	65,890,260	[MAP2K5/ LBXCOR1]	0.158	0.548	0.220 (C)	0.309 (C)	0.317 (C)	0.66 (0.49-0.88)	5.22E-03	0.136

Supplementary table 5: Stage 2b association results (continued).

dbSNP ID	Chr	Genome pos	Gene	HWE cases	HWE controls	MAF cases	MAF controls	MAF HapMap	OR (95% CI)	P_{nom}	$P_{corrected}$ (WY)
rs6500961	16p	7,406,553	A2BP1	0.111	0.772	0.386 (A)	0.330 (A)	0.283 (A)	1.27 (0.96-1.67)	8.72E-02	0.922
rs7194617	16p	7,408,352	A2BP1	0.217	1	0.476 (T)	0.383 (T)	0.317 (T)	1.46 (1.11-1.91)	5.60E-03	0.145
rs6520824	Xp	42,720,053		0.107	0.391	0.405 (C)	0.444 (C)	0.411 (C)	0.90 (0.72-1.13)	3.63E-01	1
rs1983167	Xp	42,733,328		0.096	0.479	0.357 (A)	0.384 (A)	0.356 (A)	0.92 (0.73-1.16)	4.73E-01	1
rs6610746	Xp	42,749,763		0.849	0.838	0.372 (A)	0.306 (A)	0.378 (A)	1.25 (0.99-1.59)	6.45E-02	0.844

Chr, Chromosome; HWE, P value for the deviation from Hardy-Weinberg-Equilibrium; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval; P_{nom}, nominal P value. Genome positions (Genome pos) refer to the Human March 2006 (hg18) assembly (http://genome.ucsc.edu/index.html). HapMap refers to HapMap rel21a_NCBI_Build35 (http://www.hapmap.org). Minor allele annotation refers to HapMap data. Nominal P values were calculated using logistic regression implementing Armitage trend test. $P_{corrected}$, P value corrected for multiple testing using Westfall-Young method (WY).

Supplementary Table 6: Delineation of genetic model.

dbSNP ID	Chr	Genome position	Region	Risk allele	Unrestricted model OR (95 % CI) Number of risk alleles = 1	Unrestricted model OR (95 % CI) Number of risk alleles = 2	Allele dosage model OR (95 % CI) Number of risk alleles = 1	Allele dosage model OR (95 % CI) Number of risk alleles = 2
rs2300478	2p	66,634,957	MEIS1	G	1.80 (1.43-2.27)	3.07 (2.45-3.86)	1.74 (1.57-1.92)	3.01 (2.47-3.67)
rs9296249	6p	38,473,819	BTBD9	T	1.67 (1.45-1.92)	2.85 (1.93-4.20)	1.67 (1.49-1.89)	2.80 (2.21-3.56)
rs1026732	15q	65,882,139	MAP2K5/ LBXCOR1	G	1.11 (0.87-1.42)	1.94 (1.51-2.48)	1.53 (1.39-1.70)	2.36 (1.92-2.89)

Chr, Chromosome; OR, odds ratio; CI, confidence interval. Odds ratios were obtained using logistic regression with age, sex and samples as covariates in the combined stage 1 and 2a and b samples.

Supplementary Table 7: Analysis of familial versus sporadic cases in combined 1 and stage 2a data set.

SNP	Chr	$P_{nominal}$	OR (95% CI) Familial	OR (95% CI) Sporadic
rs2300478	2p	2.23E-01	1.82 (1.55-2.14)	1.59 (1.34-1.9)
rs9296249	6p	7.31E-01	1.55 (1.27-1.89)	1.63 (1.31-2.02)
rs9357271	6p	6.81E-01	1.53 (1.25-1.86)	1.63 (1.31-2.02)
rs4626664	9p	1.02E-01	1.2 (0.97-1.49)	1.5 (1.21-1.86)
rs12593813	15q	1	1.5 (1.26-1.78)	1.5 (1.25-1.8)
rs11635424	15q	1	1.5 (1.27-1.79)	1.5 (1.25-1.8)
rs4489954	15q	7.19E-01	1.52 (1.28-1.82)	1.45 (1.21-1.75)
rs3784709	15q	8.60E-01	1.53 (1.29-1.83)	1.5 (1.25-1.81)
rs1026732	15q	9.53E-01	1.53 (1.28-1.81)	1.51 (1.26-1.82)
rs6494696	15q	1	1.52 (1.28-1.81)	1.51 (1.26-1.82)
rs1983167	Xp	1.97E-02	1.05 (0.89-1.23)	1.35 (1.13-1.61)
rs7881785	Xp	2.90E-02	1.07 (0.91-1.26)	1.37 (1.15-1.63)

Chr, Chromosome, OR, odds ratio; CI, confidence interval. Odds ratios were obtained using logistic regression with age and sex as covariates in the combined stage 1 and 2a samples. Nominal P values for differences between familial and sporadic cases were tested using Fisher`s exact test on the allele counts in familial and sporadic cases, respect.

Supplementary Table 8: Description of study subjects.

	Stage 1		Stage 2a		Stage 2b	
	Cases GER1	**Controls KORAS3/F3**	**Cases GER2**	**Controls KORAS4**	**Cases CAN**	**Controls CAN**
N individuals	393	1602	875	836	211	241
N females	287	815	645	618	133	136
N males	106	787	230	218	78	105
Mean age (SD)	60.7 (8.0)	62.6 (9.9)	60.6 (12.1)	69.9 (11.3)	53.0 (12.4)	42.2 (16.0)
Mean age at onset (SD)	33.2 (13.7)*	-	38.8 (16.9)*	-	27.7 (12.4)*	-
AaO (SD) females	32.6 (13.6)*	-	38.1 (16.5)*	-	27.0 (16.1)*	-
AaO (SD) males	34.6 (14.0)*	-	40.5 (18.0)*	-	28.8 (13.3)*	-
Positive family history	393 (100%)	-	418 (47.7%)	-	165/206 (80.1%)	-
Affymetrix 500K data	393	1602	-	-	-	-
Sequenom iPLEX data	393	-	875	836	211	241

Table includes only successfully genotyped samples. N, number; AaO= age at onset of the disease; GER, German; CAN, Canadian.

KORAS3/F3 and KORAS4, controls drawn from KORA population-based cohort study, Germany.

AaO is unknown for 13 cases (7 females, 6 males) in CAN, 21 (15 females, 6 males) in GER1 and 39 cases (30 females, 9 males) in GER2.

Supplementary Table 9: Oligonucleotide sequences for replication and finemapping.

dbSNP ID	Sequence forward PCR primer	Sequence reverse PCR primer	Sequence extension primer
rs1000756	ACG TTG GAT GGA GTT ACT TTT CTC TGT TGGC	ACG TTG GAT GAC AAC ACT AAT CAA TTT AAC	TTC TCT GTT GGC TTT TTT TTT CCA
rs10184250	ACG TTG GAT GAC GCC TTA GGC AGA AGC TC	ACG TTG GAT GTG AGG GTA TCC GAA AGG CTG	AGC TCC TCA GGA TCA CTT
rs1026731	ACG TTG GAT GCA GAA ATG GTG CTA ACA TGC	ACG TTG GAT GAT CTT CCA CCC ACG GTG AC	ATC TAA CAT GCT TTC CGC
rs1026732	ACG TTG GAT GAT GGT GCA GCT CCC TGG AAC	ACG TTG GAT GGG GTG GAA GAT GCT CTT GAC	TGG AAC CCA GGC ACC AAT A
rs10456462	ACG TTG GAT GCC AGA CTG GCA AGT AAA CAC	ACG TTG GAT GAG TGC CAT TTT AGT GGA AAG	CAG CAA ACT CAA CAC AT
rs10518744	ACG TTG GAT GCA GAC CCT ACC AGA AGA AAC	ACG TTG GAT GCA ACG ATT TTG GTT AGC AAG	GGA GGA ACC CTT GGT TCT GTG
rs10816064	ACG TTG GAT GGA ATT TTT TTG CCC TCT ACT C	ACG TTG GAT GTT TCC CAC ATT GGG CTT CAG	CCC TCT ACT CTT TAA AAT CC
rs10865353	ACG TTG GAT GCA GTG TGC ATG GTA CTC AGG	ACG TTG GAT GAC ATG GGC ACA TAC ATA CGG	GTA CTC AGG CAT ACA CA
rs10947715	ACG TTG GAT GGG GAT CCC AGG AGA AGT AAC	ACG TTG GAT GAG CCA ATA CCC AGC TAG CTC	CCC CTA AGT AAC CCC AGC CT
rs10947716	ACG TTG GAT GAC ACC AGA CAG GCA CCA AGA	ACG TTG GAT GCA GGC AGC TGG AGA CAC AT	GCA CCA AGA GAG CCT
rs10947737	ACG TTG GAT GCA AGG ATG AAT CAA GAC TTG G	ACG TTG GAT GGA CCA CAT ATG AAG ACG ATG	GGG AGG AAT CAA GAC TTG GCA ACA A
rs10947740	ACG TTG GAT GAA TAA TCC TCA AGA ACA AG	ACG TTG GAT GGC ACT CCT GGT AAT ATG TCC	CCT CAA GAA CAA GTT TGG A
rs10947749	ACG TTG GAT GGC ATC AAA AAG ACC CAA TGA C	ACG TTG GAT GAA AGA GCC TGC CTA TCA GCC	CTC CAA TGA CTT TTA TAA AAG TAG AAG
rs11071959	ACG TTG GAT GAC AGT AAA TTT GCG CTC CAG	ACG TTG GAT GTA GAG TCT CTG AAC TCC TCG	TCA CAG GAA GGG AAA G
rs11126082	ACG TTG GAT GCA GAT GCC CAC TGT GAT CTC	ACG TTG GAT GTT CCC CCA AAA CTG AAT TGC	ATC TCC TCA CCT CTC CT
rs11630854	ACG TTG GAT GGA AGT CAT CTA TCT CAT TG	ACG TTG GAT GCA ACC TGG GCA ACA AAG CAA	GGA ATC TAT CTC ATT GAT GAG ATA ATT T
rs11635286	ACG TTG GAT GCA ATG AAA GAT GTG AAT TCT C	ACG TTG GAT GTC ACT ATA GTT TTT CTA GG	TGA ATT CTC AAA TTT AAA AAA AAC CA
rs11635424	ACG TTG GAT GTG TTA GGG CTG ATG TAC TGC	ACG TTG GAT GTG AGG GCT TGG ACA AGT TAG	CGC CTG CAC CTC ACA ACA TA
rs11637445	ACG TTG GAT GCT CTT GCC TGA CTG TGA TTC	ACG TTG GAT GGA GGA AAG TAG AAT GGG AGC	CTG TGA TTC AGG CAA ATG
rs11688578	ACG TTG GAT GTA TGG TTC TGT AGC ATA GGG	ACG TTG GAT GCA AAG GAA TAT AAG ATT TTT G	TAA TCC ATG CTC CCT TT
rs11692361	ACG TTG GAT GCC CAT GCC TAG GAC ATA AAG	ACG TTG GAT GGA CGT GGA GCG ATA CTT GAA	ACA TTT CCC AGC ATC AC
rs11692504	ACG TTG GAT GGT GAA CTT TCT CTG GAT CTG	ACG TTG GAT GTA AAA ACC ACC CCT GAA ATG	AAA TAC TTG GTA TTT GTT TTC TT
rs11751154	ACG TTG GAT GGC AGC TTA CAC ACA CAA CAG	ACG TTG GAT GTG TAG CCT TTT GAA TCT TGC	CAC ACA CAA CAG CAT GTA T
rs11752799	ACG TTG GAT GAC ACT GGG AGT CCC AGT TTG	ACG TTG GAT GTA TCA GGT CAT CCT CAC TCG	CAG GAA ATC ATT AAG ATT TTA TTA TTT G
rs11757846	ACG TTG GAT GGT TAA CAA ACG TCG AGG GAG	ACG TTG GAT GAC ATC AGA ACC AGT GTG GCT	GAG GTA GTC TTT AGT TTC AGC CC
rs11757985	ACG TTG GAT GTT CCA CTA AAC ACA TCC TGC	ACG TTG GAT GCC AGT TTA CAA TGG TTC AGC	CCC GGC TTT CAC ATC ATT GTA ACA
rs11856999	ACG TTG GAT GAT GAG TTT GTC GGA GGA AGC	ACG TTG GAT GCC ACC ACA GCC CCC TTT AA	ATT GGG AAT GGC AGA ATT GTG CCA CT
rs11857017	ACG TTG GAT GTT AAA GAC ACT TTG AGG CTG	ACG TTG GAT GGA GCC TTC TGT CAT GCT AAG	ATA CAC TTT GAG GCT GAT TTT TAT TT
rs12050749	ACG TTG GAT GCT TCC CTT TTC TGG GCA TAG	ACG TTG GAT GCC CCT GTT AGT GTG GTT TTA	GGT TTC TGG GCA TAG ATT TTT TTT
rs12055513	ACG TTG GAT GGT CTG CAT TAT CTT TTC CAG	ACG TTG GAT GGA CTC ATT TGT CTA CAT TCG	CAA GTT TAT CCA ATA AGC TAT GC
rs12148363	ACG TTG GAT GGG GGC CAT CAA TTT TGA TT	ACG TTG GAT GGT ACA TCT GAG AGA GGA GCA	GAT TGA TTG TTT GGG GAC
rs12196956	ACG TTG GAT GGG TTC GAT TCT GTG AGG TTG	ACG TTG GAT GAG AGA CAA CCC ACG CAG ATG	GGA CGC AGG TCC ACA GAT AAC
rs12198616	ACG TTG GAT GTG TCG TTG CCA CTG CAC GAG	ACG TTG GAT GCT GTG AAC TTG ATT CTA CGG	CCC CGC ACA TAC ATA CAC ACG
rs12206905	ACG TTG GAT GCT GCT TTC AAA CAA CCA GG	ACG TTG GAT GGA ATG TGG GCA CAT TTA GAC	AGG TTA CAT ATG AAC TGT CAG AGC
rs12208647	ACG TTG GAT GAA ACT GCC GTT GTC AAG GTC	ACG TTG GAT GTG AAG CAA GAC AAG GAC TCG	ACC ACA GTG CTG AGC
rs12212721	ACG TTG GAT GCC TTC AAT ATG TCT GCG TTG	ACG TTG GAT GAA GTT CGG AAG TTC ACA CTG	CCC CTT TCC TGG TCA TGT GAT TCT T
rs12212820	ACG TTG GAT GCG AGA GTC GTC CTC CAC TG	ACG TTG GAT GTT CTT CCC CTG GCA AAC AGT	GAG CCA TGG GAA GTT
rs12373638	ACG TTG GAT GTT TTA CCC TTT AGC GGA GGA	ACG TTG GAT GAA GAC ATG GAG AGT CAG AGC	GGG GAG CAG TAA ATT CA
rs12441598	ACG TTG GAT GAA CAC CAG GGT TTA GCC ATC	ACG TTG GAT GCA TAT GTG CTG CTG ATC ACC	GAG CTA GGA GGT GCT
rs12469063	ACG TTG GAT GGA ACA TTC AAA AGC AAT TCA C	ACG TTG GAT GTT AAT GTC CCT ACA GAC TGC	AGC AAT TCA CTG CAT CA
rs12471916	ACG TTG GAT GAT GCT GAT GGA AGA GTG TGG	ACG TTG GAT GGG TCT TTC TTT TGA TGG GAC	CTT ACG AGT GTG GGT CAG GAA T
rs12592315	ACG TTG GAT GGA TTG CAA CAT TAA GCC CAG	ACG TTG GAT GCA AAA GGC TGT TTA CGA TTC	TTA GTG TTC ACA ATC TAG CA
rs12593664	ACG TTG GAT GTG TTA ACT ATA ACA GCA GCC	ACG TTG GAT GCA TGC ATG TGC TTA TTA GTC	CCC AAG CAG CCA AAA ATA ATG TAA
rs12593813	ACG TTG GAT GAG ACA CCA GCT ATA GCT TTC	ACG TTG GAT GTT CCA GAC AAG AGC TGC AGG	CTT TTC TCT TTT ACT CTC TGA AAT TA
rs12614369	ACG TTG GAT GTA CAG CAC TCA CCA CCT TAC	ACG TTG GAT GTG ATC TAG TAA GGC AGA ACC	CCC CAT CTA TAA ACA AGG C
rs12619205	ACG TTG GAT GCC ACT TCA GTG AGG CAT TCA	ACG TTG GAT GAG GTT CTG GAG AAT CTT CCC	ATT CAT GTG CAC CCA
rs12660215	ACG TTG GAT GGC AGG TGA ATA TGG ACT CCG	ACG TTG GAT GGC TGT CGG TAC ACT TAC TAC	CTA GGG ACT CCG GTG TTT CTG TTC AG
rs12664020	ACG TTG GAT GTC TAC CGA CCT GCA AAA AAC	ACG TTG GAT GCT GCT TTC TCG GTT ATT CTG	GAA ACA ACT ATT AGC ACC TGC TAG
rs12713566	ACG TTG GAT GGT GAG TGG CAC TAC AAA TTC	ACG TTG GAT GAT CAG TGA TTG CAT CTG ACC	GTA GGG GCA CTA CAA ATT CAC AAA AGA A
rs12713567	ACG TTG GAT GAG AGT TCA TGG TCA GTT TCC	ACG TTG GAT GGC AGG GAT TGT GAG GAA ACA	ATT CTG CCT CCC CTC
rs12898654	ACG TTG GAT GTA GAG TCC TTT GCG CTG GG	ACG TTG GAT GTC CTT TGT GTC CGC TCT GGT	GGG TTC CCA GGA TGT TCC GA
rs12901985	ACG TTG GAT GTG CAC ACA GAT GTG AAA GGC	ACG TTG GAT GCT GGC TAG TGT GGC TAT TAG	GGG TGA AAG GCA GCT TAT ATC
rs12905175	ACG TTG GAT GCT CAC TCC TTT CTT GAG AAC	ACG TTG GAT GAA TCA CCA ATG TTG GGG ATG	GAT TTA GAA CTC CAC CTC CTG TT
rs12905371	ACG TTG GAT GCC CGG CCT CAA CTA TTT CTT	ACG TTG GAT GGT GAC TTT TTT TTC TGA ACC	GAG ATT CCA TTT ATT GGT TCG
rs12917587	ACG TTG GAT GCT CCA ATG TCA CCT TCT CTC	ACG TTG GAT GGC TGT GAA AGT GTG GCA TAG	ATG AAG TGG CCA TCA TC
rs13005707	ACG TTG GAT GGA AAC AAG TAG CAA AAG AG	ACG TTG GAT GTG GCC ACA CGT CAC ACA GT	AAC ATC TCT CCT ACC TTG
rs13193103	ACG TTG GAT GGG CTT GGT ACA TAT AAA TCT C	ACG TTG GAT GCC ACC AGT ATA TGG CTA CC	CAG ATA AAT CTC AAT TAG ACA GTG TTT
rs13194038	ACG TTG GAT GCC CTT ATT CTC ATT GTG CAG	ACG TTG GAT GGC AAA ATC ACA TCT GCT AAT C	TTG TGC AGT TTC CCT AAA
rs13196708	ACG TTG GAT GCC ACT TTC TTA CAC GTA AAG C	ACG TTG GAT GGT ATA CAA TGC CTT TCT AC	CAC TTT AGC CCT TTG ATT TTC A
rs13205736	ACG TTG GAT GCT AAA ATT CAG GAA AAC AG	ACG TTG GAT GTA GAG ATA GCC TTT TCA AG	GAT TCA GGA AAA CAG ATG TTC A
rs13206817	ACG TTG GAT GAA AAA CTT CAC CTC CAA TG	ACG TTG GAT GGT TGG TCC TTT GTT TTG GTG	CAA CCT TCA CCT CCA ATG CAT TCA
rs13213112	ACG TTG GAT GTC AAC TTG GAA TGG CTC CCT	ACG TTG GAT GCT TGT ATG TGG GAG GGT AG	CCC CCG AAT GGC TCC CTC ACA GAG
rs13219887	ACG TTG GAT GAG GAC CAA GGC TTG AAA GTG	ACG TTG GAT GCA GTA GGT GTA GCG AAT AGC	TTA TAC ATC ACT ACC ACT CA
rs14429	ACG TTG GAT GTG GGC TTT CAG TGC CCT GC	ACG TTG GAT GTG GTG TCC ATG TGG GTG TG	TTG CGG GCT CCG CCT CTC CTC C
rs16890428	ACG TTG GAT GTC TGC CAT ACT GGC TTA CAC	ACG TTG GAT GGG GTT CTG GGA CAT TAT TTG	CCT CCT GGC TTA CAC ATA TTT CCA
rs16890541	ACG TTG GAT GGT TAG CCC TCA TGA GAA TAG	ACG TTG GAT GGT TTT ATA TGC TGC TGC CCC	GGA GTG AGA ATA GCT GCA TTT CTG TGA C
rs16890826	ACG TTG GAT GGA AAA TTG ATA GTC TGT CAT C	ACG TTG GAT GTC TAA TGT TGG TGA GGG GAG	CAA TTG ATA GTC TGT CAT CAA AAT C
rs16951060	ACG TTG GAT GGC TGT ATC TGC AAA GGG CAC	ACG TTG GAT GTT TCT CAG GCT TAG AAG CTC	GCC CTG AAT GGG TGA CTG
rs1699018	ACG TTG GAT GCT TAT TTC CTG TGG CTG CCT	ACG TTG GAT GGC TTT GCA GGT TAT ACA ATC	GCT GCC TTT AGG CTC
rs17032119	ACG TTG GAT GGC AGA AGA CAA ATA GTT AAA	ACG TTG GAT GGA CTT TGG ATA TGT AAG TGC	ACA CAA ATC ACT GGG A
rs17241403	ACG TTG GAT GAA CCC ACT AGG CTG CAA TAA	ACG TTG GAT GAG ACA GTC TCA TAT TCT GA	GGA TAA ATT GTT AAA CAT AGT CTT TCT

