TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Entwicklungsgenetik

Prediction of new regulatory molecular networks in neurodegenerative diseases with the help of bioinformatical methods

Regina Peis

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. E. Grill
Prüfer der Dissertation:	
	1. Univ.-Prof. Dr. W. Wurst
	2. Univ.-Prof. Dr. Dr. F. Theis

Die Dissertation wurde am 05.09.2012 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 06.11.2012 angenommen.

Contents

Abstract

The molecular mechanisms and genetic risk factors underlying neurodegenerative diseases such as Alzheimer's (AD) and Parkinson's disease (PD) as well as psychiatric diseases like stress and depression are only partly understood. However, it is strongly assumed that these diseases are highly connected either by co-occurrence or being risk factors for further neurodegeneration. To identify new factors, which may contribute to neurodegeneration, different approaches are taken including proteomics, genetics, and functional genomics. I established two bioinformatics approaches for the identification of transcription factor binding site (TFBS) modules in distinct AD-related genes as well as microRNAs regulating ADAM10 expression concerning AD.

To detect additional coregulated genes, which may potentially contribute to AD, I incorporated in the first workflow known multivariate methods like support vector machines, biclustering, and predicted TFBS modules by using in silico analysis and over 400 expression arrays from human and mouse. Two significant modules are composed of binding sites of three TF families: CTCF, SP1F, and EGRF/ZBPF, which are conserved between human and mouse APP promoter sequences.

Additionally, I applied the multivariate methods of the approach on microarray datasets of depression as well as PD mouse models to validate this part of the workflow and to unravel the underlying disease-related pathways. Furthermore, several PD related genes were analysed for TFBSs influencing PD pathogenesis and some TFBSs were verified in the lab.

The second workflow combines microRNA (miRNA) prediction software with statistical and AD disease-relevant criteria in contrast to already existing miRNA target site prediction databases. Three miRNAs (miR-103, miR-107, miR-1306) with evolutionary conserved binding sites on ADAM10 3'UTR were further analysed by occurrence in GWAS studies listed in AlzGene database, Gene Ontology analysis and literature mining. Finally, a luciferase assay verfies the potential effect of these three miRNAs on ADAM10 3'UTR in SH-SY5Y cells.

An overall network summarizing all predictions and validations of the thesis was established showing the connectivity of the neurodegenerative and psychiatric diseases on the gene regulatory level. This might explain the co-occurrence of neurodegenerative and psychiatric diseases as well as their interaction.

The specific combination of in silico promoter and multivariate analysis in the first

approach can identify regulation mechanisms of genes involved in multifactorial diseases. The second novel approach using established prediction programs and specific selection criteria with respect to AD successfully identifies miRNA:mRNA interactions and hence offers possibilities for the development of therapeutic treatments of AD.

Zusammenfassung

Die molekularen Mechanismen und genetischen Risikofaktoren, welche zu neurodegenerativen Erkrankungen wie Alzheimer und Parkinson sowie zu psychischen Erkrankungen wie Stress und Depression führen, werden bis jetzt nur zum Teil verstanden. Dennoch wird stark davon ausgegangen, dass diese Krankheiten stark miteinander verknüpft sind entweder durch gemeinsames Auftreten oder dadurch dass sie ein Risiko für weitere Neurodegeneration darstellen. Um neue Faktoren, die womöglich zu Neurodegeneration beitragen, zu identifizieren, werden verschiedene Ansätze wie z.B. Proteomiks, Genetik und funktionelle Genomik verwendet. In der vorliegenden Arbeit habe ich zwei neue bioinformtische Ansätze für die Identifikation von Modulen von Transkriptionsfaktorbindestellen in verschiedenen mit Alzheimer verknüpften Genen sowie microRNAs, welche die Expression von ADAM10 in Bezug auf Alzheimer regulieren, erstellt.

Zur Ermittlung von weiteren ko-regulierten Genen, welche sich möglicherweise auf Alzheimer auswirken, habe ich in den ersten Ansatz bekannte multivariate Methoden wie Support Vektor Maschinen als auch Biclustering aufgenommen und sage Module von Transkriptionsfaktorbindestellen vorher, indem ich eine in silico Analyse durchführe und über 400 Expressionsarrays von Mensch und Maus verwende. Zwei signifikante Module bestehen aus drei Transkriptionsfaktor-Familien: CTCF, SP1F und EGRF/ZBPF, welche zwischen der APP Promotorsequenz von Mensch und Maus konserviert sind.

Zusätzlich wende ich die multivariaten Methoden des Ansatzes auf weitere Microarray Datensätze von Mausmodellen für Depression und Parkinson an, um diesen Teil des Ansatzes zu validieren und um zugrunde liegende krankheitsrelevante Signalwege aufzudecken. Darüber hinaus suche ich in einigen Parkinson assoziierten Genen nach Transkriptionsfaktorbindestellen, welche im Krankheitsverlauf von Parkinson eine Rolle spielen können. Einige dieser Ergebnisse wurden im Labor bestätigt.

Der zweite Workflow kombiniert Software zur microRNA Vorhersage mit statistischen und Alzheimer relevanten Kriterien im Gegensatz zu bereits vorhandenen Datenbanken mit vorhergesagten Zielgenen der microRNAs. Drei microRNAs (miR-103, miR-107, miR-1306) mit evolutionär konservierten Bindestellen in der ADAM10 3'UTR wurden durch ihr Vorkommen in GWAS Studien, Gene Ontologie Analysen und Literature Mining näher charakterisiert. Schließlich zeigt ein Luciferase Assay einen möglichen Effekt dieser drei microRNA auf ADAM10 3'UTR in SH-SY5Y

Zellen.
Ein Gesamtnetzwerk fasst alle Vorhersagen und Validierungen aus meiner Arbeit zusammen und veranschaulicht die Verknüpfung der neurodegenerativen und psychischen Erkrankungen durch Interaktionen von Genen der verschiedenen Krankheiten. Interkationen von Genen, die in verschiedenen Krankheiten eine Rolle spielen, könnten eine mögliche Erklärung für das gemeinsame Auftreten und die Interaktion neurodegenerativer und psychischer Erkrankungen liefern.
Die spezifische Kombination von in silico Promotor und multivariaten Analysen im ersten Ansatz ermöglicht die Identifizierung der regulatorischen Mechanismen von Genen, die in multifaktoriellen Krankheiten beteiligt sind. Der zweite neue Ansatz, welcher bekannte Vorhersageprogramme und spezifische Auswahlkriterien in Bezug auf Alzheimer benutzt, bestimmt erfolgreich miRNA-mRNA Interaktionen und bietet somit Möglichkeiten für die Entwicklung von therapeutischen Behandlungsmöglichkeiten für Alzheimer.

Chapter 1

Introduction

1.1 Neurodegeneration

Neurodegenerative diseases lead to loss of neurons resulting in neuronal cell death. Oxidative stress and mutations in the mitochondrial DNA are incorporated in the process of aging, which is a risk factor for neurodegenerative diseases such as Alzheimer's, Parkinson's and Huntington's disease as well as Amyotrophic Lateral Sclerosis [Shukla et al., 2011]. Beside, depression co-occurs often with neurodegeneration and patients suffering from major depression develop often neurodegenerative diseases [Zunszain et al., 2011].

1.1.1 Alzheimer's disease

Alzheimer's disease (AD) is the most common form of dementia, which slowly destroys neurons and causes serious cognitive disability [Selkoe and Schenk, 2003]. Characteristics of AD are insoluble amyloid plaques and neurofibrillary tangles in the brains of AD patients, which extend progressively to neocortical brain areas during AD [Crews et al., 2010]. AD exists in a sporadic and familial (heritable) form. Mutations in amyloid-beta precursor protein (APP), presenilin 1 (PS1), and presenilin 2 (PS2) are associated with early-onset forms of familial AD, whereas sporadic AD occurs in people over the age of 65 years [Selkoe, 2001].
APP was the first gene linked to AD and is located on chromosome 21. APP is cleaved by different proteases named α-, β-, and γ-secretase (Figure 1.1). These proteases control the generation of the amyloid-β peptide (Aβ), which is considered the culprit in AD. β- and γ-secretase cleavage leads to Aβ formation. β-secretase is the aspartyl protease BACE1 [Cole and Vassar, 2008, Roßner et al., 2006]. A homolog of BACE1, BACE2, cleaves within the Aβ domain and does not contribute to Aβ generation. γ-secretase is a heterotetramer consisting of the four subunits PS1 or PS2, anterior pharynx defective 1 homolog A (APH1A), Nicastrin (NCSTN), and presenilin enhancer (PEN-2) [Steiner et al., 2008]. Aggregates of Aβ are neurotoxic and start the so-called amyloid cascade, which describes the molecular mechanisms leading to AD, including formation of plaques and tangles [Selkoe and Schenk, 2003]. The third protease, the alpha-secretase ADAM10 (a disintegrin and metalloproteinase 10)

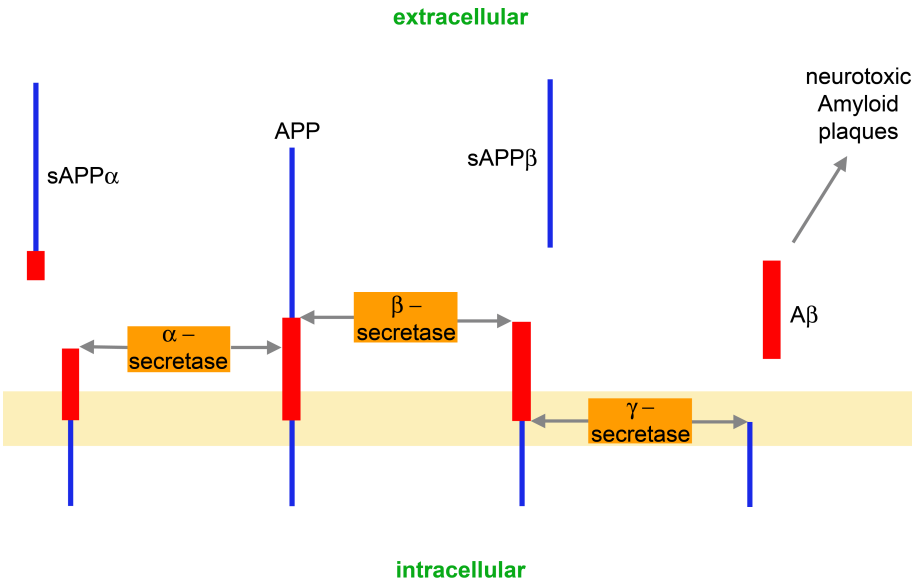


Figure 1.1: Proteolytic cleavage of APP by α-, β- and γ- secretase. The left side of the schematic representation shows the non-amyloidogenic pathway processing the soluble sAPPα fragment. Whereas the other side describes the amyloidogenic pathway resulting in insoluble and neurotoxic amyloid plaques.

[Lammich et al., 1999, Kuhn et al., 2010], avoids formation of Aβ, because it cleaves APP inside the Aβ sequence [Fahrenholz, 2007]. Additionally, α-secretase generates the soluble sAPPα, which enhances memory in normal and amnesic mice [Meziane et al., 1998]. In AD patients, the amount of sAPPα in the cerebrospinal fluid is reduced [Lannfelt et al., 1995].

1.1.2 Parkinson's disease

Parkinson's disease (PD) is the second most common neurodegeneration implicating the loss of nigrostriatal dopaminergic neurons [Braak et al., 2004]. Beside motor symptoms including resting tremor, bradykinesia and rigidity also non-motor symptoms such as depression, anxiety, olfactory and cognitive impairments occur before diagnosis, which are often not identified and poorly treated [Chaudhuri et al., 2006]. PD is classified into a sporadic and a familial form at which in the majority of cases sporadic PD is occurring [Davie, 2008]. Oxidative stress and mitochondrial dysfunction are involved in the pathogenesis of sporadic PD [Zhang et al., 2000] while several genes are known to be involved in the familial form of PD such as Parkin and DJ-1, which are autosomal recessive genes. In contrast, SNCA and MAPT are autosomal dominant PD genes and therefore risk factors for the pathogenesis of PD [Dawson et al., 2010, Wider and Wszolek, 2007].

Parkin protects cells from cell death induced by oxidative stress and is involved in the mitochondrial biogenesis especially in mitochondrial fission and fusion. DJ-1 is like Parkin neuroprotective preventing mitochondrial damage. The overexpression of autosomal recessive PD genes leads to protection of neuronal cells [Henchcliffe and Beal, 2008]. SNCA is found in Lewy bodies, which are characteristic of PD, and an overexpression of SNCA can lead to sporadic PD [Bogaerts et al., 2008]. MAPT was discovered in genome-wide association studies (GWAS) to be associated with PD [Edwards et al., 2010] and is involved in the microtubule network stabilisation [Garcia and Cleveland, 2001].
In addition to PD-related genes mitochondria dysfunction is strongly associated to PD. Usually, mitochondrial function plays a role in energy metabolism as well as in different cellular processes such as stress response and cell death pathways. Mutations in the mitochondrial DNA implicate an increase of cellular stress [Winklhofer and Haass, 2010]. Several mouse models with decreased mitochondrial DNA expression studied by Ekstrand et al. [2007] showed more cell death but no change in oxidative stress. Thus, mitochondrial dysfunction weakens the cellular energy supply, which possibly leads to cell death, and is associated to aging and neurodegeneration such as PD. Furthermore, several autosomal recessive as well as dominant PD genes are associated with mitochondria, which links sporadic and familial PD form [Winklhofer and Haass, 2010].

1.1.3 Stress and depression

Basically, stress and depression are not classified as neurodegenerative diseases, but their exists a strong association between the diseases. Neurodegeneration as well as inflammation seems to be increased in depressed brains, while neurogenesis is decreased. Inflammation causes an increase of oxidative stress leading to neurodegeneration and as a consequence to depression [Maes et al., 2009].
The symptoms of patients suffering from depression are manifold. Beside weight changes according to loss of appetite [Maxwell and Cole, 2009], sleep disturbances, fatigue as well as insomnia occur in the pathogenesis of depression [Franzen and Buysse, 2008]. Furthermore, the dysregulation of hypothalamic-pituitary-adrenocortical (HPA) axis is associated to this psychiatric disorder.

Normally, stress activates the hypothalamus (Figure 1.2) to secrete corticotropin-releasing hormone (CRH) leading to adrenocorticotropic hormone (ACTH) release from the pituitary gland. As a consequence the glucocorticoid cortisol is secreted from the adrenal cortex and decreases the levels of CRH and ACTH by a negative feedback loop [de Kloet et al., 2005, Herman and Cullinan, 1997]. In depression the HPA axis is dysregulated leading to a HPA axis hyperactivity by abnormal function

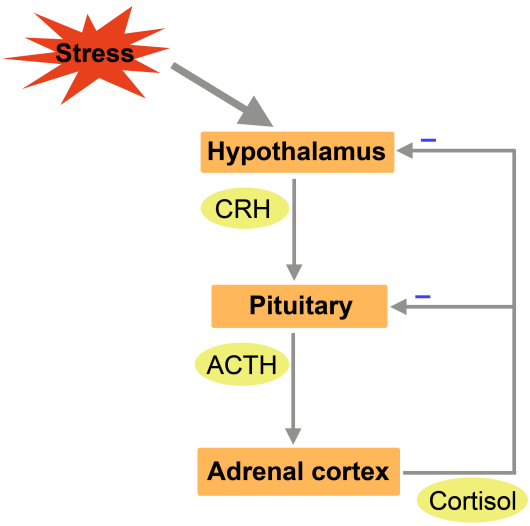


Figure 1.2: HPA-axis. The HPA-axis responses to stress by releasing CRH, ACTH and Cortisol, which decreases stress reaction by negative feedback.

of the glucocorticoid receptor (GR). The reason could be the impaired negative feedback loop leading to high levels of CRH and ACTH, but also increased inflammation is involved in depression reducing GR activity [Pariante and Lightman, 2008].

1.2 Regulatory mechanisms

Different regulatory mechanisms in the cell such as activators and repressors control metabolic processes and gene expression. Transcription factors and microRNAs bind to regulatory sites in the DNA sequence and are able to activate and repress genes. Additionally, SNPs observed in several GWASs alter DNA sequence and as a consequence can induce an increase or decrease in gene expression by modifying binding sites of activators or repressors.

1.2.1 Transcription factor binding sites

Transcription factors (TFs) bind predominantly to the upstream region of genes regulating transcription either directly or indirectly. Each TF has its own specific binding motif, whereas binding sites are 10 to 30 nucleotides long. Usually, the TF binding sites (TFBSs) are defined by position weight matrices established from several validated binding sites of the corresponding TF. The position weight matrix describes the preference of the different nucleotides at each position of the TF binding site [Werner et al., 2003].
TFs with a similar binding motif are grouped to a TF-family. Combinations of TFs in a defined order, distance range, and orientation are known as TFBSs modules

[Döhr et al., 2005]. Modules common to a set of genes, which act together in the same biological context, are able to control the expression of these gene products [Cohen et al., 2006]. Conversely, the finding that the expression of different genes is coregulated in a certain biological process may indicate that they are functionally linked in this process. This may be applicable to the identification of new disease-linked genes, for example, in neurodegeneration.
Several free available as well as commercial programs have been established for the prediction of TF binding sites. All TF binding site predictions in this thesis were performed by the software from Genomatix incorporating regularly updated TF weight matrices and a whole genome database of distinct species [Cartharius et al., 2005].

1.2.2 Single nucleotide polymorphisms

Single nucleotide polymorphism (SNP) is a variation in the DNA sequence of just a single nucleotide. About 90% of all polymorphisms in the human genome are SNPs [Brookes, 1999]. Most of the SNPs are located in non-coding regions and do not influence gene function. On average 6 SNPs are located in one gene, which are more likely to manipulate the role of a gene [Collins et al., 1998].
In contrast to synonymous SNPs whereby the amino acid is not changed, nonsynonymous SNPs alter the amino acid sequence and nonsense SNPs introduce a stop codon. In the genome nonsynonymous SNPs occur much more often than nonsense SNPs, leading to the assumption that introducing a stop codon is much more deleterious than altering one amino acid [Yamaguchi-Kabata et al., 2008].
SNPs, which cause a change in regulation, occur e.g. in TF binding sites implicating different consequences for TFs. Most probably the SNP does not affect the TF binding site, because a TF requires not 100% binding site sequence complementarity. However, in some cases increased or decreased binding, no binding as well as binding of a novel TF is possible, which alters gene expression and leads to a disease outbreak [Chorley et al., 2008].

1.2.3 MicroRNAs

MicroRNAs (miRNAs) are on average 22 nucleotides long and play a pivotal role in gene regulation. These small RNAs regulate the gene expression post-transcriptionally by suppression of mRNA translation, stimulation of mRNA deadenylation and degradation or induction of target mRNA cleavage, but have also the potential to activate translation [Chekulaeva and Filipowicz, 2009, Vasudevan and Steitz, 2007]. Over half of the mammalian protein coding-genes are regulated by miRNAs and most human mRNAs have binding sites for miRNAs [Friedman et al.,

2009]. The interaction of miRNA and target mRNA requires base pairing between the seed sequence (positions 2-8) of the miRNA at the 5' end and a sequence most frequently found in the 3' untranslated region (UTR) of the target mRNA [Fabian et al., 2010].

Two enzymes are essentially involved in the miRNA biogenesis (Figure 1.3) from

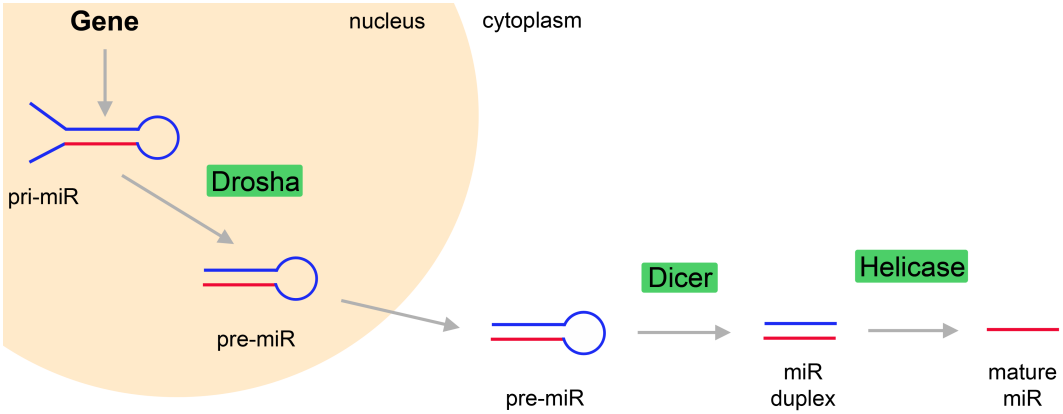


Figure 1.3: MicroRNA biogenesis by Drosha and Dicer.

the transcription to the mature miRNA. After transcription, the long primary miRNA (pri-miR) is cleaved by nuclear RNase III Drosha generating the about 70-nucleotide long pre-miRNA [Lee et al., 2003]. The pre-miRNA is transported from the nucleus to the cytoplasm in order to cut off the loop of the pre-miRNA by Dicer. The generated duplex is processed by helicase developing a single-stranded 22-nucleotide long mature miRNA [Bartel, 2004].

MiRNAs are involved in neuronal functions like neurite outgrowth and brain development [Hébert and De Strooper, 2009]. They were recently described to play a role in human neurodegenerative diseases [Satoh, 2010]. Changes in miRNA expression profiles or miRNA target sequences could contribute to the development of PD and AD [Hébert and De Strooper, 2009, Satoh, 2010].

1.3 TFBS promoter analysis (Genomatix)

TFBSs are defined by weight matrices with length of 5 - 29 bp. Similar binding site descriptions are grouped into one TF-family in order to reduce redundancy and minimize redundant matches. The TFBS matrices recorded in the matrix library originate from publications with validated binding sites or nucleotide distribution matrix passing the quality threshold. An essential property of the TFBS matrix is the core sequence defined by the highest conserved stretch of four nucleotides. Addi-

tionally, optimized matrix thresholds for each individual matrix are defined, since the matrices differ in length and conservation profile. The optimized matrix threshold is the matrix similarity threshold, which permits maximal three TFBSs of this matrix in a non-regulatory test sequence with the length of 10000 bp.
The MatInspector identifies TFBSs in DNA sequences by a three step algorithm. First, sequences matching the core positions of matrices are identified and the core similarity is calculated. If the computed core similarity reaches the threshold, the similarity to the whole matrix is checked and has to reach the optimized matrix threshold. Finally, overlapping matches of one family are filtered out for the matches with the highest score.
Several tools are based on MatInspector such as DiAlignTF and FrameWorker. DiAlignTF is a combination of MatInspector and DiAlgin, which establishes multiple sequence alignments. The identification of functional TFBSs is supported by multiple alignment considering conservation. TFBSs in the local alignment of evolutionary related sequences are highlighted.
After identification of single TFBSs the user can search by FrameWorker for functional units consisting of at least two TFBSs (modules). The program identifies shared TFBS patterns by multiple sequence alignment. The TFBSs of a module are arranged in a defined order, distance range and orientation. Finally, sequences of whole genomes can be scanned for TFBS modules defined by FrameWorker [Cartharius et al., 2005].

1.4 Microarray analysis

Microarray studies measure the changes in expression level for thousands of genes simultaneously, and therefore are an unbiased approach to identify genes with an altered expression, for example, in diseases such as AD [Blalock et al., 2004, Miller et al., 2010].
However, microarray experiments are expensive and the user has to be aware of noise incorporated in the measured expression values and less robust data analysis platforms [Iida and Nishimura, 2002, Verducci et al., 2006]. On the other hand microarrays are used in drug discovery and clinical applications such as developing biomarkers [Wang and Cheng, 2006]. Aims of the analysis of microarray datasets include the discovery of gene function [Pan, 2006], getting insights into human disease progression [Tenenbaum et al., 2008], and prediction of gene regulatory elements like TFBSs of coregulated genes [Park et al., 2002].
A microarray experiment can be divided into several steps. At the beginning, the biological problem is constituted as well as which biologically conclusions should be received from the data contributing to a successful microarray experiment. First,

the experimental design is established by deciding which array technology is used, whether single- or two-color arrays are appropriate as well as how many biological replicates of each group have to be measured. Next, data preparation is performed by visualizing and normalizing raw data in order to remove technical variation between arrays so that only biological variation is remaining. Finally, data is analysed by software such as R, which is free available and includes a huge amount of packages suitable for gene expression analysis. The main goal of the microarray analysis is to find single genes or gene sets, which are differentially regulated among groups of arrays, associated to a biological pathway or function [Slonim and Yanai, 2009].

1.5 MicroRNA target site prediction and databases

A vast majority of miRNA prediction programs and databases have been established up to date, but miRNA prediction is still a great challenge. Numerous available computational methods predict a large number of genes targeted by miRNAs regulating gene expression, but only few have been validated experimentally. Many computational predictions are false positives and therefore have to be filtered out [Watanabe et al., 2007].
MiRNA binding sites of targets are classified into 5'-dominant canonical, 5'-dominant seed only or 3'-compensatory target sites [Brennecke et al., 2005]. Each prediction program favours specific base pairing rules. Additionally, conservation varies between different prediction algorithms in the number and category of conserved species as well as in the length of conservation i.e. only seed or entire binding site conservation is required [Sethupathy et al., 2006]. The requirement of target-site conservation in different species including far related species is a potential way to reduce the false positive rate [Lewis et al., 2005].
Another considered feature of miRNA prediction programs is thermodynamic stability of the mRNA-miRNA duplex. There is no universally valid threshold, hence the prediction programs define different limiting values for accepting binding sites as well as various calculation programs are used leading to different free energy results [Min and Yoon, 2010]. Finally, comparing the output of various prediction programs the user should pay attention to different 3'UTR sequences used by distinct prediction algorithms [Sethupathy et al., 2006].

1.6 Bioinformatics analysis

Computational exploration such as enrichment as well as microarray analysis by bioinformatics programs facilitates research with an increasing amount of data. Different statistical and machine learning functions are applied in this thesis with the help of the R software for statistical computing.

1.6.1 Statistical methods

At the beginning of statistical analysis such as evaluation of microarray datasets often univariate methods are applied analysing large datasets. Univariate methods consider variables individually and not in relation to each other [Naduvilath and Dandona, 1998]. Various methods such as mean, p-value as well as foldchange computation are available for describing variables.
While the mean describes average expression level of a group, the foldchange displays the ratio of two group means and t-test is applied for calculating p-values in order to e.g. identify differentially expressed genes. Additionally, after p-value computation usually the values are corrected for multiple comparisons, which is also called p-value adjustment. Microarray datasets measure several thousands of probesets simultaneously, whereas the p-value is calculated for each probeset separately, hence false positive p-values increase with the number of probesets [Zhang and Cao, 2009, Reiner et al., 2003].
Multivariate methods are adapted considering at least two variables simultaneously [Naduvilath and Dandona, 1998]. Multivariate analysis such as contingency tables, support vector machines and clustering procedures are more complex than univariate methods.
Contingency tables (Figure 1.4), which consider two variables, are appropriate for enrichment analysis to calculate p-values according to Fisher's test. The value r reports the number of individuals/observations fulfilling both variables. R and n describe the quantity of individuals conform to variable B and A, respectively. N is the overall number of individuals analysed for both variables. After setting the four values the Fisher's test is applied computing a p-value. Finally, the p-value reports the significance of the number of individuals fulfilling both variables [Sahai and Khurshid, 1995].
Fisher's test can be used for large as well as small sample sizes and is categorized as exact test. The p-value is calculated by the formula of the hypergeometric distribution [Agresti, 2007].

	variable A	¬ variable A	
variable B	r	R - r	R
¬ variable B	n - r	N - R - n + r	N - R
	n	N - n	N

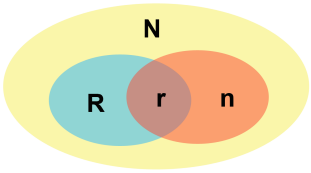


Figure 1.4: Contingency table considering two variables A and B as well as the proportion of the four values r, n, R and N in the contingency table. ¬ = not.

1.6.2 Support vector machine and recursive feature elimination

Support vector machines (SVMs) belong to the supervised learning methods and are applied in classification and regression analysis. SVM is a large margin classifier

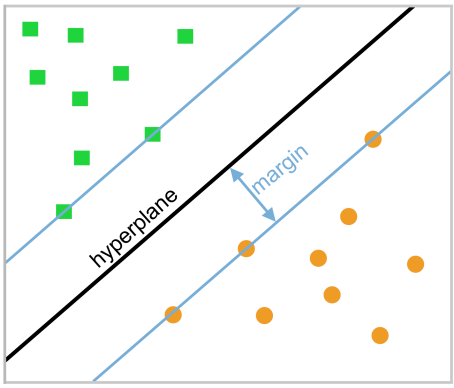


Figure 1.5: Support vector machine divides objects into two classes (orange and green) in order to get the broadest possible margin around the hyperplane, which is free of objects. Support vectors are located on the margin boundary (blue lines).

(Figure 1.5) dividing set of objects (i.e. biological replicates) into classes (i.e. disease, control) that around the class boundary (hyperplane) the broadest possible range remains, which is free of objects.
Additionally, nonseparable classes are analysed by soft margin SVM by adjusting a constant $C > 0$ (see section 2.1.3). Soft margin classifiers are suitable for noisy data such as microarray datasets because a minimal number of misclassifications is permitted [Domeniconi et al., 2005, Cortes and Vapnik, 1995]. If the value of C is small more misclassifications close to the hyperplane are allowed increasing the

margin. Whereas a larger value of C implicates a more accurate model permitting less errors and decreases the margin. Therefore SVM with a larger value of C tends to overfit [Ben-Hur et al., 2008].
SVMs are appropriate to handle high-dimensional data such as microarray datasets being accurate with a low error rate [Brown et al., 2000]. Mostly, more than two classes have to be analysed by SVM. This is easily done by dividing multi classes into multiple binary classes and analysing each by either one-versus-all (one class and the rest) or one-versus-one (every pair of classes) method. One-versus-all determines the class, whose classifier function has the highest value. In the one-versus-one method each binary classifier chooses one class, whose vote is increased by one, and in the end the class with most votes is assigned [Hsu and Lin, 2002].
Recursive Feature Elimination (RFE) in combination with SVM is a suitable tool to reduce large number of genes such as on microarrays to a applicable amount for further analysis (Figure 1.6). SVM-RFE procedure was first established by Guyon

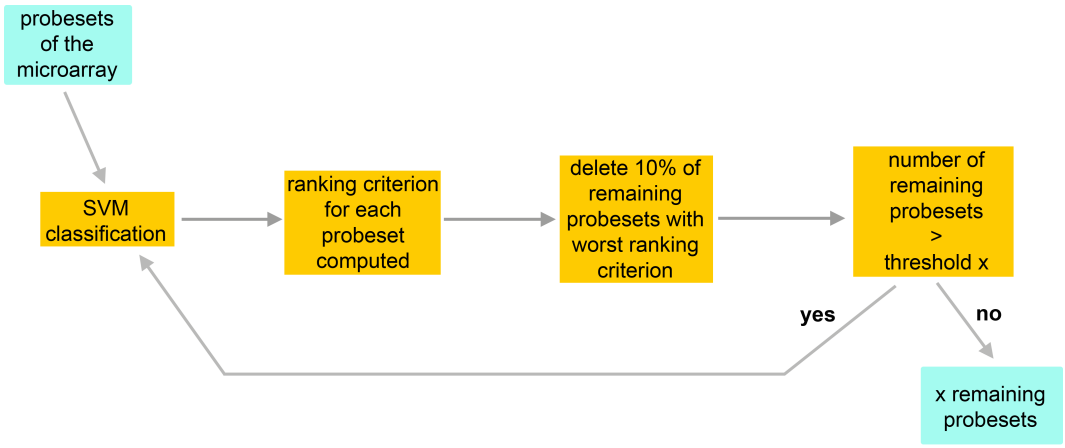


Figure 1.6: The recursive feature elimination reduces the number of microarray probesets. Probesets worst in discriminating between classes according to SVM are removed until a suitable number of probesets is reached.

et al. [2002]. The feature selection eliminates recursively in each step a certain amount of features, which are not good enough in differentiating between classes. SVM classification is applied to rank features. In each elimination step weight values for each feature e.g. probesets in the case of a microarray analysis are computed from coefficients and support vectors followed by calculation of the ranking criterion. Finally, 10 % of the remaining features with worst ranking criterion are deleted. This iterative procedure is repeated until a certain amount of features is remaining, which is suitable for further analysis [Guyon et al., 2002].

1.6.3 Biclustering

In general, cluster analysis belongs to unsupervised learning methods in data mining. The main goal of clustering is to group objects into clusters. In principal, objects belonging to the same cluster are more closely related than objects assigned to different clusters.
Clustering methods are discriminated by several various aspects. Clustering methods are divided into partitional and hierarchical procedures depending on whether objects are assigned to one non-overlapping cluster or clusters have subclusters, which are often illustrated as a tree. Another distinction of clusters is exclusive, overlapping and fuzzy, which indicates each object belongs to one cluster, objects can be assigned to more than one cluster and each object belongs to all clusters with assigned probability for membership, respectively. Furthermore, the clustering procedure is either complete or partial i.e. every object is assigned to a cluster or some objects are not grouped into clusters [Hastie et al., Tan et al., 2005].
Especially suitable for analysing gene expression data from microarray datasets are biclustering procedures. Thereby, simultaneous clustering of conditions and genes is executed to find subgroups of genes as well as conditions. In other words, biclustering is appropriate to identify coregulated genes under different conditions in microarray experiments. Interesting are biclusters with coherent values on both rows and columns, which are characterized by an additive model such as in the Plaid model [Madeira and Oliveira, 2004]. An example is shown in Figure 1.7. Coherent values in

1	8	4	5
6	13	9	10
3	10	6	7
7	14	10	11

1+0+0	1+0+7	1+0+3	1+0+4
1+5+0	1+5+7	1+5+3	1+5+4
1+2+0	1+2+7	1+2+3	1+2+4
1+6+0	1+6+7	1+6+3	1+6+4

Figure 1.7: Example of an additive bicluster with coherent values. Beside the bicluster the derivation of the coherent values is shown on the right side.

a bicluster are obtained by considering a typical value of the bicluster as well as row and column offset [Kerr et al., 2008]. The method Plaid performs two-sided clustering and considers also overlapping biclusters [Lazzeroni and Owen, 2000]. Allowing overlapping biclusters (Figure 1.8) is necessary for gene expression datasets, because genes have usually several biological functions incorporated in more than one pathway [Kerr et al., 2008].

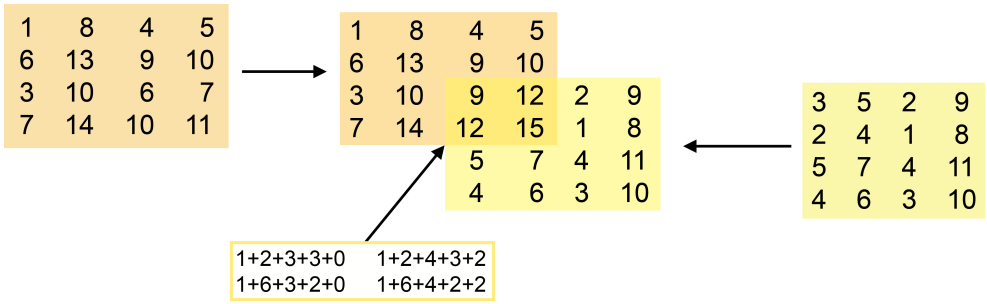


Figure 1.8: Two additive overlapping biclusters are illustrated in the middle of the figure. The overlapping values 9, 12, 12 and 15 are derived by combining the corresponding values of the two single biclusters like shown in the box below.

1.7 Aim of the thesis

The aim of the work is the identification of regulatory molecular networks in neurodegeneration. Different computational workflows are developed and evaluated. Finally, common regulatory mechanisms and disease related genes of neurodegenerative diseases and depression are discovered.

For the identification of regulatory molecular mechanisms in AD two approaches are developed to identify TFBSs modules and predict miRNA target sites. The purpose is to identify possibly regulatory TFBSs modules in AD related genes as well as key genes and miRNAs regulating ADAM10 expression.

The first approach identifies modules of TFBSs in the promoters of coregulated AD-related genes as well as in AD key genes. I hypothesized that the genes from β- and γ-secretase complex, which are responsible for Aβ formation, are coregulated. Thus, I started by analyzing TF binding modules in the genes for β- and γ- secretase by an in silico promoter approach. I included in the workflow three already existing microarray datasets, which were established under different conditions and in summary contain over 400 arrays, for analysis by using state-of-the-art bioinformatics tools focussing on multivariate methods. Multivariate variable selection was performed because variables (transcripts) contribute only in combination with other variables to the discrimination of input dataset rather than in isolation, which help to identify highly correlated genes (i.e., interaction networks) [Trümbach et al., 2010, Trevino and Falciani, 2006].

I established a second approach to identify miRNAs regulating ADAM10 expression which therefore might influence the progression of AD. The three programs RNA22, RNAhybrid and miRanda predict potential miRNA binding sites to ADAM10 3'UTR. I sought to identify the most interesting miRNAs possibly binding to ADAM10 3'UTR with additional selection criteria in particular whether they play

a role in AD. Additionally, the most interesting miRNAs were experimentally verified by a luciferase assay. The results show that miR-103, miR-107 and miR-1306 influence the expression of ADAM10. These miRNAs could play a role in AD and therefore are interesting candidates to be further analysed in relation to AD and could become potential therapeutic targets.