Supplementary Table 9 continued.

dbSNP ID	Sequence forward PCR primer	Sequence reverse PCR primer	Sequence extension primer
rs17244601	ACG TTG GAT GTC CCT GCG GCT TTT CCC TAC	ACG TTG GAT GGA CAC AAT ACA TGC TGA AGG	GCT TTT CCC TAC TCT TCG
rs17300363	ACG TTG GAT GGC TAG CTA TAG AGA TTA TGG	ACG TTG GAT GTA CCT TAT TCA GGT CTT GGG	ATG GTT AAA AAG AGA CTG CTT ATA T
rs17542411	ACG TTG GAT GTG AAG CTC TCC TGC TAG TCG	ACG TTG GAT GGA ATA AAA AGT CTG TCT CAA G	ATG CTG TCC AGG GAT A
rs17543178	ACG TTG GAT GCT AGG TTA TAA TTT TGA TAG	ACG TTG GAT GAA ACA CCA AGA CAT CCT CAG	TTG AAC CTA AGA ATA TGT CTC TG
rs17614684	ACG TTG GAT GGC CTT AAT TTG GCC TAC TGG	ACG TTG GAT GGC CAG CAG GGT ATA TGA AAC	GAC TTG GCA GGC TCC
rs17620389	ACG TTG GAT GTC GTT GAC AGC TGG TTA GTG	ACG TTG GAT GCA ATA GAC TAC AAT GAC ACT G	GGT TAG TGC TGA TTC TCA
rs17757272	ACG TTG GAT GGG AAA AGG TAG GAG GGA AAC	ACG TTG GAT GCC AAA AAA TAT ATT TAA TGT	AGG GAA ACA TTT CAG TCT ACA
rs1931762	ACG TTG GAT GCT TCT GTT AGC TGC TTT AC	ACG TTG GAT GAC ATA TAG ATG GTT TAA CTA	CTT GTG ATA GTT ATT TTT TGT TCT
rs1983167	ACG TTG GAT GGC ACC TTT GAG ATT TAC TGC	ACG TTG GAT GTT GTG CTT AGT CTG CCA CTG	AAA CTG AGA TTT ACT GCA GTG TAT
rs2061845	ACG TTG GAT GCA GCC TCC TTC ACT TCC TTG	ACG TTG GAT GCT GGA TAT TAA TCC AAC GGC	GTT GGG CCA GGG CCG C
rs2110974	ACG TTG GAT GGA AAA TAA ACA TGC ACA GAT G	ACG TTG GAT GGA AAA TAT TTA TAG AAA TCA C	TAA TGC ACA GAT GAT AAT TGA TA
rs2139246	ACG TTG GAT GGG AAC TGG CTA ATT CAA AGG	ACG TTG GAT GGT ACT TCA TGT TTT CCC AAG	CGG TGT TAT GAG GAA CAA G
rs2192954	ACG TTG GAT GCA AAT CAA ACA TGG TAT TGT C	ACG TTG GAT GGT CCC TCA AGT AAC TAA AAG	ATG GTA TTG TCA TTA TTT GTC AAA C
rs2246023	ACG TTG GAT GAA TGC CAC TGT TTA GAC CTC	ACG TTG GAT GAA GCC TGC TGA GAA AAG ATG	GTC ACT GTT TAG ACC TCT AGT TTG
rs2280334	ACG TTG GAT GAA GTA GTT GAG ACT CAA CGC	ACG TTG GAT GGT ACC TCC AAT CCA GAG AAC	CCC CGA CTC AAC GCT TCC CTC
rs228181	ACG TTG GAT GGA TAA CAC AGC AAG TGA GAC	ACG TTG GAT GGC TTT TGT ATC TAG CAT GAG G	AGT TAA CAA TTG CAA ACA GT
rs2300477	ACG TTG GAT GGG GAG ACC ATC AAT TTT GCG	ACG TTG GAT GAA AAC TGT CTC AAC ATC AGC	GCG TGC ACA TCT AGA TTC AT
rs2300478	ACG TTG GAT GGC ATT TCT CTG ACC AGA TAC	ACG TTG GAT GGC ACA ACT TGT TGC AAA TCC	CCC CAT GAC CAG ATA CTT ACA GAC
rs2300480	ACG TTG GAT GGA AAT GCA GGG ACA GGT CG	ACG TTG GAT GAG GAC ACT GTC GCC CAA ATC	GGA CAG GTC GCT GCA ATC AAA ACA
rs2300483	ACG TTG GAT GAA ATA AAG AGT TCC AAC AG	ACG TTG GAT GAT CTC CCA TGT AAC TCT CTG	AAG AGT TCC AAC AGA AAA ATT TC
rs2395694	ACG TTG GAT GTT CAG CCG TGA GGC AGT GAG	ACG TTG GAT GGA TTT CTC TCC TTA TAG GGC	TTG AAG GAG GCG AGG ACC G
rs2589985	ACG TTG GAT GCC TTT CCT CAA CCA AAA AAC	ACG TTG GAT GTG AGC CAC TGT ACC CAC GAA	AAC CAA AAA ACA GAG CTA ATA A
rs2748153	ACG TTG GAT GAC AAT ATT GCT ACA GTC CCC	ACG TTG GAT GCA CTC ATT AAA TCA AAA GGC	TAA AGT CCC CTC CAG TAC
rs2748160	ACG TTG GAT GGG TAT TGA CAA TAT GAC ATG G	ACG TTG GAT GTG AAG GTC TTC CAC CCA ATC	GAG GAC ATG GAA TTT TAT GCT ATG
rs2748169	ACG TTG GAT GAT GCA CGC GTT TCC CAG TTC	ACG TTG GAT GCA CAC ACA CAT CCA GAA TAC	GGT GCC CAG TTC TGA ACA GAT GAT
rs2748174	ACG TTG GAT GTT GTG TGC ACA CAG GTG TTT	ACG TTG GAT GAG AGT ATG GAA ATG CCA CTG	GTG GGC ACA CAG GTG TTT TCA TAA
rs28580436	ACG TTG GAT GTG CTT CTG AAG TTC CAG TGT	ACG TTG GAT GGA AGA TTA AAA AAT ATT AAT AC	GTA ATA TCT AGT AGT ATG ACC TAA ATA
rs28730807	ACG TTG GAT GGA ACT GAA AAA AAT ACT AGC C	ACG TTG GAT GCT GGA CGG TAA TAT GTT ATT	AAA AAT ACT AGC CAA TGG CC
rs28730811	ACG TTG GAT GAG TCC GAG GAC TGC CTG TCA	ACG TTG GAT GTT GAA CTG CAC GAT GAA CGG	GGG ATG TCA CAG CCT CCT CCT CTT CCC
rs2901863	ACG TTG GAT GTA CCT GGT GCT CTA ATT CCC	ACG TTG GAT GAT AAT CCT GAT GGA CAG CCC	CAT GAC CGT CCA TTA CGA
rs34118838	ACG TTG GAT GCT CTC AAG CAC AGC GTA CAC	ACG TTG GAT GTG GTG GAG AAA GAT ATC GAG	AAT GTT TCA CAA ATT TAT CCT CAC CT
rs34841627	ACG TTG GAT GTG GAG ATT CCA TTG GAT TAG	ACG TTG GAT GCT GAT TAT AAA AGC CAG AC	CCC TTT CAA TTA CCT TGT AGT GAT AA
rs35067867	ACG TTG GAT GCA ACT ACT TAA CCT GCT GCC	ACG TTG GAT GCT CCT CTG TCA ATG ACG CTT	CCT ACC CAT AAA TGG CA
rs35101671	ACG TTG GAT GTC ACC TTG GCC GGC GCG G	ACG TTG GAT GGC GGC TAC GAG CTG CGA GA	GGG GTG CCG GCG CGG GGC CTC CTA GGG
rs3784692	ACG TTG GAT GGT CAA TCC AAG CTA ACA TTT C	ACG TTG GAT GTT GTT TAT GGA TGA ATG TG	CTC ACA TTT TCA ATT AAC AGA T
rs3784709	ACG TTG GAT GTT CCT CAT TGG CCA TGA CTC	ACG TTG GAT GGG AGA GGA GGC TCT TTT AGG	CCA TTG GCC ATG ACT CAG CTC A
rs3784711	ACG TTG GAT GCT CTC AAC ACT AGC AGC TC	ACG TTG GAT GAT CTA GGT TAG GAT CCA GGC	GGA GCA GCT CAT CAG AGT
rs3890755	ACG TTG GAT GCC CTA TCA AAA ATT AGC TC	ACG TTG GAT GTG GCA GTT ACA TGT AAG GG	TAT CAA AAA TTA GCT CTT TTA CAT A
rs3923809	ACG TTG GAT GGT CCT ACT GAA TTG CAG ATG	ACG TTG GAT GTG ACA GAA TGC CAT GTC TTC	ACT GAA TTG CAG ATG GAT AAA
rs41306690	ACG TTG GAT GCA AGA TCC CAA ATA GTG GCG	ACG TTG GAT GGC ACA TCC CTG AAG AGT AAC	GAC TGG ACA GTG CAC TC
rs4131034	ACG TTG GAT GGG AAA TCC AAG GGC ATG GTG	ACG TTG GAT GTT GCC CTT CTG CTT TTC ACC	GAG ATT CTG GCA AGG GCT TTT TTA CC
rs4140443	ACG TTG GAT GGT TCT ATC TTT ATT TCA CCC	ACG TTG GAT GAA TTG TAA ACA TAT ATG AAA	TCA CCC ATA TTC TTG AAA AT
rs4236060	ACG TTG GAT GCA TGG AAT ATG AAT AAC AC	ACG TTG GAT GGA AGA GAA TAC ACC ATG GAA	GGT GCC TTC ATT TTG CTA CCC A
rs4300815	ACG TTG GAT GCT CCT GAT GTG TGA GCA CTT	ACG TTG GAT GTA GCC AAG TTG CCC ACA CTC	CCC GTC ACT TCA GTA TTG CTC A
rs4302899	ACG TTG GAT GTT CAG GGC ACT TCT TTG AGC	ACG TTG GAT GTG GCC ATG TTC TCC AAA CTC	ACT TCT TTG AGC TCA CTG CAT CA
rs4430927	ACG TTG GAT GAG GCC ATG GTC TAG TGA AAG	ACG TTG GAT GTG TCC GTA GAC GAA TTG TAG	GAC AGG GTC TAG TGA AAG AGG CCA AC
rs4489954	ACG TTG GAT GTG TCT CTA ATG CCT CTT TCG	ACG TTG GAT GGC TTC ACT GTG CCT TGA AAC	GTG TTT TAT TGG ACT GTC ATC
rs4537967	ACG TTG GAT GTC AGC AGC ACT GAC ACT GAG	ACG TTG GAT GAT GAG CAG AGG GTA AAG TGG	CCG GGG GAA AGA CAA GTC TTG AAG C
rs4544423	ACG TTG GAT GCA CCA GCT CAT AAA GAA ACC	ACG TTG GAT GTC TGC CTC CCA GCT TAA ATT	GAA ACC ACA GCA GAG C
rs4605359	ACG TTG GAT GAG GCT CGG GCA TTA TAA GAC	ACG TTG GAT GCC AGG TAT TGC CAA TTA AGG	CCC AAA TCT GCT TTG ATA CGA TTA TC
rs4623233	ACG TTG GAT GCA TTG TTT CAT GGA ACG ACC	ACG TTG GAT GAG TGC CTG GCA TAG AGT TAC	ACC CTA GTC ACA CTG ATA
rs4626419	ACG TTG GAT GCC CAA GAT TTA AAA TTG TCG	ACG TTG GAT GCC TGA CTT TCG GGA CAT TTG	AAG ATT TAA AAT TGT CGT ATT GCT ATA
rs4626664	ACG TTG GAT GGG TTG TGA ATC AAG GCA CTG	ACG TTG GAT GGA CTT TTC CAA TGA TCT TAC	AAT GGA AAT AAT AAA TCA ATT TTG AA
rs4671730	ACG TTG GAT GTC CCC ACA CAC TTG CTA ATC	ACG TTG GAT GGT CCA GGT AGA TTT CTT TGC	CAT TTC ACA ATC TTC TTC CA
rs4671737	ACG TTG GAT GCT TTT CCT TTC TTC GCT TTC	ACG TTG GAT GAA GTT TCA GAA ATC TAC CCG	CGC TTT CTT TTT TCT CTT TTT CTT TT
rs4711549	ACG TTG GAT GTA CCA TGG CAT AAA CGA CCC	ACG TTG GAT GAA TGA GAT GGT GCT GAA GAG	AGA CCC AAC TGG TCT C
rs4711550	ACG TTG GAT GGC TGG CTT CTA GAA TAA CCC	ACG TTG GAT GTT TCA TGC AGA GAG AGC GTC	CCC ATC CCA GAG CTA A
rs4714146	ACG TTG GAT GTG AAT TGC TGA TGC CAT CTC	ACG TTG GAT GCC AAA TGC ACG GAT TCT CTC	GAT TCC ATC TCT CCC CCA GTA GCC ACA
rs4714156	ACG TTG GAT GTT TGT GGA AAG GCA GCA AGC	ACG TTG GAT GGG TTC TGC ATA TTC TGA GGG	CAG CAA GCA AAA GGG AGA CTT GTC
rs4714165	ACG TTG GAT GAT CAG CAT TGA TTC TTT CC	ACG TTG GAT GTC TGT CAC CCT AGT GTT ATC	GCA TTG ATT CTT TCC TAT TCC C
rs4776970	ACG TTG GAT GAA GTG CTT ATG TGC TTC ACC	ACG TTG GAT GGG AAA GGG GAT AAT AGT GAC	CAC ATT ATT TGC CTT TTT ACA AAA ACA G
rs4776982	ACG TTG GAT GAA CTG GCT TAG CAG CGT TGA	ACG TTG GAT GAG CCA AGT CAT CGT TGG GAG	GGC AGC GTT GAG AGC GT
rs6494696	ACG TTG GAT GTG AAG GTC TGA GAG GCC TG	ACG TTG GAT GTC GCC CAC TCA CTT TCT AAC	AGG CTG CCT CCA GTG AGG GTT T
rs6500961	ACG TTG GAT GGG GCC TTG TAG ACA AGA TG	ACG TTG GAT GCA GAA GAG GAT CTT GCA CT	AAC AGA AGA TCG GGG CT
rs6520824	ACG TTG GAT GTC TCT CCA TTT TCC CAT GCG	ACG TTG GAT GGT TAG TTT GTC TCA TAG AGG G	GTC CCA TGC GAG AAG CCT GTA
rs6546232	ACG TTG GAT GGA CGA CTG CAT GCT TTA ACC	ACG TTG GAT GAG GAA AAT TTG TGT GTA GG	CAT GAT GTG CCA GAT TAC C
rs6610746	ACG TTG GAT GTG TAG GTT CAG CTA ACT TGG	ACG TTG GAT GCA GAT GGC TTG TTT ACT GC	CAC ACA AGG TCC ACT CTC
rs6705285	ACG TTG GAT GCT CAC ACC ACC ACC CTT ATG	ACG TTG GAT GTG TCA GGC CTT AGG TTA TTC	TGG TGG AGC AGA GAT T
rs6705647	ACG TTG GAT GGA AAC AAA CAC TGG ACT ACC	ACG TTG GAT GTT GTG CTA AGC TCT GGG ATG	GGC ATT GAG GAC AGG GAT TGG TCT TAC T
rs6710341	ACG TTG GAT GGG TAA ATG ATA TAC AAT TGG C	ACG TTG GAT GGT ATA TGG TGA TGA GTG TTC	TTG GCT TCA TGA AAT AAA ATG GT
rs6711787	ACG TTG GAT GAT GTT GCC TAC GGA TGG AAG	ACG TTG GAT GAC TGA AAA CTT ACA TGC ACG	TGG AAG TTA ATG CAC CAT

Supplementary Table 9 continued.