Moreover, TFBSs in PD associated genes such as Parkin, OPA1, MAPT, SNCA and AMPK are predicted possibly influencing PD pathogenesis. Additional information concerning the function of the TFs as well as of the target genes in PD and species conservation helps to limit the number of predicted TFBSs in the promoter sequences.

An implemented mSVM-RFE method is applied on PD and depression microarray datasets. For the purpose of evaluating the implemented method but also to identify PD and depression related genes. Finally, common mechanisms, genes and pathways of both neurodegenerative diseases and depression are revealed and discussed.

In the end, the validated and predicted interactions of neurodegenerative disease key genes and miRNAs as well as TFs of this thesis are summarized in a network to provide a model for the connectivity of the diseases AD, PD as well as stress and depression.

Chapter 2

Material and methods

2.1 Bioinformatics identification of modules of transcription factor binding sites

2.1.1 In silico promoter analysis

Promoter analysis was done with Genomatix software (Munich/Germany). All promoter sequences are derived from the promoter sequence retrieval database ElDorado (Release 4.9, Human Genome NCBI build 37/hg19). The DiAlignTF task of GEMS Launcher was used to check for conserved TFBSs, which are placed at the same position in the alignment, between the human and mouse APP promoter sequence (Matrix Family Library, Version 8.0, Vertebrates; Genomatix: 690 matrices from 162 families). The FrameWorker tool (GEMS Launcher) searches for all modules composed of two or more TFBSs in aligned promoter sequences. A module is defined as a set of two or more TFBSs with a defined order, distance range between the individual TFBSs, and strand orientation. A total of 727 matrices from 170 families (Matrix Family Library, Version 8.2, Vertebrates; Genomatix) were used for the analysis. The ModelInspector searches for all determined modules of TFBSs in the human promoter library (first approach: ElDorado 07–2009: 93372 promoter regions; second approach: ElDorado 02–2010: 97259 promoter regions; Genomatix Promoter Database).

2.1.2 Microarray datasets

I used three microarray datasets downloaded from the Gene Expression Omnibus [Edgar et al., 2002] in this study. The dataset of Blalock et al. [2004] consists of hippocampal probes: 9 controls and 22 AD patients with different severity (GSE1297). Gene expression was measured using GPL96: Affymetrix Human Genome U133A Array (`http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL96`) covering 22283 probesets.
The second dataset used in our analysis consists of total RNA of brains from five-month-old double-transgenic (6 ADAM10/APP, 6 dnADAM10/APP, 6 mono-transgenic APP control) mice (GSE10908) from Prinzen et al. [2009]. Gene ex-

pression was measured using GPL1261: Affymetrix Mouse Genome 430 2.0 Array (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL1261) covering 45101 probesets.
The third dataset from Webster et al. [2009] consists of human cortical samples: 187 controls and 176 patients with diagnosis of late onset AD (LOAD) (GSE15222). Gene expression was measured using GPL2700: Sentrix HumanRef-8 Expression BeadChip (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL2700) covering 24354 probesets.

2.1.3 Multivariate analysis

Statistical analysis was performed with R statistical software (R version 2.8.0, http://www.rproject.org/). Background correction and normalization of the microarray datasets (GSE1297, GSE10908) was done with the R function expresso from the R package affy. The parameter setting was as follows: bgcorrect.method (background adjustment method) = "mas", normalize.method (normalization method) = "quantiles", pmcorrect.method (perfect matches and mismatches adjustment) = "mas", and summary.method (computation of expression values) = "mas". The dataset GSE15222 is already rank-invariant normalized. I applied multiclass support vector machines with recursive feature elimination (mSVM-RFE). A SVM considers a set of objects (e.g. biological replicates) as classes, so that around the class boundaries the broadest possible range remains, which is free of data points. I used the svm function from the e1071 package in R for SVM prediction and developed an algorithm for multiclass gene selection with recursive feature elimination according to Zhou and Tuck [2007] and Guyon et al. [2002]. I implemented our own mSVM-RFE method, as described in the following, because such a specific combination of mSVM and RFE is not available in R until now. First, the samples of a microarray dataset are randomly grouped (drawing without replacement) into stratified four folds, which are four equally sized folds (except the number of samples is not a multiple of the number of folds, then a variation of one exists), such that each class is uniformly distributed among the four folds, and all combinations of three folds are used for mSVM-RFE (four combinations for mSVM-RFE). Stratified cross validation has smaller bias and variance than regular cross validation [Kohavi]. mSVM-RFE starts with all the features of a microarray, in our case gene expression values, and recursively eliminates 10% of the remaining expression values, which are not good enough for classifying according to the cost function of the SVM classifier (based on the coefficients and support vectors), until a given number of expression values are reached. By starting mSVM-RFE, this given number (stop condition for iterations) is turned over to the program as parameter and in our case this parameter, is 400

in the first part and 1000 in the second part of the workflow. This grouping into folds was done three times (to obtain stable results), and mSVM-RFE algorithm was applied twelve times on different subsets of the original dataset (in total, I did three groupings into four folds with four times mSVM-RFE per grouping, because of four combinations of three folds per grouping). I got twelve different gene selections and computed the frequency of each gene occurring in all the gene selections to identify the most important genes. The mSVM-RFE output in the first part of the workflow is restricted to genes occurring at least in 2 out of 12 gene selections, and in the second part of the workflow genes occurring at least in 1 out of 12 gene selections are taken. The gene selection in the second part is less stringent than in the first part of the workflow, since a rigorous restriction of the number of input genes for the subsequent biclustering analysis reduces the number of genes in the resulting clusters. The function svm was used with default settings except the parameters type = "C-classification", kernel = "linear", cost = 0.1. The cost parameter also called 'C' constant displays the cost if misclassifications occur. The default is one and the smaller the value the more misclassifications are allowed.
The dataset GSE15222 was not filtered by mSVM-RFE but by the illumina detection score [Illumina, 2008]. Transcripts that have a detection score ≥ 0.99 in less than 90% of cases or 90% of controls are excluded, and 8650 probesets remain [Webster et al., 2009]. Furthermore, a two-sample t-test was performed for the expression values of the remaining 8650 probesets by the R function t-test with default settings, and afterwards FDR (false discovery rate) correction [Benjamini and Hochberg, 1995] was applied by the R function p.adjust with method = "fdr". Significantly regulated genes were considered if the FDR value is equal to or below 0.05.
Pearson's chi-squared tests were performed with the R function chisq.test from the R package stats to show whether the overlap between two different genesets is significant. The function chisq.test was used with default settings.
Next, I applied biclustering by the biclust function from the biclust package in R to the output of the mSVM-RFE and the 8650 probesets of the third microarray study and used the method BCPlaid according to Turner et al. [2005] to group coregulated genes into clusters. The method allows a gene to belong to more than one cluster, and each cluster is defined with regard to some, but not necessarily all, samples. In principle, I used default parameters except the following ones. Parameter setting for AD patients: cluster = "r" (to cluster rows (probesets)), row.release (threshold to prune rows in the clusters depending on row homogeneity) = 0.1, col.release (as before, with columns) = 0.2, shuffle (before cluster is added, its statistical significance is compared against random clusters defined by this parameter) = 10, back.fit (after a cluster is added, additional iterations can be done to refine the fitting of the

cluster) = 10, max.layers (maximum number of clusters) = 10, iter.startup (number of iterations to find starting values) = 80, and iter.layer (number of iterations to find each cluster) = 80. Parameter setting for double transgenic mice: cluster = "r", row.release = 0.3, col.release = 0.5, shuffle = 100, back.fit = 500, max.layers = 100, iter.startup = 1000, and iter.layer = 1000. Parameter setting for LOAD patients: cluster = "r", row.release = 0.3, col.release = 0.5, shuffle = 10, back.fit = 100, max.layers = 20, iter.startup = 100, and iter.layer = 100.
The expression profiles were established with the R function matplot from the R package graphics. The function was used with default values except col (color of lines in plot) = c(1), type (type of plot) = "l", and lty (line type) = "solid". The colored lines of the described genes are added to the plot by matplot function with the parameter add = TRUE (if TRUE, plots are added to current one), and the width of the colored lines is enlarged by the parameter lwd = 4.

2.1.4 Enrichment analysis

Each cluster of coregulated genes was explored for enrichment of genes in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (version 6.07.2010: 210 pathways and 5368 genes(human); 206 pathways and 6060 genes(mouse)). After retrieving the number of coregulated genes in each pathway, the p-value was computed by the R function fisher.test, and afterwards FDR correction [Benjamini and Hochberg, 1995] was applied by the R function p.adjust with method = "fdr". For the pathway analysis, I report the p-value and FDR value. Additionally, I analyzed the enrichment of the modules in the corresponding cluster of coregulated genes in contrast to the whole set of human promoters. After searching for each module in human promoters, once again the p-value computation was performed, and afterwards FDR correction [Benjamini and Hochberg, 1995] was applied by the R function p.adjust with method = "fdr". The results are indicated to be significant if the FDR value is equal to or below 0.05.

2.1.5 Literature mining

Literature search by PubMed was done to extract information about the target genes of the TF modules identified and their relation to AD. Target genes are genes which are regulated by TF modules [Cartharius et al., 2005]. To verify the modules, searches were performed for TFBS-target gene interactions in all PubMed abstracts with the help of two text mining programs Pathway Studio 7.1 (Ariadne Genomics) and EXCERBT (MIPS, Helmholtz Zentrum München; `http://tinyurl.com/excerbt/`) [Barnickel et al., 2009] based on the natural language processing (NLP) technology. Additionally, information about the TFs of the modules was collected from the

BIOBASE Biological Databases (Wolfenbüttel/Germany): TRANSPATH. Mouse Genome Informatics (MGI) database (Mouse Genome Database, The Jackson Laboratory, Bar Harbor, Maine; `http://www.informatics.jax.org/`) [Bult et al., 2008] was searched for expression tissue of genes, and gene ontology (GO) was used for the functional annotation and classification of the target genes [Ashburner et al., 2000].

2.2 MSVM-RFE and biclust application

2.2.1 Depression: high and low reactivity

High and low reactivity mice microarray

The dataset established from the mouse model of Touma et al. [2008] consists of hippocampal probes of adult mice: seven high reactivity (HR) and nine low reactivity (LR) mice. Gene expression was measured using an Illumina gene expression beadchip array (type MouseWG-6_V2_0_R2_11278593_A; Illumina, San Diego, CA, USA). HR mice with a hyperreactivity of the HPA axis show symptoms of melancholic depression and atypical depressive indications were observed in LR mice with HPA axis hyporeactivity [Touma et al., 2008]. Dataset provided by Regina Widner (MPI Munich).

MSVM-RFE and biclust parameters

I applied mSVM-RFE to the whole HR LR microarray dataset. For detailed description of mSVM-RFE see section 2.1.3. The stop condition for iterations is set to 1000 and the grouping into folds was done ten times. While four mSVM-RFE runs per iteration are done, in total 40 mSVM-RFE runs were performed resulting in 40 different gene selections in the end. Genes occurring at least in 1 out of 40 gene selections are taken for further biclustering analysis. The function svm was used with default settings except the parameters type = "C-classification", kernel = "linear", cost = 0.1.
Biclustering was applied by the biclust function from the biclust package in R to the output of the mSVM-RFE (for more detail see section 2.1.3). In principle, default parameters were used except the following ones: method = BCPlaid() cluster = "r", row.release = 0.4, col.release = 0.5, shuffle = 10, back.fit = 10, max.layers = 30, iter.startup = 10, and iter.layer = 40.
The method mSVM-RFE was evaluated by sensitivity and specificity [Lalkhen and McCluskey, 2008]:

$$sensitivity = \frac{true\ positives}{true\ positives + false\ negatives} \qquad (1)$$

$$specificity = \frac{true\ negatives}{true\ negatives + false\ positives} \quad (2)$$

ROC curve was established by calculating sensitivity and specificity for several mSVM-RFE frequencies (0, 2, 4, 5, 19, 23, 37, 38, 40).

Serial analysis of gene expression

In contrast to the microarray analysis an alternative transcriptome analysis was performed the serial analysis of gene expression (SAGE). The same two RNA-pools from HR and LR mice as in case of the microarray analysis were starting point of the expressional profiling. The library was established according to the standard protocol of the Applied Biosystems SOLiD 3 System kit (Invitrogen, Paisley, UK). The preparation involves RNA binding to beads, cDNA synthesis, digestion steps as well as adapter ligation and generates tags of 60 bp (basepairs) composed of 33 bp adapter sequence and 27 bp of a specific sequence unit from a RNA transcript. The SOLiD 3 Systems Templated Bead Preparation full-scale protocol was used for the preparation of template beads. SAGE tag templates, which were extracted from the gel running with upscale PCR products and purified, were used in the emulsion PCR to amplify the tag library onto beads resulting in a clonal bead population. Slides of 41 million beads per pool were sequenced accordingly the Applied Biosystems SOLiD Analysis Tool v3.5 (Applied Biosystems, Carlsbad, CA, USA). The statistical analysis of the SAGE experiment comprises a quality control step, the identification of the adaptor sequence and the restriction enzyme binding site as well as the adaptor sequence cutting off. The alignments were computed by the BWA (Burrows-Wheeler Alignment) algorithm [Li and Durbin, 2009] and mapped to a reduced reference sequence database. Finally, for each gene symbol the reads were summed up and Z-scores (logarithm of the HR vs. LR counts) were calculated. Z-score ≥ 2 indicates a significantly differential expression. Dataset provided by Regina Widner (MPI Munich).

Quantitative real-time PCR

Quantitative real-time PCR experiments were done by Regina Widner (MPI Munich). Quantitative real-time PCR is a kinetic analysis method allowing the simultaneous amplification and measurement of PCR products during each annealing/extension phase [Higuchi et al., 1993]. The fluorescent reporter signal used in quantitative real-time PCR (qPCR) was Sybr Green. Beta-2 microglobulin served as housekeeping gene for standardization of the fluorescent signal of a target gene [Becker et al., 2004]. The analysis of the qPCR runs was performed by the absolute

quantification point function of the LightCycler Software 4.05 and the LightCycler 2.0 instrument (Roche Diagnostic, Mannheim). The $2^{-\Delta\Delta CT}$ algorithm transformed the crossing points into fold changes [Livak and Schmittgen, 2001]. Intron-spanning qPCR primers were designed for the amplification of cDNA only. For non-parametric statistics the Mann-Whitney-U test by the SPSS software (version 16.0) was performed.

Enrichment analysis and literature mining

GO [Ashburner et al., 2000] analysis and literature mining for gene interactions in all PubMed abstracts were performed by the software Pathway Studio 8.0 (Ariadne Genomics) based on database ResNet 8.0.
The enrichment of clustered genes and genes occurring in all gene selections after mSVM-RFE in KEGG pathways (version 7.4.2011: 226 pathways and 6662 genes (mouse)) was performed according to section 2.1.4.

2.2.2 Parkinson's disease: DJ-1 knockout

DJ-1 microarray

The dataset from Ulrich Hafen (Institute of Developmental Genetics; HMGU Munich) consists of microglial probes of adult mice: five DJ-1 knock-out mice and 5 wildtyp mice. All samples were stimulated with lipopolysaccharide (LPS) for 6 hours. Gen expression was measured using Affymetrix Mouse Gene ST 1.0 arrays covering 28853 probesets.

Statistical analysis

The statistical analysis was performed by the R package limma. First a linear model was fitted to the expression data for each probeset by the function lmFit with default settings. Next, empirical bayes statistics for differential expression on the microarray dataset was computed by the function eBayes with default settings. Afterwards Benjamini & Hochberg correction [Benjamini and Hochberg, 1995] was applied by the R function topTable with adjust="BH".

MSVM-RFE and biclust parameters

MSVM-RFE was applied to the whole gene expression dataset of DJ-1 ko and wild-type mice stimulated with LPS. For detailed description of mSVM-RFE see section 2.1.3. The stop condition for iterations is set to 500 and the grouping into five folds was done ten times. While five mSVM-RFE runs per iteration are done, in total

50 mSVM-RFE runs were performed resulting in 50 different gene selections in the end. Genes occurring at least in 1 out of 50 gene selections are taken for further biclustering analysis. The function svm was used with default settings except the parameters type = "C-classification", kernel = "linear", cost = 0.1.
The output of the mSVM-RFE was used for biclustering by the biclust function from the biclust package in R (for more detail see section 2.1.3). Default parameters were used except the following ones: method = BCPlaid() cluster = "r", row.release = 0.4, col.release = 0.5, shuffle = 10, back.fit = 10, max.layers = 30, iter.startup = 100, and iter.layer = 100.

Enrichment analysis and literature mining

GO [Ashburner et al., 2000] analysis and literature mining for gene interactions in all PubMed abstracts were performed by the software Pathway Studio 8.0 (Ariadne Genomics) based on database ResNet 8.0. Shortest path analysis of differentially expressed genes by the Pathway Studio was performed to get indirect interactions of regulated genes. The resulting network contains also genes, which are not regulated on the array but needed to connect differentially expressed genes of the microarray analysis.
The enrichment of clustered genes and genes occurring in all gene selections after mSVM-RFE in KEGG pathways (version 7.4.2011: 226 pathways and 6662 genes (mouse)) was performed according to section 2.1.4.
PDGene (Version 21.06.2011) database (`http://www.pdgene.org/`) [Lill et al., 2012], which is an aggregation of all published genetic association studies, was used to extract PD relevant genes from the output of the mSVM-RFE and biclustering analysis. Additional PD related genes were extracted from the PDbase database (`http://bioportal.kobic.re.kr/PDbase/index.jsp`) based on expression studies of the substantia nigra [Yang et al., 2009].
The collection of NFKB target genes (`http://www.bu.edu/nf-kb/gene-resources/target-genes/`) by the Biology Department of the Boston University was searched for mSVM-RFE and biclustering output genes.

2.3 TFBS in key genes of Parkinson's disease

2.3.1 TFBS prediction in Parkin and OPA1

Bioinformatics prediction of CREB/ATF sites in Parkin promoter and of NFKB sites in OPA1 promoter. All sequences of Parkin and OPA1 are derived from the promoter sequence retrieval database ElDorado 02-2010 and 08-2011 (Genomatix, Munich/Germany), respectively. Promoter sequences of Parkin from four different

mammalian species and of OPA1 from six different mammalian species were aligned with the DiAlignTF program [Cartharius et al., 2005] in the Genomatix software suite GEMS Launcher to evaluate overall promoter similarity and to identify conserved CREB/ATF binding sites (BSs) in Parkin and conserved NFKB BSs in OPA1. The promoter sequences of Parkin were defined as in ElDorado and elongated at the 3' end of the promoter (downstream) by 150 base pairs. The promoter sequences of OPA1 were defined as in ElDorado. For promoter analyses position weight and matrices of CREB and NFKB were used according to Matrix Family Library Version 8.2 (January 2010) and Version 8.4 (June 2011), respectively. BSs were considered as "conserved BSs" if the promoter sequences of human and the orthologs can be aligned in the region of CREB/ATF or NFKB BSs with the help of the DiAlignTF program (using default settings).

2.3.2 MAPT and SNCA - SNPs in TFBSs

All SNPs of MAPT and SNCA are extracted from PDGene (Version 19.01.2011) database (`http://www.pdgene.org/`) [Lill et al., 2012]. MAPT and SNCA are the two top ranked genes in the PDGene database. The amount of SNPs is restricted to such ones lying in the promoter or first intron.
All sequences of MAPT and SNCA are derived from promoter sequence retrieval database ElDorado 12-2010. Promoter and first intron sequences (defined as in ElDorado) of MAPT and SNCA from in each case five different mammalian species were analysed for conserved TFBSs by DiAlignTF (Matrix Family Library Version 8.3 (October 2010)). BSs were defined as conserved if the sequences of human and orthologs can be aligned in the region of TFBSs by DiAlignTF program (using default settings).
Furthermore, TRANSPATH (BIOBASE Biological Databases) was searched for information about the TFs.

2.3.3 TFBSs prediction in PRKAA1, PRKAA2 and PRKAG1

At first, specific TF-families for promoter analysis were determined by MatBase (Genomatix, Munich/Germany) according to GO search terms containing "stress", "metabolic" and "biosynthetic".
Promoter sequences of PRKAA1, PRKAA2 and PRKAG1 from twelve, nine and nine different mammalian species, respectively, are derived from promoter sequence retrieval database ElDorado 8-2011 (promoter defined as in ElDorado). Human promoter sequences of PRKAA1, PRKAA2 and PRKAG1 were analysed by MatInspector (default settings) for BSs of the determined TF-families according to Matrix Family Library Version 8.4 (June 2011). DiAlignTF (default settings) was used for

the identification of conserved TFBSs, which potentially bind to a region perfectly aligned between human and orthologs, of the determined TF-families in orthologous promoters of PRKAA1, PRKAA2 and PRKAG1.

2.4 MicroRNAs binding to ADAM10

2.4.1 MiRNA target site prediction databases

MiRNA binding sites to target genes were downloaded from seven different databases: miRBase, 5-Nov-2007, `http://www.mirbase.org/` [Griffiths-Jones et al., 2008]; microRNA, September 2008 Release, `http://www.microrna.org/microrna/home.do` [Betel et al., 2008]; PicTar via UCSC Table Browser, assembly = May 2004 (NCBI35/hg17), group = Regulation, track = PicTar miRNA, `http://genome.ucsc.edu/` [Krek et al., 2005]; PITA, version 6 (31-Aug-2008), `http://genie.weizmann.ac.il/pubs/mir07/index.html` [Kertesz et al., 2007]; RNA22, March 2007, `http://cbcsrv.watson.ibm.com/rna22.html` [Miranda et al., 2006]; TarBase, June 2008, `http://diana.cslab.ece.ntua.gr/tarbase/` [Papadopoulos et al., 2009]; TargetScan, Release 5, `http://www.targetscan.org/` [Friedman et al., 2009]. I established a workflow considering all miRNA target site predictions downloaded.

2.4.2 MiRNA target prediction

I used three prediction programs RNA22, RNAhybrid, miRanda and predicted all binding sites of the miRNA sequences to the 3'UTR sequence of human ADAM10.
RNA22 is a pattern-based method for the identification of miRNA-target sites. The method has high sensitivity, is resilient to noise, can be applied to the analysis of any genome without requiring genome-specific retraining and does not rely upon cross-species conservation. Focusing on novel features of miRNA-mRNA interaction RNA22 first finds putative miRNA binding sites in the sequence of interest then identifies the targeting miRNA and hence allows to identify sites targeted by yet undiscovered miRNAs. An implementation of RNA22 (19-May-2008) is available online at `http://cbcsrv.watson.ibm.com/rna22.html` [Miranda et al., 2006, Witkos et al., 2011].
The second program RNAhybrid is an extension of the classical RNA secondary structure prediction algorithm from Zuker and Stiegler [1981]. It finds the energetically most favorable hybridization sites of a small RNA in a large RNA incorporating 'seed-match speed-up', which first searches for seed matches in the candidate targets and only upon finding such matches the complete hybridization around the seed-match is calculated. The user can define the position and length of the seed region with the option to allow for G:U wobble base pairs in the seed pair-

ing. Intramolecular base pairings and branching structures are forbidden and statistical significance of predicted targets is assessed with an extreme value statistics of length normalized minimum free energies, a Poisson approximation of multiple binding sites, and the calculation of effective numbers of orthologous targets in comparative studies of multiple organisms. RNAhybrid, Version 2.1, is available online at `http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/` [Rehmsmeier et al., 2004, Krüger and Rehmsmeier, 2006].
The miRanda algorithm is similar to the Smith-Waterman algorithm, but scores based on the complementarity of nucleotides (A=U or G≡C) and one G:U wobble pair is allowed in the seed region but has to be compensated by matches in the 3' end of miRNA. In order to estimate the thermodynamic properties of a predicted pairing between miRNA and 3'UTR sequence, the algorithm uses folding routines from the Vienna 1.3 RNA secondary structure programming library (RNAlib) [Wuchty et al., 1999]. A conservation filter is used and optionally some rudimentary statistics about each target site can be generated. MiRanda, September 2008 Release, is available online at `http://www.microrna.org/microrna/home.do` [Witkos et al., 2011, Enright et al., 2003].
The parameter setting for RNA22 is: maximum number of "UN-paired" bases within the extent of the seed = 0, extent of seed in nucleotides = 6, minimum number of paired-up bases that you want to see in any reported heteroduplex = 14, maximum value for the folding energy in any reported heteroduplex = -25 kcal/mol. The parameter setting for RNAhybrid is: "-s 3utr_human" ("-s" tells RNAhybrid to quickly estimate statistical parameters from "minimal duplex energies" under the assumption that the target sequences are human 3'UTR sequences). The parameter setting for miRanda is the default parameter setting: gap open penalty = -8, gap extend = -2, score threshold = 50, energy threshold = -20 kcal/mol, scaling parameter = 4.
I retrieved the 3'UTR sequence of ADAM10 (human ADAM10 3'UTR based on transcript NM_001110 (chr15:58888510-58889745)) from NCBI `http://www.ncbi.nlm.nih.gov/`. I downloaded 703 mature miRNA sequences for Homo sapiens from miRBase, version 13.0 `http://www.mirbase.org/` [Griffiths-Jones et al., 2008].

2.4.3 Extraction of best miRNA predictions

The extraction of miRNAs was applied according to the following selection criteria. I checked for each miRNA how many programs predicted the miRNA to bind to human ADAM10 3'UTR. The regulation of miRNAs in AD was verified by the publication of Cogswell et al. [2008], which provides a list of miRNAs expressed in the tissues hippocampus, cerebellum and medial frontal gyrus. Another possibility to check the expression of miRNAs in the brain is the Mouse Genome Infor-

matics (MGI) database (Mouse Genome Database, The Jackson Laboratory, Bar Harbor, Maine; `http://www.informatics.jax.org/`) [Bult et al., 2008]. Literature search by PubMed was done as an additional approval, to search for already described target genes of the miRNAs, especially for target genes involved in AD. Mouse ADAM10 3'UTR based on transcript NM_007399 (chr9:70625902-70628036) from NCBI `http://www.ncbi.nlm.nih.gov/` was used for binding site search of mouse miRNAs from miRBase, version 13.0 `http://www.mirbase.org/` [Griffiths-Jones et al., 2008]. The parameter setting for RNA22 and miRanda is the same as for human miRNA binding site prediction at the human ADAM10 3'UTR. The parameter setting for RNAhybrid is "-d 1.9,0.28" (1.9 is the location parameter and 0.28 the shape parameter of the assumed extreme value distribution). Additionally, I searched by TargetScan database `http://www.targetscan.org/` [Friedman et al., 2009] and microRNA database `http://www.microrna.org/microrna/home.do` [Betel et al., 2008] for miRNAs binding to human ADAM10 3'UTR and compared the TargetScan and microRNA predictions to our list of miRNAs for equal miRNAs. ADAM10 3'UTR sequences from ten different species were analysed for conserved regions. The following sequences where taken: human ADAM10 3'UTR from transcript NM_001110 (chr15:58888510-58889745), mouse ADAM10 3'UTR from transcript NM_007399 (chr9:70625902-70628036), horse ADAM10 3'UTR from transcript XM_001498169.1 (chr1:132875124-132876868), dog ADAM10 3'UTR from transcript XM_858910 (chr30:26596273-26598436), chimp ADAM10 3'UTR from transcript XM_001172393.1 (chr15:55942343-55944774), chicken ADAM10 3'UTR from transcript ENSGALT00000034458 (chr10:7949768-7951846), rhesus monkey ADAM10 3'UTR from transcript XM_001096908 (chr7:36929437-36932008), zebra fish ADAM10 3'UTR from transcript NM_001159314 (chr7:31745579-31747655), opossum ADAM10 3'UTR from transcript ENSMODT00000011088 (chr1:162230000-162230183), zebra finch ADAM10 3'UTR from transcript XR_054746 (chr10:6638729-6639273). For multiple sequence alignment of the ten ADAM10 3'UTR sequences I applied ClustalW Version 2.1 from the European Bioinformatics Institute (EBI) `http://www.ebi.ac.uk/` [Larkin et al., 2007, Goujon et al., 2010]. I used default parameters except: DNA Weight Matrix = 'ClustalW', Clustering = 'UPGMA'. After extraction of the conserved regions between at least seven species I looked for miRNA binding sites localized in these conserved regions. Additionally, I determined the conservation (given in percentage) of the miRNA binding site sequence from human to each species.

2.4.4 Statistical analysis

Statistical analysis was performed with R statistical software (R 2.8.0, `http://www.r-project.org/`). The p-value was computed by the R function fisher.test with default settings. The Fisher's exact test is used to examine the significance of the association (contingency) between the two kinds of classification. Significantly regulated genes were considered, if the p-value is equal or below 0.05. I generated Venn diagrams to see the overlap between target genes of miR-103 and miR-107 common in 4 out of 6 databases as well as genes in the AlzGene database (`http://www.alzgene.org/`; Version: 20.06.2011) [Bertram et al., 2007]. Each set of target genes of miR-103 and miR-107 common in 4 out of 6 databases as well as the set of target genes of miR-1306 in the database PITA was explored for enrichment in gene ontology [Ashburner et al., 2000] by the software Pathway Studio 8.0 (Ariadne Genomics) based on database ResNet 8.0.

2.4.5 Literature mining and pathway analysis

Literature search by PubMed was done to extract information about the target genes of the miRNAs resulting from Pathway Studio analysis and their relation to AD. To verify the miRNAs searches were performed for miRNA interactions in all PubMed abstracts with the help of the text mining program Pathway Studio 8.0 (Ariadne Genomics) based on the natural language processing (NLP) Technology. Pathway analysis was done with the software Ingenuity Systems IPA 9.0 (`http://www.ingenuity.com/`) especially with the Path Designer.

2.4.6 Experimental validation

Experiments were performed by Kristina Endres and Sven Reinhardt from the Johannes Gutenberg-University Mainz [Augustin et al., 2012].

Material

Mature miRNAs and the inactive negative control were from Invitrogen (No. PM11012, PM13206, PM10632, PM10056). All RNA species were dissolved to 5 pmol/μl in nuclease-free water upon arrival, aliquoted and stored at $-20\,^{\circ}$C.

Cloning of the ADAM10 3'UTR luciferase reporter construct

(Figure 2.1) The 3'UTR of human ADAM10 was amplified from THP-1 chromosomal DNA using the FailSafe PCR kit (Epicentre) and the following primers:
AD10_3UTR_for 5'GCGGCCGCGCCCATTCAGCAACCCCAG 3'

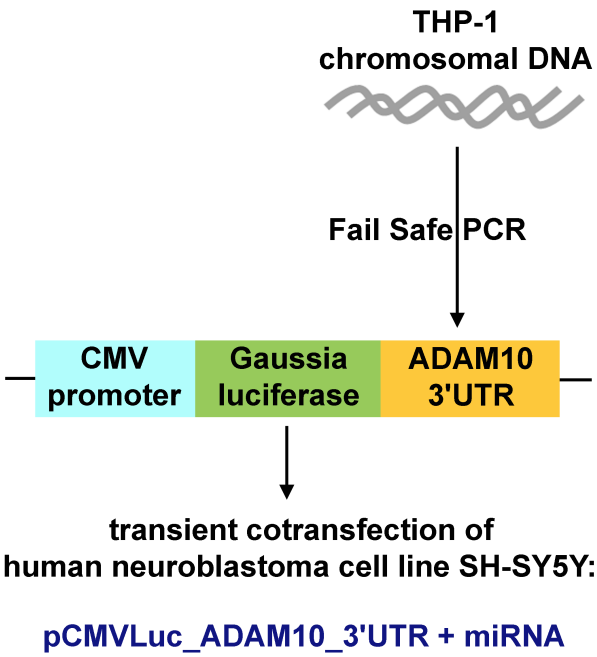


Figure 2.1: The ADAM10 3'UTR sequence was amplified from THP-1 chromosomal DNA with Fail Safe PCR and cloned into the NotI site of the pCMV-GLuc vector. SH-SY5Y cells were transiently cotransfected with pCMVLuc_ADAM10_3'UTR together with respective miRNA.

AD10_3UTR_rev 5'GCGGCCGCCACTTGTGCCCGTAGCAGCC 3'.
The obtained DNA fragment was verified by restriction digestion and sequencing. The 3'UTR was subsequently cloned into the NotI site of the pCMV-GLuc vector (NEB), which allows to monitor regulated Gaussia luciferase expression in the cell supernatant.

Cell culture

SH-SY5Y cells were cultivated in phenol red-free DMEM/F12, supplemented with 10% FCS and 1% glutamine at 37 °C, 95% air moisture, 5% CO_2 and passaged twice a week with a splitting rate of $\frac{1}{2}$ to $\frac{1}{4}$.

3'UTR luciferase reporter assay

Retro-transfection was performed using 0.005 μl Lipofectamine 2000 (Invitrogen) per μl OptiMEM-medium and 0.1 pmol/μl miRNA (Invitrogen) or negative control. For combination of miRNA 1306 together with miRNA 103 or 107 a concentration of 0.05 pmol/μl each was used. 2 ng/μl endotoxin-free plasmid DNA of the 3'UTR-reporter vector were added to 45.000 cells per well in 96 well format. Control cells

were mock-treated with nuclease-free water instead of RNA molecules.
5 hrs after transfection, the cell supernatant was exchanged to 200 μl culture medium per well. In a preliminary experiment, 10 μl cell supernatant were collected at various time points over a 72 hour period; 48 hours were determined to be the optimal incubation time (data not shown). Therefore, 10 μl cell supernatant were aspirated 48 hours after transfection and stored at $-20\,^{\circ}\mathrm{C}$ until samples were measured. Secreted Gaussia luciferase was quantitatively analyzed (Renilla-Luciferase assay, Promega) using the FluostarOptima luminometer (BMG). Cell densities were checked by quantitation of protein content in the cell lysate by NanoQuant assay (Roth).

Chapter 3

Results

3.1 Bioinformatics identification of modules of transcription factor binding sites

3.1.1 Workflow

The workflow consists of two different approaches combining in silico TFBS analysis with multivariate analysis of microarray datasets. Each part of the workflow the theoretical prediction of TFBS modules as well as the establishment of TFBS modules based on experimental data should result in similar modules in the end.
The first approach (Figure 3.1) starts with an alignment between human and mouse

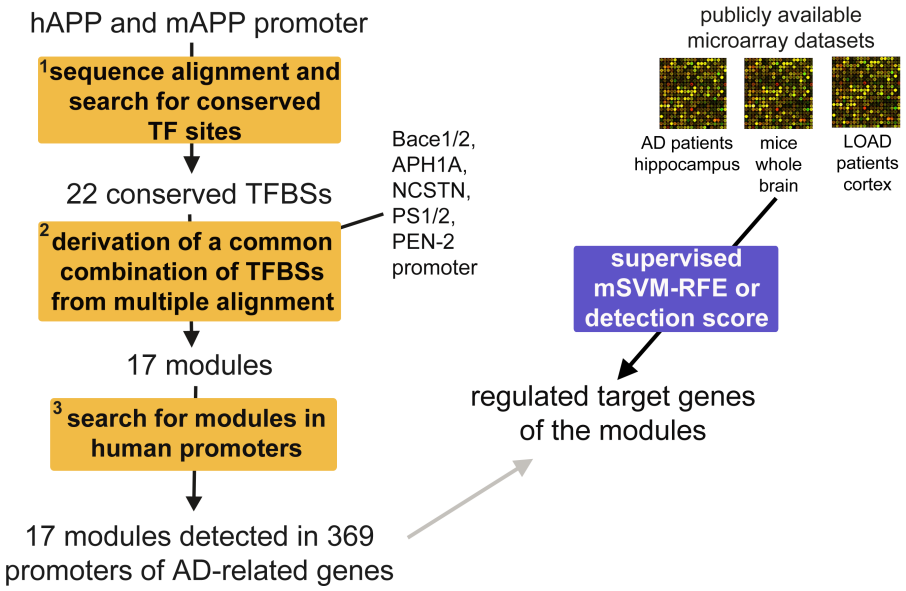


Figure 3.1: Workflow of bioinformatics analysis of promoter sequences and gene expression data to identify modules of TFBSs in AD-related genes. The colored boxes describe the methods which were used. The yellow boxes represent tools of the Genomatix software ([1]DiAlignTF, [2]FrameWorker, [3]ModelInspector), and the blue box indicates mSVM-RFE or filtering by illumina detection score. The beginning and the end of the arrow specify input and output of the methods, respectively. The gray arrow denotes the comparison of the target genes of the modules with genes differentially regulated in microarray analyses.

APP promoter sequences to search for conserved TFBSs. The result yielded 22 TFBSs (Table 3.1), which are conserved between human and mouse promoter sequences.