dbSNP ID	Sequence forward PCR primer	Sequence reverse PCR primer	Sequence extension primer
rs6721499	ACG TTG GAT GCA TCC TGA GGA TCC CAT TTG	ACG TTG GAT GGG CAC ATA AGT GTG TCT AAC	CAT TTG GTT AAG CCT TAC TA
rs6727352	ACG TTG GAT GTA CAT GAT GGC AGA GAC GTG	ACG TTG GAT GTC TGA TTA TGG TTA GGC GGG	AAG GAA GAG ACG TGT GGC TTT C
rs6917654	ACG TTG GAT GGT CTA CTC TTT GCC AGT TAC	ACG TTG GAT GGT CTC TAA AAA TCC TGA AAA C	ACT CTT TGC CAG TTA CTA TTT T
rs6923737	ACG TTG GAT GTC ACT AGC TAC TAA GCT CTC	ACG TTG GAT GGG TGC GTC AAG TAG TAC TTT	TAC TAA GCT CTC TCT CTT CT
rs6932235	ACG TTG GAT GAG GCC AAG AGC ACA ATC TAC	ACG TTG GAT GCT TTA GAG GAA TAC TGT GTC	GGT GTA ACC TGA TAA AAG GGA
rs7162980	ACG TTG GAT GGC TCC AGA TTC AAT TAT GAG	ACG TTG GAT GAG GTG CAG TAC TAG GAA GAC	GTT TAA CAA CTA GAC TCT AAG T
rs7180716	ACG TTG GAT GAA CCC TAT GTG GGA AAG GTG	ACG TTG GAT GGC ACA GTC TTA TTA CTA CTG C	GGT GAA GCC AAG ACA C
rs7181869	ACG TTG GAT GGG CGG CGG CGC CAT GTT CT	ACG TTG GAT GAC TGC CGC TGC GTC CTT G	GAT GTT CTG GGG GCA TCA
rs7194617	ACG TTG GAT GTC ATG CCC AAT TTT CAC AGG	ACG TTG GAT GAT GGG ACA TGG GCA AGT ATC	CCC CAA TTT TCA CAG GGC AAA GCA TC
rs726160	ACG TTG GAT GTA AAT CCA GCT TCT TGC CAC	ACG TTG GAT GTC GGT ACA GTC CTC TTT AGC	GAA CTC CAG CTT CTT GCC ACT ATA CCA
rs737172	ACG TTG GAT GGT CTA GGG CCG AGG CTT TG	ACG TTG GAT GGG GAA CGA GAT AGC AGC TTG	ATT TTC AGA TGC GAG GC
rs745213	ACG TTG GAT GCA TGT TTC TTT TCT TGT AGC C	ACG TTG GAT GGG TCA AGG CAG CCA AAA AAC	TTT CCT TGT AGC CTC TTC GG
rs7497457	ACG TTG GAT GGG GAT CAG CGT TTG AGT AAG	ACG TTG GAT GTT TTG CCT GCG CGT CTT TCC	TTG AAC CCT CCG CGC C
rs7563565	ACG TTG GAT GCA CAG ATA CAT TTT AAT GC	ACG TTG GAT GCA GGT TGA TTC TTA ATA TAG	TTC ATG CAA TAA AAT CCT AAA GTG TG
rs7579466	ACG TTG GAT GTC CAT TTG TAC CTT GTA AC	ACG TTG GAT GAA AAT GCC ATG TTT AAT AAG	TTT GAA CAG TTA TAT GGG TTA A
rs7586211	ACG TTG GAT GCA AAT GGC TGG CTA GTG TTC	ACG TTG GAT GAA ACA AAA GTG CCC ATG GAC	TGG ATG GCT AGT GTT CAG TTA GGT G
rs7603236	ACG TTG GAT GTT GCC AAG TTT GAA CCT CAC	ACG TTG GAT GAG AAT GAG GGC AAC ATT ACC	CTC ACC TCA CAA TTA CTG GT
rs7740763	ACG TTG GAT GAC TAC TGA AGA AAG AAA ATT	ACG TTG GAT GAC TGA ACA CTA CAG CTG AC	GGA GAA TAA ACT TTG TGT AGC CTA A
rs7745176	ACG TTG GAT GTT ACT GAT ACT CAG TAC AT	ACG TTG GAT GCC ATC CCT TGA AAT GGA ATG	TTA CTG ATA CTC AGT ACA TAA TTT A
rs7763775	ACG TTG GAT GTG ACC TTT TCC TTT AAG GAG	ACG TTG GAT GCA GGC ATT ACG TGC TTC TTG	CCC TTT TTC CTT TAA GGA GAT ACT AAG A
rs7769186	ACG TTG GAT GAA ACC AGG ACC TTC CAC TAA	ACG TTG GAT GGA AGA ATC CGA AGA AGC AGG	CAG GAC CTT CCA CTA ATA TTC C
rs7872553	ACG TTG GAT GAA CAC CTC AAA TTG CTT CAG	ACG TTG GAT GAC TCC AGC AGA CTC TAA ACC	ATC AAA TTG CTT CAG TTT GAG TGA TT
rs7881785	ACG TTG GAT GCA TTC ATA ATA GCC AAA CAC	ACG TTG GAT GCG TGA ATA TAC TCA ATT CAT	CAT CAC TCA GAT ATT CAT TAG TC
rs8025526	ACG TTG GAT GCA CTT AAG TTA GAG CAT TC	ACG TTG GAT GGG GAA TAA GAA TAA TCC TT	TTA TCT CAT TTG ACC TTC AC
rs8025790	ACG TTG GAT GAT GTT TTA TGT TTA GAT CC	ACG TTG GAT GGA ATC CAG AAA GGA TCC TGT	GTT ATG TTT AGA TCC ATT TCT CT
rs884202	ACG TTG GAT GAT TAA ATG AGC TCA CCC TCG	ACG TTG GAT GGC ACA GGT AGG AAT TGC TGA	CCT CGT GGG AAG TCT CCT G
rs909997	ACG TTG GAT GCC GTT TAT TGT TAT TTT ATC	ACG TTG GAT GGA AAC TTT AGA AGT AAA TCC	CGT TTA TTG TTA TTT TAT CAG TAA AAG T
rs910516	ACG TTG GAT GTA CCA CAA CAT GTC TGA CTC	ACG TTG GAT GGT AGT GAT AAC TGC AGT GTC	GCT GTC TTG CTT AAT TCT GA
rs915161	ACG TTG GAT GAT TAG CCT ACC AGA TCC ACG	ACG TTG GAT GAG AGT TGG TGT TCT CGG AGC	CCC CAG ATC CAC GCT CAA AC
rs922493	ACG TTG GAT GTC CCA CGG GAA TGT TGT GTC	ACG TTG GAT GTT CAT TTG CCA AAC ATG CTG	GGG CCT GGC ACT TAG
rs926564	ACG TTG GAT GCC CTT AAT GTT ATA TTG GGC	ACG TTG GAT GGA TAA AGT TCA TGG CAA TGT C	TTA TAT TGG GCA AAA TTA TAT ATA AAT G
rs9296249	ACG TTG GAT GAG TGG GCA GAT CAT GAA AGG	ACG TTG GAT GTC TCA GGG CTC CTC TTC ACC	GCT GTG GAT CTT GGA CTT TA
rs9302245	ACG TTG GAT GGC AAA TAA GTG TTG TAT TAC	ACG TTG GAT GCC ATG TTT GTG ATG CAT CTG	CCT GTT GTA TTA CAC ATA CTA ATT TAT G
rs9349073	ACG TTG GAT GCT TAA GCA ATT CAA TCC AGG	ACG TTG GAT GGC TAT ATC ACC TTA GCT GCC	GGC CCG GAG ATT TAA AAA ACT GCA ATA
rs9357271	ACG TTG GAT GGT GGA TCT TGG ACT TTA TGC	ACG TTG GAT GCG AAC GAA GTC ATG TCA CTC	GGA GCC CTG AGA AGT TT
rs9366950	ACG TTG GAT GCT GGA TGT GGG TCA TTT GTC	ACG TTG GAT GTA AAT CGA GGA GAA CTG GGC	CCC CGT CAT TCA TCC TGC TCA GTT
rs9369064	ACG TTG GAT GTA TAT GAG CAT CCA ACA CTG	ACG TTG GAT GGC CTA CAG TTG GGC AAA ATC	CCT CGC ATC CAA CAC TGT TTA TGG ATC A
rs9380739	ACG TTG GAT GCT TTT TAG TGT GTC TAA CAG G	ACG TTG GAT GCG TCA ATC TAG TTG GTT TAG	TCG TGT GTC TAA CAG GAT AAA ATG
rs9380755	ACG TTG GAT GGC TTT CCA AGT AAC CTC CTC	ACG TTG GAT GGA GGA AGT TTG TAG TCT CCG	CTC CTC TTT AAC TGA TGA TG
rs9394492	ACG TTG GAT GCT AAC ATA TCA TAC ACT GG	ACG TTG GAT GTT TCT CAA GAT CTA CAG GGC	AAC TCA ACC AAC TAG ATT GAC GAA
rs9394507	ACG TTG GAT GTC CTC CCT GTC TCT TGA CC	ACG TTG GAT GCA AGG ATC ATC TGT GTA ATG	CTA TTG ACC CCT CAG ACA
rs9462409	ACG TTG GAT GAG ATG CGC ATC TGA TGA ACG	ACG TTG GAT GTA AAA TGG GGC CAT GAT GGG	CTC TGC CAA CAG CCC T
rs9462426	ACG TTG GAT GTT AGC TCT AGC TGA GGA ATC	ACG TTG GAT GAA TCC CCC TTG CAC TGA ATC	GCA GTA TTT GCC TGC TGC A
rs9462433	ACG TTG GAT GAA ACA AGT GAG TTG TAT GG	ACG TTG GAT GGC ACT ATT TTT TGT AAC AGC	TGA GTT GTA TGG TAT ATA AAC G
rs9470822	ACG TTG GAT GGA AAA CTA TAA AAG TAA ACC	ACG TTG GAT GCA ATT GTT TAT GAC CTG GTG	ACT ATA AAA GTA AAC CAA CCG AT
rs9470888	ACG TTG GAT GTT GTG ATT CGT GAG AGG TGG	ACG TTG GAT GTG CAG CAA TCC AGT CAT ATC	TAT GCG TGA GAG GTG GTC AAA ATA

Appendix 3

Brief Communication in *Nature Genetics*

PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome

Barbara Schormair*, David Kemlink*, Darina Roeske, Gertrud Eckstein, Lan Xiong, Peter Lichtner Stephan Ripke, Claudia Trenkwalder, Alexander Zimprich, Karin Stiasny-Kolster, Wolfgang Oertel, Cornelius G Bachmann, Walter Paulus, Birgit Högl, Birgit Frauscher, Viola Gschliesser, Werner Poewe Ines Peglau, Pavel Vodicka, Jana Vavrova, Karel Sonka, Sona Nevsimalova, Jacques Montplaisir, Gustavo Turecki, Guy Rouleau, Christian Gieger, Thomas Illig, H-Erich Wichmann, Florian Holsboer, Bertram Müller-Myhsok, Thomas Meitinger & Juliane Winkelmann

Nature Genetics 40 (8): 946-948, 2008

* joint first authorship

Contributions:

I selected the German case sample for both stage 1 and 2 of the study from the patient database. I also performed part of the quality control of selected DNA samples prior to genotyping, designed the iPLEX assays for the replication phase, and carried out part of the genotyping of the replication phase and the genomic control experiment. I prepared the genotype and phenotype data necessary for the statistical analysis of the replication phase and the genomic control experiment. In addition, I performed all experiments of the mutation screening of *PTPRD* (sequencing of all coding and non-coding exons and screening for deletions of complete exons). I participated in the statistical analysis and writing of manuscript and supplementary information and designed the tables of the supplementary information.

nature genetics

© 2008 Nature Publishing Group http://www.nature.com/naturegenetics

PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome

Barbara Schormair[1,2,19], David Kemlink[1,3,19], Darina Roeske[4], Gertrud Eckstein[1,3], Lan Xiong[5], Peter Lichtner[1,2], Stephan Ripke[4], Claudia Trenkwalder[6], Alexander Zimprich[7], Karin Stiasny-Kolster[8], Wolfgang Oertel[8], Cornelius G Bachmann[9], Walter Paulus[9], Birgit Högl[10], Birgit Frauscher[10], Viola Gschliesser[10], Werner Poewe[10], Ines Peglau[11], Pavel Vodicka[12], Jana Vávrová[3], Karel Sonka[3], Sona Nevsimalova[3], Jacques Montplaisir[13,14], Gustavo Turecki[15], Guy Rouleau[5], Christian Gieger[16], Thomas Illig[16], H-Erich Wichmann[16,17], Florian Holsboer[4], Bertram Müller-Myhsok[4], Thomas Meitinger[1,2] & Juliane Winkelmann[1,2,4,18]

We identified association of restless legs syndrome (RLS) with *PTPRD* at 9p23–24 in 2,458 affected individuals and 4,749 controls from Germany, Austria, Czechia and Canada. Two independent SNPs in the 5′ UTR of splice variants expressed predominantly in the central nervous system showed highly significant *P* values (rs4626664, $P_{\text{nominal}/\lambda\,\text{corrected}} = 5.91 \times 10^{-10}$, odds ratio (OR) = 1.44; rs1975197, $P_{\text{nominal}/\lambda\,\text{corrected}} = 5.81 \times 10^{-9}$, OR = 1.31). This work identifies *PTPRD* as the fourth genome-wide significant locus for RLS.

Restless legs syndrome (RLS) is a frequent neurological phenotype characterized by a diurnal occurrence of an urge to move, usually accompanied by uncomfortable sensations in the lower limbs. Symptoms manifest at rest and improve with walking. RLS can lead to severe sleep disturbances and impaired quality of life[1]. Dopaminergics provide effective treatment, but their use is limited because of side effects[1]. A genome-wide association study (GWAS) with German and Canadian RLS cases revealed association with variants in *MEIS1*, *BTBD9* and a locus comprising *MAP2K5* and *LBXCOR1*, with ORs above 2 (ref. 2). Another GWAS conducted with Icelandic and US RLS cases showed association of *BTBD9* variants with periodic limb movements in sleep (PLMS), an associated motor feature of RLS[3]. The association with *MEIS1* and *BTBD9* was also confirmed in an independent case-control study in the US population[4]. None of these genes is located in any of the previously described linkage regions for RLS (RLS1–RLS5)[5]. Analysis of these loci in our GWAS data[2] showed nominally significant signals in RLS3 on 9p23–24. Despite criticism of the statistical analysis concerning the original linkage finding[6] and variation in the precise definition of the disease-containing interval, this is the most robust RLS linkage region, having been identified in two US families and replicated in two German families[7–10]. We therefore carried out an association study with 3,270 SNPs from this 31-Mb region (9p, 0.5–31.5 Mb).

For the exploratory genome-wide scan (stage 1), we genotyped 628 RLS cases and 1,644 population-based controls from the KORA-S3/F3 survey using Affymetrix Mapping 500K array sets (401 cases and 1,644 controls)[2] and Affymetrix Genome-Wide Human SNP 5.0 arrays (227 cases). Application of stringent quality control criteria yielded 208,733 SNPs throughout the genome for analysis. Eigenvalue-based analysis and genomic control showed minimal population substructure ($\lambda = 1.07$). Of 3,270 SNPs analyzed in RLS3, 8 SNPs with a nominal *P* value corrected for $\lambda < 10^{-3}$ passed our criterion for replication (**Fig. 1**, **Supplementary Methods** and **Supplementary Table 1** online).

In the replication phase (stage 2), we genotyped these SNPs in German (1,271 cases, 1,901 controls), Czech (279 cases, 368 controls) and Canadian (285 cases, 842 controls) samples using multiplex mass spectrometry. Part of the German and Canadian samples had been used in the replication stage of our previous GWAS[2]. Details of demographic data, recruitment, diagnosis for subjects and genotyping of both stages are shown in **Supplementary Methods** and **Supplementary Table 2** online. Genomic control analysis resulted in inflation factors of 1.10 in the German, 1.23 in the Czech and 1.26 in the Canadian sample. Separate analysis of stage 2 samples showed significantly different minor allele frequencies (MAFs) across samples but comparable ORs with unidirectional allelic association (**Supplementary Table 3** online). Heterogeneity with respect to MAFs

[1]Institute of Human Genetics, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg 85764, Germany. [2]Institute of Human Genetics, Technische Universität München, Munich 81675, Germany. [3]Department of Neurology, 1st Medical Faculty, Charles University, Prague 120 00, Czech Republic. [4]Max Planck Institute of Psychiatry, Munich 80804, Germany. [5]Centre d'excellence en neuromique, Hôpital Notre-Dame, Université de Montréal, Montréal H2L 4M1, Québec, Canada. [6]Paracelsus-Elena-Hospital, Kassel 34128, Germany. [7]Department of Neurology, Medical University of Vienna, Vienna 1090, Austria. [8]Department of Neurology, Center of Nervous Diseases, Philipps University, Marburg 35039, Germany. [9]Department of Clinical Neurophysiology, University of Göttingen, Göttingen 37070, Germany. [10]Department of Neurology, Innsbruck Medical University, Innsbruck 6020, Austria. [11]Neurologische Praxis, Berlin 10969, Germany. [12]Institute of Experimental Medicine, Czech Academy of Sciences, Prague 140 00, Czech Republic. [13]Centre d'étude du sommeil, Hôpital du Sacré-Cœur de Montréal, Montréal, Québec H4J 1C5, Canada. [14]Centre de recherche en sciences neurologiques, Université de Montréal, Montréal, Québec H4J 1C5, Canada. [15]Departments of Psychiatry and Human Genetics, McGill University, Douglas Hospital, Montreal, Quebec H4H 1R3, Canada. [16]Institute of Epidemiology, Helmholtz Zentrum München-German Research Center for Environmental Health, Munich 85764, Germany. [17]Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich 81377, Germany. [18]Neurologische Klinik und Poliklinik, Technische Universität München, Ismaninger Strasse 22, 81675 München, Germany. [19]These authors contributed equally to this work. Correspondence should be addressed to J.W. (winkelmann@lrz.tu-muenchen.de) or T.M. (meitinger@helmholtz-muenchen.de).