TF family	**description**	**TFs**
AP1F	AP1, Activating protein 1	BATF, BATF3, FOS, FOSB, FOSL1, FOSL2, JDP2, JUN, JUNB, JUND
AP1R	MAF and AP1 related factors	BACH1, BACH2, MAF, MAFA, MAFB, MAFF, MAFG, MAFK, NFE2, NFE2L1, NFE2L2, NFE2L3, NRL
AP4R	AP4 and related proteins	TFAP4
CHRE	Carbohydrate response elements, consist of two E box motifs separated by 5 bp	MLX, MLXIPL
CTCF	CTCF and BORIS gene family, transcriptional regulators with 11 highly conserved zinc finger domains	CTCF, CTCFL
E2FF	E2F-myc activator/cell cycle regulator	E2F1, E2F2, E2F3, E2F4, E2F5, E2F6, E2F7, E2F8, TFDP1, TFDP2, TFPD3
EGRF	EGR/nerve growth factor-induced protein C and related factors	EGR1, EGR2, EGR3, EGR4, WT1, ZBTB7A, ZBTB7B
GATA	GATA binding factors	GATA1, GATA2, GATA3, GATA4, GATA5, GATA6, TRPS1
GLIF	GLI zinc finger family	GLI1, GLI2, GLI3, GLIS1, GLIS2, GLIS3, ZIC1, ZIC2, ZIC3, ZIC4, ZIC5
HAND	Twist subfamily of class B bHLH transcription factors	HAND1, HAND2, LYL1, MESP1, MESP2, NHLH1, NHLH2, SCXA, SCXB, TAL1, TAL2, TCF12, TCF15, TCF3, TWIST1, TWIST2
NF1F	Nuclear factor 1	NFIA, NFIB, NFIC, NFIX
NFAT	Nuclear factor of activated T-cells	ILF2, ILF3, NFAT5, NFATC1, NFATC2, NFATC3, NFATC4
NOLF	Neuron-specific olfactory factor	EBF1, EBF2, EBF3
NRF1	Nuclear respiratory factor 1	NRF1
PAX3	PAX-3 binding sites	PAX3, PAX7
PAX9	PAX-9 binding sites	PAX9
RXRF	RXR heterodimer binding sites	NR1H2, NR1H3, NR1I2, NR1I3, RARA, RARB, RARG, RXRA, RXRB, RXRG, THRA, THRB, VDR
SP1F	GC-Box factors SP1/GC	KLF10, KLF11, KLF16, KLF5, SP1, SP2, SP3, SP4, SP5, SP6, SP7, SP8

TF family	description	TFs
XBBF	X-box binding factors	NFX1, RFX1, RFX2, RFX3, RFX4, RFX5
ZBPF	Zinc binding protein factors	ZNF148, ZNF202, ZNF219, ZNF281, ZNF300
ZF5F	ZF5 POZ domain zinc finger	ZFP161
ZFHX	Two-handed zinc finger homeodomain transcription factors	ZEB1, ZEB2

Table 3.1: Conserved TFs between human and mouse APP promoter sequence. Matrix family library version 8.1 (June 2009)

In the next step of the workflow, the conserved TFBSs and the promoter sequences of genes involved in Aβ formation (APP, BACE1, PS1/2, PEN-2, APH1A, and NCSTN) and the homolog of BACE1, BACE2, were used as input to search for all modules of TFBSs, which occur in a multiple alignment of promoter sequences of AD key genes. With this first approach, I got 17 modules composed of two or more TFBSs families (Table: 3.2) occurring in a subset of AD key genes.

Module	AD key genes - targets of the module
CTCF-E2FF-SP1F	hAPP, mApp, BACE1, NCSTN, APH1A
CTCF-SP1F	hAPP, mApp, BACE1, PS2, NCSTN, APH1A
E2FF-E2FF-EGRF	hAPP, mApp, BACE1, BACE2, PEN-2, APH1A
CTCF-E2FF-EGRF	hAPP, BACE2, PEN-2, NCSTN, APH1A
CTCF-E2FF-EGRF	hAPP, mApp, BACE1, BACE2, PEN-2, APH1A
CTCF-HAND-SP1F	hAPP, mApp, BACE2, PS2
CTCF-SP1F-SP1F	hAPP, mApp, BACE1, BACE2
CTCF-NRF1-SP1F	hAPP, mApp, BACE1, APH1A
CTCF-EGRF-NRF1	hAPP, BACE2, PS2, PEN-2
CTCF-SP1F-ZBPF	hAPP, mApp, BACE2
CTCF-EGRF-ZBPF	hAPP, BACE2, PS2
CTCF-NRF1	hAPP, mApp, BACE1, BACE2, PEN-2
CTCF-EGRF-SP1F	hAPP, BACE2, PS2
NRF1-ZBPF	hAPP, mApp, BACE2, PEN-2, APH1A
CTCF-EGRF	hAPP, mApp, BACE1, BACE2, PS1, PEN-2, APH1A
CTCF-E2FF	hAPP, mApp, BACE1, BACE2, NCSTN, APH1A
SP1F-ZBPF-ZBPF	BACE1, PS1, PEN-2, APH1A

Table 3.2: Modules identified by the first approach. 17 modules composed of two or more TFBSs families. The TFBSs families consist of several TFs (Table 3.1). The second column specifies the key genes of AD the TFs of the module putatively bind to, according to the search of the module in all human promoters by ModelInspector. Human and mouse APPs are indicated by hAPP and mApp, respectively.

An analysis of three microarray datasets should verify these modules identified in the first part of the workflow by searching for significantly regulated genes. The first microarray dataset of AD patients is composed of data from AD patients at different stages of severity (incipient, moderate and severe AD) and control. Additionally, I took a second dataset from a transgenic mouse model of AD to affirm the results of the AD patients dataset. The probes of the dataset are extracted from brain. The amount of plaques in the mice was controlled by the active as well as the dominant-negative (dn) form of α-secretase ADAM10 in order to imitate the situation of plaque formation in AD brain. The three mouse lines show different amounts of plaques in the brain. Fewer plaques are found in the brains of ADAM10/APP mice, medium plaques occur in brains of monotransgenic APP control mice, and most plaque formation appears in dnADAM10/APP mice. The reason for different plaque formation is that neurotoxic Aβ peptide levels are increased, and neuroprotective sAPPα is drastically decreased in dnADAM10/APP mice, and in ADAM10/APP mice the levels of the APP fragments are vice versa. The different mouse lines show different stages of plaque formation just like AD patients at different stages of severity [Postina et al., 2004]. Thus, this AD mouse model and its microarray dataset are appropriate to be included in this analysis to verify modules of TFBSs [Prinzen et al., 2009]. The third dataset, which I used, was established of cortex samples from LOAD patients and controls.
For microarray analysis I applied mSVM-RFE, because it is very accurate and fast in classification and has a low error rate [Zhou and Tuck, 2007]. RFE is applied due to the large number of the gene expression values on the microarray and helps to reduce the search space and avoids overfitting. Each of the TFBSs families is represented by several TFs (Table 3.1). After searching for these 17 modules in all human promoters, I limited the result to genes, which have been tested for genetic association with AD according to GWAS studies collected in the AlzGene database (`http://www.alzgene.org/`; Version: 12.05.2010) [Bertram et al., 2007] in order to get only those target genes possibly involved in AD. AlzGene database is a regularly updated aggregation of all published genetic association studies including GWAS performed on AD phenotypes. It is an important resource for AD candidate genes and contains all considerable genetic association studies and key genes of AD. I detected the 17 modules in the promoters of 369 putative AD-risk genes.
Subsequently, the 369 putative AD-risk genes were compared to the output from the mSVM-RFE of microarray datasets. I obtained for the AD patients dataset in the end 948 genes after mSVM-RFE (frequency ≥ 2) and an overlap with the putative AD-risk genes of 31 genes. The double-transgenic mice dataset was reduced

by mSVM-RFE (frequency ≥ 2) to 878 genes with an overlap of 26 to the 369 AD-related genes. A chi-squared test with p-value = 0.05391 shows a statistical trend. The third dataset from LOAD patients cortex samples was not filtered by mSVM-RFE, since it is already reduced by illumina detection score and therefore consists of only 8650 normalized expression values, which is a suitable number to apply biclustering. Comparing the 8457 genes of the GSE15222 dataset with the 369 putative AD-risk genes, I got an overlap of 199 genes between these two genesets. Starting point of the second approach (Figure 3.2) are the three microarray datasets

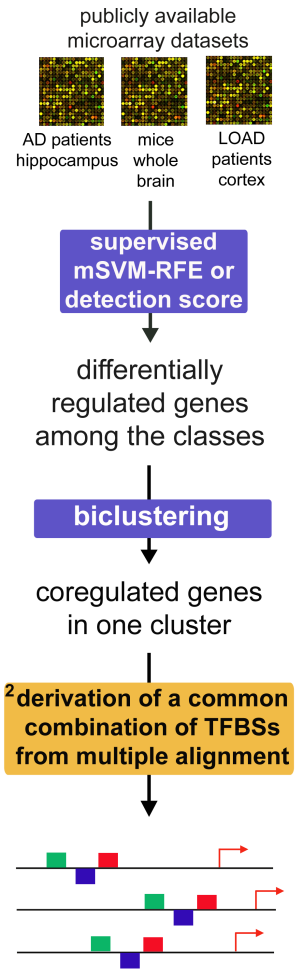


Figure 3.2: Workflow of bioinformatics analysis of promoter sequences and gene expression data to identify modules of TFBSs in AD-related genes. The colored boxes describe the methods which were used. The yellow box represents the tool [2]FrameWorker of the Genomatix software, and the blue boxes indicate multivariate methods or filtering by illumina detection score. The beginning and the end of the arrow specify input and output of the methods, respectively. The scheme at the end indicates a module composed of three TFBSs (blue, red, and green), which is common to three promoter sequences with transcription start site at the red arrow.

established from AD patients at three different stages of severity [Blalock et al., 2004], double-transgenic ADAM10/APP, dominant-negative ADAM10/APP as well as APP control mice [Prinzen et al., 2009], and AD patients with late onset AD (LOAD) [Webster et al., 2009]. Not differentially regulated genes are excluded by mSVM-RFE, and regulated genes of these microarrays may potentially play a role in AD. By mSVM-RFE (frequency ≥ 1), I reduced the dataset of the AD patients at three different stages of severity from 22283 probesets to 4844 probesets, and then after biclustering these 4844 probesets, I got five and eight clusters of coregulated genes from two biclustering runs with the same parameter setting. The double-transgenic mice dataset was reduced by mSVM-RFE (frequency ≥ 1) from 45101 to 5198 probesets, and after biclustering, I obtained 13 clusters of coregulated genes. The third dataset of LOAD patients is already reduced by illumina detection score to 8650 probesets, and therefore, I did not apply mSVM-RFE. By biclustering, I got 18 clusters of coregulated genes. By grouping the regulated genes into clusters of coregulated genes and searching for modules in the promoters of these coregulated genes, I got modules possibly responsible for the common regulation of these genes and also putatively playing an important role in the modification of AD. At the end, I obtained several modules for each cluster of coregulated genes. The target genes of three selected modules, which are described in more detail in section 3.1.3, 3.1.4, and 3.1.5, are listed in the supplemental Tables A.1 - A.8. Target genes are activated or repressed by TF modules and have corresponding TFBSs in their promoter sequences.

After the whole analysis composed of the first and second approach, I got four different sets of TFBSs modules. I obtained one set of 17 TFBSs modules from the first approach and three sets (one set for each microarray study) of on average five different TFBSs modules per cluster from the second approach. I compared these four sets with regard to similar modules and found two modules in common: CTCF-EGRF-SP1F as well as CTCF-SP1F-ZBPF, which are illustrated together with target genes in Figure 3.3. According to the first module, the target genes VAPA and EIF5 overlap between AD patients and double-transgenic mice dataset, and the target genes REEP5 and SYP overlap between double-transgenic mice and LOAD patients dataset. The overlapping target gene ADD3 (adducin 3 (gamma)) between AD and LOAD patients dataset of module CTCF-SP1F-ZBPF is also target gene of the module KLFS-SP1F-ZBPF, which is common to the three microarray datasets in the second approach. The third module has additionally overlapping target genes between AD and LOAD patients dataset: CLU and NUCKS1. The TF family KLFS includes the TFs: KLF1, KLF2, KLF3, KLF4, KLF6, KLF7, KLF8, KLF9, KLF12, KLF13, KLF15.

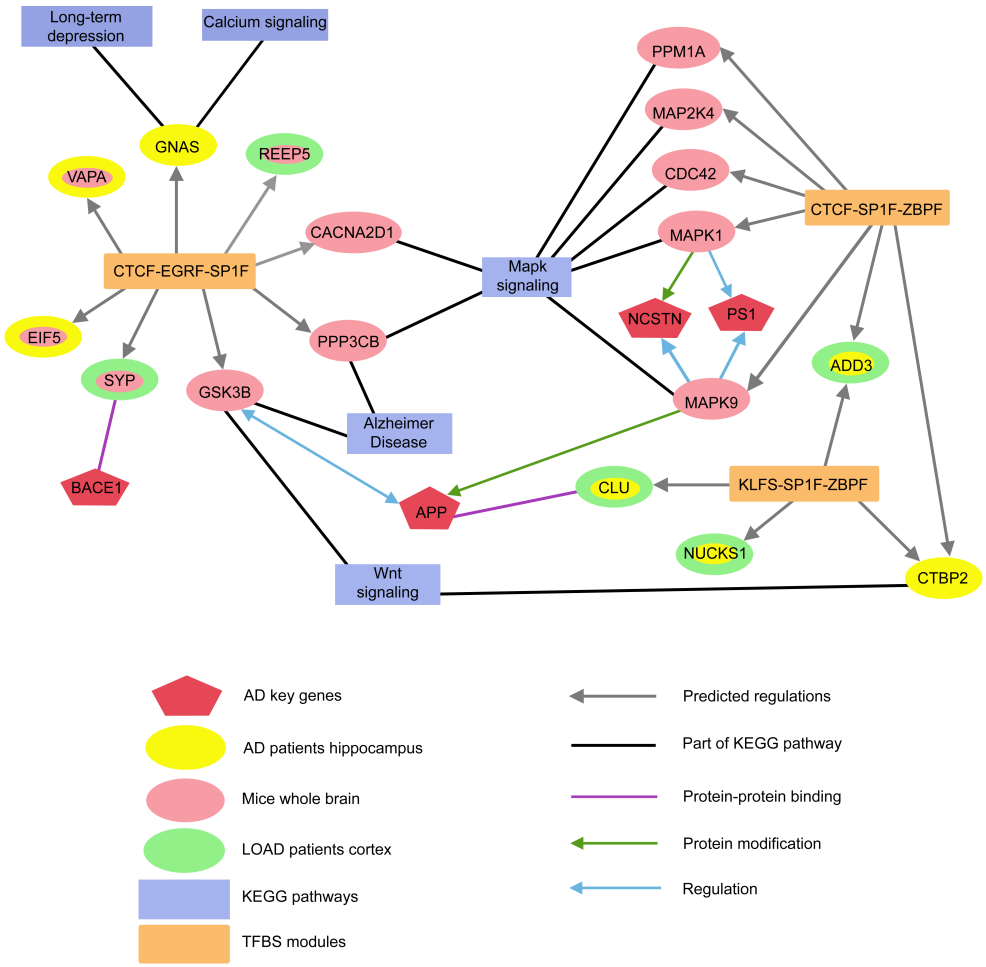


Figure 3.3: Relations of predicted target genes of three TFBSs modules. This picture summarizes important target genes of the modules, the relation of target genes to KEGG pathways playing a role in AD (blue rectangle), and the relation of the target genes to some AD key genes (red pentagon). The target genes are colored according to their membership to microarray studies, and some target genes with two colors are derived from analysis of two different microarray studies. The gray arrows are the predicted regulations of the target genes by the modules (orange rectangle), and the black lines indicate that the target gene is part of the corresponding KEGG pathway. Additionally, three different relations of the target genes to AD key genes are shown by purple, green, and blue lines, which indicate protein-protein binding, protein modification, and regulation, respectively.

In general, TFBSs frequently occurring in modules of both approaches are CTCF, EGRF, SP1F, and ZBPF, but the composition of the TFBSs for a module is slightly different between the first and second approach. The modules of the second approach mostly contain one TFBS, which is not conserved between human and

mouse APP promoter.

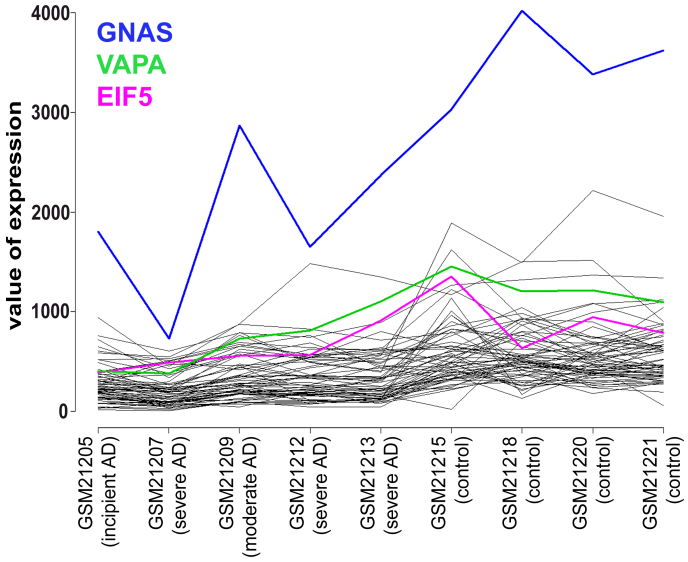
3.1.2 Modules and confirmations of TFBSs and AD-risk genes

The first common module is composed of the binding sites of the three TF families: CTCF, EGRF, and SP1F, and the second module consists of CTCF, SP1F, and ZBPF, which are all conserved between human and mouse APP promoter sequences. TFs representing both modules are predicted to bind to the promoter sequences of significantly frequent target genes in the corresponding cluster of the AD patients dataset (FDR (CTCF-EGRF-SP1F) = 0.0003; FDR (CTCF-SP1F-ZBPF) = 0.0003), of the transgenic mice dataset (FDR (CTCF-EGRF-SP1F) = 4.2×10^{-7}; FDR (CTCF-SP1F-ZBPF) = 1.3×10^{-12}) and of the LOAD patients dataset (FDR (CTCF-EGRF-SP1F) = 0.0139; FDR (CTCF-SP1F-ZBPF) = 0.0139), compared to the incidence in the whole set of human promoters (supplemental Table A.9). The expression profiles of the coregulated genes from the three datasets in each module show similar expression patterns among a subset of microarray samples (Figure 3.4).

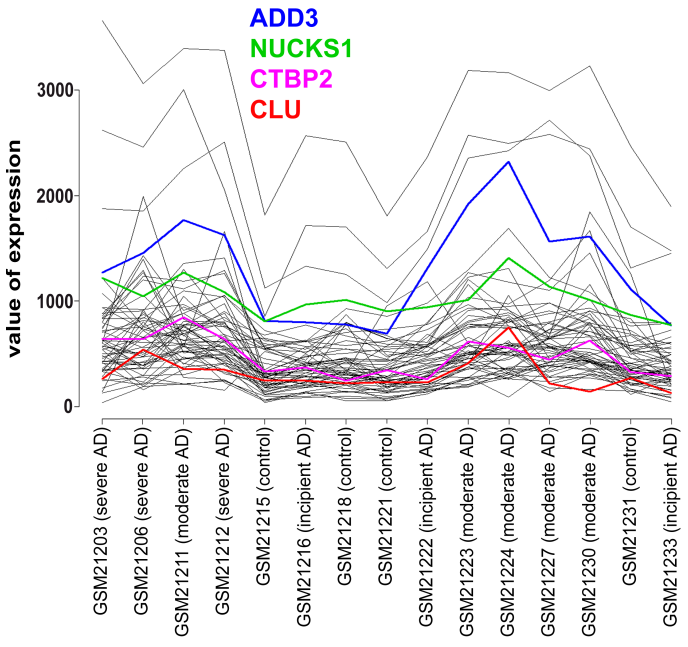
Additionally, a third significant module was detected established from a set of coregulated genes of the AD (Figure 3.4(b)) and LOAD patient's dataset (Figure 3.4(e)). The corresponding motif of this module consists of KLFS, SP1F, and ZBPF binding sites and occurs in the promoter sequences of several interesting genes in particular to Clusterin (CLU/APOJ), which is according to AlzGene database the second most strongly associated gene to AD [Guerreiro et al., 2010]. The enrichment of the module in the cluster of coregulated genes from the AD patients dataset (FDR (KLFS-SP1F-ZBPF) = 0.0003) and the LOAD patients dataset (FDR (KLFS-SP1F-ZBPF) = 0.0139) is significant compared to the occurrence in all human promoters (supplemental Table A.9).

While most of the transcription factors of all families in the resulting modules play a role in apoptosis, the transcription factor families have additional different main functions according to TRANSPATH database. The CTCF zinc finger proteins are involved in chromatin remodelling, the early growth response transcription factors (EGRFs) in learning and memory and brain development, the GC-Box factors (SP1Fs) in chromatin silencing as well as embryonic development, the zinc binding protein factors (ZBPFs) in lipid metabolism and the Krueppel-like transcription factors (KLFSs) in nervous system development and response to stress. Most of the TFs of the families are expressed in whole brain, hippocampus, or cortex.

To evaluate the importance of the detected modules, I incorporated information of the AlzGene database. Some target genes of the modules are already mentioned in AlzGene database to be associated to AD like GOT1 (glutamic-oxaloacetic transam-



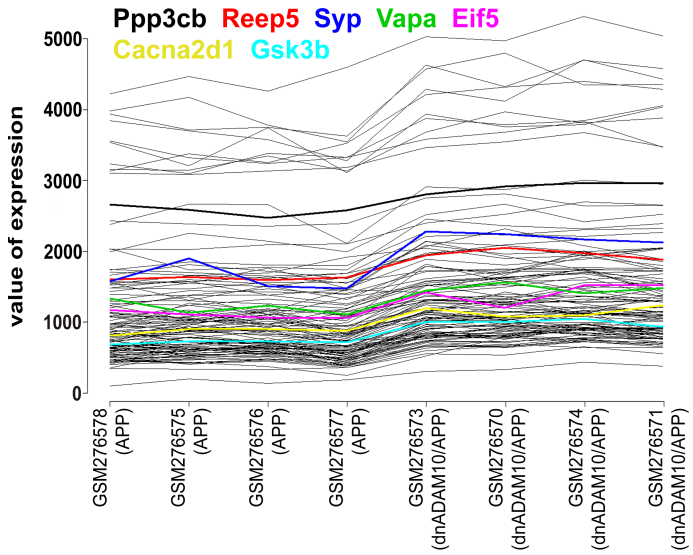
(a)



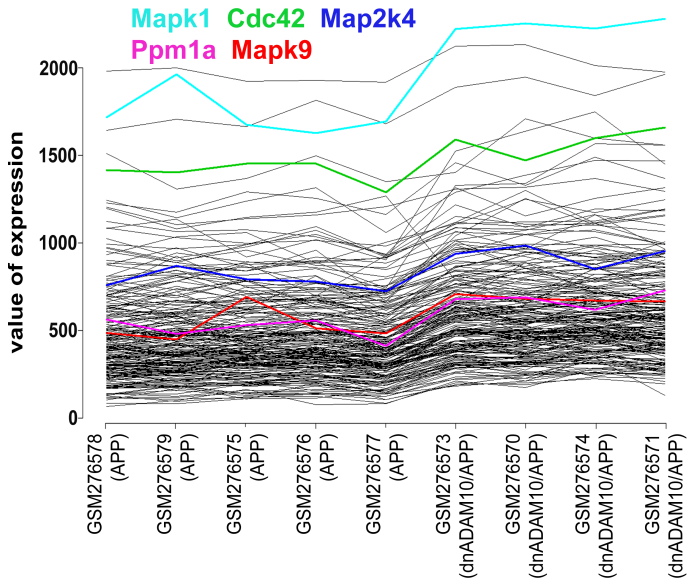
(b)

Figure 3.4: Clusters of coregulated genes of the AD patients dataset.

inase 1), Gsk3b (CTCF-EGRF-SP1F), Col25a1 (collagen, type XXV, alpha 1), Il33 (interleukin 33), and Tanc2 (tetratricopeptide repeat, ankyrin repeat and coiled-coil containing 2) (CTCF-SP1F-ZBPF), and CLU (KLFS-SP1F-ZBPF).

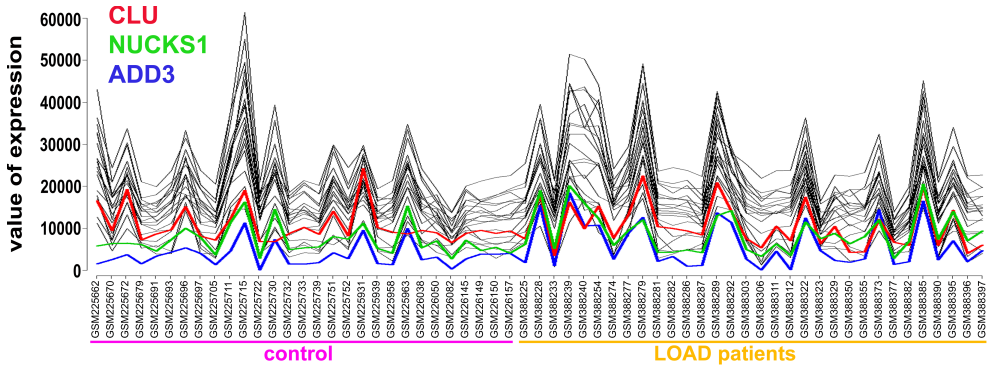


(c)

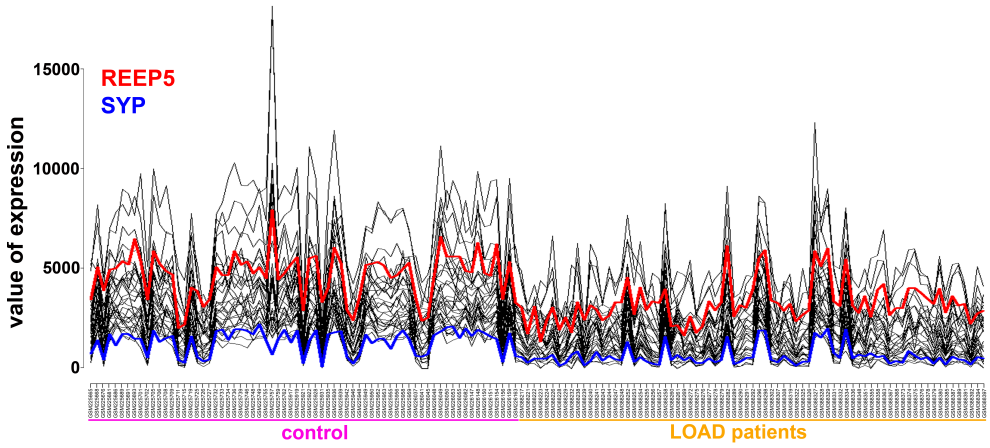


(d)

Figure 3.4: Clusters of coregulated genes of the double transgenic mice dataset.



(e)



(f)

Figure 3.4: The first two profiles (a) and (b) are clusters of coregulated genes of the AD patients dataset. Profile (c) and (d) correspond to coregulated genes of the double-transgenic mice dataset. The five lines in profile (d) correspond to genes, which are involved in the MAPK signaling pathway. The last two profiles (e) and (f) correspond to coregulated genes of the LOAD patients dataset. On the x-axis, the sample IDs (specified by accession numbers of GEO/NCBI) incorporated in the cluster are given, and y-axis indicates values of expression. One gene corresponds to a single line in the profile, and the target genes of the modules as mentioned in the text are colored. The target genes of the profiles (a) and (c) were used for the establishment of the module CTCF-EGRF-SP1F and the profiles (b) and (d) for the module CTCF-SP1F-ZBPF, at which (b) was also used for the module KLFS-SP1F-ZBPF. The target genes of the profile (e) were used for the establishment of the modules CTCF-SP1F-ZBPF and KLFS-SP1F-ZBPF. The target genes of the profile (f) were used to establish the module CTCF-EGRF-SP1F.

3.1.3 First module CTCF-EGRF-SP1F

Literature mining revealed known relations between the target genes of the first module (supplemental Table A.1, A.2 and A.3) and AD. GNAS is incorporated in long-term depression and calcium signaling, which occur significantly often in the cluster of the coregulated genes of the AD patients dataset (p-value = 0.0404, FDR = 0.0952 (human long-term depression); p-value = 0.0476, FDR = 0.0952 (human calcium signaling); supplemental Table A.10).
Cacna2d1, a target gene of the double-transgenic mice dataset, is involved in the mitogen-activated protein kinase (MAPK) signaling in the mouse, and this pathway is significantly overrepresented among the coregulated genes of the double-transgenic mice dataset (p-value = 0.0024, FDR = 0.0121; supplemental Table A.10). The target gene Gsk3b is involved in AD pathway and Wnt signaling (enrichment analysis: p-value = 0.0021, FDR = 0.0121 (mouse AD pathway); p-value = 0.0206, FDR = 0.0497 (mouse Wnt signaling); supplemental Table A.10). Another target gene Ppp3cb (protein phosphatase 3, catalytic subunit, beta isoform) is also involved in AD and MAPK signaling in the mouse according to KEGG pathways with a significant enrichment as mentioned before.
Comparing the target genes of this module from the AD patients and double-transgenic mice microarray study, two overlapping genes are found: EIF5 and VAPA. A family member of EIF5, EIF2AK2, is listed in AlzGene database. Furthermore, two overlapping target genes from the double transgenic mice and LOAD patients dataset were found: SYP, REEP5. SYP expressed in hippocampus and cortex according to TRANSPATH database. REEP5 is expressed in the brain and central nervous system according to MGI database.

3.1.4 Second module CTCF-SP1F-ZBPF

The AD and LOAD patients target genes of the second module (supplemental Tables A.4, A.5) partially overlap with the AD and LOAD patients target genes of the third module KLFS-SP1F-ZBPF (supplemental Tables A.6, A.7) such as ADD3 and CTBP2. Additional hints for the importance of the target genes putatively activated or repressed by a module were also found for the double transgenic mice target genes of the second module (supplemental Table A.8).
Coregulated genes of the double transgenic mice are significantly enriched in the MAPK pathway (enrichment analysis: p-value = 0.0010, FDR = 0.0242; supplemental Table A.10), incorporating the target genes Cdc42, Map2k4, Mapk1, Mapk9, and Ppm1a of the second module (Figure 3.4 (d)).
The gene ADD3 overlaps between the target genes of the AD and LOAD patients

dataset. According to TRANSPATH database, it is known to be upregulated in amyotrophic lateral sclerosis (ALS) and to play a role in apoptosis.

3.1.5 Additional information of the third module KLFS-SP1F-ZBPF

The AD patients target gene CTBP2 (C-terminal binding protein) is involved in the Wnt signaling pathway. In this study, one cluster of coregulated genes is significantly enriched in the human Wnt pathway (p-value = 0.0149, FDR = 0.0464; supplemental Table A.10).
Comparing the AD and LOAD patients target genes from the third module, three genes were found in common: ADD3, CLU, and NUCKS1. ADD3 is also in common for the second module described above.

3.2 MSVM-RFE and biclust application to other microarray datasets

3.2.1 HR versus LR microarray

The dataset of the HR and LR mice was reduced by mSVM-RFE from 45281 probesets to 6964 probesets, and then after biclustering these 6964 probesets, I got ten clusters of coregulated genes.
Comparing the genes from the microarray showing an adjusted p-value < 0.05 with differentially expressed genes in SAGE (Z-Score > 2) yields 25 genes in common, at which 76% of the overlapping genes occur in all gene selections of 40 mSVM-RFE runs. The remaining genes occur at least in 18 gene selections. All genes occurring in at least 18 gene selections of 40 mSVM-RFE runs compared with the differentially expressed genes in SAGE results in 48 equal genes, at which 27 genes arise in 40 gene selections.

Genesymbol	**p-value qPCR**	**Significance**	**mSVM-RFE**	**foldchange HR vs LR**	**outcome**
Accn2	0.374	n. s.	0	1	TN
Acot1	0.329	n. s.	37	1.29	TN
Acot11	0.002	**	40	1.21	TP
Acot13	0.283	n. s.	19	0.74	TN
Adam11	0.018	*	38	1.21	TP
Adam22	0.143	n. s.	40	2.21	FP
Aldh1l1	0.015	*	40	2.04	TP
Ash1l	0.028	*	40	2.22	TP
Cntn6	0.711	n. s.	5	0.78	TN
Crtac1	0.006	**	40	1.3	TP
Cyp4f13	0.689	n. s.	23	0.75	TN
Cyp4f14	0.001	***	40	2.3	TP
Cyp4f15	0.002	**	40	2.51	TP
Cyp4f16	0.329	n. s.	40	1.66	FP
Dap3	0.297	n. s.	40	$1.94^{\#}$	FP
Fam116b	0.178	n. s.	0	1	TN
Fam123a	0.085	T	2	0.89	FN
Gabrg2	0.028	*	40	1.54	TP
Hsp90ab1	0.229	n. s.	40	3.12	FP
Mthfd1	0.027	*	40	0.38	TP
Prodh	0.027	*	40	$1.32^{\#}$	TP
Pvalb	0.596	n. s.	0	1.02	TN
Rgl1	0.631	n. s.	40	4.26	FP
Sh3gl2	0.045	*	40	1.13	TP

Genesymbol	p-value qPCR	Significance	mSVM-RFE	foldchange HR vs LR	outcome
Ssh1	0.033	*	0	0.97#	FN
Tesk1	0.159	n. s.	4	2.08	TN
Timp4	0.410	n. s.	40	1.34	FP
Tmem106c	0.916	n. s.	0	0.98	TN
Tmem132d	0.051	T	40	0.44	TP
Ttbk1	0.006	**	40	6.18	TP
Usp38	0.283	n. s.	40	0.44	FP

Table 3.3: Comparison of qPCR and mSVM-RFE result. Genes tested in qPCR are listed in alphabetical order with appropriate p-value of the qPCR (column two) and beside its significance (n. s. not significant, T p-value < 0.1, * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001) is given [personal communication with Regina Widner from MPI Munich]. Column four shows the frequency of each gene occurring in all gene selections of the mSVM-RFE. Beside the foldchange HR versus LR of the microarray is given (# average foldchange due to more than one probe). The last column shows the accordance of the qPCR result with the mSVM-RFE result (TP true positives, TN true negatives, FP false positives, FN false negatives).

31 genes were determined for qPCR analysis [personal communication with Regina Widner from MPI Munich]. 18 genes were selected according to their adjusted p-value < 0.05 in the microarray analysis. Four genes were chosen, which showed no expressional differences in the microarray study, in order to test the reliability of the microarray experiment. Nine genes involved in neurogenesis and neurodevelopment were also further analysed by qPCR.

The qPCR detected 13 differentially expressed genes and two genes showing a trend in differential regulation. 16 genes were confirmed to be not regulated. The results of the qPCR experiment are shown in Table 3.3. Additionally, the frequency of the genes occurring in the mSVM-RFE gene selections as well as the foldchange on the microarray is shown.

To choose the best cut-off point of the mSVM-RFE frequency for best sensitivity as well as best selectivity a receiver operator characteristic (ROC) curve was established [Lalkhen and McCluskey, 2008] (Figure 3.5). Best cut-off point is 38, i.e. all genes occurring in at least 38 mSVM-RFE gene selections are classified to be differentially expressed in the microarray (Positives) and the remaining genes are Negatives. According to the prediction of the mSVM-RFE and the result of the qPCR the genes are signified as true positive TP, false positive FP, true negative TN or false negative FN (Table 3.3). Finally, a sensitivity of 87% by (1) and a specificity of 56% by (2) was calculated for the mSVM-RFE method.

KEGG enrichment analysis of genes determined by mSVM-RFE occurring in all gene

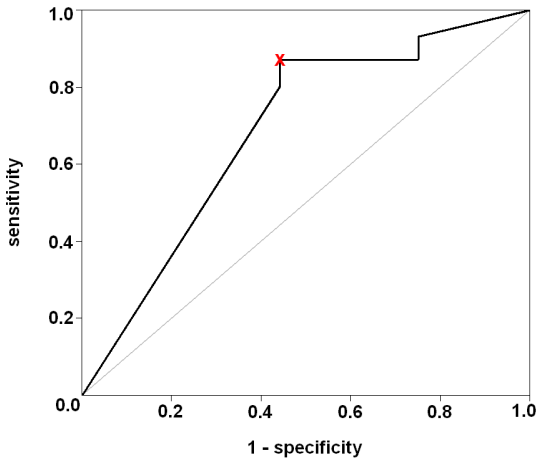


Figure 3.5: ROC curve. The x-axis describes the false positive rate (1 - specificity), while the y-axis represents the true positive rate (sensitivity). The gray and black line represent the line of zero discrimination and the ROC curve of mSVM-RFE, respectively. The best cut-off point for best sensitivity as well as best selectivity is marked by a red cross.

selections (frequency = 40) reveals three pathways incorporating qPCR validated genes (supplemental Table A.11). Aldh1l1 is involved in the pathway 'one carbon pool by folate' in mouse with a significant overrepresentation of the genes detected by mSVM-RFE with frequency = 40 (enrichment analysis: p-value = 0.0088, FDR = 0.0372). Other pathways are 'arachidonic acid metabolism' and 'lysine degradation' including the genes Cyp4f14, Cyp4f15 (cytochrome P450, family 4, subfamily f, polypeptide 14/15) and Ash1l (absent small and homeotic disks protein 1 homolog), respectively (enrichment analysis: p-value = 0.0344, FDR = 0.0799 (mouse arachidonic acid metabolism); p-value = 0.0495, FDR = 0.0799 (mouse lysine degradation)).

Additionally, each cluster was analysed for pathway enrichment. Gabrg2 (gamma-aminobutyric acid (GABA) A receptor, gamma 2) and Sh3gl2, both genes validated by qPCR and occurring in different clusters, are participating in 'neuroactive ligand-receptor interaction' and 'endocytosis', respectively (enrichment analysis: p-value/FDR = 0.0409 (mouse neuroactive ligand-receptor interaction); p-value/FDR = 0.041 (mouse endocytosis); supplemental Table A.11).

Name	**Genes**	**p-value**	**FDR**
leukotriene metabolic process	ALOX5AP, CYP4F2, CYP4F16	5.57×10^{-5}	4.57×10^{-3}
ER-associated protein catabolic process	NPLOC4, FBXO2, FBXO6	0.000213	0.008735

Name	Genes	p-value	FDR
glycoprotein catabolic process	FBXO2, FBXO6	0.000748	0.012273
response to lithium ion	ACTA1, GSTM5	0.002282	0.020789
glutathione metabolic process	GSTK1, GSTM5	0.007217	0.039451
negative regulation of cell adhesion	ADAM22, MYO1F	0.010863	0.046881
response to glucocorticoid stimulus	UGT1A10, SPARC, TYMS	0.013517	0.047892
cell adhesion	ITGAE, CD97, ADAM22, EMB, PVRL3, COL20A1, CHST10	0.014601	0.047892
mitochondrial matrix	SDHAF1, PRODH, DAP3, ACAA2, GSTK1, GCDH	0.000455	0.022275
mitochondrion	TTC19, SLC25A34, ABCB1B, RPL10A, SDHAF1, HSP90AB1, ELAC2, GSTZ1, PRODH, DAP3, ALDH1L1, MFF, ACAA2, GSTK1, CCBL2, GCDH, NT5DC3, MRPL10	0.001958	0.047980
endoplasmic reticulum	GOLT1B, TMTC2, UGT1A10, STX8, ALOX5AP, NPLOC4, SPAST, CYP4F2, CNPY3, G6PC3, CYP4F12, CYP4F16, SEC22C	0.005571	0.080511
glutathione transferase activity	GSTZ1, GSTK1, GSTM5	0.000526	0.011221
glutathione peroxidase activity	GSTZ1, GSTK1	0.002322	0.027435
glycoprotein binding	FBXO2, FBXO6	0.014066	0.050013

Table 3.4: Selected GO terms (column one) of genes determined by mSVM-RFE occurring in all gene selections with dedicated genes (column two) are shown. p-value and FDR are given in column three and four, respectively.

Furthermore, GO enrichment analysis was performed for all genes identified by mSVM-RFE with frequency = 40 and selected significant GO terms are listed in Table 3.4 with dedicated genes in column two. The first eight terms in the table are biological processes followed by three cellular component terms and three molecular function terms.

3.2.2 DJ-1 knockout microarray

The microarray dataset established from DJ-1 knockout and wildtype mice revealed several differentially expressed genes related to PD.

22 probesets with an adjusted p-value < 0.35 are differentially expressed between the different mouse models. Different probesets can correlate to one common gene and some probesets are not yet assigned to a gene. In total 12 different genes (supplemental Table A.13) correlate with these 22 probesets.
MSVM-RFE reduced the DJ-1 knockout mice dataset from 28853 probesets to 3481 probesets. Biclustering analysis revealed eight biclusters of coregulated genes by entering the 3481 probesets obtained by mSVM-RFE.

Literature mining displays indirect interactions of DJ-1/PARK7 to eight genes (PLOD1, DHRS3, RERE, CLSTN1, UBIAD1, MTHFR, ACOT7, HNMT), which occur in all gene selections and have an adjusted p-value of < 0.35. Indirect interaction means the eight genes are not directly linked to DJ-1 but at least one gene is connecting both genes via a short path. The associations to DJ-1 are described in the following.
The genes PLOD1 and PARK7 have a common regulator TGFB1 (transforming growth factor, beta 1) [Knippenberg et al., 2009, Altraja et al., 2009]. The expression of DHRS3 is influenced by PARK7 via RET (ret proto-oncogene) [Foti et al., 2010, Oppenheimer et al., 2007]. The differentially expressed gene RERE is controlled by PARK7 via HDAC1 (Histone deacetylase 1) [Zhong and Xu, 2008, Wang et al., 2006]. CLSTN1 is influenced by PARK7 via interacting with MAPK1 [Gu et al., 2009, Vagnoni et al., 2011]. UBIAD1 binds to APOE [Fredericks et al., 2011], which is upregulated by PARK7 by inhibiting MAPK8 [Pocivavsek and Rebeck, 2009, Mo et al., 2008]. The activation of MTHFR is reduced by a dominant-negative form of MAPK8 [Leclerc and Rozen, 2008], at which MAPK8 is inhibited by PARK7 [Mo et al., 2008]. ACOT7 is controlled by SREBF2 (sterol regulatory element binding transcription factor 2) [Takagi et al., 2005], which is influenced by PARK7 via IL8 (interleukin 8) [McNally et al., 2011, Yao et al., 2006], and ALB (Albumin) [Pérez-Gil et al., 1990], which regulates the PARK7 interacting gene SNCA [McLaughlin et al., 2006, Savitt et al., 2006]. The last gene HNMT participates in the same histamine-metabolizing pathway as its regulator MAOA (monoamine oxidase A) [Boudíková-Girard et al., 1993], which is controlled by PARK7 over NFE2L2 (nuclear factor (erythroid-derived 2)-like 2) [Thimmulappa et al., 2002, Clements et al., 2006]. Additionally, HNMT is also regulated by IL1B (interleukin 1, beta) [Sakata et al., 1995], which is directed by PARK7 via SNCA [Roodveldt et al., 2010, Savitt et al., 2006]. Furthermore MTHFR and HNMT are incorporated in the PDGene database and MTHFR is a NFKB target gene.

GO enrichment analysis of genes of all mSVM-RFE gene selections reveals biological processes like 'response to interleukin-1', 'vesicle-mediated transport', 'response to stress' and 'cell death' as well as the molecular function 'metal ion binding' (Table 3.5).

Name	**Genes**	**p-value**	**FDR**
response to interleukin-1	MTHFR, HNMT	0.000501	0.032715
vesicle-mediated transport	Bet3l, RAB6B, INTS4	0.001545	0.032715
response to stress	Ahsa2, PARK7, HNMT	0.001799	0.032715
cell death	OPTN, PARK7	0.008953	0.047536
metal ion binding	PDZRN4, ZSCAN2, LOC388559, RERE, POMT2, PRIM1, POLR3B, PLOD1	0.042089	0.097025

Table 3.5: Selected GO terms (Name) of genes determined by mSVM-RFE occurring in all gene selections. In column two the genes enriched in the GO term are shown following by p-value and FDR.

The mSVM-RFE output was further restricted to genes, which are transcriptionally regulated by NFKB known to play a significant role in PD [Cassarino et al., 2000]. 29 different NFKB target genes are obtained, at which 20 genes show direct interactions to each other. GO enrichment analysis shows biological process categories like 'response to lipopolysaccharide', 'inflammatory response', 'positive regulation of interleukin-6 biosynthetic process', 'positive regulation of apoptosis', 'anti-apoptosis', 'response to interleukin-1', 'aging' and 'learning or memory' (Table 3.6).

Name	**Genes**	**p-value**	**FDR**
response to lipopolysaccharide	FOS, IL1B, SELE, SELP, ASS1, DIO2	1.9×10^{-8}	5.63×10^{-6}
inflammatory response	FOS, IL1B, SELE, SELP, CD44, TNFRSF4	4.22×10^{-7}	6.26×10^{-5}
positive regulation of interleukin-6 biosynthetic process	IL1B, IFNG	1.91×10^{-5}	0.000944
positive regulation of apoptosis	IL1B, SOD1, IL2RA, GRIN2A	3.66×10^{-5}	0.001550
anti-apoptosis	IL1B, FN1, SOD1, IER3	6.39×10^{-5}	0.002229
response to interleukin-1	SELE, MTHFR	0.000501	0.007424

Name	Genes	p-value	FDR
aging	FOS, IL1B, SOD1	0.000524	0.007491
learning or memory	IL1B, GRIN2A	0.001083	0.009586

Table 3.6: Selected GO terms (Name) of genes determined by mSVM-RFE occurring in at least one gene selection, which are NFKB targets. In column two the genes enriched in the GO term are shown following by p-value and FDR.

Biclustering of all probesets obtained from mSVM-RFE reveals one interesting cluster of 41 coregulated genes incorporating NFKB target genes Ccnd2, Cd44 (CD44 antigen) as well as S100a4, from PDGene Ndufs2 (NADH dehydrogenase (ubiquinone) Fe-S protein 2) as well as Wnk1 (WNK lysine deficient protein kinase 1) and Kif1b (kinesin family member 1B) listed in PDbase.

KEGG pathway	Genes	Fisher-Test	FDR
Huntington's disease	Polr2f, Ep300, Ndufs2, Sdha, Cox6c	6.05×10^{-5}	0.000605
Citrate cycle (TCA cycle)	Idh3g, Sdha	0.002415	0.010405
Oxidative phosphorylation	Ndufs2, Sdha, Cox6c	0.004162	0.010405
Parkinson's disease	Ndufs2, Sdha, Cox6c	0.004162	0.010405
Alzheimer's disease	Ndufs2, Sdha, Cox6c	0.008381	0.016762
Cell cycle	Ccnd2, Ep300	0.035299	0.058832

Table 3.7: Coregulated genes (column two) enriched in KEGG pathways (column one) are shown with corresponding p-value and FDR in column three and four, respectively.

Pathway enrichment analysis of the 41 coregulated genes reveals several neurodegenerative diseases like Huntington, PD as well as AD and additional pathways like 'citrate cycle', 'oxidative phosphorylation' and 'cell cycle' (Table 3.7 and supplemental Table A.12).

Name	Genes	p-value	FDR
microtubule	MAP4, MID1, DYNC1H1, KIF1B, TUBB, CEP170	1.0×10^{-5}	0.00355
negative regulation of microtubule depolymerization	MAP4, MID1	0.000238	0.012092

Name	Genes	p-value	FDR
microtubule-based movement	DYNC1H1, KIF1B, TUBB	0.000334	0.014805
iron ion binding	NDUFS2, KDM3B, KDM5A	0.008378	0.043738
mitochondrial inner membrane	NDUFS2, SDHA, COX6C	0.022862	0.061955
cell cycle	EP300, CCND2, LATS1	0.026087	0.065602
mitochondrial matrix	IDH3G, OXCT1	0.035876	0.078919
magnesium ion binding	IDH3G, WNK1, LATS1	0.041236	0.084617
mitochondrion	IDH3G, NDUFS2, KIF1B, SDHA, OXCT1, COX6C	0.042429	0.085490

Table 3.8: Enrichment of coregulated genes (column two) in selected GO terms (column one). P-value and FDR is shown in column three and four, respectively.

GO enrichment analysis of all 41 coregulated genes was performed and results in several significant terms. Selected GO terms especially involved in 'mitochondrion', 'microtubule' and 'cell cycle' are shown in Table 3.8.
Literature mining of the 41 coregulated genes from the cluster shows some direct interactions and various indirect interactions between coregulated genes linked by interleukins and MAP kinases like MAPK8. Furthermore, key genes of neurodegenerative diseases such as SNCA, several TFs and genes like TNF, INS (insulin), BCL2 connect as well coregulated genes.

3.3 TFBSs in key genes of Parkinson's disease

3.3.1 ATF4 binding site in Parkin promoter

The Parkin promoter sequence, which was elongated downstream of the transcription start site by 150 bp, of human was analysed for conserved ATF4 TFBSs by Genomatix software. This CREB/ATF (activating transcription factor 4) TFBS shown in Figure 3.6 is conserved among the species human, mouse, cow and horse. An additional binding site is located downstream of the TSS, in the first intron of Parkin, in human, mouse and cow.

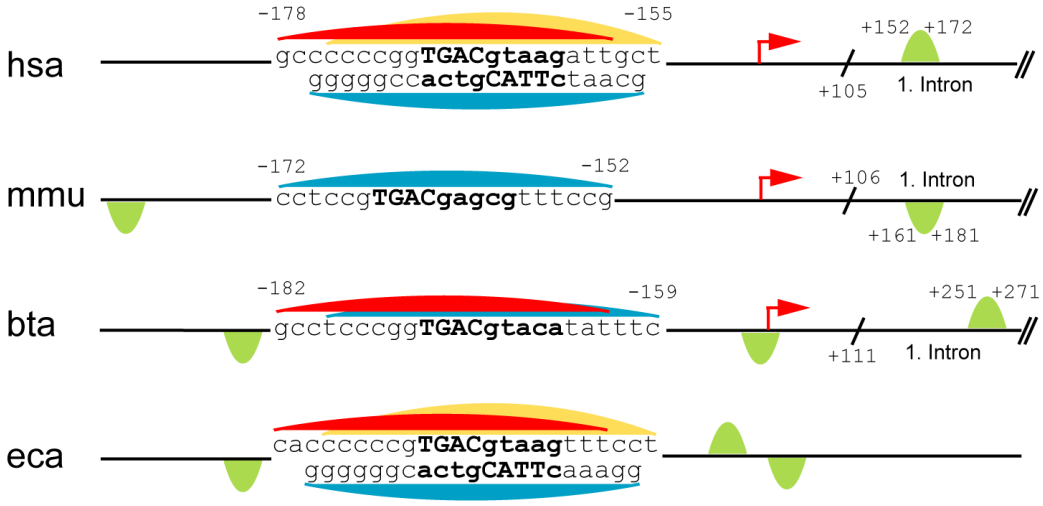


Figure 3.6: Conserved CREB/ATF binding sites in Parkin. Human, mouse, cow and horse promoter sequences of Parkin, which are elongated downstream of the transcription start site by 150 bp. Red arrow indicates transcription start site (TSS) and positions are denoted relative to the TSS. The CREB/ATF binding sites are indicated by semicircles. Red, yellow and blue semicircles are predicted by three different binding motifs, which correspond to a Genomatix-defined family of 14 matrices describing the CREB/ATF binding site. The red and yellow colored binding site is conserved between human, cow and horse, as well as human and horse, respectively, whereas the blue binding site is conserved across all four species. The green semicircles (not conserved) are additional binding sites. Downstream of the TSS, in the first intron of Parkin, an additional CREB/ATF binding site is located in human, mouse and cow. The consensus ATF4 binding site is written in bold letters. hsa, Homo sapiens (human); mmu, Mus musculus (mouse); bta, Bos taurus (cow); eca, Equus caballus (horse).

3.3.2 NFKB binding sites in the promoter of OPA1

I searched for Nuclear factor kappa B/c-rel (NFKB) binding sites in the human OPA1 promoter by Genomatix software and discovered four binding sites, which are conserved among the species human, chimp, rhesus monkey, mouse, rat and horse (Figure 3.7). One NFKB binding site in the human OPA1 promoter is located

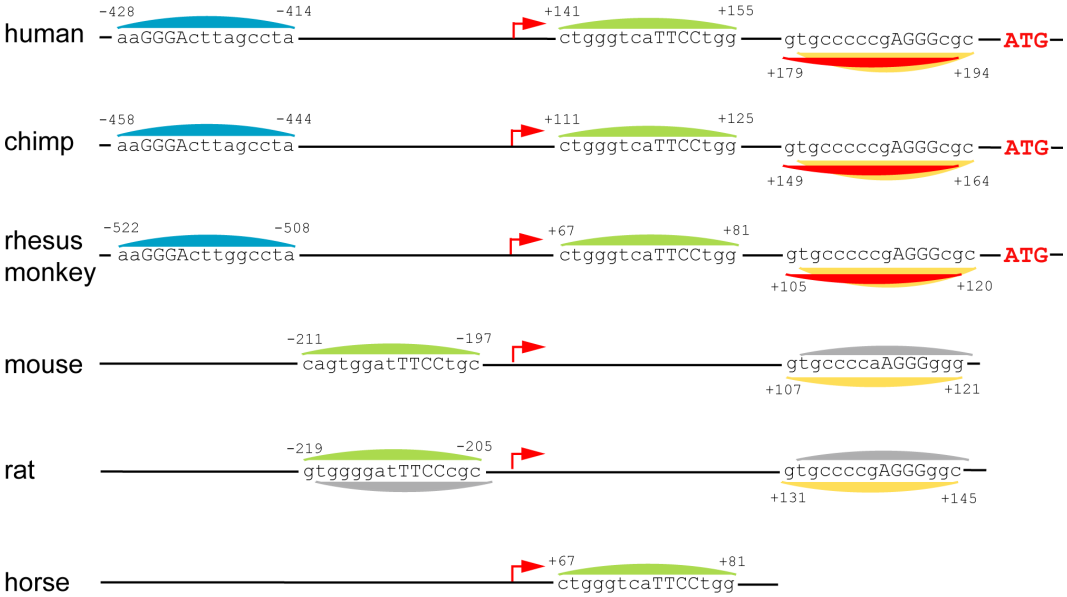
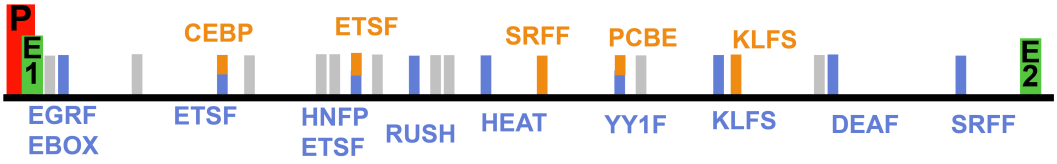


Figure 3.7: Conserved NFKB binding sites in OPA1 promoter of human, chimp, rhesus monkey, mouse, rat and horse. Red arrow indicates transcription start site (TSS) and positions are denoted relative to the TSS. The translation start site is shown by a red ATG. The NFKB binding sites are indicated by semicircles. Blue (V$NFKAPPAB.02), green (V$NFKAPPAB65.01), red (V$NFKAPPAB.01) and yellow (V$NFKAPPAB50.01) binding sites (semicircles) are predicted by four different binding motifs, which correspond to a Genomatix-defined family of seven matrices describing the NFKB binding site. The blue and green colored binding site upstream of the TSS and on the plus-strand is conserved between human, chimp and rhesus monkey as well as mouse and rat, respectively. The green colored binding site downstream of the TSS and on the plus-strand is conserved between human, chimp, rhesus monkey and horse. The red and yellow colored binding site on the minus-strand is conserved between human, chimp and rhesus monkey as well as human, chimp, rhesus monkey, mouse and rat, respectively. The gray semicircles (not conserved) are additional binding sites. Homo sapiens (human); Pan troglodytes (chimp); Macaca mulatta (rhesus monkey); Mus musculus (mouse); Rattus norvegicus (rat); Equus caballus (horse).

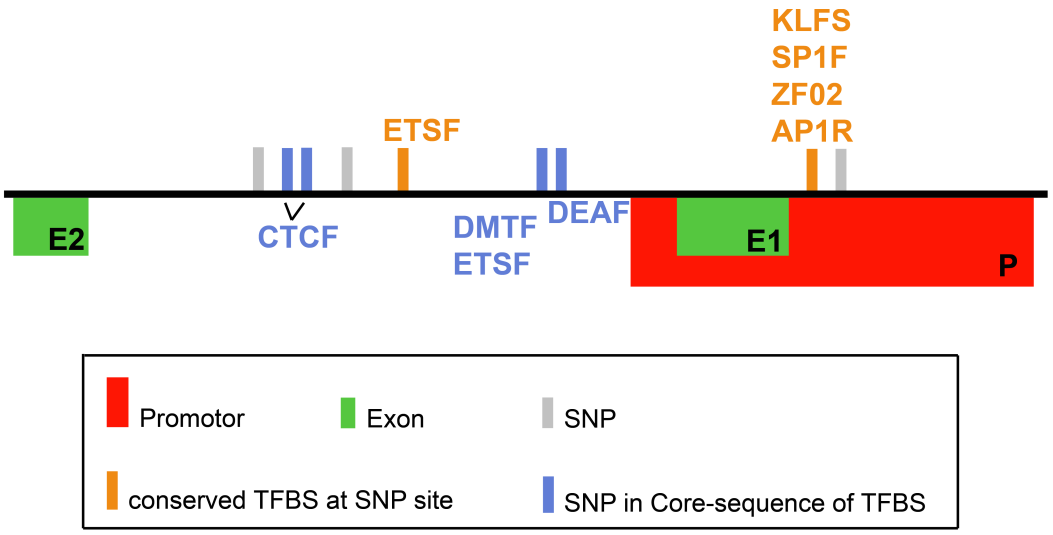
upstream of the TSS and three sites are arranged downstream of the TSS, whereas two sites with a similar motif are overlapping.

3.3.3 SNPs in TFBSs of MAPT and SNCA promoter and first intron

I analysed MAPT and SNCA promoter and first intron sequences for TFBSs by Genomatix software. PD associated SNPs are used to filter TFBS predictions in order to get TFs influencing PD pathogenesis. PDGene database is an aggregation of all published genetic association studies of PD phenotypes and contains 58 and 96 SNPs of MAPT and SNCA, respectively. MAPT and SNCA are the two top ranked genes in the PDGene database. According to PDGene database SNPs have not been found in the promoter of MAPT but 21 SNPs (supplemental Table A.14) have been found in the first intron. SNCA contains two and seven SNPs (supplemental Table A.15) in the promoter and first intron, respectively. The Figures 3.8 (a) and (b) show the SNPs in the promoter and first intron of MAPT and SNCA and TFBSs according to the two criteria conservation and core-sequence are colored orange and blue. The core-sequence is defined as the highest conserved, consecutive positions (usually 4) in the TFBS sequence. Some TFBSs match even both criteria highlighted



(a)



(b)

Figure 3.8: SNPs of MAPT (a) and SNCA (b) from promoter till second exon. The drawings show the region from promoter till second exon of MAPT and SNCA. SNPs in this region are marked by gray colored stripes. SNPs located in a conserved TFBS are colored orange and SNPs being part of the TFBS core-sequence are colored blue. The corresponding TF-family of the colored SNPs are written above or below.

by a bicolored stripe like CEBP, ETSF and PCBE. The TFBSs ETSF and PCBE in MAPT are conserved between human, chimp and cow. The SRFF (serum response element binding factor) BS lies in a region conserved between human, chimp and rat and the KLFS BS is additionally conserved to mouse. The strongest conserved TFBS of MAPT is CEBP in the species human, chimp, mouse, rat and cow. The ETSF BS of SNCA is conserved in human, chimp and dog and the four TFs AP1R (MAF and AP1 related factors), KLFS, SP1F and ZF02 are binding to a region additionally conserved to mouse and cow. Both genes have the TFBSs DEAF, ETSF and KLFS in common. The association of another TF YY1F and ETSF to PD are discussed in section 4.3.2 (Discussion).

3.3.4 TFBSs in PRKAA1, PRKAA2 and PRKAG1

The kinases PRKAA1, PRKAA2 and PRKAG1 belonging to the AMP-activated protein kinase family are possibly regulated by similar TFs at which the AMPK alpha genes (PRKAA1 and PRKAA2) are supposed to be regulated by the same TFs. In total 38 TF-families were selected according to GO terms including 'stress', 'metabolic' and 'biosynthetic', because TFs regulating metabolic, bio-energetically and stress pathways are most interesting in the context of PD and AMPK (PRKAA1, PRKAA2, PRKAG1). I searched for these 38 TF-families in AMPK genes by the Genomatix software.
The TFBSs analysis of the human promoters yields several common TFs (Figure 3.9). ZF02 and ETSF are in common for all three genes PRKAA1, PRKAA2 and

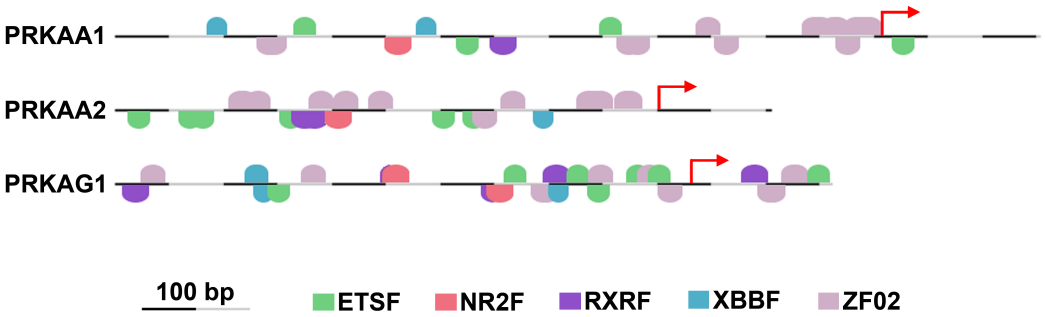


Figure 3.9: Common TFBSs in PRKAA1, PRKAA2 and PRKAG1. The picture shows the promoter of the three genes with the BSs of five common TFs marked by different colors. Red arrow indicates TSS.

PRKAG1 and additionally the corresponding TFs have at least four BSs predicted in the promoter sequences of all three genes. Furthermore, RXRF, XBBF (x-box binding factors) and NR2F occur in the promoters of the three genes, but partially only one site in the promoter is predicted. Additionally, PRKAA1 and PRKAA2

have the following TFs in common with at least two binding sites in each promoter: EBOX, NRF1, STAT (signal transducer and activator of transcription). The TFs IRFF (interferon regulatory factors), HOXF (paralog hox genes 1-8 from the four hox clusters A, B, C, D) and YY1F are also detected in the promoter sequences of PRKAA1 and PRKAA2.

The promoter sequences of PRKAA1 from rhesus monkey, chimp, human, mouse, rat, cow, rabbit, chicken, zebra finch, western clawed frog, zebra fish and opossum were searched for conserved TFs of the initially defined TF group.

Name	**Species related promoter sequence**
YY1F	rhesus monkey, chimp, human, mouse, rat, cow, rabbit, chicken, zebra finch, western clawed frog, zebra fish, opossum
ZF02	rhesus monkey, chimp, human, mouse, rat, cow, rabbit, zebra finch, opossum
STAT	rhesus monkey, chimp, human, mouse, rat, cow, opossum
EBOX, ETSF, NRF1	rhesus monkey, chimp, human, mouse, rat, cow
IRFF	rhesus monkey, human, cow, rabbit, zebra fish, opossum

Table 3.9: Conserved TFBSs in PRKAA1 promoter of 12 species. The first column describes the conserved TF-family among the species in column two.

Seven TFs were discovered to be conserved in at least six species including also far-related animals. YY1F is conserved among all 12 species and the mostly conserved other TFs are common to at least six species (Table 3.9).
Chimp, human, mouse, rat, cow, chicken, western clawed frog, zebra fish and opossum promoter sequences of PRKAA2 were scanned for the initially defined TFs conserved in at least three sequences. Eight TFs were found and the strongest conserved TF between eight species is like in PRKAA1 YY1F. The other conserved TFs in different species are listed in Table 3.10. The five TFs YY1F, ZF02, EBOX, ETSF and NRF1 are conserved in PRKAA1 as well as in PRKAA2.

Name	**Species related promoter sequence**
YY1F	chimp, human, mouse, rat, cow, chicken, western clawed frog, zebra fish
ZF02	chimp, human, mouse, rat, cow, opossum
EBOX	chimp, human, rat
ETSF, RXRF, MYOD	chimp, human, cow

Name	Species related promoter sequence
NR2F	chimp, human, western clawed frog
NRF1	chimp, human, mouse

Table 3.10: Conserved TFBSs in PRKAA2 promoter of nine species. The first column describes the conserved TF-family among the species in column two.

Nine species (chimp, human, mouse, rat, dog, cow, pig, horse and rabbit) of PRKAG1 have conserved BSs of ETSF, FKHD (fork head domain factors), NR2F, P53F (p53 tumor suppressor), PERO (peroxisome proliferator-activated receptor), RXRF, SF1F (vertebrate steroidogenic factor) and ZF02. The TFs ETSF and ZF02 are additionally conserved in PRKAA1 and PRKAA2.

3.4 Computational identification and experimental validation of microRNAs binding to ADAM10

3.4.1 MicroRNA prediction

The established workflow (Figure 3.10) obtains miRNAs with possible binding sites to the human ADAM10 3'UTR. The human miRNAs with its binding sites to human

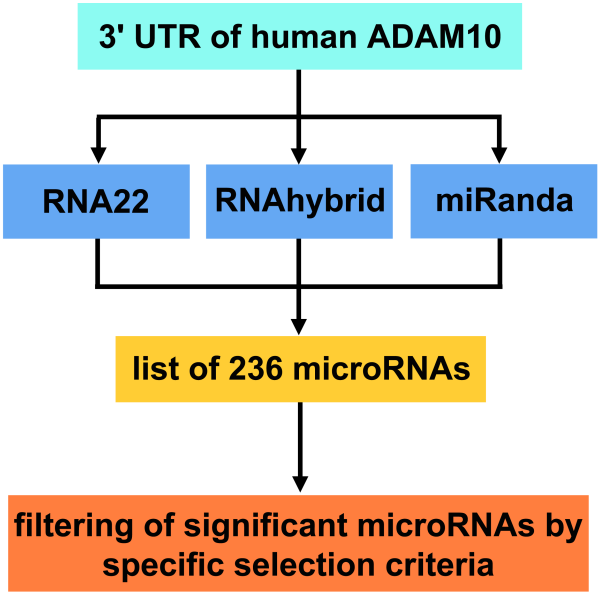


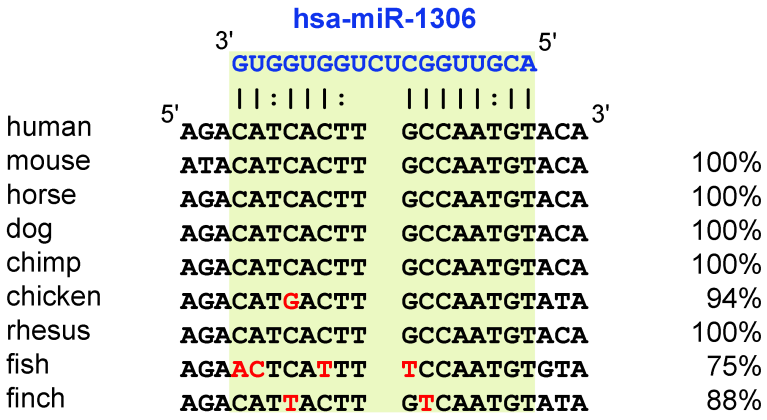
Figure 3.10: Computational prediction of miRNAs binding to ADAM10 3'UTR. Three programs RNA22, RNAhybrid and miRanda are used for the prediction of miRNAs binding to the human ADAM10 3'UTR. After retrieving a list of 236 miRNAs by RNA22, RNAhybrid or miRanda I extracted the relevant miRNAs according to selection criteria: prediction by at least two programs, differential regulation in AD patients [Cogswell et al., 2008], tissue-specific expression (MGI), binding to AD key genes, corresponding mouse miRNA binding to the mouse ADAM10 3'UTR sequence predicted by at least two of the three programs, additional prediction by TargetScan and microRNA and evolutionary conservation.

ADAM10 3'UTR were predicted by the three programs, RNA22, RNAhybrid and miRanda.
122 miRNAs are predicted by at least two programs to bind to human ADAM10 3'UTR sequence and 52 of them are significant according to expression and the following selection criteria:

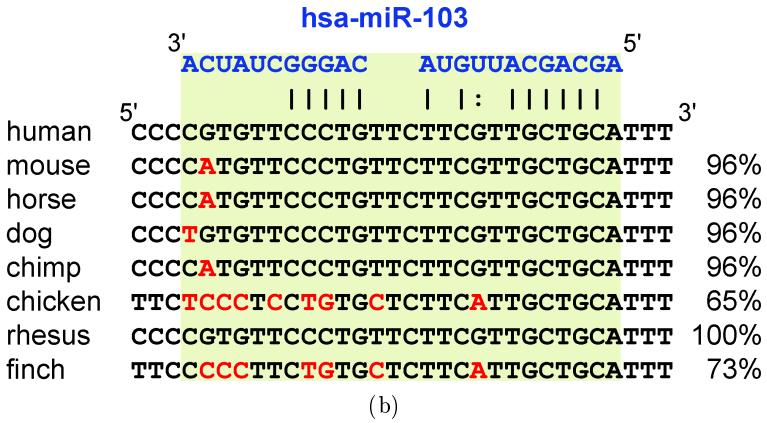
- at least two programs predict the miRNA binding site
- expression of the miRNA in brain provided by the MGI database

- regulation of the miRNA in AD as described by Cogswell and colleagues (2008) in the tissues hippocampus, cerebellum and medial frontal gyrus [Cogswell et al., 2008]
- miRNA binding to a gene, which is involved in AD revealed by literature search
- corresponding mouse miRNA predicted to bind to the mouse ADAM10 3'UTR sequence by at least two of the three programs
- prediction of the miRNA by other webtools such as TargetScan and microRNA

The 52 miRNAs incorporate none-conserved as well as conserved miRNAs or rather miRNA binding sites. Filtering according to conservation helps to find functionally active miRNA binding sites, additionally reduces the number of miRNAs and improves the selection of candidates for further experimental validations. In the analysis eleven miRNA binding sites are conserved across at least seven species, at which four of these miRNAs are also conserved in the evolutionarily far related species zebra fish. The only miRNA binding site miR-1306 in the human ADAM10 3'UTR predicted by all three programs is also conserved in the evolutionarily far related zebra fish as well as in mouse, horse, dog, chimp, chicken, rhesus monkey and zebra finch (Figure 3.11 (a)). This binding site is located on chromosome 15 positions 58889309–58889324 and the programs RNA22, RNAhybrid and miRanda predicted the binding energy -32.29, -25.7 and -22.55 kcal/mol, respectively. The conservation of the miRNA binding site sequence between human and the species mouse, horse, dog, chimp as well as rhesus monkey is 100%. The conservation of chicken, zebra finch and zebra fish to human in this binding region is 94%, 88% and 75%, respectively.
Other miRNAs predicted to bind to human ADAM10 3'UTR are miR-103 as well as miR-107 both having the same binding site located on chromosome 15 positions 58889443–58889468. This site is predicted by the two programs RNAhybrid and miRanda with binding energy -27.9 and -23.66 kcal/mol for miR-103, respectively, as well as -26.2 and -22.28 kcal/mol for miR-107, respectively. The conservation of the miRNA binding site sequence between human and the species mouse, horse, dog and chimp is 96%, while the conservation of chicken, rhesus monkey and finch to human is 65%, 100% and 73%, respectively (Figure 3.11 (b) and (c)). Additionally, miR-202, miR-423-5p, miR-503, miR-184 and miR-922 bind also to the conserved binding region chromosome 15 positions 58889443-5889473 and miR-330-5p (chr15:58889149-58889178), miR-671-5p (chr15:58889720-58889745) and miR-432 (chr15:58889688-58889718) bind to a region with good conservation also to the evolutionarily far related species zebra fish.



(a)

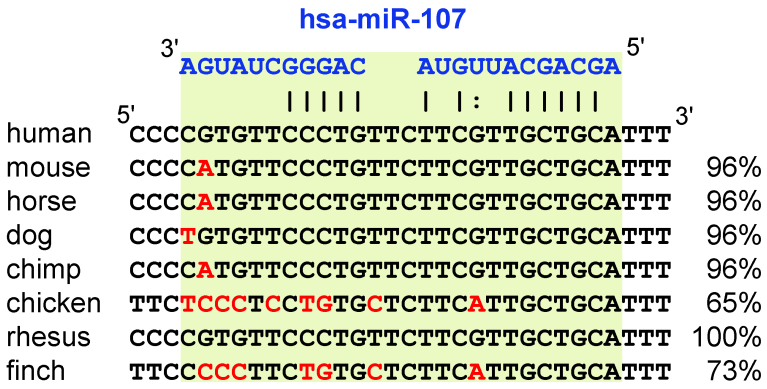


(b)

Figure 3.11: Conservation of the miR-1306 and miR-103 binding site within the ADAM10 3'UTR

I choose the three most interesting miRNAs 1306, 107 and 103 according to predicted binding energy, selection criteria and conservation as well as relation to AD. Additional analyses were performed with the AlzGene database, further miRNA target site prediction databases, GO, literature mining and validation experiments to identify the involvement in AD.

Twelve target genes of miR-1306 (Figure 3.12) were predicted concerning their function in AD, brain, nervous system or other neurodegenerative diseases and graphically interrelated by Path Designer (Ingenuity). The predictions rely on TargetScan while the dedicated functions are Ingenuity Expert Findings or from GO.



(c)

Figure 3.11: Conservation of the miR-107 binding site within the ADAM10 3'UTR. The figure shows the conservation of the miR-1306 (a), miR-103 (b) and miR-107 (c) binding region (light green) between different species. The blue sequence represents the miRNA binding to the DNA of different species as listed on the left side. On the right side the conservation of the miRNA binding region (light green) from different species to human sequence is given. Nucleotide mismatches in the binding region to human binding region are marked in red. The lines and colons below the miRNA sequence show perfect nucleotide matches and G:U/T wobble pairs, respectively.