Received 15 February; accepted 3 June; published online 27 July 2008; doi:10.1038/ng.190

© 2008 Nature Publishing Group http://www.nature.com/naturegenetics

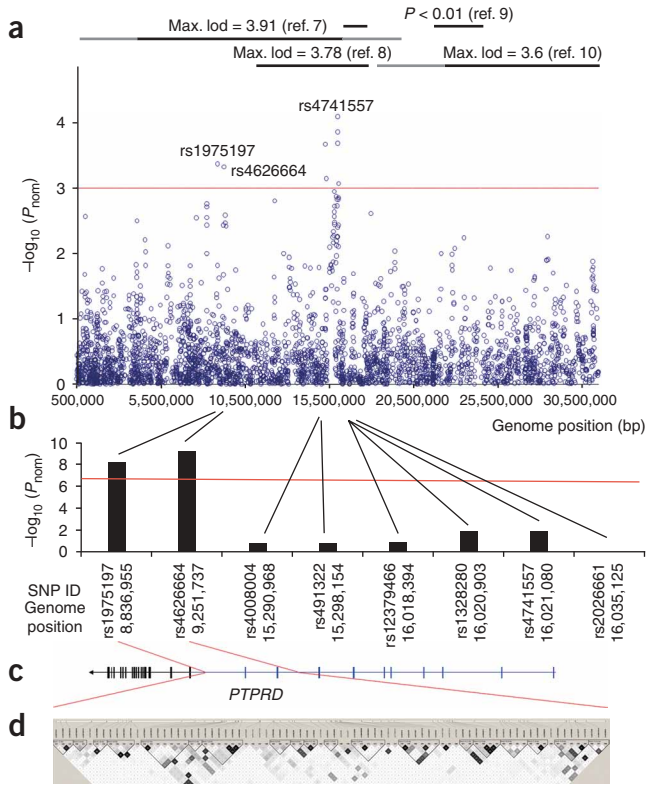


Figure 1 Association results for stages 1 and 2 over the chromosomal segment analyzed. (**a**) Results of stage 1 ($-\log_{10}$ of P nominal (P_{nom}) corrected for λ) for chromosome 9p, 0.5–31.5 Mb. The red line indicates the cut-off for selection of SNPs for replication. Position and extent of linkage signals[7–10] are shown as horizontal bars. Black bars represent the narrowest suggested region as defined by intrafamilial recombination events; gray bars extend to the maximum size. Maximum multi-point lod scores[7,8,10] and the P value from nonparametric linkage analysis[9] are denoted above the bars. Genomic positions refer to the UCSC Genome Browser Human March 2006 assembly (http://genome.ucsc.edu/). (**b**) Results of joint analysis of stages 1 and 2 for the eight SNPs within RLS3 selected for replication, given as $-\log_{10}$ of P nominal (P_{nom}) corrected for λ. Red line represents the cut-off for genome-wide significance after correction for multiple testing ($-\log_{10}(P_{nom}) = 6.62$, $P_{nom} < 2.4 \times 10^{-7}$). (**c**) Position of associated SNPs with genome-wide significance in *PTPRD*. Exons are depicted as bars, introns as lines. The noncoding 5′ UTR is highlighted in blue. Position of SNPs is indicated by red lines. (**d**) LD structure of region between rs1975197 and rs4626664. Gray shading indicates extent of LD (dark gray, high LD; light gray, low LD). Haploview 4.0 (http://www.broad.mit.edu/mpg/haploview/) and data from 1,639 KORA controls were used for visualization.

necessitated inclusion of country of origin as a covariate, allowing joint analysis of all stage 1 and 2 samples. This resulted in two SNPs with genome-wide significance after Bonferroni (B) correction for multiple testing: rs4626664, with $P_{nominal/\lambda\,corrected} = 5.91 \times 10^{-10}$, $P_{corrected(B)} = 0.00012$, OR = 1.44 and rs1975197, with $P_{nominal/\lambda\,corrected} = 5.81 \times 10^{-9}$, $P_{corrected(B)} = 0.0012$, OR = 1.31 (**Table 1** and **Supplementary Table 4** online). Both SNPs were also significant after Bonferroni correction in the German subsample and in the combined analysis of all stage 2 samples. In the Canadian subsample, both SNPs were nominally significant (rs4626664, $P_{nominal/\lambda\,corrected} = 0.018$; rs1975197, $P_{nominal/\lambda\,corrected} = 0.024$), whereas the Czech sample showed only a trend for association for the stronger signal (rs4626664, $P_{nominal/\lambda\,corrected} = 0.075$), most likely explainable by lack of power due to the smaller sample size (**Table 1** and **Supplementary Table 3**). Because cases and controls were not perfectly matched for age and sex, we used these factors as covariates in all analyses.

The association signals are located 0.41 Mb apart and map to introns 8 and 10 of *PTPRD*, within two separate linkage disequilibrium (LD) blocks. Logistic regression did not show any significant interaction between these SNPs ($P = 0.986$), as also evidenced by the lack of LD between them ($r^2 = 0$). They are separated by 17 haplotype boundaries, indicating a hot spot of recombination between them[11]. There is no significant interaction with risk alleles in *MEIS1* (rs4626664, $P = 0.463$; rs1975197, $P = 0.957$), *BTBD9* (rs4626664, $P = 0.487$; rs1975197, $P = 0.246$) and *LBXCOR1-MAP2K5* (rs4626664, $P = 0.510$; rs1975197, $P = 0.859$), and therefore no evidence for epistasis. Haplotype analysis showed no increase in significance compared to single SNP analysis. Power for the joint analysis was 77.4% and 99.4% to detect an allelic association with an OR of 1.3 and 1.4 with genome-wide significance level $\alpha = 0.05$ and a MAF of 0.17 (**Supplementary Methods**).

Sequence analysis revealed no mutations in 35 coding and 10 noncoding exons of *PTPRD* among nine affected individuals from

Table 1 Association results for rs1975197 and rs4626664

dbSNP ID	Genome position	Gene	Risk allele[a]	MAF stage 1 GER ca. (623) co. (1,639)	MAF stage 2 GER ca. (1,271) co. (1,900)	MAF stage 2 CZ ca. (279) co. (368)	MAF stage 2 CAN ca. (285) co. (842)	Stage 1 $P_{nom/\lambda\,corrected}$	Stage 2 $P_{corrected\,(B)}$ GER CZ CAN	Stage 2 combined analysis $P_{corrected\,(B)}$	Joint analysis stage 1 & 2 $P_{nom/\lambda\,corrected}$	Joint analysis stage 1 & 2 OR (95% CI)
rs1975197	Chr 9p: 8,836,955	*PTPRD*	T	0.216 0.164	0.196 0.157	0.158 0.136	0.203 0.156	4.42E–04	1.55E–03 1 1.81E–01	3.29E–05	5.81E–09	1.31 (1.20–1.44)
rs4626664	Chr 9p: 9,251,737	*PTPRD*	A	0.175 0.133	0.167 0.117	0.196 0.146	0.159 0.117	4.73E–04	6.88E–05 7.47E–01 2.44E–01	7.53E–07	5.91E–10	1.44 (1.31–1.59)

SNPs with genome-wide significant association located in *PTPRD*. Genome positions refer to the Human March 2006 (hg18) assembly. ca, cases; co, controls; numbers in parentheses denote successfully genotyped sample numbers; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval; P_{nom}, nominal P value. P_{nom} values in stage 1 were calculated using logistic regression with age, sex and the first four components from the MDS analysis of the IBS matrix as covariates. P_{nom} values in the individual analysis of stage 2 were calculated using logistic regression with age and sex as covariates. In the combined stage 2 analysis and the joint analysis of stage 1 and 2, country of origin was included as an additional covariate. P_{nom} in all analyses were corrected for population stratification by dividing the corresponding χ^2 by the inflation factor λ ($P_{nom/\lambda\,corrected}$). $P_{corrected\,(B)}$, λ–corrected P value corrected for multiple testing according to Bonferroni correcting for 208,733 SNPs in stage 1 and the joint analysis of stage 1 and 2, and 10 SNPs in the stage 2 analyses.
[a]Unidirectional association in all samples.

© 2008 Nature Publishing Group http://www.nature.com/naturegenetics

an RLS3-linked family, three index cases from families with RLS in which linkage to RLS3 was not excluded and one control compared to the reference sequence (NM_002839). We also did not find any exon deletions or duplications using quantitative real-time PCR. Among eight nonsynonymous coding SNPs genotyped in replication samples, only rs10977171 and rs35929428 were polymorphic, and these did not show any association (**Supplementary Tables 5** and **6** online). The familial relative risk figures estimated by the risk to siblings of an affected individual (λ_s) were all below 1.04 and explain only a minor portion of the original RLS3 linkage signal[7].

PTPRD belongs to the family of type IIa receptor-like protein tyrosine phosphatases. These molecules are characterized by an extracellular region containing cell adhesion motifs and an intracellular region containing two phosphatase domains[12]. Several *PTPRD* mRNA isoforms are expressed in a developmental and tissue-specific manner[13]. Both RLS-associated SNPs are located within the 5′ UTR, consisting of ten noncoding exons contained in two known long splice variants expressed predominantly in fetal and adult brain tissue[13,14]. The involvement of *PTPRD* in RLS is unknown. Studies in *Ptprd* and *Ptprs* knockout mice have shown that these proteins function in axon guidance and termination of mammalian motorneurons during embryonic development[12]. Investigations in neuroblastoma tumor tissue and cell lines have identified microdeletions and aberrant splicing patterns in the 5′ UTR of *PTPRD* that may influence mRNA stability and thereby gene expression[15].

The RLS-associated SNPs are common (MAF (CEU) > 0.13) and show weak effects (rs4626664, OR = 1.44, 95% CI = 1.3–1.6; rs1975197, OR = 1.31, 95% CI = 1.2–1.4). We failed to detect rare alleles with strong effects within this gene that could explain the linkage signal. The association of two independent signals strengthens the evidence for *PTPRD* as a gene influencing risk of RLS.

PTPRD is the fourth locus associated with RLS at a significance level that withstands correction for multiple testing in a genome-wide analysis for common variants. The two newly identified association signals on chromosome 9p add another four to the previous six risk alleles from chromosomes 2p, 6p and 15q, making a total of ten possible risk alleles (referring to homozygous carriers). Analysis of the receiver operating characteristic curve shows limited usefulness for individual risk prediction, with the area under the curve estimated at 0.624. This is in line with heritability estimates of 0.6, pointing to genetic and nongenetic effects contributing to the risk of RLS[7]. Dependent on the number of risk alleles, there is an increased risk for RLS with an empirical OR larger than 9, as we found when analyzing 309 carriers with at least 7 risk alleles (**Supplementary Methods**). This makes RLS highly amenable to association studies using common variants.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We are grateful to all individuals who participated in this study. We wish to thank J. Favor, B. Pütz and K. Oexle for discussions and R. Feldmann, J. Golic, K. Junghans and B. Schmick for technical assistance. We acknowledge L. Habersack, H. Rhese and J. Schmidt-Evers from the German RLS patient organization for supporting this study. 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 co-workers, who are responsible for the design and conduct 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 Canadian part of the study was supported by a Canadian Institutes of Health Research (CIHR) grant to G.R., J.M. and G.T. D.K. was supported by grant MSM0021620849, and J.V. and K.S. were supported by grant MSM0021620816. Recruitment of Czech controls was funded by grant IGA NR 8563-5, Ministry of Health of the Czech Republic.

AUTHOR CONTRIBUTIONS

Study design: B.M.-M., T.M., J.W. Recruitment and biobanking of RLS cases: C.T., A.Z., K.S.-K., W.O., C.G.B., W. Paulus, B.H., B.F., V.G., W. Poewe, I.P., T.M., J.W. Recruitment and biobanking of KORA controls: C.G., T.I., H.-E.W. Recruitment and biobanking of Canadian RLS cases and controls: L.X., J.M., G.T., G.R. Recruitment and biobanking of Czech RLS cases and controls: D.K., P.V., J.V., K.S., S.N. Affymetrix genotyping: B.S., G.E., P.L. Sequenom genotyping: B.S., D.K., P.L. Supervision of all markers typed: P.L., J.W. Statistical analysis: D.R., S.R., B.M.-M. Clustering of Affymetrix genotypes: D.R., S.R., B.M.-M. Manuscript writing: B.S., D.K., D.R., F.H., B.M.-M., T.M., J.W.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Allen, R.P. *et al. Sleep Med.* **4**, 101–119 (2003).
2. Winkelmann, J. *et al. Nat. Genet.* **39**, 1000–1006 (2007).
3. Stefansson, H. *et al. N. Engl. J. Med.* **357**, 639–647 (2007).
4. Vilariño-Güell, C., Farrer, M.J. & Lin, S.C. *N. Engl. J. Med.* **358**, 425–427 (2008).
5. Winkelmann, J. *et al. Mov. Disord.* **22**(Suppl. 18), S449–S258 (2007).
6. Ray, A. & Weeks, D.E. *Am. J. Hum. Genet.* **76**, 705–707 (2005).
7. Chen, S. *et al. Am. J. Hum. Genet.* **74**, 876–885 (2004).
8. Liebetanz, K.M. *et al. Neurology* **67**, 320–321 (2006).
9. Kemlink, D. *et al. Mov. Disord.* **22**, 207–212 (2007).
10. Lohmann-Hedrich, K. *et al. Neurology* **70**, 686–694 (2008).
11. Bonnen, P.E., Wang, P.J., Kimmel, M., Chakraborty, R. & Nelson, D.L. *Genome Res.* **12**, 1846–1853 (2002).
12. Uetani, N., Chagnon, M.J., Kennedy, T.E., Iwakura, Y. & Tremblay, M.L. *J. Neurosci.* **26**, 5872–5880 (2006).
13. Pulido, R., Krueger, N.X., Serra-Pagès, C., Saito, H. & Streuli, M. *J. Biol. Chem.* **270**, 6722–6728 (1995).
14. Sato, M. *et al. Genes Chromosom. Cancer* **44**, 405–414 (2005).
15. Nair, P., Depreter, K., Vandesompele, J., Speleman, F. & Stallings, R.L. *Genes Chromosom. Cancer* **47**, 197–202 (2008).

Protein-tyrosine Phosphatase Receptor Type Delta (PTPRD) is Associated with Restless Legs Syndrome

Barbara Schormair, David Kemlink, Darina Roeske, Gertrud Eckstein, Lan Xiong, Peter Lichtner, Stephan Ripke, Claudia Trenkwalder, Alexander Zimprich Karin Stiasny-Kolster, Wolfgang Oertel, Cornelius G Bachmann, Walter Paulus, Birgit Högl, Birgit Frauscher, Viola Gschliesser, Werner Poewe, Ines Peglau, Pavel Vodicka, Jana Vávrová, Karel Sonka, Sonja Nevsimalova, Jacques Montplaisir, Gustavo Turecki, Guy Rouleau, Christian Gieger, Thomas Illig, H-Erich Wichmann, Florian Holsboer, Bertram Müller-Myhsok, Thomas Meitinger, Juliane Winkelmann

Supplementary Information

Supplementary Methods

Study population and phenotype assessment

a) German samples (stage 1 and 2)

Cases were of European descent (n= 1,899, mean age 61.00 ± 12.30 years, range 6–92 years, 28.70% males) and recruited via specialized outpatient clinics for RLS in Munich, Marburg, Kassel, Göttingen (Germany, n= 1,538), Vienna, and Innsbruck (Austria, n= 361). In all cases diagnosis was made according to the diagnostic criteria of the International RLS Study Group[1] assessed in a personal interview conducted by an RLS expert. A positive family history was based on the report of at least one first degree family member affected by RLS. We excluded patients with secondary RLS due to uremia, dialysis or iron deficiency anemia. The presence of secondary RLS was determined by clinical interviews, complete physical and neurological examination, blood chemistry and nerve conduction studies in selected cases when clinically recommended.

Controls were of European descent and recruited from the KORA S3/F3 and S4 surveys. These are representative samples from the general population living in or near the city of Augsburg, Germany. KORA procedures and samples have been described.[2] Consequent to informed consent, each of the surveys sampled subjects from ten strata according to age (range 25–74 years) and sex (equal ratio) with a minimum stratum size of > 400 subjects. In the KORA S3 study 4,856 subjects were studied between 1994 and 1995, and in S4 altogether 4,261 subjects between 1999 and 2001. 3,006 individuals from S3 returned for follow up between 2003 and 2005 (S3/F3).

For stage 1, we included 1,644 subjects from S3/F3 (mean age 62.61 ± 10.01, range 35–84, 49.45% males), and for stage 2, we included 1,901 subjects from S4 (mean age 57.32 ± 10.28, range= 25–74, 39.97% males).

b) Czech samples (stage 2)

Cases (n= 279, mean age 55.79 ± 14.92, range= 12–91, 36.91% males) were of European descent and recruited in the Center for Disorders of Sleep and Wakefulness,

Department of Neurology of First Faculty of Medicine and the General Teaching Hospital, Prague. Diagnosis was made according to the diagnostic criteria of the International RLS StudyGroup[1] assessed in a personal interview conducted by an RLS expert. In 64 patients a polysomnography was conducted showing PLMS index > 5 in 82.8% of these. Controls (n= 368, mean age 45.35 ± 9.72, range= 18–61, 42.93% males) were selected randomly from the Czech blood and bone marrow donor registry.

c) Canadian samples (stage 2)

Cases and Controls were recruited through the Sleep Disorder Center at the Hôpital du Sacré-Coeur, Montréal, Canada. All subjects were exclusively of French-Canadian ancestry, which was defined as having four grandparents of French-Canadian origin. Cases (n= 285, mean age 53.62 ± 11.72, range 7–93, 37.45% males) were diagnosed according to the diagnostic criteria of the International RLS Study Group[1] assessed in a personal interview conducted by an RLS expert. In 156 patients a polysomnography was conducted showing PLMS index > 5 in 82.1% of these. Control subjects were 842 unrelated individuals from the same population (mean age 45.50 ± 12.84, range 7–89, 47.15% males). Secondary cases, defined by the same conditions as mentioned above, were excluded.

All studies were performed according to the declaration of Helsinki and were approved by the institutional review boards in Germany, Austria, Czech Republic, and Canada. Written informed consent was obtained from each participant. Demographic data of successfully genotyped samples passing population stratification analysis is available in Supplementary Table 2.

SNP selection, genotyping and quality control

Stage 1: 401 cases and 1,644 controls had been genotyped on the Affymetrix® Mapping 500K array set in a previous experiment (GWA-1).[3] We enlarged this sample by genotyping 227 additional cases using the Affymetrix® Genome-Wide Human SNP Array 5.0, (GWA-2). Both arrays have a highly overlapping SNP content (n= 500,568). Hybridization of genomic DNA was performed according to the manufacturer's standard recommendations.[4]

GWA-1 genotypes have been called using the BRLMM clustering algorithm[5] and underwent a strict quality control protocol.[3] A total of 236,758 SNPs fulfilled the quality criteria[3] which subsequently reduced available markers for the following combined 500K and 5.0 analysis.

GWA-2 cases were called separately using the BRLMM-P algorithm.[6] All genotypes for the 236,758 SNPs from GWA-1 and GWA-2 were then merged for combined statistical analysis. To minimize false positive signals we applied stringent quality control criteria before processing the genotypes to the statistical analyses. Only SNPs fulfilling the following quality were included in the analysis: Call rate > 98%, minor allele frequency (MAF) > 10%, P value for deviations from Hardy-Weinberg-Equilibrium (HWE) > 0.00001. In addition, genotypes for each SNP were compared between GWA-1 and GWA-2. For SNPs with a significant discrepancy in the allelic χ^2-test (P value < 10^{-7}) cluster quality was controlled by visual inspection (Affymetrix® SNP Signal Tool). SNPs with inconclusive clusters were not analysed further. In addition, significant discrepancies (allelic P value < 10^{-8} in the χ^2-statistics) in the genotype distribution between GWA-1-cases and GWA-2-cases as well as P values for deviation from HWE < 0.00001 in either GWA-1 or GWA-2 led to the exclusion of the respective SNP. In total, 208,733 markers fulfilled these criteria and were subjected to statistical analysis.