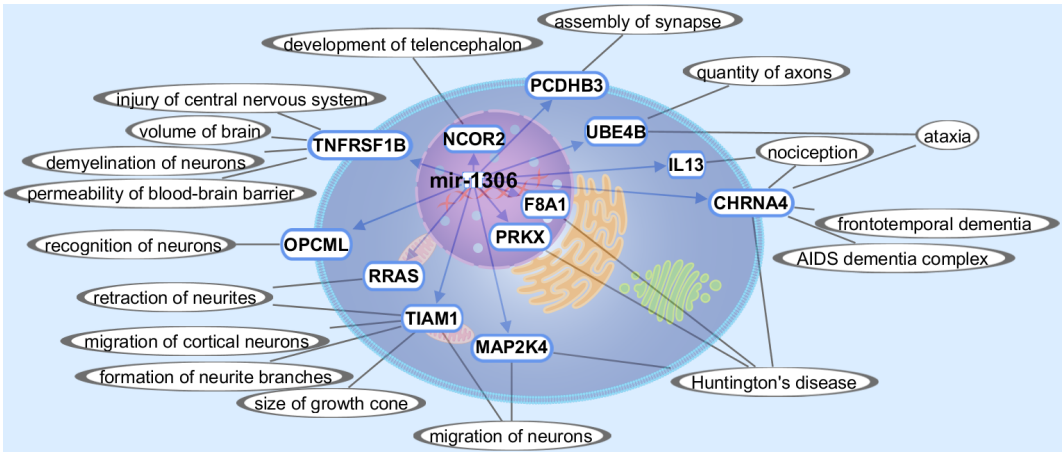


Figure 3.12: Target gene predictions for miR-1306. Twelve target genes of miR-1306 were predicted by TargetScan concerning their function in AD, brain, nervous system or other neurodegenerative diseases and graphically interrelated by Path Designer (Ingenuity). The miRNA is located in the nucleus and the predicted target genes indicated by blue arrows and blue framed ellipses are located according to their GO either in the nucleus, cytoplasm or membrane. The functions of the target genes are denoted by gray lines and gray framed ellipses and derived from GO or Ingenuity expert findings, which are substantiated by literature.

3.4.2 Prediction of miRNA target genes and their relation to AD

I searched for target gene predictions of the miR-1306, miR-103 and miR-107 in the six databases miRBase, microRNA, PicTar, PITA, RNA22, as well as TargetScan (Figure 3.13). The combination of different miRNA target site prediction databases

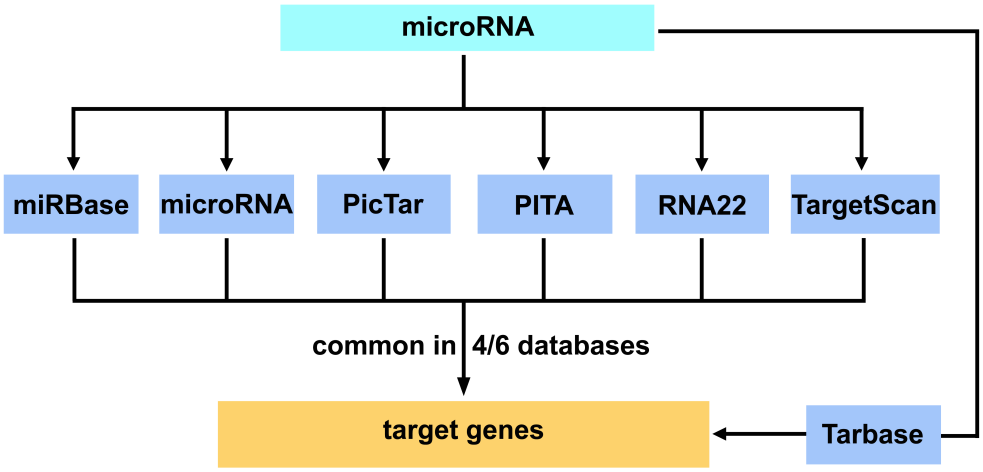


Figure 3.13: Workflow for miRNA target site prediction. Seven databases miRBase, microRNA, PicTar, PITA, RNA22, TargetScan and Tarbase, which is experimentally supported by miRNA targets, are incorporated in the prediction of miRNA target sites. The input is a miRNA and the output is a set of target genes common to four out of six databases. Tarbase contains experimentally supported miRNA targets and not in-silico predicted miRNA binding sites: all miRNA target genes from Tarbase are automatically included in the output of the analysis.

and the restriction of the output to four out of six databases in the case of miR-103 and miR-107 are important to reduce the amount of false positive miRNA target sites in the end. In the case of miR-1306 there is no restriction of the output, because only the database PITA incorporates predictions for this miRNA. Additionally, Tarbase can be included in the analysis. All miRNA target sites from Tarbase are automatically included in the output of the analysis. In this case Tarbase does not contain target genes for the miRNAs: miR-103, miR-107 and miR-1306. The six miRNA target site prediction databases miRBase, microRNA, PicTar, PITA, RNA22, and TargetScan contain in total 18915 different (according to EntrezGene) human genes at which PITA alone contains 16819 genes. The analysis resulted in 156 and 157 target genes for miR-103 and miR-107, respectively, common in four out of six databases, and 890 target genes for miR-1306 of database PITA.
AlzGene database was used to see for which of the miRNA target genes a genetic association with AD has been found. The six miRNA target site prediction databases

have in total 18915 genes and PITA contains 16819 genes, at which 636 and 591 genes are listed in AlzGene database, respectively. The overlap between AlzGene database genes and the target genes of miR-103, miR-107 as well as miR-1306 are 12, 14 (Figure 3.14) and 24 genes, respectively (supplemental Table A.16). MiR-103

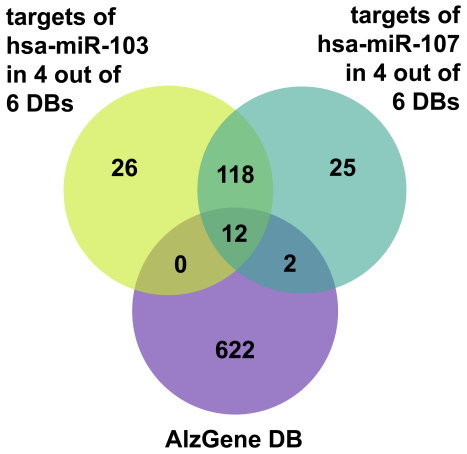


Figure 3.14: The Venn diagram shows the significant overlap of the target genes of miR-103 and miR-107 common in four out of six databases as well as the genes of AlzGene database.

and miR-107 have 130 target genes in common (Figure 3.14, supplemental Table A.16). Applying a Fisher's exact test I got a p-value of 0.0065, 0.0009 and 0.1904 for the overlap of miR-103, miR-107 and miR-1306, respectively, with the AlzGene database.

3.4.3 Gene ontology

GO analysis with the predicted target genes of the three miRNAs was done to validate the functionality of the miRNAs and their involvement to AD. With the help of the literature mining tool Pathway Studio I searched for molecular functions and biological processes common to the target genes of the three miRNAs.

Name	Overlap	p-value	FDR	GO
Wnt receptor signaling pathway	7	7.68×10^{-6}	0.001365	biological process
nervous system development	9	0.000367	0.018655	biological process
dendrite development	3	0.000613	0.022958	biological process
positive regulation of anti-apoptosis	3	0.00072	0.024377	biological process
learning	3	0.000777	0.025123	biological process
regulation of neuron apoptosis	2	0.001854	0.030698	biological process
brain development	5	0.002298	0.035521	biological process

Name	Overlap	p-value	FDR	GO
calcium ion binding	15	0.001105	0.036862	molecular function
Notch signaling pathway	3	0.003259	0.040504	biological process
cerebellum development	2	0.006388	0.049366	biological process
metal ion binding	27	0.041212	0.124150	molecular function

Table 3.11: Selected GO terms with an enrichment of miR-103 target genes. The table shows the GO entities (Name) with enrichment of predicted target genes of miR-103. Additionally a p-value and FDR value is given. Overlap is the number of target genes enriched in the GO entity (column 1) belonging to a specific ontology category (column 5).

Name	Overlap	p-value	FDR	GO
learning	4	3.33×10^{-5}	0.004544	biological process
brain development	6	0.000351	0.013933	biological process
nervous system development	9	0.00041	0.013939	biological process
dendrite development	3	0.000641	0.016504	biological process
positive regulation of anti-apoptosis	3	0.000752	0.01791	biological process
regulation of neuron apoptosis	2	0.00191	0.027435	biological process
Notch signaling pathway	3	0.003398	0.041090	biological process
calcium ion binding	14	0.002465	0.048950	molecular function
cerebellum development	2	0.006575	0.051159	biological process
metal ion binding	28	0.018689	0.079914	molecular function
Wnt receptor signaling pathway	3	0.035972	0.087148	biological process

Table 3.12: Selected GO terms with an enrichment of miR-107 target genes. The table shows the GO entities (Name) with enrichment of predicted target genes of miR-107. Additionally a p-value and FDR value is given. Overlap is the number of target genes enriched in the GO entity (column 1) belonging to a specific ontology category (column 5).

Name	Overlap	p-value	FDR	GO
metal ion binding	153	1.86×10^{-8}	6.62×10^{-6}	molecular function
nervous system development	33	6.86×10^{-8}	1.36×10^{-5}	biological process
brain development	17	7.06×10^{-6}	0.000785	biological process
calcium ion binding	56	2.08×10^{-5}	0.002116	molecular function
learning	6	0.000321	0.014353	biological process
dendrite development	4	0.010762	0.10343	biological process
Wnt receptor signaling pathway	9	0.012744	0.113187	biological process

Name	Overlap	p-value	FDR	GO
positive regulation of anti-apoptosis	4	0.012987	0.113187	biological process
Notch signaling pathway	5	0.01916	0.124447	biological process
cerebellum development	3	0.024124	0.124447	biological process
regulation of neuron apoptosis	2	0.043731	0.173128	biological process

Table 3.13: Selected GO terms with an enrichment of miR-1306 target genes. The table shows the GO entities (Name) with enrichment of predicted target genes of miR-1306. Additionally a p-value and FDR value is given. Overlap is the number of target genes enriched in the GO entity (column 1) belonging to a specific ontology category (column 5).Overlap is the number of target genes enriched in the GO entity (column 1) belonging to a specific ontology category (column 5).

The Tables 3.11, 3.12 and 3.13 show significant p-values and FDR for selected common GOs of miR-103, miR-107 and miR-1306, respectively. In total, 37, 45 and 118 significantly overrepresented molecular functions and 154, 163 and 369 significantly overrepresented biological processes were identified for miR-103, miR-107 and miR-1306, respectively.

3.4.4 Literature mining

A network (Figure 3.15) containing already published interactions of miR-103 and miR-107 with genes involved in AD or included in the AlzGene database was established using the literature mining tool Pathway Studio. Genes connected to the miRNAs are listed in the AlzGene database, involved in miRNA processing or show a connection to AD pathogenesis by being involved in processes playing a role in AD. Additionally, both miRNAs are linked by genes and linoleic acid, which is known to affect AD [Liu et al., 2004].

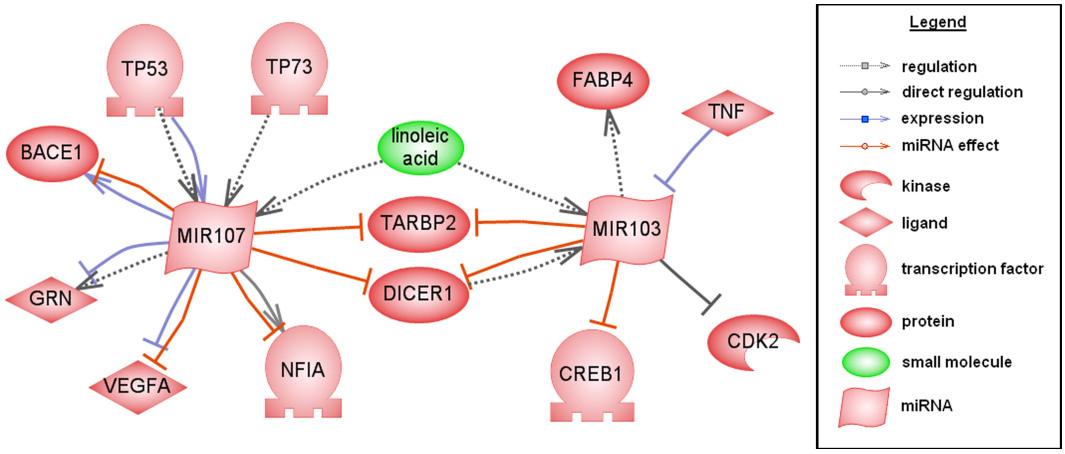


Figure 3.15: Interaction network of miR-103 and miR-107. The network (established by Pathway Studio) shows already published interactions between miR-103 as well as miR-107 and genes known to be involved in AD or neurodegenerative processes. The different types of proteins or small molecules are indicated by different symbols and the various interactions like regulation, expression and miRNA effect are also displayed by various arrows.

3.4.5 Experimental validation of bioinformatically predicted miRNAs

The following experiments were done in the context of a cooperation by Kristina Endres and Sven Reinhardt [Augustin et al., 2012]. For the purpose of validation cotransfection experiments with a Gaussia reporter vector harbouring the 3'UTR of ADAM10 downstream of the luciferase coding sequence together with the respective miRNAs were performed. As a positive control miR-122 was used. The programs RNA22, RNAhybrid and miRanda predicted a miR-122 binding site to human ADAM10 3'UTR with binding energy -31.2, -25.6 and -21.84 kcal/mol (on average: -26.21 kcal/mol), respectively, comparable to the miR-1306 binding site prediction. Time resolved measurement revealed that 48 hrs incubation period resulted in maximal effects of the miRNAs in SH-SY5Y cells (data not shown): while the negative control miRNA had no impact on luciferase activity measured in the cell supernatant, miR-122 reduced the luminescent signal to 57 % as compared to water treated control cells (Figure 3.16). The three miRNAs identified by bioinformatical approaches and integration of literature mining all showed a significant decreasing effect on the ADAM10 3'UTR-reporter construct: miR-1306 lowered the luminescent signal to 72%, miR-103 to 55% and miR-107 to 48% of control. While miR-103 and miR-107 target the same DNA sequence within the ADAM10 3'UTR (see Figure 3.11), miR-1306 has a binding site in closer proximity to the stop codon. Therefore miR-1306 and miR-103 or miR-107 were combined, respectively, and ob-

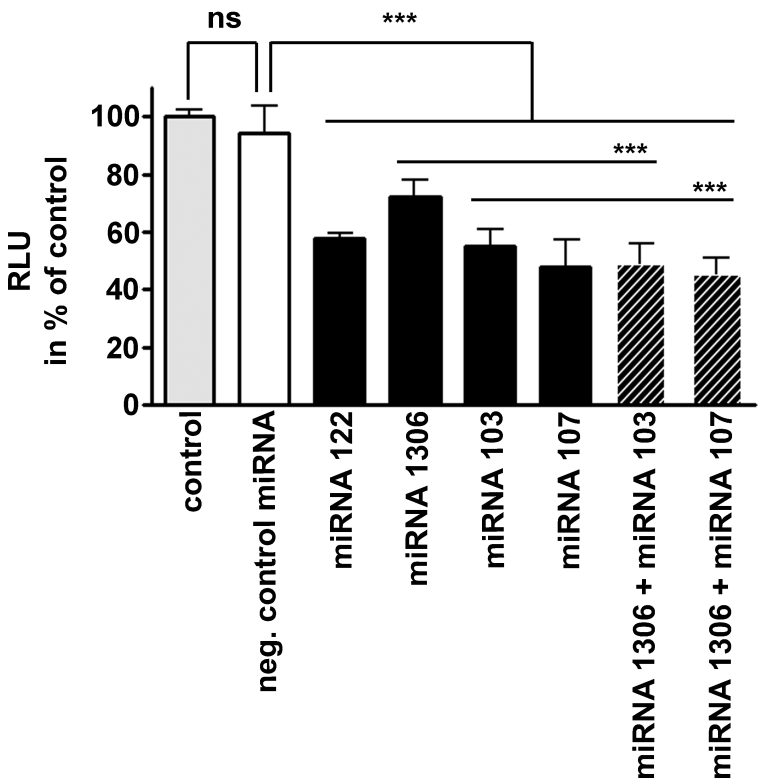


Figure 3.16: Experimental validation of selected miRNAs. SH-SY5Y cells were transiently cotransfected with the Gaussia reporter vector harbouring the 3'UTR of ADAM10 downstream of the luciferase coding sequence together with the respective miRNAs. MiR-122, which has been described by Bai et al. [2009] to target ADAM10, served as a positive control. After 48 hrs of incubation, luminescence was measured in the cell supernatant. Values obtained for control (water) treated cells were set to 100%, data represent mean ±SD of three independent experiments performed in triplicate. *** $P < 0.001$, ns = not significant, RLU = Relative light units. [Augustin et al., 2012]

served a slight but significant synergistic effect (miR-1306 vs. miR-1306/103 $p < 0.001$; miR-1306 vs. miR-1306/107 $p < 0.001$).

Chapter 4

Discussion

4.1 Evaluation of bioinformatics approach

This part of the discussion concentrates on the evaluation of two established workflows for the identification of TFBSs modules and miRNAs regulating key genes of diseases. Moreover, the implemented mSVM-RFE method was applied on different microarray datasets with evaluating the function by ROC curve.

4.1.1 Identification of TFBS modules in AD key genes and coregulated genes

The two different approaches of the workflow start at different points, but in the end similar modules are obtained. The first approach mainly consists of in silico promoter analyses, which is a commonly accepted procedure to predict TFBSs as well as TFBSs modules, with additional verification by differentially expressed AD-related genes (Figure 3.1). Whereas the second part of the workflow from the beginning focuses on experimental AD datasets to find coregulated genes differentially expressed and regulated by the same TFBSs modules.
Sequence conservation is highly predictive in identifying functional active sites [Capra and Singh, 2007]. The hypothesis is that the key genes of AD (beta- and gamma-secretase) are supposedly coregulated and therefore I suppose, that they have the same TFBSs in their promoter regions. TFBSs modules regulating key genes of AD are thought to play a role in AD. These AD key genes are predominantly responsible for APP cleavage and involved in plaque formation.
The modules obtained from the second approach mostly contain one TFBS, which is not conserved between human and mouse APP promoter and therefore these modules could not be found by the first approach focussing on conserved TFBSs. The second approach focusses on experimental results either from AD patients or AD mouse models analysed by bioinformatical methods (Figure 3.2). The aim is to get all groups of coregulated genes of the array, which are differentially regulated among the classes control and AD and therefore could play a role in AD. In the end these coregulated genes are taken and searched for TFBSs modules, which could regulate these genes. The idea is, that coregulated genes, which show the same expression profile, are regulated by the same mechanisms such as TFBSs modules. Promoter

sequences of the coregulated genes were searched for the same binding sites of the TFs of the module. Slightly different modules between the two approaches are obtained, but in the end, I have two modules in common, which connect the in silico promoter analysis with the expression data from microarrays of AD.

The in silico promoter analysis focusses on TFBSs modules and not on single TFBSs. Although TFBSs can occur almost anywhere in the promoter and do not show any pattern with respect to location, TFBSs are often grouped together and such functional modules have been described in many cases such as in Werner [1999]. The arrangement of TFBSs of a promoter module seems to be much more limited than it is suggested by the variety and distribution of TFBSs in the promoter sequence [Werner, 1999]. However, TFBSs modules found in the promoter of genes suggest functional connection but do not prove it [Gailus-Durner et al., 2001]. The common regulation of genes from different gene classes can depend on regulatory modules of TFBSs, which are often conserved regions in the promoter sequences and therefore can be identified, while other TFBSs are non conserved between the promoters [Klingenhoff et al., 2002, Werner, 2003]. Therefore, promoter sequence conservation is considered in the in silico promoter analysis by human and mouse APP as well as the used software from Genomatix is also based on sequence conservation.
The Genomatix software used for TFBS prediction has several advantages to other prediction programs. It provides a database of all promoter sequences of various species and a huge amount of all known TF matrices and TFBS modules for prediction, which are regularly updated. In addition, setting up multiple sequence alignment considering conservation and establishing TFBS modules by oneself is easily embedded in the software. The graphical output of TFBS predictions offers a comfortable overview of the analysis.
In this study, commercial software has been used as part of the bioinformatics workflow, but comparable open-source software and databases can be used as well. VISTA (`http://genome.lbl.gov/vista/index.shtml`) and JASPAR (`http://jaspar.genereg.net/`) can be used for the analysis of TFBSs. For literature mining, STRING (EMBL, `http://string-db.org/`) and EXCERBT (MIPS, `http://tinyurl.com/excerbt/`) are available, and for finding tissue expression or binding partners of the genes, the GeneCards database (Weizmann Institute of Science, `http://www.genecards.org/`) is one alternative.

4.1.2 Evaluation of mSVM-RFE method

The implemented multivariate mSVM-RFE method was applied on two different microarray datasets concerning depression and PD. The multivariate method was compared to univariate p-value computation and evaluated by qPCR results [per-

sonal communication with Regina Widner (MPI)].
Univariate DJ-1 knockout microarray analysis yields only 12 significantly regulated genes including DJ-1 and according to the adjusted p-value smaller than 0.35 (supplemental Table A.13). In many cases more significantly differential regulated genes are observed in a microarray experiment. In the case of HR versus LR the comparison of differentially expressed SAGE genes with genes determined by mSVM-RFE occurring in at least 18 gene selections yields nearly twice as much common genes than compared to the differentially expressed genes on the microarray (adjusted p-value < 0.05). Leading to the first statement, that mSVM-RFE is not as strict as the univariate t-statistics using an adjusted p-value in dividing the genes in unregulated and differentially expressed genes. Differentially expressed genes not recognized by the adjusted p-value are detected by mSVM-RFE and not excluded for further analyses and validations.

For this reason and to consider genetical interactions multivariate analysis by mSVM-RFE was adopted. MSVM-RFE is not as stringent as t-statistics with subsequent computation of the adjusted p-value considering also interactions between the genes and might detect interesting genes not considered by the adjusted p-value. MSVM-RFE restricts the DJ-1 microarray to 40 genes best differentially regulated (frequency = 50) among wildtype and DJ-1 knockout incorporating also the 12 regulated genes with an adjusted p-value < 0.35. This expanded amount of differentially regulated genes is appropriate for further GO analysis and literature mining. In addition the combination of univariate and multivariate methods as well as additional gene information help to sort the genes prior to experimental validation. The more analysis evaluate the gene to be differentially regulated in DJ-1 knockout versus wildtype the more this gene should be validated in the lab.

Due to the validation of genes from the HR versus LR microarray dataset by qPCR [personal communication with Regina Widner (MPI)] the mSVM-RFE function could be evaluated. For dividing genes in possibly differentially regulated genes and not-regulated genes the frequency threshold (frequency of a gene occurring in the gene selections) of mSVM-RFE is set more stringent to at least 38. This cut-off point of 38 is chosen, because most of the genes pass with few failures yielding the highest possible sensitivity and specificity. The ROC curve (Figure 3.5) identifies the optimal cut-off point, where the false positive rate is lowest (x-axis) and the true positive rate is highest (y-axis). This leads to a high sensitivity of 87% and a lower specificity of 56% of the mSVM-RFE analysis. This is not conflicting, because the higher one evaluation value, the lower is the second evaluation value [Lalkhen and McCluskey, 2008]. High sensitivity means nearly all differentially expressed genes are predicted by mSVM-RFE being regulated and only few interesting genes are lost. The specificity

is lower than the sensitivity, but half of the unregulated genes are correctly rejected by mSVM-RFE, which recognizes almost 87% of all differentially expressed genes.

Concluding, mSVM-RFE recognizes more differentially regulated genes by considering the interactions amongst the gene expression levels than an adjusted p-value < 0.05, which was calculated by a t-test for the expression levels of each gene separately. Therefore, interactions of the genes are considered by multivariate methods and regulatory molecular networks can be detected by the analysis.

4.1.3 Established workflow for miRNA target site prediction

Up to date numerous miRNA target site prediction databases are available incorporating plenty of false positive miRNA target predictions. A specific context such as a particular disease or miRNA expression in certain brain regions is not considered by these databases. Additionally, conservation is often not enough considered and comparing predictions of different databases is hindered by various 3'UTR sequences used in databases. The database miTALOS represents an interesting approach identifying pathways regulated by miRNAs as well as considering tissue expression [Kowarsch et al., 2011]. The workflow established in this thesis (Figure 3.10) combines specific selection criteria for the identification of miRNAs influencing the pathogenesis of AD by regulating ADAM10.
Instead of looking up miRNA target sites of ADAM10 in the miRNA target site prediction databases, I predicted the binding sites of human miRNAs to human ADAM10 3'UTR by the three programs, RNA22, RNAhybrid and miRanda, with the aim to yield a more accurate miRNA target site prediction. The same 3'UTR sequence was used for all three prediction programs to compare the results. Due to the fact that the three prediction programs focus on different aspects for miRNA target site prediction (pattern-based search, seed matching, conservation, energy or structure) various properties of the target sequence as well as of the miRNA binding behaviour are covered (section 2.4.2).
To consider different aspects of the distinct prediction algorithms at least two programs should predict a miRNA binding site. Each program focusses on a specific characteristic of miRNA binding site and thus neglects other important aspects of the binding site considered by another different miRNA prediction program. Besides, the miRNA should be expressed in the tissue brain for regulating ADAM10 with regard to AD. Moreover, miRNAs showing differential expression levels in brain regions affected by AD such as hippocampus, cerebellum and medial frontal gyrus strengthen the assumption that they play a role in the pathogenesis of AD.

An additional confirmation of a miRNA being involved in AD is a binding site to a target gene, which is involved in AD, described in the literature. Thus the

miRNA might regulate also other AD key genes possibly leading to a coregulation of the disease genes by the miRNA. Furthermore, the miRNA prediction is strengthened by the corresponding mouse miRNA binding to the mouse ADAM10 3'UTR sequence predicted by at least two of the three miRNA prediction programs RNA22, RNAhybrid and miRanda. To confirm the predicted miRNAs also other webtools with precompiled results were considered by searching for miRNAs predicted to bind to ADAM10.

Consideration of conservation across species (Figure 3.11), including those obviously not developing AD, was chosen as a filter criterion because a correlation between miRNA conservation and disease susceptibility has in general been suggested by Lu et al. [2008]. The selection procedure therefore excludes non-conserved miRNA binding sites, which also might be relevant for development of the disease but are not lost by being included in the list of 52 miRNAs according to the specific selection criteria except conservation. In regard to AD being a human disease, future analysis of non-conserved miRNAs might also represent a valuable approach which I did not follow in the context of this thesis.

4.1.4 Prediction of miRNA target genes and their relation to AD

To confirm the relationship of the three miRNAs (miR-103, miR-107 and miR-1306) to AD I searched for target gene predictions of these miRNAs in the six databases miRBase, microRNA, PicTar, PITA, RNA22, as well as TargetScan (Figure 3.13). The database Tarbase has a special position, because in contrast to the other six databases Tarbase contains experimentally supported miRNA targets and not in-silico predicted miRNA binding sites. In contrast to the other six databases Tarbase doesn't contain false positive miRNA target sites. Therefore, target genes of Tarbase are directly included in the output, while target genes of the other six databases have to be in common to at least four databases to be incorporated in the output list.
As I focus on miRNAs playing a role in AD, I used the AlzGene database to investigate which target genes of the miRNAs have a possible genetic association with AD. The number of miRNA target genes for each miRNA listed in the AlzGene database was determined. The p-value shows that 12 (p-value = 0.0065) and 14 (p-value = 0.0009) are significant high numbers of overlapping genes between the target genes of miR-103 and miR-107 and the AlzGene database, respectively (Figure 3.14). The significant high amount of AD-related target genes suggests that miR-103 and miR-107 might play a role in AD. It is not remarkable that the p-value for the overlap of miR-1306 (p-value = 0.1904) with the AlzGene database is not significant as the restriction to four out of six databases was not possible, which consequentially leads

to the inclusion of a lot of false positive target genes in the 890 target genes of miR-1306.

4.1.5 GWAS

The publicly available databases AlzGene and PDGene are regularly updated aggregations of all published genetic association studies including GWAS performed on AD and PD phenotypes, respectively. These databases contain all considerable genetic association studies and key genes of AD and PD.
Concerning the SNPs the advantages and disadvantages of GWAS are discussed in several publications. Variants in a population are grouped into two forms: rare variants of large effect and common variants of small effect. A variant can be linked to different characteristics and mutations are associated to the same disease or trait in different populations [Visscher et al., 2012]. Genotype-by-environment interactions and epigenetic influences can not be discovered by GWAS as well as heritability is not incorporated. GWAS detected less variants of moderate effect, but thousands of common variants, which comprise nearly all the genetic variance in GWASs [Gibson, 2012].

4.2 Regulatory network in Alzheimer's disease

This section focusses on AD pathogenesis particularly on coregulated genes related to AD and regulated by TFBSs modules. A hypothetical regulatory network is established. In addition, miR-103, miR-107 and miR-1306 were identified to regulate ADAM10 expression.

4.2.1 TFBSs modules and their relation to AD

Three significant modules CTCF-EGRF-SP1F, CTCF-SP1F-ZBPF and KLFS-SP1F-ZBPF were obtained from the combination of in-silico promoter analysis with statistical analysis of microarrays of patients studies and mouse models (section 3.1). First, additional information about the TFs incorporated in the TFBSs modules was collected from TRANSPATH database. Several TFs show a link to AD either by expression in appropriate tissues or involvement in AD related pathways.

Several TFs of EGRF, SP1F, ZBPF and KLFS families show a mRNA or protein expression in brain at which mRNA of KLF3 and KLF7 are expressed in midbrain and hippocampus, respectively. An occurrence of EGR1, EGR2, KLF10 and KLF6 protein is seen in hippocampus, cerebral cortex, cerebellum and hippocampus, respectively. These brain regions are associated to AD. In AD patients gray matter is reduced in the midbrain [Whitwell et al., 2007] as well as plaques and tangles are observed in the hippocampus [Rapp et al., 2006]. Increased amyloid beta-peptides and atrophy as well as neuronal loss occur in cerebral cortex and cerebellum, respectively [Schmechel et al., 1993, Sjöbeck and Englund, 2001]. This strengthens the assumption, that the TFBSs and as a consequence the TFBSs modules play a regulatory function in AD pathogenesis.

The TFs CTCF and KLF1 are associated to the chromatin remodeling process. Chromatin remodeling, which is a dynamic and highly regulated process, is linked to AD and memory function. Alterations in chromatin structure occur in many neurodegenerative diseases like AD and the inhibition of histone deacetylases could be considered as a therapeutic treatment for AD patients with memory impairment [Lee and Ryu, 2010].

Fraser et al. [2010] showed an overall reduction of fatty acids in AD especially for stearic acids in frontal and temporal cortex as well as arachidonic acids in the temporal cortex. An additional intake of n-3 (omega-3) fatty acids and consumption of fish once a week possibly reduces the risk of suffering from AD [Morris et al., 2003]. TRANSPATH database reveals the TF SP1 to be involved in the fatty acid biosynthetic process.

Lipids are involved in the regulation of APP processing. A possibility for pre-

venting AD or therapeutic treatment of neurodegeneration could be to target lipid metabolism for example by lowering cholesterol levels [Grösgen et al., 2010]. The TF ZNF202 is part of the lipid metabolism.
The AD key gene PS1 (γ-secretase) is required for neural development especially for the decision between postmitotic neurons and neural progenitor cells. The absence of PS1 causes premature differentiation of neural progenitor cells [Handler et al., 2000]. In addition to PS1 also the TF SP3 incorporated in the TF-family SP1F regulates neuron differentiation.
MAPK and Wnt receptor pathway, which incorporate the TF KLF5, play a role in AD. An activation of MAPK pathways can be observed in vulnerable neurons of AD patients by regulating cellular processes, which are affected in AD [Zhu et al., 2002]. Wnt signaling pathway plays an important role in brain development and a damage in the Wnt pathway leads to the onset and development of AD [De Ferrari and Inestrosa, 2000].
Furthermore, brain development and especially environmental influences during this process are linked to AD by predefining APP expression and regulation in older brains. An exposure to plumb during brain development leads to a change in APP expression and this may influence the amyloidogenesis later in life [Basha et al., 2005]. The TF EGR2 is linked to AD by being involved in the brain development process.
A response to oxidative stress during development could be the inhibition of DNA methylation in APP, which leads to an increase of APP and as a consequence also of Aβ [Zawia et al., 2009]. Several TFs SP1, SP3, SP4 and KLF4 are known to be involved in the pathway response to stress and possibly influence AD via this process. Learning (EGR1, EGR3, KLF9) and Memory (EGR1, EGR3, SP4) as well as the nervous system development (KLF7, KLF9) are associated with AD via the APOE (Apolipoprotein E) receptors (LDL receptor related proteins), which are AD risk genes. This receptors are required for nervous system development, influence the expression of APP as well as learning and memory by controlling synaptic functions [Herz, 2009, Herz and Chen, 2006]. This processes participating in the progression of AD link the three TF families EGRF, SP1F and KLFS with the pathogenesis of AD.

Additionally, regulations of AD key genes by the transcription factors CTCF, EGR1, SP1, and ZNF202 as known from the literature confirm the three modules.
The TF EGR1 is known to upregulate the AD key gene PS2, which has an upstream promoter P1 and a downstream promoter P2. Neuronal cells show a EGR1 repression of P1 activity by 50% and P2 is upregulated by direct binding of EGR1 [Renbaum et al., 2003].

CTCF binds to APBβ domain, a nuclear factor binding site in proximal APP promoter, as well as acts as a transcriptional activator in the promoter of APP [Vostrov and Quitschke, 1997]. APOE is repressed by ZNF202 (ZBPF) according to TRANSPATH database. The mRNA regulation of APOE and ZNF202 is inverse: APOE is upregulated while ZNF202 shows reduced expression [Langmann et al., 2003].
SP1 can regulate both β-secretase genes, BACE1 and BACE2. The overexpression of SP1 upregulates BACE2 promoter activity and the TF is required for BACE2 transcription [Sun et al., 2005]. The expression of BACE1 is controlled by SP1, which was also shown in vivo, in addition BACE1 is a downstream target of the TF SP1 [Christensen et al., 2004].
Additionally, SP1 activates the transcription of PS1 by several BSs upstream and downstream from the initiation site [Pastorcic and Das, 1999]. Two GC-elements, which are SP1 (GC-box factor) sites, in the promoter of APP regulate the expression [Pollwein et al., 1992].
In addition, I found SP1 significantly regulated on the LOAD patients microarrays (FDR = 3.4×10^{-9}), which is in agreement with the results of a dysregulation and upregulation of this transcription factor in AD [Citron et al., 2008, Santpere et al., 2006]. This strengthens the occurrence of the TF SP1F in all three modules possibly regulating AD related genes.

Taken together, all TF families of the modules are linked to AD. The TFs are either expressed in the appropriate brain regions, involved in pathways playing a role in AD or regulate AD key genes as known from literature.

Coregulated genes in AD

Literature mining revealed known relations between the coregulated target genes of the TFBSs modules and AD confirming the modules CTCF-EGRF-SP1F, CTCF-SP1F-ZBPF and KLFS-SP1F-ZBPF. Furthermore, several target genes and their linkage to AD are discussed to strengthen the significance of the three TFBSs modules.
GNAS is predicted to be the target gene of the first module CTCF-EGRF-SP1F (section 3.1.3) and involved in calcium signaling and long-term depression. An upregulation of calcium signaling leads to amyloid metabolism with neuronal cell apoptosis [Berridge, 2010]. Another target gene Gsk3b regulates the APP accumulation after Aβ formation [Takashima et al., 1995]. An over-representation of coregulated genes incorporating also Gsk3b in the AD as well as Wnt pathway was revealed. Additionally, the TF KLF3 (BKLF), a member of the KLFS TF family, which is incorporated in the third module KLFS-SP1F-ZBPF, binds the AD patients target

gene CTBP2 in vivo [Turner et al., 2003], which is also involved in the Wnt signaling pathway. Wnt pathway has been found to prevent neurodegenerative diseases like AD by inhibiting Aβ-dependent cytotoxic effects [Cerpa et al., 2009].
CTBP2 (target gene of AD patients dataset) a target gene of the module CTCF-SP1F-ZBPF (section 3.1.4) and KLFS-SP1F-ZBPF (section 3.1.5) binds ZNF219, a member of the ZBPF TF family, in vitro [Hildebrand and Soriano, 2002]. The target gene Cdc42 of the module CTCF-SP1F-ZBPF (section 3.1.4) seems also be involved in AD by regulating synapse formation in neurons and Cdc42-GTPase activity is increased in neurons stimulated with Aβ_{1-42}. Additionally, Cdc42 is upregulated in neuronal populations of AD brains in comparison to controls [Mendoza-Naranjo et al., 2007]. In the MAPK pathway the coregulated genes of the double transgenic mice dataset are significantly enriched incorporating the target genes Cdc42, Map2k4, Mapk1, Mapk9, and Ppm1a. Some studies indicate that Mapk9 possibly interacts with SP1 and the JNK pathway targets SP1 [Chuang et al., 2008, Higuchi et al., 2004].

Overlapping target genes of the module between different microarray studies provide more evidence for the functionality of the module. Additional text mining revealed a strong connection of seven genes to AD and neurodegeneration.
The target gene EIF5 (eukaryotic translation initiation factor 5) (section 3.1.3) is not mentioned in AlzGene database, but in the EIF2 regulation pathway, it is downstream from its family member EIF2AK2. Another member of the gene family, EIF2alpha, is also linked to AD. The phosphorylation of EIF2alpha leads to termination of global protein translation and induces apoptosis. In addition, degenerative neurons in AD brain show high immunoreactivity for phosphorylated EIF2alpha concluding that phosphorylation of EIF2alpha is associated with the degeneration of neurons in AD [Chang et al., 2002]. Additionally, GSK3B, which is involved in AD pathway, EIF5, EIF2AK2, and EIF2alpha play a role in the regulation of EIF2 according to BioCarta (`http://www.biocarta.com/pathfiles/h_eif2Pathway.asp`). Thus, the possibility arises, that EIF5 is also involved in AD and neurodegeneration.
Another overlapping target gene VAPA, a vesicle-associated membrane protein, (section 3.1.3) interacts with its family member VAPB through the transmembrane domain [Hamamoto et al., 2005], and both are reduced in human ALS patients, another neurodegenerative disease. Additionally, both genes interact with lipid binding proteins; especially VAPA is involved in lipid export [Wyles et al., 2002] and neurite outgrowth [Saita et al., 2009]. The mutation VAPB-P56S, which forms stable aggregates that are continuous with the endoplasmic reticulum (ER) and mitochondria and impairs normal VAP function, may result in abnormal lipid transport and biosynthesis and induce slow degeneration of neurons [Teuling et al., 2007]. Approx-

imately 30% of ALS patients with dementia have AD [Hamilton and Bowser, 2004]. A strong connection between VAPA and AD exists via a second neurodegenerative disease ALS as well as by abnormal lipid processes.

SYP, a synaptic vesicle marker and a common target gene of the module from the double transgenic mice and LOAD patients dataset (section 3.1.3), was colocalized with the reactivity of APP and PS1, two key genes of AD. SYP is also localized in the synaptosomal vesicles, where also an association of N- and C-terminal PS1 fragments and APP was detected [Beher et al., 1999]. Other diseases associated with SYP are schizophrenia, ALS, and dementia according to TRANSPATH database. The promoter region of SYP contains four SP1 binding sites located within 100 bp from the transcription start point [Bargou and Leube, 1991]. An abnormally elevated SYP level in the frontal cortex and hippocampal molecular layer exists in old mice lacking BACE1. Studies demonstrate that the absence of BACE1 eliminates plaque pathology. Additionally, SYP deficits correlate with levels of soluble Aβ, and the loss of the associated presynaptic protein SYP is a key pathological feature of AD [McConlogue et al., 2007]. SYP seems to be involved in several distinct neurodegenerative diseases and has several links to key genes of AD as well as to TFs from TFBSs modules.

The fourth overlapping gene REEP5 (receptor accessory protein 5) (section 3.1.3) induces apoptosis according to GO. A common characteristic in the brains of patients suffering from neurodegenerative diseases like AD is massive neuronal death due to apoptosis, and furthermore apoptotic cell death has been found in neurons and glial cells in AD [Shimohama, 2000]. Several studies have shown the direct effect of REEP5 on shaping ER tubules and propose that this protein is involved in the stabilization of highly curved ER membrane tubules. The peripheral ER consists of a network of membrane tubules and in the study by Voeltz et al. [2006] IP_3 (inositol trisphosphate) receptor was a candidate to be involved in ER network formation as well as rapid Ca^{2+} efflux correlating with ER network formation [Voeltz et al., 2006]. Moreover, there exist indications that ER stress is involved in AD pathogenesis [Salminen et al., 2009]. The ER can release stored Ca^{2+} through ER membrane receptor channels like IP_3, and some findings suggest that perturbed ER Ca^{2+} homeostasis contributes to the dysfunction and degeneration of neurons that occur in AD [Mattson, 2010]. REEP5 plays a role in AD by several ER mechanisms.

ADD3, which overlaps between the target genes of the AD and LOAD patients dataset (section 3.1.4) and is a common target gene of TFBSs module CTCF-SP1F-ZBP and KLFS-SP1F-ZBPF (section 3.1.5), plays a role in apoptosis, which leads to neuronal death in AD. Furthermore it belongs to the adducin family of proteins, which is involved in postsynaptic changes in the actin cytoskeleton that occur as a

result of synaptic activation. A study from 2005 suggests that adducin is involved in setting synaptic strength, as well as synaptic plasticity underlying learning and memory [Rabenstein et al., 2005, Porro et al., 2010]. Memory loss is the primary and one of the earliest clinical symptoms of AD, and studies suggest that aging supports the formation of soluble Aβ assemblies, which implicate negative effects on memory [Ashe, 2001]. ADD3 is possibly a new detected gene involved in AD by influencing synpatic functions leading to learning and memory deficits.
Genome-wide association studies in AD have detected CLU to be involved in developing AD by finding a strong association for an intronic single nucleotide polymorphism [Guerreiro et al., 2010]. Biologically, CLU seems to be involved in the pathogenesis of AD by interacting with different molecules like lipids or amyloid proteins [Jones and Jomary, 2002], but also in brains of AD patients the level of CLU mRNA is significantly higher than in control brains [Oda et al., 1994]. Furthermore, APP/PS1 transgenic mice showed increased plasma CLU, age-dependent increase in brain CLU, and amyloid and CLU colocalization in plaques [Thambisetty et al., 2010]. Another disease associated with CLU is ALS and according to GO, CLU, which has increased protein levels in frontal cortex and hippocampus in AD [Lidström et al., 1998], is involved in lipid transporter activity, apoptosis, and neuron development. Additionally, CLU is determined by KLF4 overexpression [Dang et al., 2003] and contains binding sites for SP1 [Jin and Howe, 1997]. Both TFs are incorporated in the third module (section 3.1.5) and in combination with the AD-related target gene CLU verify the importance of the module for AD pathogenesis.
The overlapping target gene NUCKS1 (nuclear casein kinase and cyclin-dependent kinase substrate 1) (section 3.1.5) is known to be a strongly associated gene to PD [Satake et al., 2009], another neurodegenerative disease. A study by Wilson et al. [2003] showed that progression of classical symptoms of PD in old person is associated with eight times as likely to develop AD as well. The results of this study suggest a strong link between progressive motor impairment and the development of AD. Interestingly, pathologic findings like Lewy bodies are similar to PD and dementia, and perhaps there is a connection in how the two diseases progress over time [Wilson et al., 2003]. Furthermore, NUCKS1 may play a role in cell proliferation [Grundt et al., 2007]. The proliferation of neural progenitor cells is reduced in mice transgenic for a mutated form of amyloid precursor protein, transgenic mouse model of AD, that causes early onset familial AD, and it was shown that Aβ can affect the proliferation of neural progenitor cells [Haughey et al., 2002]. NUCKS1 is involved in PD and in AD by cell proliferation connecting both neurodegenerative diseases.
Concluding, literature search and mining revealed several interesting target genes showing a strong connection to AD such as SYP and CLU strengthening the as-

sumption that the TFBSs module may be part of neurodegenerative processes as observed in AD. Furthermore, a possible new gene involved in AD, ADD3, was identified.

Regulatory network in AD

Taken together, all relations of the target genes to the modules, AD key genes and KEGG pathways as described in 4.2.1 are shown in Figure 3.3. The three modules composed of TFs, which are mainly expressed in brain and mostly associated with AD by AD related pathways, are connected to each other by common target genes or KEGG pathways playing a role in AD. Some target genes have known relations to AD key genes further verifying the relation of the modules with their target genes to AD. Promising is the fact that three microarray studies with different platforms, stages of AD in patients and varying amounts of plaques in an AD mouse model have in the end two modules in common. Additionally, one module common to two microarray studies includes the target gene CLU, which is the second most strongly associated gene to AD according to AlzGene database. Although the datasets are different, I got in the end significantly regulated target genes of the modules, which are even the same between datasets of different species and brain tissues, again verifying the significance of the three modules for AD.

In order to verify the modules the TFBSs target interactions could be validated by a knockout (or knockdown) of single or combinations of TFs in mice and measuring the expression compared to wildtype mice by qPCR of the best downstream candidate genes. Another possible experiment would be the screening for SNPs in the sequences of the candidate genes from postmortem AD patient's DNA.

4.2.2 MicroRNAs predicted to regulate ADAM10

Several miRNAs were predicted to bind to human ADAM10 3'UTR and eleven miRNAs were selected according to the specific selection criteria as described in section 4.1.3. The most significant miRNA concerning the specific selection criteria is miR-1306, the exclusive miRNA predicted by all three miRNA prediction programs. The second most significant miRNAs possibly binding to human ADAM10 3'UTR are miR-103 as well as miR-107. Additionally, eight miRNAs with a highly conserved binding region were predicted, but these eight miRNAs have no indication to be involved in AD (Table 4.1).

miRNA	**∅ kcal/mol**	**Specific selection criteria**	**tissue**	**Conservation zebra fish**
1306	-26.85	predicted by 3 programs, mouse ADAM10		+
107	-24.24	targets BACE1, predicted by TargetScan, literature for AD, mouse ADAM10		
103	-25.78	predicted by TargetScan, literature for AD, mouse ADAM10	hippocampus, cerebellum	
330-5p	-27.20	predicted by microRNA, mouse ADAM10	hippocampus	+
432	-22.81	predicted by microRNA	cerebellum	+
423-5p	-22.1	mouse ADAM10	hippocampus, medial frontal gyrus	
671-5p	-27.61	mouse ADAM10		+
922	-27.99	predicted by microRNA		
503	-25.41	predicted by microRNA, mouse ADAM10		
202	-25.33	predicted by microRNA, mouse ADAM10		
184	-23.33	mouse ADAM10		

Table 4.1: List of predicted miRNAs binding to a conserved region of human ADAM10 3'UTR. The best eleven predicted miRNAs, which bind to a conserved region of human ADAM10 3'UTR, are shown with the average predicted binding energy in kcal/mol (column 2). Column 3 describes additional verifications of the miRNAs from literature, if the miRNA is also predicted to bind to the mouse ADAM10 3'UTR or other prediction programs predicting the miRNA. Column 4 specifies in which tissue the miRNA is regulated in AD according to Cogswell et al. [2008] and the conservation of the binding region in zebra fish is marked with + in the last column.

Table 4.1 shows a list of the highest ranking miRNA binding site predictions according to the specific selection criteria. The three most significant miRNAs 1306, 107 and 103 were selected for further analyses.

MiR-107 and miR-103 are downregulated with age [Noren Hooten et al., 2010] as well as in AD gray matter [Wang et al., 2011] and repress the translation of cofilin, an intracellular actin-modulating protein. In brains of a transgenic mouse model of AD the level of miR-103 and miR-107 is decreased while the cofilin protein level is increased which results in the formation of rod-like structures [Yao et al., 2010]. Furthermore, miR-107 expression is decreased even in the earliest stages of AD. As

miR-107 regulates beta-site APP-cleaving enzyme 1 (BACE1) it might be involved in accelerated disease progression [Wang et al., 2008a]. The downregulation of miR-103 and miR-107 with age could concern a protective effect against plaque formation because reduced levels of these miRNAs would lead to an increased level of the predicted target ADAM10 and its neuroprotective product sAPPα in brains of AD patients. Experimental observations of strong inhibition ($> 44\%$) of ADAM10 expression in the reporter assay upon application of miR-103 and miR-107 would coincide with such a possible protective influence on amyloid pathology (section "Experimental validation of bioinformatically predicted miRNAs"). According to the publication from Cogswell et al. [2008] miR-103 is differentially expressed in hippocampus and cerebellum in AD. In addition, the program TargetScan verifies the same binding site of miR-103 and miR-107 to human ADAM10.
The most significant miR-1306 is further analysed due to its good conservation to the far related species zebra fish (Figure 3.11a) with only one mismatch in the seed region. It is the only miRNA whose binding site to human ADAM10 is predicted by all three programs RNA22, RNAhybrid and miRanda, which strengthens the assumption that this binding site is functionally active. Additionally, as shown in Figure 3.12 the hypothesis is verified that miR-1306 is associated to AD: Twelve predicted target genes of miR-1306 are involved in processes and functions playing a role in AD, the nervous system and other neurodegenerative diseases. MiR-1306 possibly regulates genes like the cholinergic receptor, nicotinic, alpha 4 (CHRNA4), tumor necrosis factor receptor superfamily member 1B (TNFRSF1B) and mitogen-activated protein kinase kinase 4 (MAP2K4), which are associated to AD by the functions frontotemporal dementia, demyelination of neurons and Huntington's disease, respectively.
The miR-1306 predicted target gene MAP2K4 has been found to be involved in AD and is putatively regulated by modules of transcription factor binding sites [Augustin et al., 2011]. Furthermore, miR-1306 is located on chromosome 22 within the second exon of DiGeorge syndrome critical region gene 8 (DGCR8), which is essential for miRNA biogenesis by being a subunit of the microprocessor complex [Wang et al., 2007]. Evers et al. [2006] presents a case of a DiGeorge syndrome patient with the typical deletion in chromosome band 22q11.2, which contains DGCR8, suffering from dementia.

Additionally, a putative miR-1306 target gene common in four out of six databases, RXRA (retinoid X receptor, alpha) (supplemental Table A.16), is mentioned in AlzGene database. Seven SNPs were identified in RXRA acting as risk factor for AD [Kölsch et al., 2009]. Furthermore, ADAM10 is regulated by retinoic acid, at which the alpha-isotype of RAR massively participates influencing ADAM10 gene expres-

sion [Tippmann et al., 2009].
Concluding, several strong associations of the three most significant miRNAs to AD pathogenesis were already found in literature.

Molecular functions and processes of predicted miRNA target genes

Gene Ontology analysis (Table 3.11-3.13) represents an overview of the molecular functions and processes with an enrichment of the predicted target genes of miR-103, miR-107 and miR-1306. The molecular function 'calcium ion binding' is significant in all three miRNA analyses and is considered to be involved in AD (Table 3.11-3.13). Calcium ions are found in an elevated level in tangle-bearing neurons of AD patients compared to healthier neurons [Nixon et al., 1994]. Further, an abnormal increase of intracellular Calcium ion levels in neurites associated with Aβ deposits was demonstrated in a mouse model of AD [Mattson, 2010].
An additional evidence is given by the significant enrichment of genes, which are involved in biological processes 'learning', 'brain development' and 'nervous system development' (Table 3.11-3.13), that the three miRNAs are involved in AD. Mouse models with an overexpression of ADAM10 showed a positive effect of the α-secretase on learning and memory and mice with a dominant-negative mutant form of ADAM10 had learning deficiencies [Postina et al., 2004]. Environmental influences occurring during brain development predefine the expression and regulation of APP. As a consequence levels of APP and Aβ are increased causing AD later in life [Zawia et al., 2009]. The Aβ fragments forming plaques are of varying length depending on the site of cleavage. The Aβ_{42} fragment is a ligand for the cellular prion protein (PRNP), which is essential for the development of the nervous system [Kim and Tsai, 2009].

Genes linked to miRNAs and their relation to AD

Further evidence for the association of miR-103 and miR-107 to AD was provided by the established network of genes being involved in AD pathogenesis and directly interacting with miR-103 as well as miR-107. This network (Figure 3.15) allowed confirming the relation of the two miRNAs to AD by genes linked to both miRNAs and AD.
The genes dicer 1, ribonuclease type III (Dicer1) and TAR (HIV-1) RNA binding protein 2 (TARBP2) targeted by miRNA-103 and miRNA-107 are components of the miRNA-processing complex [Boominathan, 2010]. Besides those two genes, a link between the two miRNAs is provided by linoleic acid [Parra et al., 2010], which probably affects AD by increasing the expression of PS1 and Aβ [Liu et al., 2004].
Another target gene in the network, Granulin (GRN), is regulated by miR-107 [Wang

et al., 2010] which regulates BACE1 as well [Wang et al., 2008a]. Therefore both genes might be involved in neurodegenerative diseases especially AD. The tumor suppressors TP53 as well as TP73 appear to regulate the processing of miR-107 [Boominathan, 2010]. MiR-103 increases the expression level of fatty acid binding protein 4 (FABP4) while its expression is reduced by TNF [Xie et al., 2009]. All six genes linked to the miRNAs 103 and 107 GRN, BACE1, TP53, TP73, FABP4 and TNF are included in the AlzGene database, hence putatively playing a role in AD.
The four remaining target genes cyclin-dependent kinase 2 (CDK2), cAMP responsive element binding protein 1 (CREB1), nuclear factor I/A (NFIA) and vascular endothelial growth factor A (VEGFA) of the network are mentioned in literature to be involved directly or in processes developing AD. MiR-103 directly binds and represses CDK2 and CREB1 through 3'UTR binding [Liao and Lönnerdal, 2010]. CDK2 is a key regulator in neuronal differentiation with the downregulation of CDK2 as fundamental event [Dobashi et al., 1995] and neuronal differentiation is regulated by PS1, a major key gene of AD [Wines-Samuelson et al., 2005]. The transcription factor CREB1 is involved in learning and memory. A direct involvement in AD is seen in some mouse models, where its activity is impaired [Puzzo et al., 2005].
NFIA is negatively regulated by miR-107 [Garzon et al., 2007] and plays an important role in the formation of the corpus callosum in the developing brain. The disruption of NFIA results in agenesis of the corpus callosum [das Neves et al., 1999], whereas the size of the corpus callosum is significantly reduced in AD patients [Teipel et al., 2002]. The expression of the hypoxia-regulated gene VEGFA is decreased by miR-107 [Yamakuchi et al., 2010]. Additionally, it is known, that SNPs within the VEGFA promoter region are associated with increased risk for AD, by reducing the neuroprotective effect of VEGFA [Del Bo et al., 2005].
These findings of the literature mining and the involvement in AlzGene database confirm the biological role of the genes in neurodegenerative processes especially in AD and hence the involvement of miR-103 and miR-107 in AD is verified.

Experimental validation of bioinformatically predicted miRNAs

To demonstrate that the selected miRNAs 1306, 103 and 107 directly regulate ADAM10 expression by interaction with the 3'UTR of the human gene, validation experiments were performed (Figure 3.16) [Augustin et al., 2012]. Mir-122 has been identified and validated as an important regulator of ADAM10 in hepatocellular carcinoma by an experimental approach [Bai et al., 2009] and therefore served as positive control for the validation experiments. The reduction by miR-122 is even higher than the one observed in the initial publication from Bai et al. [2009] but might be due to the different reporter enzymes or cell lines used. The selected miRNAs 1306, 103 and

107 reduced the luminescent signal to 72%, 55% and 48%, respectively, and the positive control miR-122 showed a similar reduction to 57%. These experimental results suggest an influence of miR-103, miR-107 and miR-1306 on ADAM10 expression. Nevertheless, the biological impact of either miRNA has to be elucidated further, e.g. by mRNA and protein measurements. Assessing the effect of the selected miRNAs on pathological features in AD mouse models would also help to understand their distinct role in pathogenesis. However, this experiment shows that the computational workflow developed in this work (section 4.1.3) consisting of miRNA binding site prediction programs and the ranking of the results by specific selection criteria is a suitable tool for the identification of miRNAs influencing key genes of diseases such as AD.

4.3 Regulatory network in Parkinson's disease

This part of the discussion concentrates on PD pathogenesis especially on genes related to PD and their regulation by TFBSs as well as their relation to DJ-1. Autosomal dominant (MAPT, SNCA, OPA1) and autosomal recessive (Parkin, DJ-1) PD genes are analysed as well as genes of AMP-activated protein kinase. Finally, the NFKB pathway and its relation to PD and PD related genes is discussed.

4.3.1 CREB/ATF binding site in Parkin promoter

The conservation of the CREB/ATF binding site in the Parkin promoter region among the species human, mouse, cow and horse (Figure 3.6) supports the possibility, that this site might be functional relevant [Cohen et al., 2006, Zhang and Gerstein, 2003]. An additional binding site located in the first intron of Parkin might also play a role in the regulation of Parkin. Bouman et al. [2011] validated the binding of ATF4 to the conserved binding site in the Parkin promoter. Endoplasmic reticulum stress causes an upregulation of ATF4 and Parkin. Whereas the TF c-Jun is binding to the same TF site in the Parkin promoter, but inhibiting Parkin expression. This leads to a competition of both TFs for one binding site in Parkin promoter. Concluding, the upregulation of Parkin by ATF4 after mitochondrial or ER stress inhibits stress-induced cell death [Bouman et al., 2011] and thus possibly prevents the loss of dopaminergic neurons and the pathogenesis of PD.

4.3.2 SNPs matching TFBSs in MAPT and SNCA regulatory region

The focus of this analysis lies on the identification of functionally active TFBSs with the help of PD related SNPs falling in the core-sequence of the BS or lying in a conserved BS (Figure 3.8). Conservation of a BS suggests a functional active TFBS and a SNP located in the core-sequence of a TFBS has a powerful effect on the binding site [Klingenhoff et al., 2002, Werner, 2003]. Hong et al. [2011] observed several SNPs located in TFBSs of genes e.g. in AKAP13 (a kinase (PRKA) anchor protein 13) a regulatory SNP is associated with blood pressure and matches a GATA3 (GATA binding protein 3) TFBS. SNPs located in these binding sites hit a potentially functional TFBS and the change of only one nucleotide in especially the core-sequence leads to deactivation or modification of the binding site. Thus, TFs can no more bind or other TFs bind to the site possibly leading to a misregulation of the gene.

The detected TFBSs (Figure 3.8) possibly regulating MAPT and SNCA rely on SNPs located in the BS. There is the hypothesis that both autosomal dominant genes, MAPT and SNCA, are commonly regulated for example by TFBSs, which are in common for both genes such as the detected TFs DEAF, ETSF and KLFS

(section 3.3.3). The TF Elk-1 (ETS domain-containing protein) belonging to the ETSF family interacts with SNCA by forming a complex. Additionally, both genes bind to the MAP kinase ERK-2/MAPK1 [Iwata et al., 2001]. ETSF is predicted to bind to both MAPT and SNCA.
YY1F predicted to bind to MAPT by Genomatix (section 3.3.3) is related to PD. Arawaka et al. [2006] already reported that YY1 belonging to the YY1F family regulates the PD-related gene G-protein-coupled receptor kinase 5 (GRK5). GRK5 is accumulated in Lewy Bodies and phosphorylates SNCA possibly influencing the development of PD [Arawaka et al., 2006].

Finally, the detected TFBSs supposedly influencing the expression of MAPT and SNCA seem to be more probably to have an involvement on PD like other TFBSs due to PD-related SNPs located in TFBSs.

4.3.3 Common TFBSs in PRKAA1, PRKAA2 and PRKAG1

The kinases PRKAA1, PRKAA2 and PRKAG1 belonging to the AMP-activated protein kinase family are possibly regulated by similar TFs at which the AMPK alpha genes (PRKAA1 and PRKAA2) are supposed to be regulated by the same TFs. Furthermore PRKAA2 is already listed in PDGene database being associated to PD.
The initial restriction of the TFs to those being involved in cellular or oxidative stress, metabolic and bioenergetic incorporating biosynthetic pathways reduces the search space in order to get functionally TFs in the right context. All three genes are supposed to be involved in stress, metabolic and biosynthetic pathways in the context of PD. AMPK is activated by metabolic stress and regulates energy balance in organisms by altering cellular metabolism in order to produce ATP. It seems to be an optimal therapeutic target for several metabolic diseases due to its involvement in essential cellular mechanisms [Ronnett et al., 2009]. Furthermore, AMPK is activated in PD and seems to increase cell viability in PD and prevents cell death [Choi et al., 2010].
ETSF, common predicted to all three genes (Figure 3.9), is known to form a complex with the PD-related gene SNCA [Iwata et al., 2001] and the conservation strengthens the assumption that TFs from ETSF family are functionally active in regulating AMPK subunits. The strongest conserved TF YY1F in the genes of the AMPK alpha subunit (Table 3.9, 3.10) was already described in section 4.3.2 to be involved in PD by regulating the PD-related gene GRK5 [Arawaka et al., 2006]. YY1 incorporated in the YY1F family is expressed in adult rodent brain especially in the hippocampus as well as in neurons [Rylski et al., 2008]. Beside, both TFs have predicted binding sites in MAPT and SNCA (section 4.3.2). NRF1, which is in common predicted

for PRKAA1 and PRKAA2, is downregulated in PD substantia nigra and striatum [Shin et al., 2011]. Additionally, NRF1 regulates the mitochondrial activity, while dysfunction of mitochondria plays a central role in the pathogenesis of PD [Piao et al., 2012].

4.3.4 PD related genes identified in a DJ-1 knockout mouse model

Microglia from DJ-1 knockout and wildtype mouse model was stimulated with LPS to study the role of DJ-1 under neuroinflammation conditions in the etiopathology of PD [personal communication Ulrich Hafen (IDG, HMGU)]. Systemic LPS causes chronic neuroinflammation and progressive neurodegeneration by loss of dopaminergic neurons like observed in PD [Qin et al., 2007]. LPS induces microglial activation and microglial activation plays a key role in the initiation and progression of PD [Liu and Bing, 2011].

Literature mining reveals a great network of indirect interactions between significantly regulated genes on the microarray dataset and DJ-1 (section 3.2.2). This network indicates, that the genes differentially expressed between wildtype and DJ-1 knockout mice are influenced by DJ-1. Further GO enrichment analysis of the significantly expressed genes (Table 3.5) reveals several GO terms associated to PD e.g. 'response to stress'. This strengthens the assumption, that these differentially expressed genes caused by the DJ-1 knockout are involved in PD pathogenesis. The GO terms comprise 'cell death', 'metal ion binding', 'inflammatory response', 'aging', 'citrate cycle', 'microtubule' etc. and their relation to PD is described in the following.

Several significant expressed genes identified by mSVM-RFE in all gene selections are linked to GO terms related to PD (Table 3.5). According to the microarray dataset these genes change their expression level after knocking out DJ-1 resulting in processes influencing PD as well as neurodegeneration. Baulac et al. [2009] shows that oxidative stress, which can cause changes in the mechanisms of cellular signaling, leads to an upregulation of DJ-1 in a morpholino-based zebrafish model, which shows responsiveness of DJ-1 to cellular stress. Additionally, DJ-1 expression is also increased in AD brains [Baulac et al., 2009]. Moreover, mutations of DJ-1 cause cell death induced by oxidative stress, which is characteristic for PD [Taira et al., 2004]. Besides, metal ions are incorporated in the oxidative stress process and were already discussed for drug targets treating neurodegeneration [Gaeta and Hider, 2005]. Several publications report an influence of interleukin-1 mutations in distinct populations on PD. In Taiwan population a polymorphism in interleukin-1α is associated to late-onset sporadic PD [Wu et al., 2007]. Whereas in Turkish patients a mutation in interleukin-1β is linked to sporadic PD [Arman et al., 2010]. A Finnish

population shows once again an influence of the interleukin-1β polymorphism on PD [Mattila et al., 2002]. Vesicle-mediated transport seems also to be involved in PD. The type of transport by synaptic vesicles is altered in neurodegeneration and may be involved in PD pathogenesis [Esposito et al., 2011].

Besides, all NFKB target genes identified by mSVM-RFE method were analysed for enrichment in GO terms playing a role in PD and neurodegeneration (Table 3.6). Inflammatory response is linked to the LPS stimulation as a consequence and fundamental process in PD brains. Chronic inflammation is associated to various neurodegenerative diseases like AD and PD. Moreover, microglia is involved in inflammation by phagocytosing damaged cells and also intact cells in the neighbourhood leading to the degeneration of dopaminergic neurons [Kim and Joh, 2006]. Apoptosis was already very early discussed to be a consequence and not a cause of inflammation leading to the resolution of inflammation [Haanen and Vermes, 1995]. Collier et al. [2011] describes another factor aging to be the major risk of PD. The same mechanisms in the cell can be observed during aging compared to dopaminergic neuron loss in PD. Furthermore they suggest that aging promotes a state prior to PD and additional genetic as well as environmental parameters increase neuronal cell death in PD compared to normal aging [Collier et al., 2011]. The GO term 'learning and memory' is a typical characteristic of AD, but in PD patients the risk increases to develop AD as well [Wilson et al., 2003], which was already described in section 4.2.1.

After identifying the significantly expressed genes on the microarray dataset these genes were grouped into groups of coregulated genes possibly regulated by the same mechanisms or involved in the same biological processes or molecular functions. Hence KEGG and GO analysis of the coregulated genes was performed to search for common pathways, functions as well as cellular components.
One interesting cluster of coregulated genes with enrichment in neurodegenerative KEGG pathways and 'citrate cycle', 'cell cycle' and 'oxidative phosphorylation' was established (Table 3.7). The study of Folch et al. [2011] describes patients suffering from neurodegenerative diseases, which show an upregulation of cell cycle proteins in the brain. Cell cycle is involved in cell death by increasing the expression of E2F1 (E2F transcription factor 1), which might cause neuronal apoptosis [Folch et al., 2011]. Moreover, abnormal cell cycle processes were observed in dopaminergic neurons in PD, which might be a therapeutic treatment target [Höglinger et al., 2007].

In addition, GO analysis reveals several interesting GO terms with an enrichment of coregulated genes incorporating 'microtubule' and 'mitochondria' as well as 'cell cycle' (Table 3.8). The PD-related gene Parkin influences microtubule stabilization leading to a protection of dopaminergic neurons [Feng, 2006] and Lrrk2 is also interacting with microtubules by PD mutations enhancing Lrrk2-microtubule association

[Kett et al., 2012]. Lrrk2 is involved in PD and close associated with the cytoskeleton including microtubules [Parisiadou and Cai, 2010]. Due to this enrichment of coregulated genes in microtubule associated GO terms (Table 3.8) I would suggest DJ-1 also to be involved in microtubule stabilization.
The coregulated genes of this interesting cluster were also overrepresented in the GO term 'magnesium ion binding' (Table 3.8). Hashimoto et al. [2008] showed in a rat PD model that magnesium stops the decline of dopaminergic neurons in the substantia nigra caused by a toxic substance and implicates a positive effect on the length of neurites. Possibly, magnesium prevents the degeneration of dopaminergic neurons [Hashimoto et al., 2008].

Significantly regulated genes of the DJ-1 knockout microarray have still to be validated in the lab, but the above described associations of the gene functions to PD help to decide which genes to validate in detail. In the end, I suggest several genes as candidates for further validations as described in the following with decreasing significance. The genes are ordered according to specific selection criteria.
Two genes worthwhile for further analysis are Mthfr and Hnmt: both genes have an adjusted p-value below 0.35 (supplemental Table A.13) and occur in all mSVM-RFE gene selections. They are linked to DJ-1 by indirect interactions as well as incor-

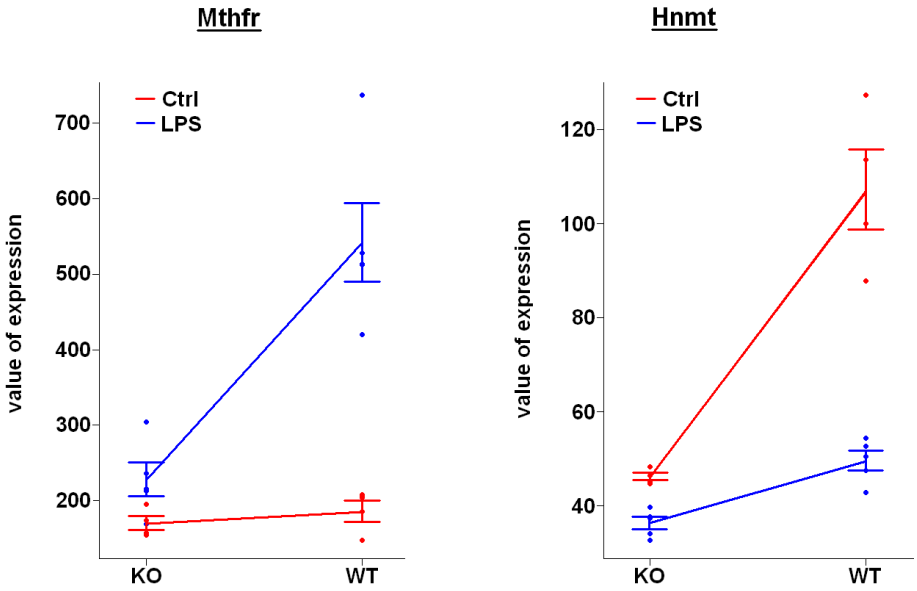


Figure 4.1: Expression values of Mthfr and Hnmt on microarray. The blue and red line show mean and standard error of the mean in the case of LPS stimulation and control treatment, respectively. The x-axis describes knockout (KO) as well as wildtype (WT) mice and the y-axis shows the values of expression.

porated in PDGene database. Mthfr is regulated by Tnfrsf1a (tumor necrosis factor receptor superfamily, member 1A) and Hnmt polymorphism is already associated

with PD [Palada et al., 2012].
Hnmt is required for degradation of histamine, which causes inflammatory processes leading to dopaminergic degeneration [Haas et al., 2008]. Accordingly, the Hnmt expression of wildtype mice not stimulated with LPS is significantly increased compared to knockout mice (Figure 4.1) leading to the assumption that the level of histamine is decreased and as a consequence inflammation and loss of dopaminergic neurons is prevented under wildtype conditions.
Mthfr, methylenetetrahydrofolate reductase, expression is also significantly increased in wildtype mice stimulated with LPS (Figure 4.1). A SNP detected in Mthfr was observed to increase the levels of homocysteine as well as to be responsible for PD pathogenesis [Alatab et al., 2011]. An elevated level of homocysteine leads to oxidative stress and mitochondrial dysfunction as well as neuronal degeneration such as in PD [Duan et al., 2002]. This strengthens the hypothesis that increased (wildtype mice) and decreased (knockout mice) Mthfr expression implicates decreased and increased homocysteine levels, respectively. Whereas high levels of homocysteine results in PD.
The next gene Plod1 is also significantly regulated according to adjusted p-value (supplemental Table A.13) and mSVM-RFE as well as interacts indirectly with DJ-1. In addition, it forms a complex with DJ-1 and AR (Androgen receptor) [Tillman, 2007].
Further genes showing the same significances like Plod1, but which do not have known relationship to PD, are Acot7, Dhrs3, Ubiad1, Clstn1 and Rere. Ankrd33b (ankyrin repeat domain 33B) and Rab6b (ras-related protein Rab-6B) are significantly expressed according to mSVM-RFE and the adjusted p-value (supplemental Table A.13). Additionally, several NFKB target genes from the mSVM-RFE output are Ccnd2, S100a4, Ass1, Miip, Fn1, Fos and Serpine1, which could be validated by qRT-PCR as well as in-situ hybridization.

4.3.5 NFKB pathway related to PD

In mitochondria of PD patients an increased NFKB activity is observed and the activation of NFKB supports cell survival and performs antiapoptotic, which may lead to a protection of dopaminergic neurons [Cassarino et al., 2000]. Furthermore, it was shown that the PD-related gene Parkin activates signaling by NFKB pathway, which leads to neuroprotection. Hence a disruption in this signaling pathway may lead to the pathogenesis of PD and thus would be a therapeutic treatment target of PD [Henn et al., 2007].
The NFKB pathway is known to be associated to PD. Thereby genes significantly regulated in the DJ-1 knockout microarray were searched for NFKB target genes

(section 3.2.2), possibly also involved in PD. Several significant genes identified by mSVM-RFE are NFKB target genes such as Ccnd2, S100a4, Ass1, Miip, Fn1, Fos and Serpine1 and further GO analysis shows association of these genes to PD (Table 3.6) suggesting NFKB pathway to be influenced by the DJ-1 knockout.
DJ-1 knockout and wildtype probes were treated with LPS for 6 hours. This fact might explain the significant enrichment of the NFKB target genes derived from the mSVM-RFE analysis in the GO term 'response to lipopolysaccharide' (3% of genes from this category). There have been established already a lot of in vivo and vitro LPS PD models, because LPS causes inflammatory dopaminergic neurodegeneration [Dutta et al., 2008].

Conserved NFKB binding sites were predicted by Genomatix in the autosomal dominant PD gene OPA1 (section 3.3.2). OPA1 regulates the mitochondrial function and mitochondrial dysfunction is associated with PD [Winklhofer and Haass, 2010]. The conservation of four NFKB binding sites in OPA1 promoter (Figure 3.7) support the possibility, that these sites might be functional relevant [Cohen et al., 2006]. Müller-Rischart et al. [2013] has shown that NFKB regulates OPA1. The association of OPA1 and NFKB and the involvement of OPA1 in PD was examined more closely by Müller-Rischart et al. [2013]

NFKB binds to the PD-related gene OPA1 influencing PD pathogenesis. Additionally, several significantly regulated genes in the DJ-1 knockout mouse model are also NFKB target genes verifying the involvement of these genes in the pathogenesis of PD.