Stage 2: The threshold for selection of a SNP for replication in stage 2 was a λ-corrected P nominal < 10^{-3} in the RLS3 linkage region. In total, 8 SNPs fulfilled this criterion.

Samples were analysed with the MassARRAY® system (Sequenom Inc, San Diego, CA, USA) using the iPLEX Gold chemistry (www.sequenom.com). SNP assays were

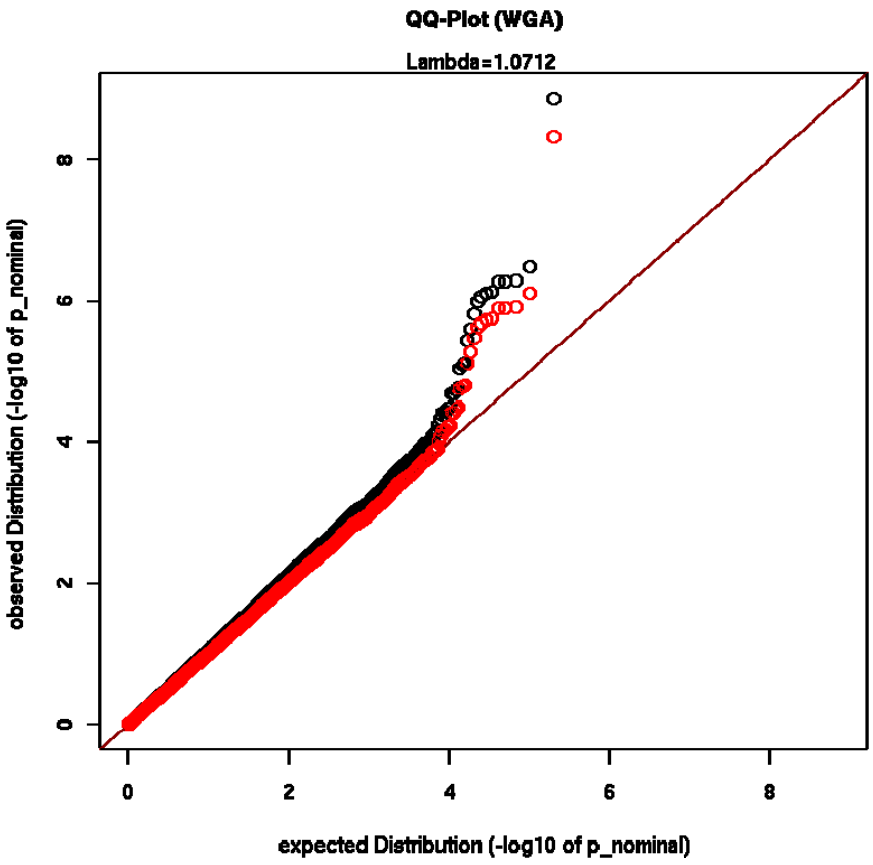
designed using AssayDesign 3.1.2.2 with iPLEX Gold default parameters. Automated genotype calling was done with SpectroTYPER 3.4 software and genotype clustering was visually checked by an experienced evaluator with no knowledge of subject phenotype. Quality control criteria leading to inclusion of a SNP in statistical analysis were call rate > 95%, MAF > 10% and P > 0.0001 for deviations from HWE in controls. By genotyping 880 (rs1026732, rs2300478, rs9296249) or 2,944 and 2,955 samples (rs1975197 and rs4626664, respectively) in duplicate, we calculated concordance rates above 99.7%.

Genomic control (GC) experiment in stage 2: Selection of GC SNPs was based on a SNP set used in a previous study on German population substructure.[7] The original set consisted of 144 intergenic and intronic SNPs and 68 coding SNPs. We replaced the coding SNPs with intergenic or intronic SNPs in their vicinity to create a marker set more neutral to selection forces. Of these 212 SNPs, 205 could be converted into iPLEX assays. Genotyping, genotype calling, and quality control were performed as described above. A total of 176 SNPs passed quality control and were used for statistical analysis.

Population stratification

Stage 1: Possible effects of population stratification were tested using EIGENSTRAT[8] (http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm). This analysis identified 8 outliers (4 cases and 4 controls). These individuals were removed and not taken into further analysis. In the remaining data set we applied the method of genomic control.[9] We used logistic regression with age, sex, and identical by state (IBS) vectors obtained from a multi dimensional scaling (MDS) analysis as implemented in PLINK 1.02 in the stage 1 data as covariates. We compared the resulting distribution of P values versus the expected χ^2 distribution with one degree of freedom. The inflation factor λ which is the ratio of the observed distribution's median and the expected distribution's median is 1.0712. This implies only minimal remaining stratification effects.

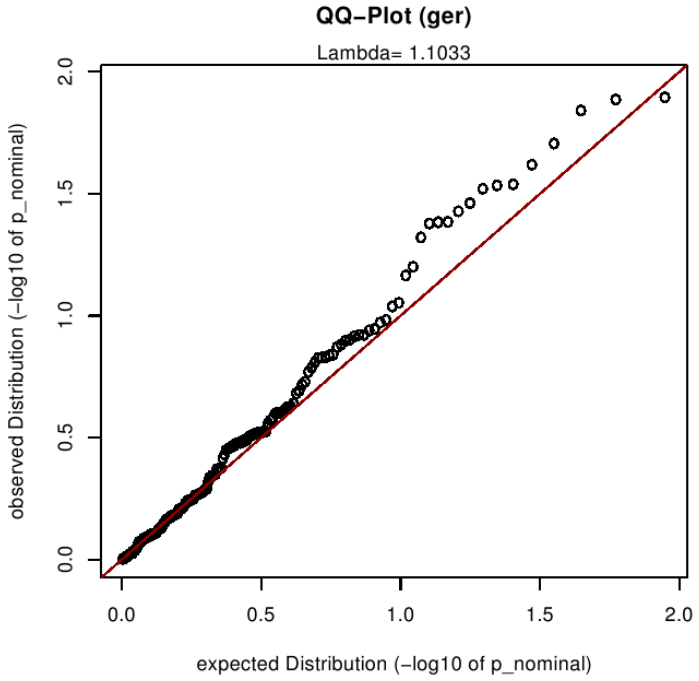
Extent of population stratification**:**

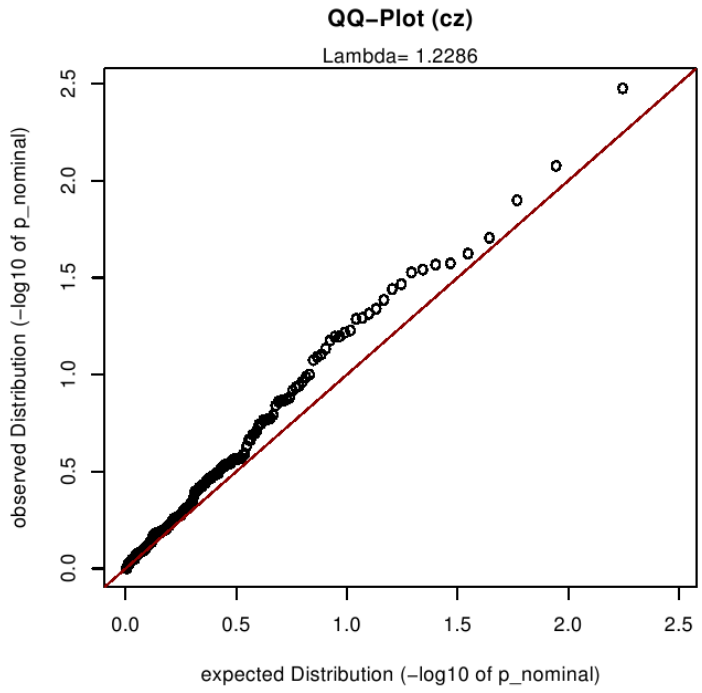


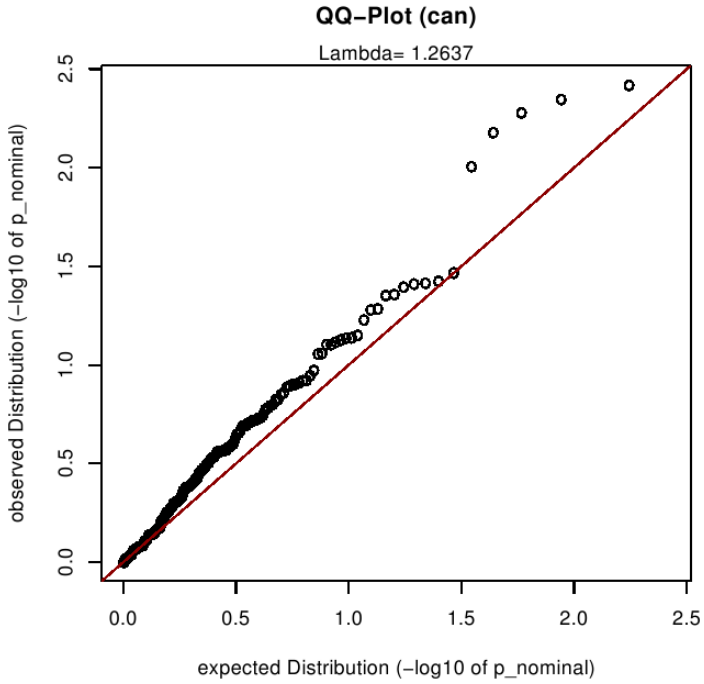
The distribution of expected (under the null hypothesis) vs. observed $-\log_{10} P_{nom}$ values from stage 1 analysis, before (black) and after (red) correction by division of the corresponding χ^2 values with $\lambda = 1.0712$. Adherence to the diagonal indicates lack of inflation of the statistic. The uncorrected Q-Q-plot indicates that more significant P values than expected were observed. After correction, the adherence to the diagonal in the lower parts of the distribution is nearly perfect.

Stage 2: Population stratification in the individual stage 2 subsamples was tested using the expanded genomic control method GCF[10] on 176 SNPs from intergenic and intronic regions, genotyped in all stage 2 samples. The obtained inflation factors were 1.10 for the German, 1.23 for the Czech, and 1.26 for the Canadian sample.

Extent of population stratification in the individual subsamples:







The distribution of expected (under the null hypothesis) vs. observed $-\log_{10} P_{nom}$ values, before correction.

Statistical Analysis

Statistical analysis was implemented in PLINK 1.02 http://pngu.mgh.harvard.edu/~purcell/plink/ and R 2.7.0 (http://www.r-project.org/). Stage 1 data were analysed using logistic regression with age, sex and the first four components from the MDS analysis of the IBS matrix as covariates. Resulting χ^2 values were divided by $\lambda = 1.07$ to correct for population stratification. Detailed results are shown in Supplementary Table 1.

In stage 2, individual subsamples were analysed using logistic regression with age and sex as covariates. Resulting χ^2 values were divided by the respective λ values to correct for population stratification. The combined analysis of all stage 2 samples was performed using logistic regression with country of origin set as a factorial covariate with three levels (German, Czech, and Canadian) as an additional covariate. Using the Cochran-Mantel-Haenszel test as implemented in PLINK 1.02 gave similar results. Because of the different λ values in the individual subsamples, correction for population stratification was done by weighted linear combinations of the inflation factors with weights obtained following ref. 11. Detailed results are shown in Supplementary Table 3.

The joint analysis of stage 1 and 2 data was done using logistic regression as described for combined analysis in stage 2. Detailed results are shown in Supplementary Table 4.

Correction for multiple testing was done according to Bonferroni in all analyses. Significance level was set at 0.05/10 for stage 2 (8 SNPs tested in replication stage + two nonsynonymous coding SNPs) and at 0.05/208,733 for stage 1 and the joint analysis of stage 1 and 2 for genome-wide significance.

Possible interaction of the significant SNPs rs1975197 and rs4626664 was tested using logistic regression as described above. The P value for interaction between the SNPs was not significant (P= 0.986). The SNPs can therefore be treated as independent. P values for interaction of SNPs and country were also not significant (rs1975197, P= 0.873; rs4626664, P= 0.944). Thus we see no evidence for heterogeneity in the effects between the SNPs and country of origin. We therefore used the combined sample in the analysis.

To test for gene-gene interaction we performed a logistic regression analysis, both on the allelic and the genotypic level among the SNPs considered (*MEIS1*: rs2300478,

BTBD9: rs9296249, *MAP2K5/LBXCOR1*: rs1026732, *PTPRD*: rs1975197, rs4626664). These tests show no significant interaction and therefore no evidence for epistasis.

Haplotypes were analysed using UNPHASED 3.0.5 (http://www.mrc-bsu.cam.ac.uk/personal/frank/software/unphased/). No significance increase of the association results could be found.

For the calculation of the sibling risk attributable to these loci we used the power calculator CATS for genome-wide association studies.[12]

Power Calculation

Power was calculated using the web-based Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/) with prevalence set at 0.08 and unselected controls modus. The significance level was set at 0.05 for stage 2 analysis and at 0.05/208,733 for genome-wide significance in the joint analysis of stage 1 and 2. Power was calculated for ORs of 1.2 to 1.4. MAFs and number of samples used for calculation are denoted in Tables a) and b).

a) significance level alpha = 0.05

		Sample	GER stage 2	CZ stage 2	CAN stage 2	ALL stage 2	ALL stage 1&2
		N cases	1271	279	285	1835	2458
		N controls	1900	368	842	3110	4749
		OR	power	power	power	power	power
MAF	0.17	1.2	0.7899	0.2429	0.3144	0.9266	0.9817
		1.3	0.9811	0.4520	0.5801	0.9987	1.0000
		1.4	0.9995	0.6653	0.8045	1.0000	1.0000
	0.20	1.2	0.8353	0.2676	0.3465	0.9515	0.9903
		1.3	0.9898	0.4958	0.6286	0.9995	1.0000
		1.4	0.9998	0.7145	0.8452	1.0000	1.0000

b) significance level alpha = 2.4×10^{-7} (= 0.05/208,733)

		Sample	GER stage 2	CZ stage 2	CAN stage 2	ALL stage 2	ALL stage 1&2
		N cases	1271	279	285	1835	2458
		N controls	1900	368	842	3110	4749
		OR	power	power	power	power	power
MAF	0.17	1.2	0.0082	4.732E-05	0.0001	0.0399	0.1321
		1.3	0.1295	0.0004	0.0013	0.4272	0.7744
		1.4	0.5308	0.0027	0.0094	0.9047	0.9943
	0.20	1.2	0.0129	6.520E-05	0.0002	0.0610	0.1922
		1.3	0.1873	0.0006	0.0020	0.5447	0.8650
		1.4	0.6487	0.0042	0.0143	0.9538	0.9985

Mutation Screening of PTPRD

All coding and non-coding exons including adjacent splice sites of *PTPRD* (reference sequence NM_002839) were sequenced in nine affected family members carrying the RLS3 haplotype of family RLS-0001 linked to RLS3 (individuals II4, II7, II14, III6, III8, III14, III18, III20, IV2, see ref. 13), three index cases of RLS families in which linkage to RLS3 could not be excluded (family 0038, 0054, 0331, see ref. 14) and one healthy control. Sequencing was performed with the BigDye terminator chemistry 3.1 (ABI) on an ABI3730 sequencer. Analysis was performed with the software Pregap and Gap from the Staden package.[15] Detailed results are available in Supplementary Table 5.

Deletions and duplications of exons were screened using real-time quantitative PCR on an ABI7900HT real-time PCR system with SYBR Green I as the detection dye. We confirmed the specificity of the PCR by running a melting-curve analysis for each amplicon after PCR. Gene dosage was determined using the $\Delta\Delta C_T$ method.[16] To control for variation of DNA concentration and PCR efficiency, we used *BNC1* as a reference gene for normalisation of samples. The mean gene dosage of four healthy controls (two males, two females) was used as the calibrator. All samples were run in duplicate. Primers were designed using ExonPrimer (http://ihg.gsf.de) or Primer3 (http://frodo.wi.mit.edu/cgibin/primer3/primer3_www.cgi). All primer sequences are available on request.

In addition, known nonsynonymous coding SNPs in *PTPRD* (n= 8, according to NCBI dbSNP, Build 128) were genotyped in stage 2 samples using MassARRAY® system and iPLEX Gold chemistry (Sequenom Inc, San Diego, CA, USA). Detailed results are available in Supplementary Table 6.

Calculation of ORs and receiver operating characteristic (ROC) curve over all RLS genes.

For *MEIS1*, *BTBD9* and *MAP2K5/LBXCOR1*, one representative SNP was selected for OR calculation, as SNPs in these genes are in high LD with each other. For *PTPRD*, both associated SNPs were included, as they are independent signals. For these SNPs we compared the number of risk alleles for cases and controls implicitly modeling a multiplicative effect. ORs were calculated using Fisher´s exact test.[17] Analysis was performed using data from 2,386 cases and 4,473 controls from stage 1 and 2, where genotypes for all selected SNPs were available. To assess false-positive and true-positive rates for a test based on varying numbers of the risk alleles, we present the ROC curve and area under the curve as calculated using the R-package ROCR.

SNP selection:

Gene	SNP
MEIS1	rs2300478
BTBD9	rs9296249
MAP2K5/LBXCOR1	rs1026732
PTPRD	rs1975197, rs4626664

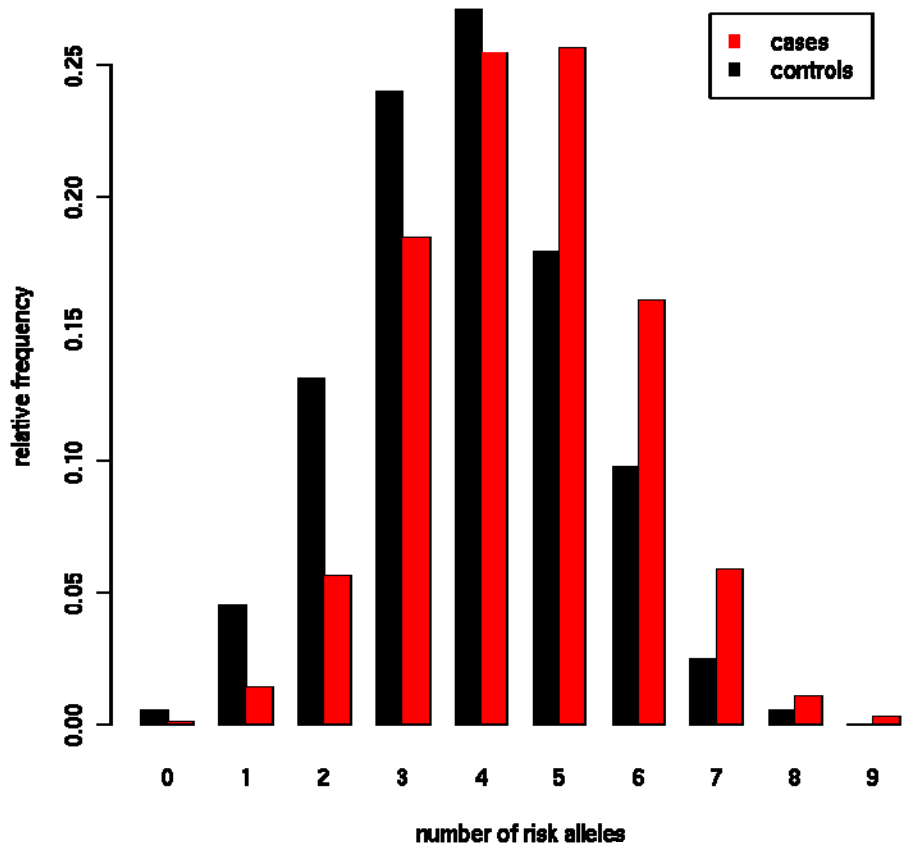
Distribution of risk alleles (n = number):

n risk alleles	0	1	2	3	4	5	6	7	8	9
n controls	24	201	587	1073	1214	802	436	112	23	1
n cases	3	34	134	439	608	612	383	140	26	7

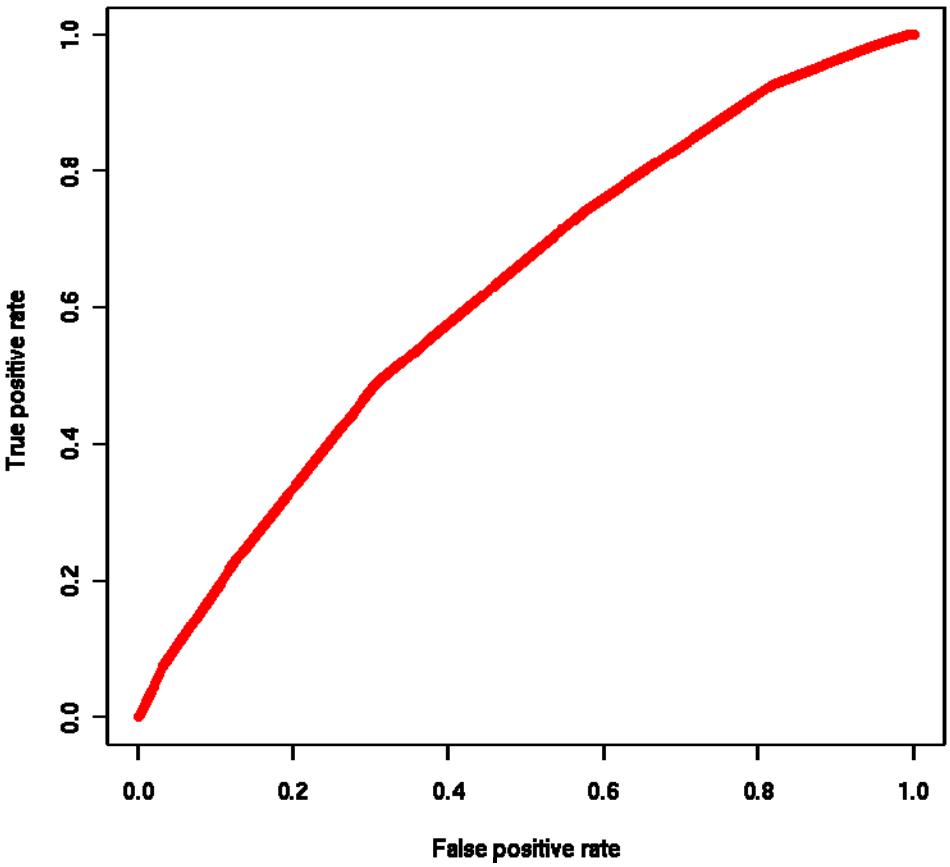
Calculation of ORs (cases vs. controls):

n risk alleles (x)	OR for 0 vs. x risk alleles (95% CI)
1	1.35 (0.38-7.39)
2	1.82 (0.54-9.60)
3	3.27 (0.98-17.06)
4	4.00 (1.21-20.85)
5	6.10 (1.84-31.79)
6	7.02 (2.11-36.68)
7	9.93 (2.90-52.86)
8	8.79 (2.23-51.56)
9	1.13 (0.02-17.03)

Distribution of risk alleles in cases and controls:



ROC curve:



Plot of the ROC of the number of risk alleles as a predictor. The area under the curve is 0.624.

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 Med.* **4**, 101–119 (2003).
2. Wichmann, H.E., Gieger, C., Illig, T. & MONICA/KORA Study Group. KORA-gen-resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67**, S26–S30 (2005).
3. Winkelmann, J. *et al*. Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. *Nat. Genet.* **39**, 1000–1006 (2007). Epub 2007 Jul 18.
4. Affymetrix Inc. Affymetrix® Genome-Wide Human SNP Nsp/Sty Assay 5.0.*http://www.affymetrix.com/support/downloads/manuals/genome_wide_ snp _5_0_manual.pdf* (2007).
5. Affymetrix Inc. BRLMM: an Improved Genotype Calling Method for the GeneChip® Human Mapping 500K Array Set. *http://www. affymetrix. com/support/technical/whitepapers/brlmm_whitepaper. pdf* (2006).
6. Affymetrix Inc. BRLMM-P: a Genotype Calling Method for the SNP 5.0 Array.*http://www.affymetrix.com/support/technical/whitepapers/brlmmp_ whitepaper.pdf* (2007).
7. Steffens, M. *et al*. SNP-based analysis of genetic substructure in the German population. *Hum. Hered*. **62**, 20–29 (2006).
8. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
9. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
10. Devlin, B., Bacanu, S.A. & Roeder, K., Genomic controls to the extreme. *Nat. Genet.* **36**, 1129–1130 (2004).

11. Freedman, M.L. *et al*. Assessing the impact of population stratification on genetic association studies. *Nat. Genet*. **36**, 388–393 (2004).

12. Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association. *Nat. Genet*. **38**, 209–213 (2006).

13. Liebetanz, K.M. *et al*. RLS3: fine-mapping of an autosomal dominant locus in a family with intrafamilial heterogeneity. *Neurology* **67**, 320–321 (2006).

14. Winkelmann, J. *et al*. Evidence for further genetic locus heterogeneity and confirmation of RLS-1 in restless legs syndrome. *Mov. Disord*. **21**, 28–33 (2006).

15. Staden, R., Beal, K.F. & Bonfield, J.K. The Staden Package, 1998. *Methods Mol. Biol.* **132**, 115–130 (2000).

16. Livak, K.J. & Schmittgen, T.D. Analysis of Relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods* **25**, 402–408 (2001).

17. Fisher, R.A. On the interpretation of χ2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society* **85**, 87–94 (1922).

18. Wigginton, J.E., Cutler, D.J. & Abecasis, G.R. A note on exact tests of Hardy-Weinberg equilibrium. *Am. J. Hum. Genet*. **76**, 887–893 (2005).

Supplementary table 1: Stage 1 results for SNP with $P_{nominal/\lambda\text{-corrected}} < 10^{-3}$ on chromosome 9p (0.5–31.5Mb).

SNP ID	Genome pos	Gene	SNP type	risk allele	MAF ca	MAF co	HWE ca	HWE co	OR (95% CI)	P_{nom}	$P_{nom/\lambda\text{-corrected}}$	$P_{corrected}$ (B)
rs1975197	8,836,955	PTPRD	intronic	T	0.216	0.164	0.723	0.528	1.38 (1.16-1.64)	2.70E-04	4.42E-04	1
rs4626664	9,251,737	PTPRD	intronic	A	0.175	0.133	0.889	0.162	1.43 (1.17-1.72)	2.97E-04	4.73E-04	1
rs4008004	15,290,968	C9ORF52	intronic	T	0.254	0.207	0.034	0.499	1.36 (1.16-1.60)	1.26E-04	2.12E-04	1
rs491322	15,298,154	-	intergenic	C	0.461	0.413	0.107	1.000	1.28 (1.11-1.46)	4.74E-04	7.34E-04	1
rs12379466	16,018,394	C9ORF93 isoform AL832443	intronic	G	0.387	0.453	0.499	0.014	1.32 (1.15-1.51)	8.13E-05	1.41E-04	1
rs1328280	16,020,903	C9ORF93 isoform AL832443	intronic	C	0.383	0.449	0.233	0.014	1.31 (1.14-1.50)	1.26E-04	2.11E-04	1
rs4741557	16,021,080	C9ORF93 isoform AL832443	intronic	C	0.385	0.453	0.174	0.014	1.33 (1.16-1.52)	4.60E-05	8.24E-05	1
rs2026661	16,035,125	C9ORF93 isoform AL832443	intronic	C	0.462	0.407	0.359	0.036	1.27 (1.10-1.46)	5.54E-04	8.49E-04	1

Supplementary table 1: Stage 1 results for SNP with $P_{nominal/\lambda\text{-corrected}} < 10^{-3}$ on chromosome 9p (0.5–31.5Mb), continued.

These 8 SNPs fulfilled the criterion for replication of a nominal P value $< 10^{-3}$ after correction for λ using logistic regression with age, sex and the first four components from the MDS analysis of the IBS matrix as covariates. Genome positions (Genome pos) refer to the Human March 2006 (hg18) assembly (http://genome.ucsc.edu). chr, chromosome; ca, cases; co, controls; MAF, minor allele frequency; HWE, P value for the deviation from Hardy-Weinberg-Equilibrium[18]; OR, odds ratio; CI, confidence interval; P_{nom}, nominal P value. $P_{nom/\lambda\text{-corrected}}$ nominal P value corrected for population stratification by dividing the resulting χ^2 by the inflation factor λ (= 1.07). $P_{corrected}$ (B), λ-corrected P value corrected for multiple testing using Bonferroni on 208,733 tests.

Supplementary table 2: Demographic data of stage 1 and 2 samples.

	Stage 1			Stage 2					
	GWA-1 Cases	**GWA-2 Cases-**	**Controls KORAS3/F3**	**Cases GER**	**Controls KORAS4**	**Cases CZ**	**Controls CZ**	**Cases CAN**	**Controls CAN**
N individuals	399	224	1639	1271	1900	279	368	285	842
N females	287	151	830	916	1140	176	210	187	445
N males	112	73	809	355	760	103	158	107	397
Mean age (SD)	60.7 (6.4)	57.4 (9.5)	62.6 (7.8)	61.05 (13.2)	57.3 (10.3)	55.8 (14.9)	45.4 (9.7)	53.6 (11.72)	45.5 (12.8)
Mean age at onset (SD)*	33.1 (11.7)	35.6 (14.4)	-	40.5 (17.5)	-	38.4 (18.0)	-	27.8 (12.2.)	-
Positive family history	399 (100%)	155 (69.2%)	-	595 (46.8%)	-	107 (38.4%)	-	182 (63.9%)	-

Table includes only samples that were successfully genotyped and not removed due to population stratification. N, number; GER, German; CZ, Czech; CAN, Canadian. KORAS3/F3 and KORAS4, controls drawn from KORA population-based cohort study, Germany.

*Age at onset is unknown for 21 individuals in GWA-1 cases, 131 in GWA-2 cases, 224 in GER, and 51 in CAN.

Supplementary table 3: Stage 2 association results for individual and combined analysis.

Sample	GER			CZ			CAN			Combined analysis GER/CZ/CAN			
dbSNP ID	MAF ca co	OR (95% CI) LR	$P_{corrected}$ (B) LR	MAF ca co	OR (95% CI) LR	$P_{corrected}$ (B) LR	MAF ca co	OR (95% CI) LR	$P_{corrected}$ (B) LR	OR (95% CI) LR	$P_{nom/\lambda\text{-corrected}}$ CMH	$P_{nom/\lambda\text{-corrected}}$ LR	$P_{corrected}$ (B) LR
rs1975197	0.196 0.157	1.31 (1.15-1.49)	1.55 E-03	0.158 0.136	1.17 (0.87-1.58)	1	0.203 0.156	1.38 (1.08-1.76)	1.81 E-01	1.29 (1.16-1.44)	1.66 E-06	3.29 E-06	3.29 E-05
rs4626664	0.167 0.117	1.51 (1.30-1.75)	6.88 E-05	0.196 0.146	1.41 (1.05-1.90)	7.47 E-01	0.159 0.117	1.40 (1.07-1.82)	2.44 E-01	1.48 (1.32-1.67)	1.55 E-09	7.53 E-08	7.53 E-07
rs4008004	0.219 0.211	1.05 (0.93-1.18)	1	0.195 0.238	1.27 (0.97-1.66)	1	0.199 0.206	1.05 (0.82-1.34)	1	1.00 (0.91-1.11)	8.67 E-01	7.92 E-01	1
rs491322	0.425 0.422	1.01 (0.91-1.12)	1	0.413 0.431	1.07 (0.86-1.34)	1	0.399 0.409	1.04 (0.86-1.28)	1	1.00 (0.92-1.09)	5.94 E-01	1.00 E-01	1
rs12379466	0.422 0.422	1.00 (0.90-1.11)	1	0.434 0.440	1.02 (0.82-1.28)	1	0.423 0.448	1.10 (0.91-1.33)	1	1.03 (0.95-1.12)	7.23 E-01	7.04 E-01	1
rs1328280	0.424 0.425	1.01 (0.91-1.14)	1	0.436 0.441	1.02 (0.82-1.28)	1	0.425 0.448	1.09 0.91-1.32)	1	1.03 (0.95-1.12)	7.18 E-01	6.74 E-01	1
rs4741557	0.423 0.426	1.01 (0.91-1.12)	1	0.435 0.439	1.02 (0.81-1.28)	1	0.424 0.446	1.09 (0.91-1.32)	1	1.03 (0.95-1.12)	7.58 E-01	6.57 E-01	1
rs2026661	0.439 0.435	1.01 (0.92-1.12)	1	0.445 0.458	1.05 (0.84-1.32)	1	0.433 0.460	1.11 (0.92-1.34)	1	1.02 (0.94-1.11)	9.26 E-01	7.71 E-01	1

GER, German; CZ, Czech; CAN, Canadian; ca, cases; co, controls; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval; $P_{nom/\lambda\text{-corrected}}$, nominal P value corrected for population stratification by dividing the resulting χ^2 by the respective inflation factor λ. $P_{corrected}$ (B), λ-corrected P value corrected for multiple testing using Bonferroni on 10 tests. LR, logistic regression; CMH, Cochran-Mantel-Haenszel test.

Supplementary table 4: Joint analysis stage 1 and 2: genome-wide significance data.

dbSNP ID	Genome position on chr 9p	Gene	SNP type	Risk allele	OR (95% CI)	P_{nom}	$P_{nom/\lambda\text{-corrected}}$	$P_{corrected}$ (B)
rs1975197	8,836,955	PTPRD	intronic	T	1.31 (1.20-1.44)	6.46E-10	5.81E-09	1.21E-03
rs4626664	9,251,737	PTPRD	intronic	A	1.44 (1.31-1.59)	4.96E-11	5.91E-10	1.23E-04
rs4008004	15,290,968	C9ORF52	intronic	T	1.07 (0.99-1.16)	1.22E-01	1.45E-01	1
rs491322	15,298,154	-	intergenic	C	1.05 (0.98-1.13)	1.21E-01	1.44E-01	1
rs12379466	16,018,394	C9ORF93 isoform AL832443	intronic	G	1.10 (1.03-1.18)	7.89E-02	1.23E-01	1
rs1328280	16,020,903	C9ORF93 isoform AL832443	intronic	C	1.10 (1.03-1.18)	8.02E-03	1.25E-02	1
rs4741557	16,021,080	C9ORF93 isoform AL832443	intronic	C	1.10 (1.03-1.18)	6.72E-03	1.07E-02	1
rs2026661	16,035,125	C9ORF93 isoform AL832443	intronic	G*	1.00 (0.94-1.08)	9.05E-01	9.11E-01	1

chr, chromosome; OR, odds ratio; CI, confidence interval; P_{nom}, nominal P value obtained from logistic regression with age, sex and country of origin as covariates; $P_{nom/\lambda\text{-corrected}}$, nominal P value corrected for population stratification; $P_{corrected}$ (B), P value corrected for multiple testing using Bonferroni on 208,733 tests. Genome positions (Genome pos) refer to the Human March 2006 (hg18) assembly (http://genome.ucsc.edu).

* minor allele varies between stage 1 and 2 due to chance variation (MAF close to 0.5)

Supplementary table 5: Results from sequencing of PTPRD: coding and splice-site SNPs.

			Family ID	0001									0331	0054	0038	
			Patient ID	III:6	III:18	III:8	II:14	II:4	IV:2	III:20	III:14	II:7	index	index	Index	control
Exon	dbSNP ID	Alleles	Function													
2 (nc)	rs41265268	C/G	-	G/G	G/G	G/G	GG	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
3 (nc)	rs12006140	G/T	-	T/T	T/T	T/T	TT	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	rs9886850	A/G	-	G/G	A/G	G/G	GG	G/G	G/G	A/G	G/G	A/G	A/G	A/G	A/G	G/G
14	rs1061345	A/C	Met>Ile	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
20	rs59927586	A/G	Thr>Thr	G/G	A/G	G/G	A/G	A/G	G/G	G/G	A/G	G/G	G/G	G/G	G/G	G/G
	rs57597995	C/T	Ala>Ala	T/T	C/T	T/T	C/T	C/T	T/T	T/T	C/T	T/T	T/T	T/T	T/T	T/T
	rs10977171	C/G	Glu>Gln	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
24	rs35278543	-/G	frameshift	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-
26	rs12346849	A/T	Gly>Gly	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	rs34983381	-/C	frameshift	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-
27	rs3824417	C/T	Ile>Ile	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	rs35929428	A/G	Cys>Arg	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
29	rs12344148	G/T	Arg>Arg	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
	rs7869444	A/C	Asp>Glu	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
30	rs7865681	A/G	Arg>Arg	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
31	rs58686491	A/G	Ala>Ala	A/G	A/G	A/G	A/G	A/G	A/G	A/A	A/A	A/G	A/G	A/A	G/G	A/A
	rs41281787	G/T	Pro>Pro	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
36	rs2279776	C/G	Gly>Gly	C/G	C/G	C/G	C/G	C/G	C/C	C/G	C/G	C/C	C/G	C/C	C/C	C/C
37	rs2133788	A/T	Val>Asp	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
38	Not annotated SNP	C/T	Thr>Thr	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
39	rs41281783	A/G	Met>Thr	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
40	rs12351899	A/G	Asn>Asn	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
43	rs3215098	-/ATTCCTGAAC TGTAACTTACC	frameshift	INS/INS	INS/INS	INS/INS	INS/-	INS/-	INS/-	INS/-	-/-	-/-	INS/-	INS/-	INS/-	-/-
	rs35096262	-/AACTTACCAT TCTTGAACTGT	intronic	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	INS/INS	INS/-	INS/-	INS/-	INS/INS
45	rs35081204	-/G	-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-

Supplementary table 5: Results from sequencing of PTPRD: coding and splice-site SNPs, continued.