4.3.6 Regulatory network of PD genes as well as predicted and validated TFBSs

As a conclusion from the results of section 4.3 a hypothetical network of all PD genes analysed for TFBSs was established (Figure 4.2) incorporating known interactions of the PD genes from literature and predicted as well as validated TFBSs to PD genes. This network shows a possible regulation of PD genes by TFBSs and the relations of the PD genes to each other. Parkin is known to bind to SNCA [Choi et al., 2001] and physically interacts with MAPT [Moussa, 2009]. Additionally, there is a direct interaction between SNCA and MAPT by SNCA possibly favouring MAPT phosphorylation [Lei et al., 2010]. OPA1 is an indirect target for Parkin, PRKAA1 as well as SNCA connected by BCL2. Furthermore, OPA1 is also regulated by CASP3 (caspase 3), which is controlled by SNCA, MAPT as well as PRKAG1 via KCNN4 (potassium intermediate/small conductance calcium-activated channel). PRKAA1 and PRKAA2 are linked by being subunits of AMPK alpha. OPA1 is regulated by NFKB [Müller-Rischart et al., 2013] and Parkin by ATF4 [Bouman et al., 2011],

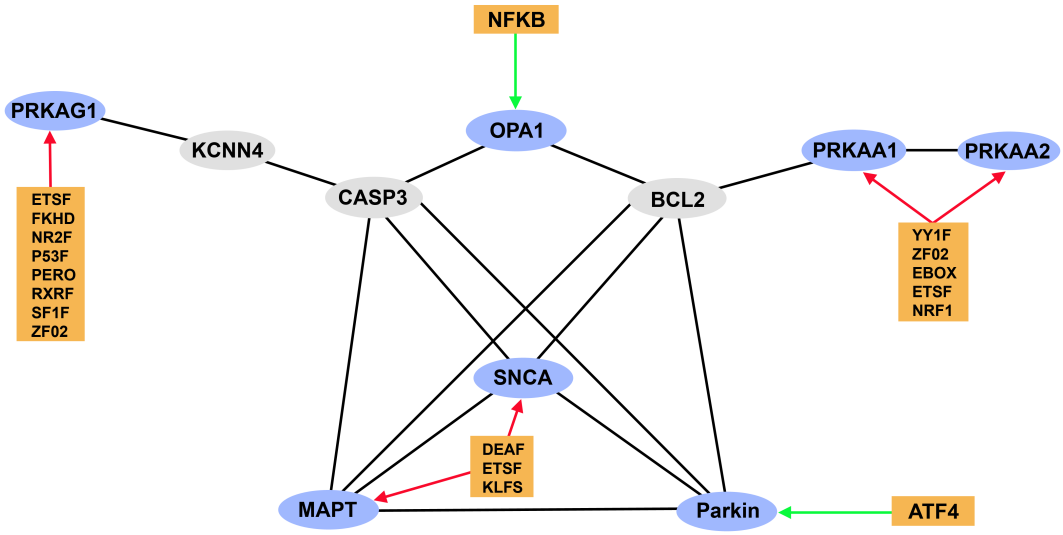


Figure 4.2: Concluding hypothetical PD network with predicted and validated TFBS. Analysed PD genes are shown in blue ellipses, whereas genes connecting PD genes are marked by gray ellipses. TFs are grouped in orange boxes. NFKB and ATF4 have already been validated to regulate OPA1 and Parkin, respectively, which is expressed by a green arrow. Not yet validated TFBSs but predicted ones are shown by red arrows. Several conserved TFs can possibly bind to PRKAG1. SNCA and MAPT are putatively regulated by one of the three common TFs, while PRKAA1 and PRKAA2 are supposedly controlled by one of the five common and conserved TFs.

the control of the other five genes (shown by blue ellipses in Figure 4.2) through a predicted TF have still to be confirmed.

4.4 Regulatory network in stress and neurodegeneration

This section focusses on the association of AD and depression. Additionally, the relation of depression to PD and AD by mitochondrial processes as well as MAPK pathway is discussed, respectively.

4.4.1 Depression related pathways identified in HR/LR mouse models

HR and LR mice established by Touma et al. [2008] with a hyper- and a hypo-reactivity of the HPA axis, respectively, were used to reveal genes related to depression. The LR mice behave in an untypical depressive manner whereas the HR mice show a melancholic depressive like behaviour [Touma et al., 2008].
KEGG enrichment analysis (section 3.2.1) reveals three pathways incorporating genes, which occur in all mSVM-RFE gene selections. These pathways play a role in depression, stress or anxiety and are described in the following.
The pathway 'one carbon pool by folate' is linked to neurodegneration by folate, which is involved in the one-carbon-metabolism and lack of folate leads to e.g. changes in DNA methylation. Disorder in one-carbon-metabolism is associated with neurodegeneration like AD and neuropsychiatric diseases like depression [Kronenberg et al., 2009]. Depression is linked to low levels of folate and a therapeutic treatment of the disease could be the ingestion of folic acid and vitamin B_{12} [Coppen and Bolander-Gouaille, 2005]. Furthermore, low levels of folate are also linked to cognitive impairment [Kim et al., 2008] connecting dementia with depression.
The differentially expressed genes are overrepresented in another KEGG pathway 'arachidonic acid metabolism'. Arachidonic acid belongs to the fatty acids and is associated with bipolar disorder (manic depression). Additionally, it was shown that the signaling of arachidonic acid is increased in bipolar disorders and several anti-manic drugs target and as a consequence decrease arachidonic acid cascade kinetics and enzymes [Rao and Rapoport, 2009]. A significantly increased concentration of arachidonic acid was also observed in brain regions of a rat model for depression [Green et al., 2005].
The third pathway is 'lysine degradation'. Smriga et al. [2002] showed in rats, that a lack of lysine leads to an increase in stress-induced anxiety. Additionally, a supplement with lysine in people consuming lysine deficient diet results in a decrease of anxiety and stress [Smriga et al., 2004]. Anxiety is linked to psychiatric disorders like depression [Gross and Hen, 2004].
Other KEGG pathways in two different biclusters with an enrichment of coregulated genes are 'neuroactive ligand-receptor interaction' and 'endocytosis' (section 3.2.1). Postsynaptic AMPAR (α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid recep-

tor) endocytosis is a final step in the expression of various forms of long term depression [Wang, 2008].

GO analysis of genes determined by mSVM-RFE and occurring in all gene selections revealed several GO terms associated to depression such as the cellular component 'endoplasmic reticulum' (Table 3.4). Endoplasmic reticulum stress is supposed to be connected to stress-related depression disorders [Ishisaka et al., 2011a]. Restraint stress leads to an upregulation of endoplasmic reticulum related genes in mouse brain regions via increase of corticosterone [Ishisaka et al., 2011b], which belongs to the group of glucocorticoids. Glucocorticoids regulate irreversible neuronal cell death in the hippocampus and an increased glucocorticoid level is connected to a shrinkage of the hippocampus and to depression [Sapolsky, 2000].
The alkali metal lithium implicates a protective effect against bipolar disorders [Wood et al., 2004]. The treatment of depressive disorders by lithium was discovered very early by Cade [1949] and lithium is probably the most important psychotropic drug used decades ago [Mitchell and Hadzi-Pavlovic, 2000]. A long-term treatment of bipolar and other major affective disorder patients with the mood stabilizer lithium reduces the risk of completed and attempted suicide [Baldessarini et al., 2006]. Lithium is able to protect the hippocampus from negative consequences of chronic stress and hence would indirectly prevent the break out of depression [Wood et al., 2004]. The GO term 'response to lithium ion' (Table 3.4) comprises genes differentially expressed on the microarray dataset: genes such as ACTA1 (actin, alpha 1) and GSTM5 (glutathione S-transferase, mu 5) are possibly targets for treatment with lithium to prevent depression pathogenesis.

4.4.2 The link between depression and AD

The HR versus LR microarray dataset analysis as well as the search for TFBSs modules in coregulated genes related to AD revealed associations between the neurodegeneration AD and depression.
Two significantly expressed genes of the HR versus LR microarray dataset, which have been validated by qPCR (Table 3.3) [personal communication with Regina Widner from MPI Munich], show a relation to AD pathogenesis. Ttbk1 (Tau-tubulin kinase-1) is known to be involved in AD. Sato et al. [2008] discovered a higher expression of Ttbk1 in AD patients brain compared to control brain and Ttbk1 transgenic mice showed learning and memory deficits in addition to an accumulation of tau and an increased neurofilament phosphorylation. Ttbk1 has been suggested to be a fundamental gene in AD leading to memory dysfunction [Sato et al., 2008]. The upregulation of Ttbk1 in HR mice (foldchange (HR versus LR) = 6.18), which show symptoms of melancholic depression, leads to a direct link between AD and depres-

sion as well as stress by Ttbk1.
The AD related gene Sh3gl2 (endophilin 1) was validated by qPCR (Table 3.3) [personal communication with Regina Widner from MPI Munich]. Ren et al. [2008] showed an elevated expression of Sh3gl2 in brains of AD patients leading to an increase of JNK and as a consequence to neuronal cell death. Moreover, endophilin 1 reflects the interplay of ABAD (amyloid binding alcohol dehydrogenase) and Aβ and hence detects the development of AD in the patient [Ren et al., 2008]. Concluding, the expression of Ttbk1 as well as Sh3gl2 is increased in depression as well as in AD, suggesting a co-occurrence of both diseases.

An additional performed literature search reveals several interactions between depression and AD. Depression accelerates the cognitive decline in AD patients, which possess more plaque and tangle formation in the hippocampus brain region than AD patients without being depressed in their lifetime [Rapp et al., 2006]. Lack of folic acid, which is essential for the nervous system, is linked to depression as well as increases the risk of AD [Reynolds, 2002]. Additionally, Solas et al. [2010] revealed that old rats exposed to stress early in life show elevated levels of Aβ and BACE1 indicating an ascending amyloidogenic processing of APP. The combination of aging and stress influences cognition and develops AD [Solas et al., 2010]. In AD patients depression may appear earlier than typical AD symptoms like memory deficits [Geerlings et al., 2000].

Furthermore, coregulated genes of the AD patients microarray dataset are enriched in the KEGG pathway 'long-term depression' (section 3.1.3). An enhancement of long-term depression leads to loss of memory initiated by long-term potentiation [Berridge, 2010]. The study from Zhang et al. [2010] shows that the gene expression of APP and GNAS, which is incorporated in the cluster of coregulated genes as well as in long-term depression, is significantly upregulated in patients with endogenous depression. Interestingly, clinically significant depression develops in at least 40% of all demented patients [Wragg and Jeste, 1989], and depressive symptoms like significant loss of appetite, insomnia, and fatigue occur commonly in the course of AD [Tune, 1998].

It seems reasonable to assume that the exposure to depression and stress even before AD leads to an acceleration and exacerbation of AD pathogenesis, which is also visible in a stronger plaque formation. But also typical depressive symptoms occur in the course of AD. These aspects make it difficult to decide which disease occurs first and consequently leads to the second disease. Additionally, both diseases have seemingly the same risk genes identified with the HR/LR mice and AD patients microarray dataset analysis as well as common risk factors such as folic acid. Key genes of AD show a significant high regulation in depression and even APP levels

are increased in mice stressed by forced swim test suggesting APP to be involved in stress-responsive pathways [Tsolakidou et al., 2010].

4.4.3 The role of mitochondria in depression and PD

The study of the PD (DJ-1 knockout mice) and depression (HR and LR mice) microarray dataset as well as the OPA1 promoter analysis revealed the mitochondrium to be involved in PD and depression.

GO analysis of genes determined by mSVM-RFE occurring in all gene selections from the HR versus LR microarray dataset (Table 3.4) revealed several categories, which play a role in stress and depression. Interestingly, GO analysis concerning cellular components revealed the 'mitochondrial matrix' as well as the 'mitochondrion' to be the best two hits. Genes, which were detected to be significantly enriched in the mitochondrion and hence are functionally active in this cellular compartment, are according to the mSVM-RFE analysis differentially expressed between the two mouse models HR and LR. Furthermore, two enriched genes, Prodh (proline dehydrogenase (oxidase) 1) and Aldh1l1, are additionally validated by qPCR (Table 3.3) [personal communication with Regina Widner (MPI)] to be upregulated in HR compared to LR mice. The subcellular location of the products of the differentially expressed genes in the mitochondrion leads to the assumption that mitochondrial functions are also changed or even disturbed in depression leading to alteration in the cellular energy supply and stress response, which are typical mitochondrial functions [Winklhofer and Haass, 2010]. Shao et al. [2008] already revealed, that psychiatric diseases like schizophrenia and major depression show a dysfunction of the mitochondrion by diminished gene expression and cumulative impairment in general.

The mitochondrion in conjunction with oxidative stress is a central component in neurodegenerative diseases [Shukla et al., 2011]. A reduced mitochondrial function is observed in aged people and mitochondrial malfunction emerges in neurodegenerative diseases, at which the electron transfer is decreased and impairment through oxidation occurs [Navarro and Boveris, 2010].

One interesting cluster of coregulated genes from the DJ-1 knockout experiment with enrichment in neurodegenerative KEGG pathways was established (Table 3.7). In addition, GO enrichment analysis reveals several interesting GO terms incorporating once again 'mitochondrion' as well as 'mitochondrial matrix' (Table 3.8). These mitochondrial enriched genes are differentially expressed between DJ-1 knockout and wildtype according to mSVM-RFE.
An additional TFBS analysis of the OPA1 promoter revealed NFKB binding sites as described in section 4.3.5 and NFKB activity is known to be increased in PD in mitochondria [Cassarino et al., 2000]. The PD-related gene OPA1 plays a central

role in mitochondria by being involved in inner membrane fusion in the mitochondria (Mitofussion) [Winklhofer and Haass, 2010]. Mitochondrial fragmentation and decreased ATP levels induced by PINK1 and Parkin loss of function are prevented by overexpression of OPA1 [Lutz et al., 2009]. OPA1 downregulation causes mitochondrial fragmentation followed by disorganization of the cristae structure and finally leads to apoptosis [Olichon et al., 2003]. Mitochondrial dysfunction is associated with PD and several PD-related genes like PINK1, Parkin and OPA1 are involved in regulating mitochondrial function [Winklhofer and Haass, 2010]. Glasl et al. [2012] showed that neurons in PINK1-deficiency mice, a mouse model for early phases of PD, have less fragmented mitochondria. Consequently, dysfunction of mitochondria linked to aging and neurodegeneration may be a primary cause for PD and hence mitochondria could be a drug target for therapeutic treatment of PD [Muqit et al., 2006].

Neurodegenerative diseases such as PD seem to be connected to psychiatric disorders by the cellular component mitochondria. But also literature reveals an association between PD and depression. Aarsland et al. [2012] reviewed that around 35% of PD patients suffer from depression as well as mild depression seems to be present in the whole etiopathology of PD and is a risk factor for moderate to severe depression. In both diseases genes with differential expression compared to wildtype are overrepresented in mitochondria. Hence, mitochondria possibly plays a central role in the occurrence of both diseases simultaneously or even a central role not only in PD but also in the pathology of depression.

4.4.4 MAPK pathway associated to AD and depression

The TFBSs module analysis revealed several coregulated AD related genes of the double-transgenic mice dataset to be involved in the MAPK signaling pathway linking MAPK pathway with AD. Literature search revealed an association of MAPK signaling to depression, too.

Cacna2d1 (Figure 3.4c), a target gene of the TFBSs module CTCF-EGRF-SP1F from the double-transgenic mice dataset, is involved in the MAPK signaling pathway (section 3.1.3). The MAPK signaling pathway is significantly overrepresented in a second cluster of coregulated genes from the double-transgenic mice dataset (section 3.1.4). The second cluster of coregulated genes establishing the TFBSs module CTCF-SP1F-ZBPF comprises the MAPK associated genes Cdc42, Map2k4, Mapk1, Mapk9, and Ppm1a (Figure 3.4d). The striking enrichment of coregulated genes from a microarray dataset with AD background confirms the association of MAPK signaling and AD.
Additional literature search strengthens the relation of the genes enriched in the sig-

naling pathway to AD. PS1, a AD key gene, inhibits the Map2k4 activity [Kim et al., 2001]. The phosphorylation of Mapk1, an extracellular signal-regulated kinase (Erk), is activated by Aβ, and this Erk signal is involved in repression of L-glutamate uptake in astrocytes possibly defending neurodegeneration in the pathogenesis of AD [Abe and Misawa, 2003]. The ubiquitously expressed Mapk9 (JNK2) phosphorylates APP and amyloid precursor-like protein 2 (APLP2) induced by cellular stress. The phosphorylation is involved in neural functions and AD pathogenesis [Taru and Suzuki, 2004]. Genes involved in MAPK signaling are regulated by AD key genes as well as regulate the expression of AD key genes either preventing or even promoting AD pathogenesis.
Additionally, the activation of Erk cascades through EGF (epidermal growth factor) yields an increased CTCF expression showing that CTCF is a downstream target of Erk cascades [Li and Lu, 2005]. Furthermore, CTCF is also incorporated in the TFBSs modules (section 3.1.3 and 3.1.4) established from the MAPK signaling related genes. MAPK pathway is involved in the production of proinflammatory cytokines in the hippocampus induced by Aβ and is a potential target for future therapeutics in AD [Munoz et al., 2007].

MAPK signaling plays an important role in depression and is associated to CRH, a crucial gene of the HPA axis in the stress response pathway. Dermitzaki et al. [2002] showed that CRH phosphorylates p38 MAPK by activating CRHR1 (Corticotropin releasing hormone receptor 1) transmitting signals as well as CRH and cAMP increase the activity of the MAPK pathway [Kovalovsky et al., 2002]. Furthermore, Wefers et al. [2012] demonstrated an increased depression-like behaviour in the brains of adult mice by inhibiting MAPK signaling, which is incorporated in the development of brain, long-term memory and antidepressant response. Therefore, MAPK signaling may be involved in the regulation of depression-like behaviour and MAP kinase phosphatases are possibly new targets for the development of antidepressant drugs [Wefers et al., 2012].

MAPK links AD and depression by being a common pathway of both diseases. It seems that the dysfunction of MAPK pathway promotes the pathogenesis of both diseases. But as previously discussed in section 4.4.2 if depression is a AD-associated condition, a presenting symptom of the disease, or a risk factor for AD pathogenesis has not been fully understood until now and has to be further analysed.

4.5 Hypothetical regulatory molecular network in neurodegenerative diseases

The validated and predicted interactions of neurodegenerative diseases and depression are summarized in a hypothetical regulatory molecular network. It comprises

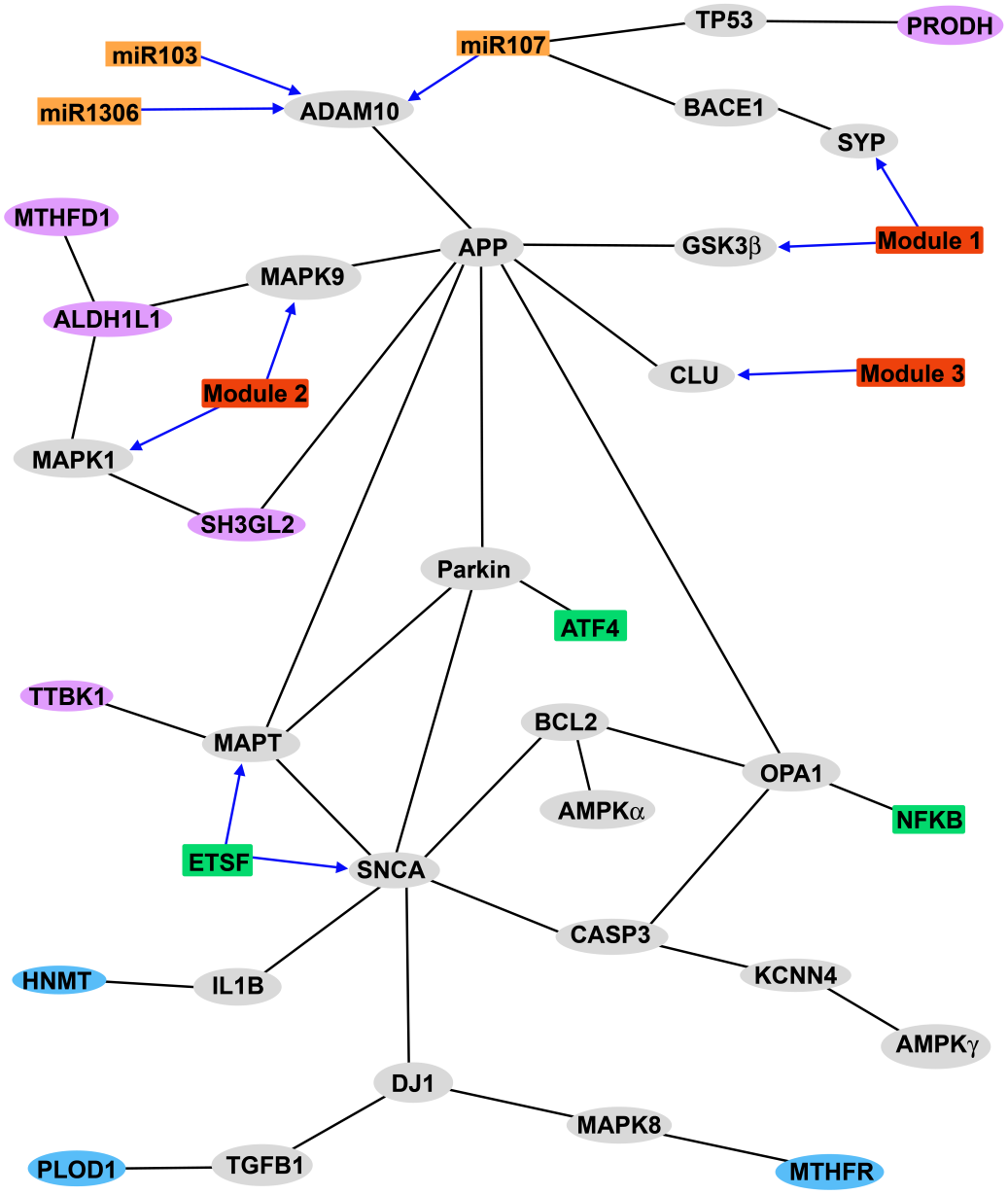


Figure 4.3: The hypothetical network summarizes all analyses of the thesis incorporating TFBSs, microRNA and microarray analysis for AD, PD and depression. The network is in detail explained in the text.

key genes and differentially regulated genes of the three discussed diseases as well as predicted microRNAs, TFs and TFBSs modules of this thesis (Figure 4.3). The red boxes correspond to the three defined TFBSs modules of AD possibly regulating the genes connected by blue arrow in the network [Augustin et al., 2011]. The miRNAs seemingly influencing the expression of ADAM10 are described in orange colored boxes [Augustin et al., 2012]. The lilac highlighted genes are derived from the HR versus LR microarray analysis and were successfully validated in the lab by qPCR [personal communication with Regina Widner (MPI)]. Furthermore, these genes show a good connectivity either directly or indirectly to AD- and PD-related genes. The green TFs are predicted to regulate PD key genes. The regulation of Parkin and OPA1 by the TFs ATF4 and NFKB, respectively, was validated in the lab [Bouman et al., 2011, Müller-Rischart et al., 2013]. ETSF possibly regulates MAPT and SNCA verified by conserved binding sites and PD associated SNPs, which are located in the TF core-sequences of both genes, as well as by literature describing an involvement of the TF in PD. The genes in the blue ellipses are the most promising candidates for further validations of the DJ-1 knockout microarray dataset. Beside, the gray colored genes are AD- and PD-related genes or at least genes showing a connection to neurodegenerative diseases as well as some genes for connecting the whole network.

Additionally, a strong link between AD and PD is given by connections of APP to MAPT, Parkin as well as OPA1. The network contains not all known relations between the genes and the diseases in order to maintain a clear representation. Definitely, more associations between AD and PD are described in the literature. While depression related genes (lilac ellipses) are distributed over the whole network being involved in both neurodegenerative diseases, AD and PD related genes are separated. But a strong relationship is clearly visible in the middle of the network connecting the upper AD part with the PD section below.

The Aβ fragments derived from APP cleavage by beta- and gamma-secretases lead to neurotoxic amyloid plaques as described in section 1.1.1. Rosen et al. [2010] showed in AD brains an interaction of Aβ with Parkin decreasing the Aβ level. Whereas the knockdown of Parkin increases the amount of Aβ [Rosen et al., 2010]. Parkin seems to have a protective effect in producing much less amyloid plaques. Moreover, the presence of Aβ fragments leads to a changed MAPT expression level producing more cerebrospinal fluid tau and hence leading to more neurodegeneration [Kauwe et al., 2008]. Additionally, Aβ levels are negatively correlated with OPA1 levels, i.e. overexpression of APP decreases OPA1 expression. Concluding, Aβ fragments contribute to mitochondrial fragmentation by influencing mitochondrial fission/fusion leading to mitochondrial and neuronal dysfunction [Wang et al., 2008b].

This network suggests a molecular connection of the diseases to each other and might explain the co-occurrence of the diseases either as accessory symptom or longterm side effect. Especially, stress and depression appear to be risk factors for AD and PD.

Chapter 5

Conclusion

Two established bioinformatics approaches for TFBSs module and miRNA identification were evaluated. The TFBSs module approach searches for common regulatory TFBSs modules in coregulated AD-related genes as well as in AD key genes influencing the pathogenesis of AD. Common TFBSs modules were obtained by combining in silico analysis with an approach relying on expression values of AD related genes.

Moreover, the multivariate methods of the established workflow were applied on depression as well as PD microarray datasets and the implemented mSVM-RFE method was evaluated by ROC curve recognizing almost 87% of all differentially expressed genes. Concluding, multivariate methods consider interactions amongst the gene expression levels and hence regulatory molecular networks can be detected.

Additionally, a computational approach for the identification of miRNAs putatively influencing the expression of ADAM10 was generated. The miRNA target site prediction approach is a combination of the prediction software and specific selection criteria for filtering out false positive miRNA predictions: disease relevance, specificity of expression and evolutionary conservation of binding sites.

Many of the coregulated genes identified to be target genes of modules of TFBSs have links to the molecular mechanisms of AD. In particular, the binding sites of the TF families: CTCF, EGRF, KLFS, SP1F, and ZBPF are proposed for further investigations and could provide potential targets for therapeutic treatment of AD and other neurodegenerative diseases. Already known regulation mechanisms of target genes by TFs confirm these modules, such as CLU, which is linked to AD. Additionally, several target genes like ADD3, which are not yet described as AD-related genes, are possibly involved in AD pathogenesis.

A potential functionality of the selected miRNAs 103, 107 and 1306 on ADAM10 3'UTR was confirmed by 3'UTR luciferase reporter assay. These experiments underline the reliability of my computational approach. The miRNA workflow can also be applied to key genes of other diseases with adjustment of the selection criteria according to the scientific research interest. This approach provides a new selection tool for identification and ranking of AD-related miRNAs, but to elucidate a profound pathological role of selected candidates further experiments have to be done.

Significantly regulated downstream targets of DJ-1 such as Hnmt and Mthfr were identified to be PD related genes and the influence of both genes on the pathogenesis of PD was discussed. Moreover, the relation of NFKB and PD was considered in the analysis by focussing on NFKB target genes, which seem to be regulated in PD. Additionally, NFKB BSs were identified in the OPA1 promoter and validated by Müller-Rischart et al. [2013]. Beside, OPA1 playing an important role in mitofussion also Parkin is associated to mitochondria being targeted by ATF4 after mitochondrial stress [Bouman et al., 2011]. Moreover, several TFBSs were predicted in autosomal recessive and dominant PD genes and verified by literature search.

Differentially expressed genes related to depression were experimentally verified by qPCR [personal communication with Regina Widner from MPI Munich] confirming the predictions obtained by the mSVM-RFE method. Regulatory mechanisms linking depression and AD as well as PD were identified. AD and depression have seemingly the same risk genes and risk factors such as deficiency of folic acid. Furthermore, AD key genes are significantly high expressed in depression. MAPK signaling pathway was discovered to be a common pathway of coregulated AD related genes and is a underlying pathway in CRH signaling. Literature search revealed also an association between PD and depression. Additionally, genes from the PD and depression microarray dataset were detected to be significantly enriched in the mitochondrion. The interaction between both diseases was discussed and even a new relationship to mitochondrial processes in depression was revealed.

Finally, a hypothetical overall network combining all predictions and validations of TFBSs, miRNAs and key genes of neurodegenerative diseases shows possible connections between the different diseases.

Bibliography

D Aarsland, S Påhlhagen, CG Ballard, U Ehrt, and P Svenningsson. Depression in parkinson disease — epidemiology, mechanisms and management. *Nature Reviews Neurology*, 8:35–47, 2012.

K Abe and M Misawa. The extracellular signal-regulated kinase cascade suppresses amyloid beta protein-induced promotion of glutamate clearance in cultured rat cortical astrocytes. *Brain Research*, 979:179–187, 2003.

A Agresti. *An introduction to categorical data analysis*. Wiley, second edition edition, 2007.

S Alatab, A Hossein-nezhad, K Mirzaei, F Mokhtari, G Shariati, and A Najmafshar. Inflammatory profile, age of onset, and the mthfr polymorphism in patients with multiple sclerosis. *Journal of molecular neuroscience*, 44:6–11, 2011.

S Altraja, J Jaama, E Valk, and A Altraja. Changes in the proteome of human bronchial epithelial cells following stimulation with leucotriene e4 and transforming growth factor-beta1. *Respirology*, 14:39–45, 2009.

S Arawaka, M Wada, S Goto, H Karube, M Sakamoto, CH Ren, S Koyama, H Nagasawa, H Kimura, T Kawanami, K Kurita, K Tajima, M Daimon, M Baba, T Kido, S Saino, K Goto, H Asao, C Kitanaka, E Takashita, S Hongo, T Nakamura, T Kayama, Y Suzuki, K Kobayashi, T Katagiri, K Kurokawa, M Kurimura, I Toyoshima, K Niizato, K Tsuchiya, T Iwatsubo, M Muramatsu, H Matsumine, and T Kato. The role of g-protein-coupled receptor kinase 5 in pathogenesis of sporadic parkinson's disease. *The Journal of Neuroscience*, 26:9227–38, 2006.

A Arman, N Isik, A Coker, F Candan, KS Becit, and EO List. Association between sporadic parkinson disease and interleukin-1 beta -511 gene polymorphisms in the turkish population. *European Cytokine Network*, 21:116–21, 2010.

M Ashburner, CA Ball, JA Blake, D Botstein, H Butler, JM Cherry, AP Davis, K Dolinski, SS Dwight, JT Eppig, MA Harris, DP Hill, L Issel-Tarver, A Kasarskis, S Lewis, JC Matese, JE Richardson, M Ringwald, GM Rubin, and G Sherlock. Gene ontology: tool for the unification of biology. *Nature Genetics*, 25:25–29, 2000.

KH Ashe. Learning and memory in transgenic mice modeling alzheimer's disease. *Learning & Memory*, 8:301–308, 2001.

R Augustin, SF Lichtenthaler, M Greeff, J Hansen, W Wurst, and D Trümbach. Bioinformatics identification of modules of transcription factor binding sites in alzheimer's disease-related genes by in silico promoter analysis and microarrays. *International Journal of Alzheimer's Disease*, 2011:154325, 2011.

R Augustin, K Endres, S Reinhardt, PH Kuhn, SF Lichtenthaler, J Hansen, W Wurst, and D Trümbach. Computational identification and experimental validation of micrornas binding to the alzheimer-related gene adam10. *BMC Medical Genetics*, 13:35, 2012.

S Bai, MW Nasser, B Wang, SH Hsu, J Datta, H Kutay, A Yadav, G Nuovo, P Kumar, and K Ghoshal. Microrna-122 inhibits tumorigenic properties of pepatocellular carcinoma cells and sensitizes these cells to sorafenib. *Journal of Biological Chemistry*, 284:32015–32027, 2009.

RJ Baldessarini, L Tondo, P Davis, M Pompili, FK Goodwin, and J Hennen. Decreased risk of suicides and attempts during long-term lithium treatment: a meta-analytic review. *Bipolar Disorders*, 8:625–39, 2006.

RC Bargou and RE Leube. The synaptophysin-encoding gene in rat and man is specifically transcribed in neuroendocrine cells. *Gene*, 99:197–204, 1991.

T Barnickel, J Weston, R Collobert, HW Mewes, and V Stümpflen. Large scale application of neural network based semantic role labeling for automated relation extraction from biomedical texts. *PLoS ONE*, 4:e6393, 2009.

DP Bartel. Micrornas: genomics, biogenesis, mechanism, and function. *Cell*, 116:281–97, 2004.

MR Basha, W Wei, SA Bakheet, N Benitez, HK Siddiqi, YW Ge, DK Lahiri, and NH Zawia. The fetal basis of amyloidogenesis: exposure to lead and latent overexpression of amyloid precursor protein and beta-amyloid in the aging brain. *The Journal of Neuroscience*, 25:823–9, 2005.

S Baulac, H Lu, J Strahle, T Yang, MS Goldberg, J Shen, MG Schlossmacher, CA Lemere, Q Lu, and W Xia. Increased dj-1 expression under oxidative stress and in alzheimer's disease brains. *Molecular neurodegeneration*, 4:12, 2009.

J Becker, P Schmidt, F Musshoff, M Fitzenreiter, and B Madea. Mor1 receptor mrna expression in human brains of drug-related fatalities-a real-time pcr quantification. *Forensic Science International*, 140:13–20, 2004.

D Beher, C Elle, J Underwood, JB Davis, R Ward, E Karran, CL Masters, K Beyreuther, and G Multhaup. Proteolytic fragments of alzheimer's disease-associated presenilin 1 are present in synaptic organelles and growth cone membranes of rat brain. *Journal of Neurochemistry*, 72: 1564–1573, 1999.

A Ben-Hur, CS Ong, S Sonnenburg, B Schölkopf, and G Rätsch. Support vector machines and kernels for computational biology. *PLoS Computational Biology*, 4:e1000173, 2008.

Y Benjamini and Y Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57:289–300, 1995.

M Berridge. Calcium hypothesis of alzheimer's disease. *Pflügers Archiv European Journal of Physiology*, 459:441–449, 2010.

L Bertram, MB McQueen, K Mullin, D Blacker, and RE Tanzi. Systematic meta-analyses of alzheimer disease genetic association studies: the alzgene database. *Nature Genetics*, 39:17–23, 2007.

D Betel, M Wilson, A Gabow, DS Marks, and C Sander. The microrna.org resource: targets and expression. *Nucleic Acids Research*, 36:D149–D153, 2008.

EM Blalock, JW Geddes, KC Chen, NM Porter, WR Markesbery, and PW Landfield. Incipient alzheimer's disease: Microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proceedings of the National Academy of Sciences of the United States of America*, 101:2173–2178, 2004.

V Bogaerts, J Theuns, and C Van Broeckhoven. Genetic findings in parkinson's disease and translation into treatment: a leading role for mitochondria? *Genes, Brain and Behavior*, 7:129–151, 2008.

L Boominathan. The tumor suppressors p53, p63, and p73 are regulators of microrna processing complex. *PLoS ONE*, 5:e10615, 2010.

B Boudíková-Girard, MC Scott, and R Weinshilboum. Histamine n-methyltransferase: inhibition by monoamine oxidase inhibitors. *Agents and actions*, 40:1–10, 1993.

L Bouman, A Schlierf, AK Lutz, J Shan, A Deinlein, J Kast, Z Galehdar, V Palmisano, N Patenge, D Berg, T Gasser, R Augustin, D Trümbach, I Irrcher, DS Park, W Wurst, MS Kilberg, J Tatzelt, and KF Winklhofer. Parkin is transcriptionally regulated by atf4: evidence for an interconnection between mitochondrial stress and er stress. *Cell Death and Differentiation*, 18:769–82, 2011.

H Braak, E Ghebremedhin, U Rüb, and K Bratzke, H und Del Tredici. Stages in the development of parkinson's disease-related pathology. *Cell and Tissue Research*, 318:121–134, 2004.

J Brennecke, A Stark, RB Russell, and SM Cohen. Principles of microrna-target recognition. *PLoS Biology*, 3:e85, 2005.

AJ Brookes. The essence of snps. *Gene*, 234:177–186, 1999.

MP Brown, WN Grundy, D Lin, N Cristianini, CW Sugnet, TS Furey, M Jr Ares, and D Haussler. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proceedings of the National of Academy Sciences U S A*, 97:262–7, 2000.

CJ Bult, JT Eppig, JA Kadin, JE Richardson, JA Blake, and the Mouse Genome Database Group. The mouse genome database (mgd): mouse biology and model systems. *Nucleic Acids Research*, 36:D724–D728, 2008.

JFJ Cade. Lithium salts in the treatment of psychotic excitement. *Medical Journal of Australia*, 2: 349–352, 1949.

JA Capra and M Singh. Predicting functionally important residues from sequence conservation. *Bioinformatics*, 23:1875–1882, 2007.

K Cartharius, K Frech, K Grote, B Klocke, M Haltmeier, A Klingenhoff, M Frisch, M Bayerlein, and T Werner. Matinspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics*, 21:2933–42, 2005.

DS Cassarino, EM Halvorsen, RH Swerdlow, NN Abramova, WD Jr Parker, TW Sturgill, and JP Jr Bennett. Interaction among mitochondria, mitogen-activated protein kinases, and nuclear factor-kappab in cellular models of parkinson's disease. *Journal of Neurochemistry*, 74:1384–92, 2000.

W Cerpa, EM Toledo, L Varela-Nallar, and NC Inestrosa. The role of wnt signaling in neuroprotection. *Drug News & Perspectives*, 22:579–91, 2009.

RC Chang, AK Wong, HK Ng, and J Hugon. Phosphorylation of eukaryotic initiation factor-2alpha (eif2alpha) is associated with neuronal degeneration in alzheimer's disease. *Neuroreport*, 13:2429–32, 2002.

KR Chaudhuri, DG Healy, AH Schapira, and National Institute for Clinical Excellence. Non-motorsymptoms of parkinson's disease: diagnosis and management. *The Lancet Neurology*, 5: 235–245, 2006.

M Chekulaeva and W Filipowicz. Mechanisms of mirna-mediated post-transcriptional regulation in animal cells. *Current Opinion in Cell Biology*, 21:452–460, 2009.

JS Choi, C Park, and JW Jeong. Amp-activated protein kinase is activated in parkinson's disease models mediated by 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine. *Biochemical and Biophysical Research Communications*, 391:147–51, 2010.

P Choi, N Golts, H Snyder, M Chong, L Petrucelli, J Hardy, D Sparkman, E Cochran, JM Lee, and B Wolozin. Co-association of parkin and alpha-synuclein. *Neuroreport*, 12:2839–43, 2001.