Genotypes of all detected variants in non-coding and coding exons, including 10bps of flanking intron sequence, in the individuals sequenced. One not annotated SNP was identified in coding Exon 38. nc, non-coding; INS, insertion.

Supplementary table 6: Nonsynonymous coding SNPs in PTPRD – association results for combined analysis in a subset of stage 2 samples (1,835 cases and 2,468 controls).

SNP ID	Genome position on chr 9p	Exon	P_{nom}	$P_{corrected}$ (B)	OR (95% CI)
rs1061345	8,518,742	14	monomorph	na	na
rs10977171	8,508,052	20	3.18E-01	1	1.07 (0.87-1.31)
rs35278543	8,489,761	24	MAF < 0.001	na	na
rs34983381	8,482,869	26	MAF < 0.001	na	na
rs35929428	8,475,834	27	4.51E-01	1	1.09 (0.93-1.27)
rs7869444	8,474,298	29	MAF < 0.001	na	na
rs2133788	8,366,656	37	monomorph	na	na
rs41281783	8,331,730	39	monomorph	na	na

chr, chromosome; OR, odds ratio; CI, confidence interval; P_{nom}, nominal P value obtained from logistic regression with age, sex and country of origin as covariates; $P_{corrected}$ (B), λ-corrected P value corrected for multiple testing using Bonferroni on 10 tests performed in stage 2.

Genome positions (Genome pos) refer to the Human March 2006 (hg18) assembly (http://genome.ucsc.edu).

Appendix 4

Manuscript submitted at: *Neurology*

Variants in *MEIS1* and *BTBD9* are associated with restless legs syndrome in end-stage renal disease

Barbara Schormair*, Jens Plag*, Darina Roeske, Nadine Groß, Walter Samtleben, Peter Lichtner, Andreas Ströhle, Bertram Müller-Myhsok, Thomas Meitinger, Uwe Heemann and Juliane Winkelmann

* joint first authorship

Contributions:

I participated in the organization of the study by designing questionnaires and information letters for patients and clinicians, arranging transport and extraction of blood samples, and collecting the clinical information. I created the database for the clinical and demographic information of the recruited patients and selected the study population from this database. In the genotyping stage, I performed the quality control and prepared the genotype and phenotype input files for statistical analysis. I also conducted the statistical analysis, wrote the manuscript and created all tables and figures in manuscript and supplement.

Title:
Variants in *MEIS1* and *BTBD9* are associated with restless legs syndrome in end-stage renal disease

Authors:
Barbara Schormair, MSc[1,2*], Jens Plag, MD,[3*]Darina Roeske, MSc,[4] Nadine Groß[5], Walter Samtleben, MD,[6] Peter Lichtner, PhD,[1,2] Andreas Ströhle, MD,[3] Bertram Müller-Myhsok, MD,[4] Thomas Meitinger, MD,[1,2] Uwe Heemann, MD,[7] Juliane Winkelmann, MD[1,2,5]

[1]Institute of Human Genetics, Helmholtz Zentrum München - German Research Center For Environmental Health, Neuherberg, Germany
[2]Institute of Human Genetics, Technische Universität München, Munich, Germany
[3]Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charité – Universitätsmedizin Berlin, Berlin, Germany
[4]Max Planck Institute of Psychiatry, Munich, Germany
[5]Department of Neurology, Klinikum Rechts der Isar, Technische Universität München, Munich, Germany
[6]Department of Internal Medicine, Nephrology Division, University of Munich - Klinikum Grosshadern, Munich, Germany
[7]Department of Nephrology, Klinikum Rechts der Isar, Technische Universität München, Munich, Germany
[*]both authors contributed equally

Abstract

Objective: Uremic restless legs syndrome (uRLS) is a common and distressing comorbidity in end-stage renal disease (ESRD) patients with so far unknown etiology. Our study was aimed at elucidating the contribution of the genetic risk factors for idiopathic RLS (iRLS) to uRLS.

Methods: We screened 10 iRLS-associated variants in four genomic loci encoding the genes *MEIS1*, *BTBD9*, *MAP2K5*/*LBXCOR1*, and *PTPRD* in 200 RLS-positive and 443 RLS-negative ESRD patients using multiplex PCR and MALDI-TOF mass spectrometry. Statistical analysis (Armitage trend test with Bonferroni correction) was done using PLINK v1.05.

Results: The association to *MEIS1* and *BTBD9* was significant with nominal P values < 0.004 and remained significant after correction for multiple testing. Odds ratios (ORs) of 1.5 (95% confidence interval of 1.2 – 2.0) were similar to those observed in iRLS. For *MAP2K5/LBXCOR1* and *PTPRD* our data allowed no final conclusion with regard to their contribution to uRLS. We estimated a higher disease prevalence (31.1%) and a lower number of cases with positive family history of RLS (19%) in comparison to the general population.

Conclusions:

We demonstrate for the first time that sequence variants in *MEIS1* and *BTBD9* are genetic susceptibility factors for RLS in ESRD patients. The identified variants cannot fully explain the higher prevalence. Further genetic and non-genetic factors involved in susceptibility to uRLS remain to be identified. Our results indicate a partial overlap of the genetic predisposing factors, suggesting common pathophysiological mechanisms in both forms of RLS.

Introduction

Uremic restless legs syndrome (uRLS) in end-stage renal disease (ESRD) is one of the most common symptomatic forms of RLS and is associated with increased mortality and reduced quality of life.[1] The prevalence of uRLS ranges between 18.4%[2] and 49%[3] in ESRD patient populations of European descent.

Both uremic and idiopathic RLS (iRLS) are diagnosed when the following four clinical diagnostic criteria are fulfilled: (1) an urge to move the legs usually accompanied by unpleasant sensations, (2) worsening of symptoms during rest and inactivity, (3) symptoms are relieved by movement, and (4) symptoms are worse in the evening or at night or occur only in the evening or at night.[4] Periodic limb movements in sleep (PLMS) further support the diagnosis and are present in 80% of RLS patients.[5]

For iRLS a strong genetic contribution has been demonstrated.[6] Genome-wide association studies (GWAS) identified associated variants in four loci encompassing the genes *MEIS1*, *BTBD9, MAP2K5/LBXCOR1*, and *PTPRD*.[7-9] Replication studies in iRLS cases from the United States (US) and Europe confirmed the association of *MEIS1* and *BTBD9*. The *MAP2K5*/*LBXCOR1* locus was replicated in the European cases but showed only a trend for association in the US cases.[10, 11]

The pathophysiology of uRLS and the effect of iRLS risk alleles is still unexplained. So far, clinical and biochemical parameters between RLS-positive and RLS-negative ESRD patients were compared but yielded ambiguous results. Increasing duration of dependence on dialysis, frequency of dialysis sessions, phosphate metabolism, and anemia have inconsistently been associated to RLS in ESRD patients.[3, 12-17]

Whether iRLS and uRLS share common pathophysiological mechanisms and genetic susceptibility is unknown. We therefore investigated the iRLS-associated variants[7-9] in a case-control association study in ESRD patients.

Subjects and Methods

Study population and recruitment procedure

ESRD patients on maintenance hemodialysis were recruited in three waves between January 2005 and August 2008 in 16 dialysis centers in Munich and the surrounding region. The recruitment procedure is given in Fig. 1. In brief, all ESRD patients who agreed to participate in the study answered a self-administered diagnostic questionnaire[18] incorporating the four essential criteria[4]. All ESRD patients who answered at least one question affirmatively were then examined by an expert clinician in a face-to-face interview. Patients were classified RLS-positive or RLS-uncertain (less than four diagnostic criteria fulfilled) according to the personal interview. Patients who answered all diagnostic questions negatively were classified as RLS-negative without personal interview. Uncertain-RLS patients or patients with incomplete data were excluded from the study. All RLS-positive patients underwent a further assessment of their clinical symptoms, e.g. family history (positive family history defined as at least one relative affected by RLS as reported by the proband) and age of onset of RLS. Data on duration of dependence on dialysis and hours of dialysis per week was collected for all ESRD patients. In the final analyses we included only patients of European descent.

SNP selection, genotyping, and quality control

We genotyped 10 of the most significant iRLS associated SNPs (single nucleotide polymorphisms)[7-9] across the four genomic regions. Genotyping was performed following the iPLEX Gold protocol using matrix-assisted laser desorption/ ionization time-of-flight (MALDI-TOF) mass spectrometry (Sequenom). Assays were designed using AssayDesign 3.1.2.2 with the default parameters for the iPLEX Gold chemistry. Cleaned extension products were analyzed by a mass spectrometer (Bruker Daltronik), and peaks were identified using SpectroTYPER RT 3.4.

Quality control criteria leading to exclusion of a SNP from further analysis were a call rate < 90%, a minor allele frequency (MAF) < 10% and $P < 0.001$ for deviations from Hardy-

Weinberg-Equilibrium (HWE) in controls. DNA samples with a call rate < 90% over all SNPs were excluded from the analysis.

Statistical analysis

Statistical analysis was performed with PLINK v1.05.[19] Association was tested in a one-sided test implementing Armitage trend test. Correction for multiple testing was done according to Bonferroni under the assumption of a one-sided test on four regions (P value corrected, P_{corr} = nominal P value, P_{nom} x 4). Homogeneity of odds ratios (ORs) was tested with Breslow-Day test. Dialysis parameters and mean age were compared between cases and controls using a two-sided Student t-test. Gender ratios were compared using a χ^2-test. Study power was calculated using the CATs power calculator[20] with disease prevalence set at 0.20, and risk allele frequencies and ORs as estimated in previous studies.[7, 8, 11]

Results

Study population

From a total of 1,617 ESRD patients 737 agreed to participate (45.6%). The relatively low response rate was due to several reasons. Many ESRD patients had severe comorbidities such as dementia or depression, and were primarily not able or interested therefore unable to answer questions reliably. Others could not participate because of language and comprehension problems. According to the self-administered diagnostic questionnaire 253 ESRD patients were considered potential uRLS cases. Of these, 53 were subsequently excluded from the study (uncertain diagnosis: n = 44; non-European descent: n = 6, incomplete data: n = 3). The final study population included 200 RLS-positive patients (prevalence of uRLS = 31.1%). Positive family history of RLS was reported by 38 of these patients (19%), 133 (66.5%) had a negative family history and for 29 (14.5%) there was no information available. Controls were selected from 484 RLS-negative ESRD patients. Of these, 41 were subsequently excluded from the study (non-European descent: n = 13,

incomplete data: n = 28), yielding 443 RLS-negative ESRD patients. Cases and controls did not differ significantly for age, gender, and the recorded dialysis parameters. Demographic data and dialysis parameters of the final study population are given in Table 1.

Association analysis

All genotyped SNPs passed quality control and were subjected to statistical analysis. A total of 8 individuals (1 case, 7 controls) were excluded due to low genotyping quality. The final analysis therefore included 199 cases and 436 controls. Armitage trend test revealed a significant association of variants in *MEIS1* and *BTBD9* to uRLS (Table 2). Within *MEIS1*, two of three SNPs were significantly associated after correction for multiple testing: rs12469063 (P_{nom} = 0.001, P_{corr} = 0.004, OR = 1.52), and rs2300478 (P_{nom} = 0.0025, P_{corr} = 0.01, OR = 1.47). In *BTBD9*, rs3923809 was associated (P_{nom} = 0.0005, P_{corr} = 0.002, OR = 1.56). For *MAP2K5*/*LBXCOR1* and *PTPRD* the nominal P values were between 0.057 and 0.3 (Table 2). Haplotype analysis confirmed the association of the known 'AG' haplotype in *MEIS1*[7] (P_{corr} = 0.048, OR = 1.57 (95% confidence interval (CI) = 1.10-2.23). Power of the uRLS sample to detect the associations was 92% for *MEIS1*, 75% for *BTBD9*, 61% for *MAP2K5*/*LBXCOR1*, and 35% for *PTPRD* (Supplementary Table 1). A subanalysis with cases stratified according to their family history revealed a trend for differences in the size of the contribution of the associated loci to familial or sporadic uRLS (Supplementary Table 2). Analysing only cases with a positive family history (n = 38), revealed a significant association both to *MEIS1* and *BTBD9* (rs12469063, P_{nom} =0.002, P_{corr} = 0.008; rs2300478, P_{nom} =0.004, P_{corr} = 0.016, and rs3923809, P_{nom} =0.003, P_{corr} = 0.012). Using only cases with a negative family history (n = 133), the two loci showed only nominally significant P values (rs12469063, P_{nom} =0.014, P_{corr} = 0.056; rs2300478, P_{nom} =0.021, P_{corr} = 0.084, and rs3923809, P_{nom} = 0.015, P_{corr} = 0.06).

Discussion

We demonstrate for the first time that sequence variants in *MEIS1* and *BTBD9* are genetic susceptibility factors for RLS in ESRD patients ($P_{corr} \leq 0.01$). The effect size of these variants is within the same range as observed in iRLS studies of comparable sample size (ORs_{uRLS} = 1.47 to 1.56, 95% CIs = 1.19-2.04; ORs_{iRLS}= 1.43 to 1.59, 95% CIs = 1.12 – 2.2).[10, 11]

Based on the present data we can neither prove nor exclude a contribution of *MAP2K5*/*LBXCOR1* and *PTPRD* to uRLS. Although they were not significantly associated in our study, their ORs showed the same direction as in iRLS[7, 8, 10, 11] and the CIs were overlapping.

Previous genotype/phenotype analysis in iRLS patients showed *BTBD9* more associated to PLMS, the motor component of RLS, than to the sensory symptoms.[9] Therefore, its strong association to uRLS is remarkable in the context that the motor symptoms seem to be more prominent in uRLS in comparison to iRLS.[21, 22]

Our result that only 19% of uRLS cases reported a positive family history is in line with previous observations showing a lower frequency of familial RLS (12%)[23] in uRLS compared to iRLS (30 - 92%).[6] The impact of the associated variants is not statistically different between either familial or sporadic uRLS. The ORs tended to be higher in the familial subgroup but the difference was not significant ($P_{Breslow\text{-}Day} > 0.3$) and 95% CIs were overlapping which is in line with iRLS studies (Supplementary Table 2).[7, 8, 10, 11] The reduced familial clustering of RLS in ESRD patients could be interpreted as an indicator of a stronger influence of non-genetic than genetic factors on developing the disease.

The prevalence of uRLS observed in our study (31.1%) is concordant with previous investigations in uRLS2, [13, 17] and is higher than in the general population at the same age[24, 25] suggesting that additional genetic and/or non-genetic risk factors must be present in ESRD patients. Complex diseases result from genetic and non-genetic or environmental factors and their interactions. Calculating the OR for ESRD as a risk factor for RLS by comparing our

ESRD patients sample to a hypothetical sample from the general population of the same size (643, of which 64 (10%) have RLS), shows an effect size of 4. Therefore reduced renal function and dependence on dialysis seem to be a strong trigger for RLS, acting independently of the genetic susceptibility. This is supported by the abolishment of RLS symptoms after renal transplantation.[17]

Finally, both iRLS and uRLS share genetic risk factors, suggesting a partial overlap in the predisposing mechanism and in the pathophysiology. It remains to be investigated if there are genetic variants specific to uRLS which are not relevant for iRLS.

Acknowledgement

We are grateful to all patients who participated in this study. We also thank the participating Kuratorium für Dialyse und Nierentransplantation e.V. (KfH) dialysis centers KfH Nierenzentrum Ebersberg (Dr. med. Kathrin Stephanie Lange), KfH Nierenzentrum Freising (Dr. med. Monika Wildgruber), KfH Nierenzentrum Germering (Dr. med. Hermann Hacker), KfH Nierenzentrum Ingolstadt (Dr. med. Friedrich Josef Lazarus), KfH Nierenzentrum Miesbach (Dr. med. Daniela Anne Soreth-Rieke), KfH Nierenzentrum München Burggrafenstrasse (Dr. med. Gabriele Schätzle), KfH Nierenzentrum München Elsenheimer Strasse (PD Dr. med. Stephan Lederer), KfH Nierenzentrum München Isoldenstrasse (Prof. Dr. med. Johannes Mann), KfH Nierenzentrum München Seybothstrasse (Prof. Dr. med. Jürgen Scherberich), KfH Nierenzentrum Neuried (Prof. Dr. med. Walter Samtleben), KfH Nierenzentrum Oberschleissheim (Dr. med. Norbert Bockreiss), KfH Nierenzentrum Pfaffenhofen (Dr. med. Axel Witte), KfH Nierenzentrum Rosenheim (Dr. med. Andreas Thiele), KfH Nierenzentrum Straubing (Prof. Dr. med. Marianne Haag-Weber), KfH Nierenzentrum Traunstein (Dr. med. Martin Gottsmann) and KfH Nierenzentrum Unterhaching (Dr. med. Klaus Hällfritzsch), for their help in conducting the study.

References

1. Winkelman JW, Chertow GM, Lazarus JM. Restless legs syndrome in end-stage renal disease. Am J Kidney Dis 1996;28:372-378.

2. Merlino G, Piani A, Dolso P, et al. Sleep disorders in patients with end-stage renal disease undergoing dialysis therapy. Nephrol Dial Transplant 2006;21:184-190.

3. Siddiqui S, Kavanagh D, Traynor J, Mak M, Deighan C, Geddes C. Risk factors for restless legs syndrome in dialysis patients. Nephron Clin Pract 2005;101:c155-160.

4. Allen RP, Picchietti D, Hening WA, Trenkwalder C, Walters AS, Montplaisi J. 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.

5. Montplaisir J, Boucher S, Poirier G, Lavigne G, Lapierre O, Lesperance P. Clinical, polysomnographic, and genetic characteristics of restless legs syndrome: a study of 133 patients diagnosed with new standard criteria. Mov Disord 1997;12:61-65.

6. Winkelmann J, Polo O, Provini F, et al. Genetics of restless legs syndrome (RLS): State-of-the-art and future directions. Mov Disord 2007;22 Suppl 18:S449-458.

7. 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.

8. 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.

9. 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.

10. Vilarino-Guell C, Farrer MJ, Lin SC. A genetic risk factor for periodic limb movements in sleep. N Engl J Med 2008;358:425-427.

11. Kemlink D, Polo O, Frauscher B, et al. Replication of Restless Legs Syndrome Loci in Three European Populations. J Med Genet 2009;10:10.

12. Collado-Seidel V, Kohnen R, Samtleben W, Hillebrand GF, Oertel WH, Trenkwalder C. Clinical and biochemical findings in uremic patients with and without restless legs syndrome. Am J Kidney Dis 1998;31:324-328.

13. Gigli GL, Adorati M, Dolso P, et al. Restless legs syndrome in end-stage renal disease. Sleep Med 2004;5:309-315.

14. Huiqi Q, Shan L, Mingcai Q. Restless legs syndrome (RLS) in uremic patients is related to the frequency of hemodialysis sessions. Nephron 2000;86:540.

15. Roger SD, Harris DC, Stewart JH. Possible relation between restless legs and anaemia in renal dialysis patients. Lancet 1991;337:1551.

16. Takaki J, Nishi T, Nangaku M, et al. Clinical and psychological aspects of restless legs syndrome in uremic patients on hemodialysis. Am J Kidney Dis 2003;41:833-839.

17. Kavanagh D, Siddiqui S, Geddes CC. Restless legs syndrome in patients on dialysis. Am J Kidney Dis 2004;43:763-771.

18. 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 2004;164:196-202.

19. 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.

20. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet 2006;38:209-213.

21. Wetter TC, Stiasny K, Kohnen R, Oertel WH, Trenkwalder C. Polysomnographic sleep measures in patients with uremic and idiopathic restless legs syndrome. Mov Disord 1998;13:820-824.

22. Enomoto M, Inoue Y, Namba K, Munezawa T, Matsuura M. Clinical characteristics of restless legs syndrome in end-stage renal failure and idiopathic RLS patients. Mov Disord 2008;23:811-816; quiz 926.

23. Winkelmann J, Wetter TC, Collado-Seidel V, et al. Clinical characteristics and frequency of the hereditary restless legs syndrome in a population of 300 patients. Sleep 2000;23:597-602.

24. Rothdach AJ, Trenkwalder C, Haberstock J, Keil U, Berger K. Prevalence and risk factors of RLS in an elderly population: the MEMO study. Memory and Morbidity in Augsburg Elderly. Neurology 2000;54:1064-1068.

25. Phillips B, Young T, Finn L, Asher K, Hening WA, Purvis C. Epidemiology of restless legs symptoms in adults. Arch Intern Med 2000;160:2137-2141.

Figure 1: Overview of recruitment procedure.

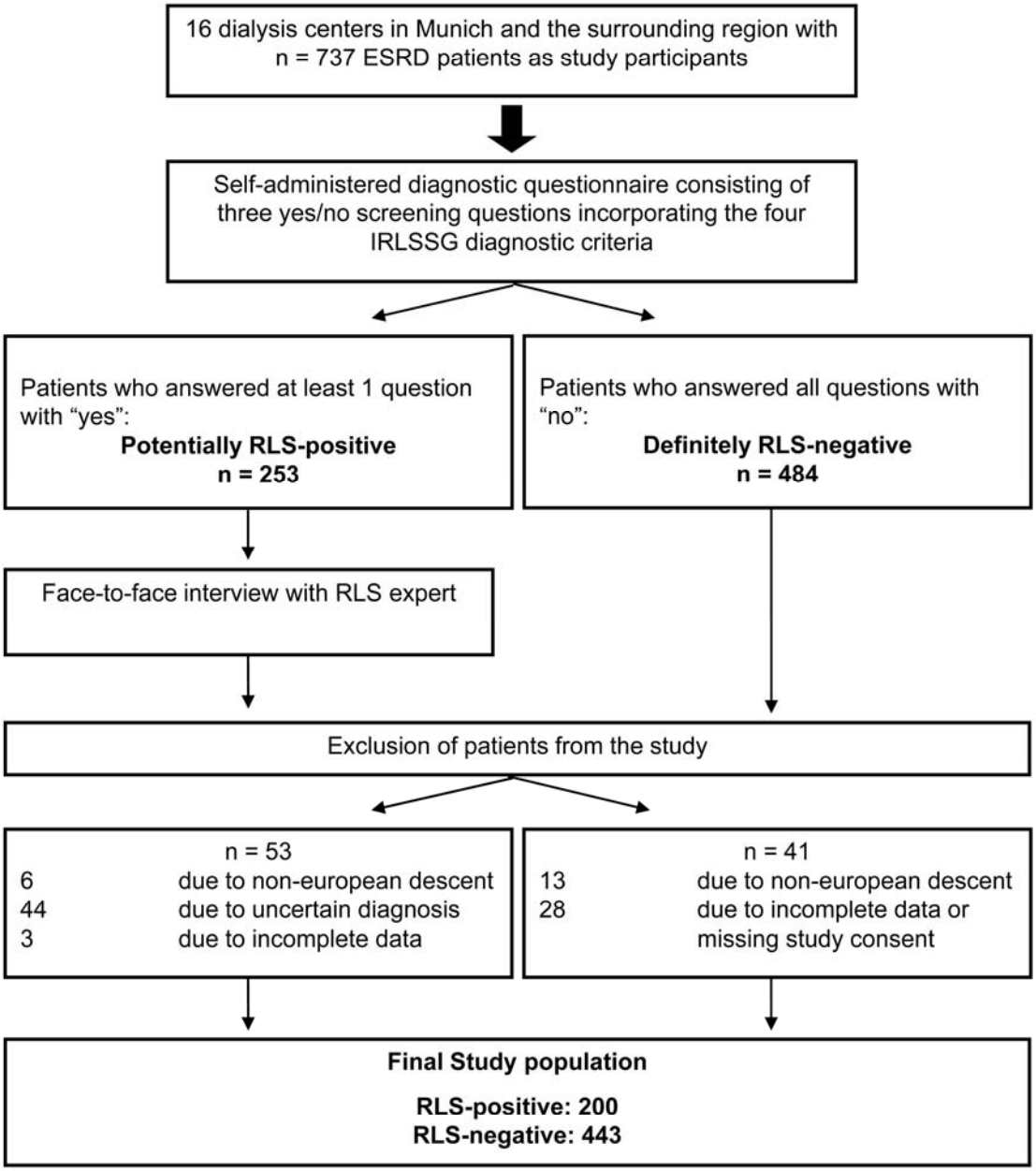


Table 1: Demographic and dialysis data of final study population.

	RLS-positive ESRD patients	RLS-negative ESRD patients	P*
mean dialysis duration (hours per week, mean ± SD)	13.6 ± 2.2 (NA for 42 patients)	13.6 ± 2.7 (NA for 138 patients)	0.82
mean dependence on dialysis (months, mean ± SD)	70.7 ± 68.5 (NA for 34 patients)	60.7 ± 62.2 (NA for 105 patients)	0.10
Age (years, mean ± SD)	64.8 ± 12.5	65 ± 13.4	0.91
Gender (% female)	43.5	35.8	0.06
Age of onset of RLS (years, mean ± SD)	54.8 ± 15.4 (NA for 149 patients)	NA	NA

*P values were obtained from two-sided student t-tests for dialysis parameters and mean age, for gender ratios from a χ^2-test.
NA = not available, SD = standard deviation.

Table 2: Association results for all RLS-positive ESRD patients vs. all RLS-negative ESRD patients.

Chr	Genomic position	SNP	Minor allele	MAF (cases)	MAF (controls)	Risk allele	P_{nom}	P_{corr}	OR (95% CI)
2p	66611925	rs6710341	G	0.168	0.144	G	0.1310	0.5240	1.21 (0.87-1.67)
	66617811	**rs12469063**	**G**	0.312	0.231	**G**	**0.0010**	**0.0040**	**1.52 (1.17-1.98)**
	66634957	**rs2300478**	**G**	0.314	0.240	**G**	**0.0025**	**0.0100**	**1.47 (1.13-1.91)**
6p	**38548947**	**rs3923809**	**C**	0.248	0.341	**T**	**0.0005**	**0.0020**	**1.56 (1.19-2.04)**
9p	8836955	rs1975197	T	0.153	0.176	T	0.1530	0.6120	0.84 (0.61-1.16)
	9251737	rs4626664	A	0.136	0.126	A	0.3240	1	1.09 (0.76-1.54)
15q	65824632	rs11635424	A	0.309	0.354	G	0.0575	0.2300	1.22 (0.95-1.59)
	65859329	rs3784709	T	0.302	0.340	C	0.0870	0.3480	1.19 (0.93-1.54)
	65882139	rs1026732	A	0.308	0.341	G	0.1220	0.4880	1.16 (0.90-1.49)
	65890260	rs6494696	C	0.306	0.343	G	0.0970	0.3880	1.18 (0.92-1.52)

Chr, Chromosome; MAF, minor allele frequency; P_{nom}, nominal P values, were obtained from Armitage trend test. Pcorr, P value corrected, for multiple testing according to Bonferroni, equals P_{nom} x 4. ORs, odds ratios, were obtained from Armitage trend test and refer to the risk allele. Significant associations are highlighted in bold.

Variants in *MEIS1* and *BTBD9* are associated with restless legs syndrome in end-stage renal disease

Barbara Schormair, MSc[1,2*], Jens Plag, MD,[3*]Darina Roeske, MSc,[4] Nadine Groß[5], Walter Samtleben, MD,[6] Peter Lichtner, PhD,[1,2] Andreas Ströhle, MD,[3] Bertram Müller-Myhsok, MD,[4] Thomas Meitinger, MD,[1,2] Uwe Heemann, MD,[7] Juliane Winkelmann, MD[1,2,5]

[1]Institute of Human Genetics, Helmholtz Zentrum München - German Research Center For Environmental Health, Neuherberg, Germany
[2]Institute of Human Genetics, Technische Universität München, Munich, Germany
[3]Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charité – Universitätsmedizin Berlin, Berlin, Germany
[4]Max Planck Institute of Psychiatry, Munich, Germany
[5]Department of Neurology, Klinikum Rechts der Isar, Technische Universität München, Munich, Germany
[6]Department of Internal Medicine, Nephrology Division, University of Munich - Klinikum Grosshadern, Munich, Germany
[7]Department of Nephrology, Klinikum Rechts der Isar, Technische Universität München, Munich, Germany
[*]both authors contributed equally

Supplementary Information

Supplementary Table 1: Study power analysis

Sample	All Cases	Cases with positive family history	Cases with negative family history
Gene (risk allele freq, OR)			
MEIS1 (0.3, 1.7)	91%	67%	78%
BTBD9 (0.7, 1.6)	75%	49%	60%
MAP2K5/LBXCOR1 (0.7, 1.5)	61%	37%	47%
PTPRD (0.2, 1.4)	35%	17%	24%

Freq, frequency; OR, odds ratio. Study power was calculated using the CATs power calculator with disease prevalence set at 0.20, and risk allele frequencies and ORs as estimated in previous studies.[1-3]

Supplementary Table 2: Association results for subgroup analysis based on family history of uRLS cases.

			Cases with positive family history			Cases with negative family history			All cases	
Chr	Genomic position	SNP	P_{nom}	P_{corr}	OR (95% CI)	P_{nom}	P_{corr}	OR (95% CI)	OR (95% CI)	$P_{Breslow-Day}$
2p	66611925	rs6710341	0.035	0.138	1.72 (0.97-3.04)	0.456	1	1.02 (0.69-1.51)	1.21 (0.87-1.67)	0.34
	66617811	rs12469063	**0.002**	**0.008**	2.07 (1.27-3.38)	**0.014**	0.056	1.42 (1.04-1.93)	1.52 (1.17-1.98)	0.43
	66634957	rs2300478	**0.004**	**0.016**	1.98 (1.21-3.22)	**0.021**	0.084	1.38 (1.02-1.87)	1.47 (1.13-1.91)	0.46
6p	38548947	rs3923809	**0.003**	**0.012**	2.44 (1.30-4.55)	**0.015**	0.060	1.41 (1.04-1.92)	1.56 (1.19-2.04)	0.31
9p	8836955	rs1975197	0.343	1	0.87 (0.46-1.65)	0.284	1	0.89 (0.62-1.30)	0.84 (0.61-1.16)	0.97
	9251737	rs4626664	0.194	0.774	0.70 (0.31-1.56)	0.244	0.976	1.15 (0.77-1.71)	1.09 (0.76-1.54)	0.54
15q	65824632	rs11635424	0.181	0.724	1.27 (0.75-2.08)	0.122	0.486	1.19 (0.88-1.59)	1.22 (0.95-1.59)	0.98
	65859329	rs3784709	0.252	1	1.19 (0.71-1.96)	0.137	0.546	1.18 (0.88-1.59)	1.19 (0.93-1.54)	1
	65882139	rs1026732	0.268	1	1.18 (0.70-1.96)	0.176	0.702	1.15 (0.85-1.54)	1.16 (0.90-1.49)	1
	65890260	rs6494696	0.234	0.934	1.20 (0.72-2.00)	0.168	0.670	1.15 (0.85-1.54)	1.18 (0.92-1.52)	0.99

Chr, Chromosome; P_{nom}, nominal P values, were obtained from Armitage trend test. Pcorr, P value corrected, for multiple testing according to Bonferroni, equals P_{nom} x 4. ORs, odds ratios, were obtained from Armitage trend test and refer to the risk allele. $P_{Breslow-Day}$, P value for homogeneity of ORs as implemented in PLINK v1.05. Significant associations are highlighted in bold.

References

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. 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
3. Kemlink D, Polo O, Frauscher B et al. Replication of Restless Legs Syndrome Loci in Three European Populations. J Med Genet. 2009;10:10

Danksagung

Zunächst danke ich Herrn Prof. Dr. Thomas Meitinger, Direktor des Instituts für Humangenetik des Helmholtz Zentrums München, für die Möglichkeit, meine Doktorarbeit an seinem Institut durchzuführen. Außerdem möchte ich Ihm für seine Unterstützung dieses Projekts durch konstruktive Kritik und anregende Diskussionen danken.

Weiterhin danke ich Herrn Prof. Dr. Wolfgang Wurst, Direktor des Instituts für Entwicklungsgenetik des Helmholtz Zentrums München, für die Übernahme der Vertretung meiner Doktorarbeit gegenüber der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München.

Herrn Prof. Dr. Fries, Leiter des Lehrstuhls für Tierzucht des WZW Weihenstephan, danke ich für die Übernahme des Vorsitzes der Prüfungskommission und sein großes Interesse an meiner Arbeit.

Ganz besonders möchte ich mich bei meiner AG-Leiterin Frau Prof. Dr. Juliane Winkelmann bedanken. Liebe Juliane, ohne Deine beständige Unterstützung, Motivation und Förderung wäre diese Arbeit niemals möglich gewesen. Ich danke Dir für Dein immer offenes Ohr in allen Doktorarbeitslagen, Deine Ideen und Ratschläge und Deine immer vorhandene Bereitschaft zu helfen.

Einen ganz großen Anteil an dieser Arbeit hat auch die AG Statistische Genetik am MPI für Psychiatrie in München unter der Leitung von Prof. Dr. Bertram Müller-Myhsok, die die statistische Analyse der Assoziationsstudien durchgeführt hat. Vielen Dank für Eure großartige Arbeit und Unterstützung bei der statistischen Analyse.

Nicht vergessen möchte ich selbstverständlich auch meine Kollegen aus dem Institut für Humangenetik und aus meiner Arbeitsgruppe, die mich über die vergangenen Jahre hinweg immer unterstützt haben. Weiterhin möchte ich mich bei allen TAs bedanken, die mich im Labor unterstützt haben. Liebe Mädels aus dem 34ger, es war eine tolle Zeit bei Euch. Vielen Dank, dass Ihr mir für den Großteil meiner Doktorandenzeit ein Dach über dem Kopf gewährt habt!

Zum Schluss möchte ich mich noch von ganzem Herzen bei meinen Eltern, meinen Brüdern, meinem Freund und meinen Freunden bedanken. Danke für Eure Unterstützung, Euren Glauben an mich und Eure immer vorhandene Bereitschaft, sich meine Sorgen und Ängste anzuhören. Ohne Euch hätte ich diese Arbeit nicht geschafft.

Conferences and Presentations

13th Annual Meeting of the German Society of Neurogenetics (München, Germany, October 11-13, 2007)

Poster presentation: "Case-control association study in uremic RLS – preliminary results

Keystone Symposia Meeting "Genetics and Biochemistry of Sleep" (Tahoe City, USA, March 7-12, 2008)

Poster presentation: "MEIS1 and BTBD9 are associated with uremic RLS – a case-control study in end-stage renal disease patients"

Attendance of the conference was supported by a Keystone Symposia scholarship awarded to me.

16th Annual Meeting of the German Sleep Society (Kassel, Germany, October 16-18, 2008)

Invited talk at the forum for young scientists: "PTPRD (protein tyrosine phosphatase type delta) on chromosome 9 is associated with restless legs syndrome (RLS)"

The RLS Foundation and International RLS Study Group Scientific Meeting 2008 (Baltimore, USA, October 27-29, 2008)

Poster presentation: "PTPRD (protein tyrosine phosphatase type delta) on chromosome 9 is associated with restless legs syndrome (RLS)"

Attendance of the conference was supported by an IRLSSG travel grant awarded to me.

European Human Genetics Conference 2009 of the European Society of Human Genetics (Vienna, Austria, May 23-26, 2009)

Poster presentation: "Shared genetic susceptibility in primary and secondary Restless Legs Syndrome: A case-control association study in ESRD patients"

Publications

Winkelmann J, Schormair B, Lichtner P, Ripke S, Xiong L, Jalilzadeh S, Fulda S, Pütz B, Eckstein G, Hauk S, Trenkwalder C, Zimprich A, Stiasny-Kolster K, Oertel W, Bachmann CG, Paulus W, Peglau I, Eisensehr I, Montplaisir J, Turecki G, Rouleau G, Gieger C, Illig T, Wichmann HE, Holsboer F, Müller-Myhsok B, Meitinger T. Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. *Nat. Genet*. **39(8)**:1000-1006, 2007.

Winkelmann J, Lichtner P, Schormair B, Uhr M, Hauk S, Stiasny-Kolster K, Trenkwalder C, Paulus W, Peglau I, Eisensehr I, Illig T, Wichmann HE, Pfister H, Golic J, Bettecken T, Pütz B, Holsboer F, Meitinger T, Müller-Myhsok B. Variants in the neuronal nitric oxide synthase (*nNOS, NOS1*) gene are associated with restless legs syndrome. *Mov. Disord*. **23(3)**:350-358. 2008.

Schormair B, Kemlink D, Roeske D, Eckstein G, Xiong L, Lichtner P, Ripke S, Trenkwalder C, Zimprich A, Stiasny-Kolster K, Oertel W, Bachmann CG, Paulus W, Högl B, Frauscher B, Gschliesser V, Poewe W, Peglau I, Vodicka P, Vávrová J, Sonka K, Nevsimalova S, Montplaisir J, Turecki G, Rouleau G, Gieger C, Illig T, Wichmann HE, Holsboer F, Müller-Myhsok B, Meitinger T, Winkelmann J. *PTPRD* (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome. *Nat. Genet*. **40(8)**:946-948. 2008.

Kemlink D, Polo O, Frauscher B, Gschliesser V, Högl B, Poewe W, Vodicka P, Vavrova J, Sonka K, Nevsimalova S, Schormair B, Lichtner P, Silander K, Peltonen L, Gieger C, Wichmann HE, Zimprich A, Roeske D, Müller-Myhsok B, Meitinger T, Winkelmann J. Replication of restless legs syndrome loci in three European populations. *J. Med. Genet*. **46(5)**:315-318. 2009.

Schormair B, Plag J, Roeske D, Groß N, Samtleben W, Lichtner P, Ströhle A, Müller-Myhsok B, Meitinger T, Heemann U, Winkelmann, J. Variants in *MEIS1* and *BTBD9* are associated with restless legs syndrome in end-stage renal disease. Manuscript submitted at *Neurology*.