BN Chorley, X Wang, MR Campbell, GS Pittman, MA Noureddine, and DA Bell. Discovery and verification of functional single nucleotide polymorphisms in regulatory genomic regions: current and developing technologies. *Mutation Research*, 659:147–57, 2008.

MA Christensen, W Zhou, H Qing, A Lehman, S Philipsen, and W Song. Transcriptional regulation of bace1, the β-amyloid precursor protein β-secretase, by sp1. *Molecular and Cellular Biology*, 24:865–874, 2004.

JY Chuang, YT Wang, SH Yeh, YW Liu, WC Chang, and JJ Hung. Phosphorylation by c-jun nh2-terminal kinase 1 regulates the stability of transcription factor sp1 during mitosis. *Molecular biology of the cell*, 19:1139–1151, 2008.

BA Citron, JS Dennis, RS Zeitlin, and V Echeverria. Transcription factor sp1 dysregulation in alzheimer's disease. *Journal of Neuroscience Research*, 86:2499–2504, 2008.

CM Clements, RS McNally, BJ Conti, TW Mak, and JP Ting. Dj-1, a cancer- and parkinson's disease-associated protein, stabilizes the antioxidant transcriptional master regulator nrf2. *Proceedings of the National of Academy Sciences U S A*, 103:15091–6, 2006.

JP Cogswell, J Ward, IA Taylor, M Waters, Y Shi, B Cannon, K Kelnar, J Kemppainen, D Brown, C Chen, RK Prinjha, JC Richardson, AM Saunders, AD Roses, and CA Richards. Identification of mirna changes in alzheimer's disease brain and csf yields putative biomarkers and insights into disease pathways. *Journal of Alzheimer's Disease*, 14:27–41, 2008.

CD Cohen, A Klingenhoff, A Boucherot, A Nitsche, A Henger, B Brunner, H Schmid, M Merkle, MA Saleem, KP Koller, T Werner, HJ Gröne, PJ Nelson, and M Kretzler. Comparative promoter analysis allows de novo identification of specialized cell junction-associated proteins. *Proceedings of the National Academy of Sciences*, 103:5682–5687, 2006.

SL Cole and R Vassar. The role of amyloid precursor protein processing by bace1, the beta-secretase, in alzheimer disease pathophysiology. *Journal of Biological Chemistry*, 283:29621–29625, 2008.

TJ Collier, NM Kanaan, and JH Kordower. Ageing as a primary risk factor for parkinson's disease: evidence from studies of non-human primates. *Nature Reviews Neuroscience*, 12:359–366, 2011.

FS Collins, LD Brooks, and A Chakravarti. A dna polymorphism discovery resource for research on human genetic variation. *Genome Research*, 8:1229–1231, 1998.

A Coppen and C Bolander-Gouaille. Treatment of depression: time to consider folic acid and vitamin b12. *Journal of Psychopharmacology*, 19:59—-65, 2005.

C Cortes and V Vapnik. Support vector networks. *Machine Learning*, 20:273–297, 1995.

L Crews, E Rockenstein, and E Masliah. App transgenic modeling of alzheimer's disease: mechanisms of neurodegeneration and aberrant neurogenesis. *Brain Structure and Function*, 214: 111–126, 2010.

DT Dang, X Chen, J Feng, M Torbenson, LH Dang, and VW Yang. Overexpression of krüppel-like factor 4 in the human colon cancer cell line rko leads to reduced tumorigenecity. *Oncogene*, 22: 3424–3430, 2003.

L das Neves, CS Duchala, F Godinho, MA Haxhiu, C Colmenares, WB Macklin, CE Campbell, KG Butz, and RM Gronostajski. Disruption of the murine nuclear factor i-a gene (nfia) results in perinatal lethality, hydrocephalus, and agenesis of the corpus callosum. *Proceedings of the National Academy of Sciences*, 96:11946–11951, 1999.

CA Davie. A review of parkinson's disease. *British Medical Bulletin*, 86:109–127, 2008.

TM Dawson, HS Ko, and VL Dawson. Genetic animal models of parkinson's disease. *Neuron*, 66: 646–61, 2010.

GV De Ferrari and NC Inestrosa. Wnt signaling function in alzheimer's disease. *Brain Research Reviews*, 33:1–12, 2000.

ER de Kloet, M Joëls, and F Holsboer. Stress and the brain: from adaptation to disease. *Nature Reviews Neuroscience*, 6:463–475, 2005.

R Del Bo, M Scarlato, S Ghezzi, F Martinelli Boneschi, C Fenoglio, S Galbiati, R Virgilio, D Galimberti, G Galimberti, M Crimi, C Ferrarese, E Scarpini, N Bresolin, and GP Comi. Vascular endothelial growth factor gene variability is associated with increased risk for ad. *Annals of Neurology*, 57:373–380, 2005.

E Dermitzaki, C Tsatsanis, A Gravanis, and AN Margioris. Corticotropin-releasing hormone induces fas ligand production and apoptosis in pc12 cells via activation of p38 mitogen-activated protein kinase. *The Journal of Biological Chemistry*, 277:12280–7, 2002.

Y Dobashi, T Kudoh, A Matsumine, K Toyoshima, and T Akiyama. Constitutive overexpression of cdk2 inhibits neuronal differentiation of rat pheochromocytoma pc12 cells. *Journal of Biological Chemistry*, 270:23031–23037, 1995.

S Döhr, A Klingenhoff, H Maier, M Hrabé de Angelis, T Werner, and R Schneider. Linking disease-associated genes to regulatory networks via promoter organization. *Nucleic Acids Research*, 33: 864–872, 2005.

C Domeniconi, D Gunopulos, and J Peng. Large margin nearest neighbor classifiers. *IEEE Transactions of Neural Networks*, 16:899–909, 2005.

W Duan, B Ladenheim, RG Cutler, II Kruman, JL Cadet, and MP Mattson. Dietary folate deficiency and elevated homocysteine levels endanger dopaminergic neurons in models of parkinson's disease. *Journal of neurochemistry*, 80:101–10, 2002.

G Dutta, P Zhang, and B Liu. The lipopolysaccharide parkinson's disease animal model: mechanistic studies and drug discovery. *Fundamental & clinical pharmacology*, 22:453–64, 2008.

R Edgar, M Domrachev, and AE Lash. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic Acids Res*, 30:207–10, 2002.

TL Edwards, WK Scott, C Almonte, A Burt, EH Powell, GW Beecham, L Wang, S Züchner, I Konidari, G Wang, C Singer, F Nahab, B Scott, JM Stajich, M Pericak-Vance, J Haines, JM Vance, and ER Martin. Genome-wide association study confirms snps in snca and the mapt region as common risk factors for parkinson disease. *Annals of human genetics*, 74:97–109, 2010.

MI Ekstrand, M Terzioglu, D Galter, S Zhu, C Hofstetter, E Lindqvist, S Thams, A Bergstrand, FS Hansson, A Trifunovic, B Hoffer, S Cullheim, AH Mohammed, L Olson, and NG Larsson. Progressive parkinsonism in mice with respiratory-chain-deficient dopamine neurons. *Proceedings of the National of Academy Sciences U S A*, 104:1325–30, 2007.

AJ Enright, B John, U Gaul, T Tuschl, C Sander, and DS Marks. Microrna targets in drosophila. *Genome Biology*, 5:R1, 2003.

G Esposito, FA Clara, and P Verstreken. Synaptic vesicle trafficking and parkinson's disease. *Developmental Neurobiology*, 72:134–144, 2011.

LJ Evers, MP Vermaak, JJ Engelen, and LM Curfs. The velocardiofacial syndrome in older age: dementia and autistic features. *Genetic counseling*, 17:333–40, 2006.

MR Fabian, N Sonenberg, and W Filipowicz. Regulation of mrna translation and stability by microrna s. *Annual Review of Biochemistry*, 79:351–379, 2010.

F Fahrenholz. Alpha-secretase as a therapeutic target. *Current Alzheimer Research*, 4:412–417, 2007.

J Feng. Microtubule: a common target for parkin and parkinson's disease toxins. *The Neuroscientist*, 12:469–76, 2006.

J Folch, F Junyent, E Verdaguer, C Auladell, JG Pizarro, C Beas-Zarate, M Pallàs, and A Camins. Role of cell cycle re-entry in neurons: A common apoptotic mechanism of neuronal cell death. *Neurotoxicity Research*, Epub ahead of print, 2011.

R Foti, S Zucchelli, M Biagioli, P Roncaglia, S Vilotti, R Calligaris, H Krmac, JE Girardini, G Del Sal, and S Gustincich. Parkinson disease-associated dj-1 is required for the expression of the glial cell line-derived neurotrophic factor receptor ret in human neuroblastoma cells. *The Journal of biological chemistry*, 285:18565–74, 2010.

PL Franzen and DJ Buysse. Sleep disturbances and depression: risk relationships for subsequent depression and therapeutic implications. *Dialogues in Clinical Neuroscience*, 10:473–81, 2008.

T Fraser, H Tayler, and S Love. Fatty acid composition of frontal, temporal and parietal neocortex in the normal human brain and in alzheimer's disease. *Neurochemical Research*, 35:503–513, 2010.

WJ Fredericks, T McGarvey, H Wang, P Lal, R Puthiyaveettil, J Tomaszewski, J Sepulveda, E Labelle, JS Weiss, ML Nickerson, HS Kruth, W Brandt, LA Wessjohann, and SB Malkowicz. The bladder tumor suppressor protein tere1 (ubiad1) modulates cell cholesterol: implications for tumor progression. *DNA and cell biology*, 30:851–64, 2011.

RC Friedman, KK Farh, CB Burge, and DP Bartel. Most mammalian mrnas are conserved targets of micrornas. *Genome Research*, 19:92–105, 2009.

A Gaeta and RC Hider. The crucial role of metal ions in neurodegeneration: the basis for a promising therapeutic strategy. *British Journal of Pharmacology*, 146:1041–59, 2005.

V Gailus-Durner, M Scherf, and T Werner. Experimental data of a single promoter can be used for in silico detection of genes with related regulation in the absence of sequence similarity. *Mammalian Genome*, 12:67–72, 2001.

ML Garcia and DW Cleveland. Going new places using an old map: tau, microtubules and human neurodegenerative disease. *Current Opinion in Cell Bilogy*, 13:41–48, 2001.

R Garzon, F Pichiorri, T Palumbo, M Visentini, R Aqeilan, A Cimmino, H Wang, H Sun, S Volinia, H Alder, GA Calin, CG Liu, M Andreeff, and CM Croce. Microrna gene expression during retinoic acid-induced differentiation of human acute promyelocytic leukemia. *Oncogene*, 26:4148–4157, 2007.

MI Geerlings, RA Schoevers, AT Beekman, C Jonker, DJ Deeg, B Schmand, HJ Adèr, LM Bouter, and W Van Tilburg. Depression and risk of cognitive decline and alzheimer's disease. results of two prospective community-based studies in the netherlands. *The British Journal of Psychiatry*, 176:568–575, 2000.

G Gibson. Rare and common variants: twenty arguments. *Nature Reviews Genetics*, 13:135–145, 2012.

L Glasl, K Kloos, F Giesert, A Roethig, B Di Benedetto, R Kühn, J Zhang, U Hafen, J Zerle, A Hofmann, MH de Angelis, KF Winklhofer, SM Hölter, DM Vogt-Weisenhorn, and W Wurst. Pink1-deficiency in mice impairs gait, olfaction and serotonergic innervation of the olfactory bulb. *Experimental neurology*, 235:214–27, 2012.

M Goujon, H McWilliam, W Li, F Valentin, S Squizzato, J Paern, and R Lopez. A new bioinformatics analysis tools framework at embl-ebi. *Nucleic Acids Research*, 38:W695–W699, 2010.

P Green, I Gispan-Herman, and G Yadid. Increased arachidonic acid concentration in the brain of flinders sensitive line rats, an animal model of depression. *The Journal of Lipid Research*, 46: 1093–1096, 2005.

S Griffiths-Jones, HK Saini, S van Dongen, and AJ Enright. mirbase: tools for microrna genomics. *Nucleic Acids Research*, 36:D154–D158, 2008.

S Grösgen, MOW Grimm, P Friess, and T Hartmann. Role of amyloid beta in lipid homeostasis. *Biochimica et Biophysica Acta*, 1801:966–974, 2010.

C Gross and R Hen. The developmental origins of anxiety. *Nature Reviews Neuroscience*, 5:545–552, 2004.

K Grundt, IV Haga, HS Huitfeldt, and AC Østvold. Identification and characterization of two putative nuclear localization signals (nls) in the dna-binding protein nucks. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1773:1398–1406, 2007.

L Gu, T Cui, C Fan, H Zhao, C Zhao, L Lu, and H Yang. Involvement of erk1/2 signaling pathway in dj-1-induced neuroprotection against oxidative stress. *Biochemical and Biophysical Research Communications*, 383:469–74, 2009.

RJ Guerreiro, J Beck, JR Gibbs, I Santana, MN Rossor, JM Schott, MA Nalls, H Ribeiro, B Santiago, NC Fox, C Oliveira, J Collinge, S Mead, A Singleton, and J Hardy. Genetic variability in clu and its association with alzheimer's disease. *PLoS ONE*, 5:e9510, 2010.

I Guyon, J Weston, S Barnhill, and V Vapnik. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46:389–422, 2002.

C Haanen and I Vermes. Apoptosis and inflammation. *Mediators of Inflammation*, 4:5–15, 1995.

HL Haas, OA Sergeeva, and O Selbach. Histamine in the nervous system. *Physiological reviews*, 88:1183–241, 2008.

I Hamamoto, Y Nishimura, T Okamoto, H Aizaki, M Liu, Y Mori, T Abe, T Suzuki, MMC Lai, T Miyamura, K Moriishi, and Y Matsuura. Human vap-b is involved in hepatitis c virus replication through interaction with ns5a and ns5b. *Journal of Virology*, 79:13473–13482, 2005.

RL Hamilton and R Bowser. Alzheimer disease pathology in amyotrophic lateral sclerosis. *Acta Neuropathologica*, 107:515–522, 2004.

M Handler, X Yang, and J Shen. Presenilin-1 regulates neuronal differentiation during neurogenesis. *Development*, 127:2593–2606, 2000.

T Hashimoto, K Nishi, J Nagasao, S Tsuji, and K Oyanagi. Magnesium exerts both preventive and ameliorating effects in an in vitro rat parkinson disease model involving 1-methyl-4-phenylpyridinium (mpp+) toxicity in dopaminergic neurons. *Brain Research*, 1197:143–151, 2008.

T Hastie, R Tibshirani, and J Friedman. *The elements of statistical learning: data mining, inference, and prediction; Chapter: 14.3 Cluster Analysis.* Springer New York.

NJ Haughey, A Nath, SL Chan, AC Borchard, MS Rao, and MP Mattson. Disruption of neurogenesis by amyloid β-peptide, and perturbed neural progenitor cell homeostasis, in models of alzheimer's disease. *Journal of Neurochemistry*, 83:1509–1524, 2002.

SS Hébert and B De Strooper. Alterations of the microrna network cause neurodegenerative disease. *Trends in Neurosciences*, 32:199–206, 2009.

C Henchcliffe and MF Beal. Mitochondrial biology and oxidative stress in parkinson disease pathogenesis. *Nature Clinical Practice Neurology*, 4:600–609, 2008.

IH Henn, L Bouman, JS Schlehe, A Schlierf, JE Schramm, E Wegener, K Nakaso, C Culmsee, B Berninger, D Krappmann, J Tatzelt, and KF Winklhofer. Parkin mediates neuroprotection through activation of ikappab kinase/nuclear factor-kappab signaling. *The Journal of Neuroscience*, 27:1868–78, 2007.

JP Herman and WE Cullinan. Neurocircuitry of stress: central control of the hypothalamo–pituitary–adrenocortical axis. *Trends in Neurosciences*, 20:78–84, 1997.

J Herz. Apolipoprotein e receptors in the nervous system. *Curr Opin Lipidol*, 20:190–6, 2009.

J Herz and Y Chen. Reelin, lipoprotein receptors and synaptic plasticity. *Nature Reviews Neuroscience*, 7:850–9, 2006.

H Higuchi, A Grambihler, A Canbay, SF Bronk, and GJ Gores. Bile acids up-regulate death receptor 5/trail-receptor 2 expression via a c-jun n-terminal kinase-dependent pathway involving sp1. *Journal of Biological Chemistry*, 279:51–60, 2004.

R Higuchi, C Fockler, G Dollinger, and R Watson. Kinetic pcr analysis: real-time monitoring of dna amplification reactions. *Nature Biotechnology*, 11:1026–1030, 1993.

JD Hildebrand and P Soriano. Overlapping and unique roles for c-terminal binding protein 1 (ctbp1) and ctbp2 during mouse development. *Molecular and Cellular Biology*, 22:5296–5307, 2002.

GU Höglinger, JJ Breunig, C Depboylu, C Rouaux, PP Michel, D Alvarez-Fischer, AL Boutillier, J Degregori, WH Oertel, P Rakic, EC Hirsch, and S Hunot. The prb/e2f cell-cycle pathway mediates cell death in parkinson's disease. *Proceedings of the National Academy of Sciences of the United States of America*, 104:3585–90, 2007.

KW Hong, JE Lim, and B Oh. A regulatory snp in akap13 is associated with blood pressure in koreans. *Journal of human genetics*, 56:205–10, 2011.

CW Hsu and CJ Lin. A comparison of methods for multiclass support vector machines. *IEEE Transaction on Neural Networks*, 13:415–25, 2002.

K Iida and I Nishimura. Gene expression profiling by dna microarray technology. *Critical Reviews in Oral Biology and Medicine*, 13:35–50, 2002.

Illumina. *BeadStudio Gene Expression Module v3.4 User Guide (11317265 Rev. A)*, July 2008.

M Ishisaka, K Kakefuda, M Yamauchi, K Tsuruma, M Shimazawa, A Tsuruta, and H Hara. Luteolin shows an antidepressant-like effect via suppressing endoplasmic reticulum stress. *Biological & Pharmaceutical Bulletin*, 34:1481–6, 2011a.

M Ishisaka, T Kudo, M Shimazawa, K Kakefuda, K Oyagi, Aand Hyakkoku, K Tsuruma, and H Hara. Restraint-induced expression of endoplasmic reticulum stress-related genes in the mouse brain. *Pharmacology & Pharmacy*, 2:10–16, 2011b.

A Iwata, S Miura, I Kanazawa, M Sawada, and N Nukina. alpha-synuclein forms a complex with transcription factor elk-1. *Journal of Neurochemistry*, 77:239–52, 2001.

G Jin and PH Howe. Regulation of clusterin gene expression by transforming growth factor beta. *Journal of Biological Chemistry*, 272:26620–6, 1997.

SE Jones and C Jomary. Clusterin. *The International Journal of Biochemistry & Cell Biology*, 34: 427–431, 2002.

JS Kauwe, C Cruchaga, K Mayo, C Fenoglio, S Bertelsen, P Nowotny, D Galimberti, E Scarpini, JC Morris, AM Fagan, DM Holtzman, and AM Goate. Variation in mapt is associated with cerebrospinal fluid tau levels in the presence of amyloid-beta deposition. *Proceedings of the National of Academy Sciences U S A*, 105:8050–4, 2008.

G Kerr, HJ Ruskin, M Crane, and P Doolan. Techniques for clustering gene expression data. *Computers in Biology and Medicine*, 38:283–93, 2008.

M Kertesz, N Iovino, U Unnerstall, U Gaul, and E Segal. The role of site accessibility in microrna target recognition. *Nature Genetics*, 39:1278–1284, 2007.

LR Kett, D Boassa, CC Ho, HJ Rideout, J Hu, M Terada, M Ellisman, and WT Dauer. Lrrk2 parkinson disease mutations enhance its microtubule association. *Human Molecular Genetics*, 21:890–899, 2012.

D Kim and LH Tsai. Bridging physiology and pathology in ad. *Cell*, 137:997–1000, 2009.

JM Kim, SW Kim, IS Shin, SJ Yang, WY Park, SJ Kim, HY Shin, and JS Yoon. Folate, vitamin b12, and homocysteine as risk factors for cognitive decline in the elderly. *Psychiatry Investigation*, 5:36–40, 2008.

JW Kim, TS Chang, JE Lee, SH Huh, SW Yeon, WS Yang, CO Joe, I Mook-Jung, RE Tanzi, TW Kim, and EJ Choi. Negative regulation of the sapk/jnk signaling pathway by presenilin 1. *The Journal of Cell Biology*, 153:457–464, 2001.

YS Kim and TH Joh. Microglia, major player in the brain inflammation: their roles in the pathogenesis of parkinson's disease. *Experimental & Molecular Medicine*, 38:333–47, 2006.

A Klingenhoff, K Frech, and T Werner. Regulatory modules shared within gene classes as well as across gene classes can be detected by the same in silico approach. *In Silico Biology*, 2:17–26, 2002.

M Knippenberg, MN Helder, BZ Doulabi, RA Bank, PI Wuisman, and J Klein-Nulend. Differential effects of bone morphogenetic protein-2 and transforming growth factor-beta1 on gene expression of collagen-modifying enzymes in human adipose tissue-derived mesenchymal stem cells. *Tissue Engineering. Part A*, 15:2213–25, 2009.

R Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2*, Montreal, Quebec, Canada.

H Kölsch, D Lütjohann, F Jessen, J Popp, F Hentschel, P Kelemen, S Friedrichs, TA Maier, and R Heun. Rxra gene variations influence alzheimer's disease risk and cholesterol metabolism. *Journal of cellular and molecular medicine*, 13:589–98, 2009.

D Kovalovsky, D Refojo, AC Liberman, D Hochbaum, MP Pereda, OA Coso, GK Stalla, F Holsboer, and E Arzt. Activation and induction of nur77/nurr1 in corticotrophs by crh/camp: involvement of calcium, protein kinase a, and mapk pathways. *Molecular Endocrinology*, 16:1638–51, 2002.

A Kowarsch, M Preusse, C Marr, and FJ Theis. mitalos: Analyzing the tissue-specific regulation of signaling pathways by human and mouse micrornas. *RNA*, 17:809–19, 2011.

A Krek, D Grün, MN Poy, R Wolf, L Rosenberg, EJ Epstein, P MacMenamin, I da Piedade, KC Gunsalus, M Stoffel, and N Rajewsky. Combinatorial microrna target predictions. *Nature Genetics*, 37:495–500, 2005.

G Kronenberg, M Colla, and M Endres. Folic acid, neurodegenerative and neuropsychiatric disease. *Current molecular medicine*, 9:315–23, 2009.

J Krüger and M Rehmsmeier. Rnahybrid: microrna target prediction easy, fast and flexible. *Nucleic Acids Research*, 34:W451–W454, 2006.

PH Kuhn, H Wang, B Dislich, A Colombo, U Zeitschel, JW Ellwart, E Kremmer, S Roszner, and SF Lichtenthaler. Adam10 is the physiologically relevant, constitutive alpha-secretase of the amyloid precursor protein in primary neurons. *The EMBO Journal*, 29:3020–3032, 2010.

AG Lalkhen and A McCluskey. Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia, Critical Care & Pain*, 8:221–223, 2008.

S Lammich, E Kojro, R Postina, S Gilbert, R Pfeiffer, M Jasionowski, C Haass, and F Fahrenholz. Constitutive and regulated alpha-secretase cleavage of alzheimer's amyloid precursor protein by a disintegrin metalloprotease. *Proceedings of the National Academy of Sciences of the United States of America*, 96:3922–3927, 1999.

T Langmann, C Schumacher, SG Morham, C Honer, S Heimerl, C Moehle, and G Schmitz. Znf202 is inversely regulated with its target genes abca1 and apoe during macrophage differentiation and foam cell formation. *Journal of Lipid Research*, 44:968–977, 2003.

L Lannfelt, H Basun, LO Wahlund, BA Rowe, and SL Wagner. Decreased alpha-secretase-cleaved amyloid precursor protein as a diagnostic marker for alzheimer's disease. *Nature Medicine*, 1: 829–832, 1995.

MA Larkin, G Blackshields, NP Brown, R Chenna, PA McGettigan, H McWilliam, F Valentin, IM Wallace, A Wilm, R Lopez, JD Thompson, TJ Gibson, and DG Higgins. Clustal w and clustal x version 2.0. *Bioinformatics*, 23:2947–2948, 2007.

L Lazzeroni and A Owen. Plaid models for gene expression data. *Statistica Sinica*, 12:61–86, 2000.

D Leclerc and R Rozen. Endoplasmic reticulum stress increases the expression of methylenetetrahydrofolate reductase through the ire1 transducer. *The journal of biological chemistry*, 283: 3151–60, 2008.

J Lee and H Ryu. Epigenetic modification is linked to alzheimer's disease: is it a maker or a marker? *BMB reports*, 43:649–655, 2010.

Y Lee, C Ahn, J Han, H Choi, J Kim, J Yim, J Lee, P Provost, O Rådmark, S Kim, and VN Kim. The nuclear rnase iii drosha initiates microrna processing. *Nature*, 425:415–9, 2003.

P Lei, S Ayton, DI Finkelstein, PA Adlard, CL Masters, and AI Bush. Tau protein: relevance to parkinson's disease. *International Journal of Biochemistry & Cell Biology*, 42:1775–8, 2010.

BP Lewis, CB Burge, and DP Bartel. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microrna targets. *Cell*, 120:15–20, 2005.

H Li and R Durbin. Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics*, 25:1754–1760, 2009.

T Li and L Lu. Epidermal growth factor-induced proliferation requires down-regulation of pax6 in corneal epithelial cells. *Journal of Biological Chemistry*, 280:12988–12995, 2005.

Y Liao and B Lönnerdal. Global microrna characterization reveals that mir-103 is involved in igf-1 stimulated mouse intestinal cell proliferation. *PLoS ONE*, 5:e12976, 2010.

AM Lidström, N Bogdanovic, C Hesse, I Volkman, P Davidsson, and K Blennow. Clusterin (apolipoprotein j) protein levels are increased in hippocampus and in frontal cortex in alzheimer's disease. *Experimental Neurology*, 154:511–521, 1998.

CM Lill, JT Roehr, MB McQueen, FK Kavvoura, S Bagade, BM Schjeide, LM Schjeide, E Meissner, U Zauft, NC Allen, T Liu, M Schilling, KJ Anderson, G Beecham, D Berg, JM Biernacka, A Brice, AL Destefano, CB Do, N Eriksson, SA Factor, MJ Farrer, T Foroud, T Gasser, T Hamza, JA Hardy, P Heutink, EM Hill-Burns, C Klein, JC Latourelle, DM Maraganore, ER Martin, M Martinez, RH Myers, MA Nalls, N Pankratz, H Payami, W Satake, WK Scott, M Sharma, AB Singleton, K Stefansson, T Toda, JY Tung, J Vance, NW Wood, CP Zabetian, 23andMe Genetic Epidemiology of Parkinson's Disease Consortium, International Parkinson's Disease Genomics Consortium, Parkinson's Disease GWAS Consortium, Wellcome Trust Case Control Consortium 2), P Young, RE Tanzi, MJ Khoury, F Zipp, H Lehrach, JP Ioannidis, and L Bertram. Comprehensive research synopsis and systematic meta-analyses in parkinson's disease genetics: The pdgene database. *PLoS Genet*, 8:e1002548, 2012.

M Liu and G Bing. Lipopolysaccharide animal models for parkinson's disease. *Parkinson's Disease*, 2011:327089, 2011.

Y Liu, L Yang, K Conde-Knape, D Beher, MS Shearman, and NS Shachter. Fatty acids increase presenilin-1 levels and gamma-secretase activity in pswt-1 cells. *Journal of Lipid Research*, 45: 2368–2376, 2004.

KJ Livak and TD Schmittgen. Analysis of relative gene expression data using real-time quantitative pcr and the $2^{-\Delta\Delta CT}$ method. *Methods*, 25:402–408, 2001.

M Lu, Q Zhang, M Deng, J Miao, Y Guo, W Gao, and Q Cui. An analysis of human microrna and disease associations. *PLoS One*, 3:e3420, 2008.

AK Lutz, N Exner, ME Fett, JS Schlehe, K Kloos, K Lämmermann, B Brunner, A Kurz-Drexler, F Vogel, AS Reichert, L Bouman, D Vogt-Weisenhorn, W Wurst, J Tatzelt, C Haass, and KF Winklhofer. Loss of parkin or pink1 function increases drp1-dependent mitochondrial fragmentation. *The Journal of biological chemistry*, 284:22938–51, 2009.

SC Madeira and AL Oliveira. Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 1:24–45, 2004.

M Maes, R Yirmyia, J Noraberg, S Brene, J Hibbeln, G Perini, M Kubera, P Bob, B Lerer, and M Maj. The inflammatory & neurodegenerative (i&nd) hypothesis of depression: leads for future research and new drug developments in depression. *Metabolic Brain Disease*, 24:27–53, 2009.

KM Mattila, JO Rinne, T Lehtimäki, M Röyttä, JP Ahonen, and M Hurme. Association of an interleukin 1b gene polymorphism (-511) with parkinson's disease in finnish patients. *Journal of Medical Genetics*, 39:400–2, 2002.

MP Mattson. Er calcium and alzheimer's disease: in a state of flux. *Science Signaling*, 3:pe10, 2010.

MA Maxwell and DA Cole. Weight change and appetite disturbance as symptoms of adolescent depression: toward an integrative biopsychosocial model. *Clinical Psychology Review*, 29:260–73, 2009.

L McConlogue, M Buttini, JP Anderson, EF Brigham, KS Chen, SB Freedman, D Games, K Johnson-Wood, M Lee, M Zeller, W Liu, R Motter, and S Sinha. Partial reduction of bace1 has dramatic effects on alzheimer plaque and synaptic pathology in app transgenic mice. *Journal of Biological Chemistry*, 282:26326–26334, 2007.

RW McLaughlin, JK De Stigter, LA Sikkink, EM Baden, and M Ramirez-Alvarado. The effects of sodium sulfate, glycosaminoglycans, and congo red on the structure, stability, and amyloid formation of an immunoglobulin light-chain protein. *Protein science*, 15:1710–22, 2006.

RS McNally, BK Davis, CM Clements, MA Accavitti-Loper, TW Mak, and JP Ting. Dj-1 enhances cell survival through the binding of cezanne, a negative regulator of nf-kappab. *The journal of biological chemistry*, 286:4098–106, 2011.

A Mendoza-Naranjo, C Gonzalez-Billault, and RB Maccioni. Abeta1-42 stimulates actin polymerization in hippocampal neurons through rac1 and cdc42 rho gtpases. *Journal of Cell Science*, 120:279–288, 2007.

H Meziane, JC Dodart, C Mathis, S Little, J Clemens, SM Paul, and A Ungerer. Memory-enhancing effects of secreted forms of the beta-amyloid precursor protein in normal and amnestic mice. *Proceedings of the National Academy of Sciences of the United States of America*, 95:12683–12688, 1998.

JA Miller, S Horvath, and DH Geschwind. Divergence of human and mouse brain transcriptome highlights alzheimer disease pathways. *Proceedings of the National Academy of Sciences*, 107: 12698–12703, 2010.

H Min and S Yoon. Got target? computational methods for microrna target prediction and their extension. *Experimental & Molecular Medicine*, 42:233–44, 2010.

KC Miranda, T Huynh, Y Tay, YS Ang, WL Tam, AM Thomson, B Lim, and I Rigoutsos. A pattern-based method for the identification of microrna binding sites and their corresponding heteroduplexes. *Cell*, 126:1203–1217, 2006.

PB Mitchell and D Hadzi-Pavlovic. Lithium treatment for bipolar disorder. *Bulletin of the World Health Organization*, 78:515–7, 2000.

JS Mo, MY Kim, EJ Ann, JA Hong, and HS Park. Dj-1 modulates uv-induced oxidative stress signaling through the suppression of mekk1 and cell death. *Cell death and differentiation*, 15: 1030–41, 2008.

MC Morris, DA Evans, JL Bienias, CC Tangney, DA Bennett, RS Wilson, N Aggarwal, and J Schneider. Consumption of fish and n-3 fatty acids and risk of incident alzheimer disease. *Archives of Neurology*, 60:940–6, 2003.

CE Moussa. Parkin attenuates wild-type tau modification in the presence of beta-amyloid and alpha-synuclein. *Journal of Molecular Neuroscience*, 37:25–36, 2009.

AK Müller-Rischart, A Pilsl, P Beaudette, M Patra, K Hadian, M Funke, R Peis, A Deinlein, C Schweimer, PH Kuhn, SF Lichtenthaler, E Motori, S Hrelia, W Wurst, D Trümbach, T Langer, D Krappmann, G Dittmar, J Tatzelt, and KF Winklhofer. The e3 ligase parkin maintains mitochondrial integrity by increasing linear ubiquitination of nemo. *Molecular Cell*, 49:908–921, 2013.

L Munoz, H Ranaivo, S Roy, W Hu, J Craft, L McNamara, L Chico, L Van Eldik, and DM Watterson. A novel p38 alpha mapk inhibitor suppresses brain proinflammatory cytokine up-regulation and attenuates synaptic dysfunction and behavioral deficits in an alzheimer's disease mouse model. *Journal of Neuroinflammation*, 4:21, 2007.

MM Muqit, S Gandhi, and NW Wood. Mitochondria in parkinson disease: back in fashion with a little help from genetics. *Archives of neurology*, 63:649–54, 2006.

TJ Naduvilath and L Dandona. Statistical analysis: the need, the concept, and the usage. *Indian Journal of Ophthalmology*, 46:51–8, 1998.

A Navarro and A Boveris. Brain mitochondrial dysfunction in aging, neurodegeneration, and parkinson's disease. *Frontiers in Aging Neuroscience*, 2:34, 2010.

RA Nixon, KI Saito, F Grynspan, WR Griffin, S Katayama, T Honda, PS Mohan, TB Shea, and M Beermann. Calcium-activated neutral proteinase (calpain) system in aging and alzheimer's disease. *Annals of the New York Academy of Sciences*, 747:77–91, 1994.

N Noren Hooten, K Abdelmohsen, M Gorospe, N Ejiogu, AB Zonderman, and MK Evans. microrna expression patterns reveal differential expression of target genes with age. *PLoS ONE*, 5:e10724, 2010.

T Oda, GM Pasinetti, HH Osterburg, C Anderson, SA Johnson, and CE Finch. Purification and characterization of brain clusterin. *Biochemical and Biophysical Research Communications*, 204: 1131–6, 1994.

A Olichon, L Baricault, N Gas, E Guillou, A Valette, P Belenguer, and G Lenaers. Loss of opa1 perturbates the mitochondrial inner membrane structure and integrity, leading to cytochrome c release and apoptosis. *The Journal of biological chemistry*, 278:7743–6, 2003.

O Oppenheimer, NK Cheung, and WL Gerald. The ret oncogene is a critical component of transcriptional programs associated with retinoic acid-induced differentiation in neuroblastoma. *Molecular cancer therapeutics*, 6:1300–9, 2007.

V Palada, J Terzić, J Mazzulli, G Bwala, J Hagenah, B Peterlin, AY Hung, C Klein, and D Krainc. Histamine n-methyltransferase thr105ile polymorphism is associated with parkinson's disease. *Neurobiology of Aging*, 33:836.e1–3, 2012.

W Pan. Incorporating gene functions as priors in model-based clustering of microarray gene expression data. *Bioinformatics*, 22:795–801, 2006.

GL Papadopoulos, M Reczko, VA Simossis, P Sethupathy, and AG Hatzigeorgiou. The database of experimentally supported targets: a functional update of tarbase. *Nucleic Acids Research*, 37: D155–D158, 2009.

CM Pariante and SL Lightman. The hpa axis in major depression: classical theories and new developments. *Trends in Neurosciences*, 31:464–468, 2008.

L Parisiadou and H Cai. Lrrk2 function on actin and microtubule dynamics in parkinson disease. *Communicative & Integrative Biology*, 3:396–400, 2010.

PJ Park, AJ Butte, and IS Kohane. Comparing expression profiles of genes with similar promoter regions. *Bioinformatics*, 18:1576–1584, 2002.

P Parra, F Serra, and A Palou. Expression of adipose microrna s is sensitive to dietary conjugated linoleic acid treatment in mice. *PLoS ONE*, 5:e13005, 2010.

M Pastorcic and HK Das. An upstream element containing an ets binding site is crucial for transcription of the human presenilin-1 gene. *Journal of Biological Chemistry*, 274:24297–24307, 1999.

J Pérez-Gil, P Estrada, C Acebal, and R Arche. Effect of albumin on acyl-coa: lysolecithin acyltransferase, lysolecithin: lysolecithin acyltransferase and acyl-coa hydrolase from rabbit lung. *Molecular and cell biochemistry*, 94:167–73, 1990.

Y Piao, HG Kim, MS Oh, and YK Pak. Overexpression of tfam, nrf-1 and myr-akt protects the mpp(+)-induced mitochondrial dysfunctions in neuronal cells. *Biochimica et Biophysica Acta*, pages 577–85, 2012.

A Pocivavsek and GW Rebeck. Inhibition of c-jun n-terminal kinase increases apoe expression in vitro and in vivo. *Biochemical and Biophysical Research Communications*, 387:516–20, 2009.

P Pollwein, CL Masters, and K Beyreuther. The expression of the amyloid precursor protein (app) is regulated by two gc-elements in the promoter. *Nucleic Acids Research*, 20:63–68, 1992.

F Porro, M Rosato-Siri, E Leone, L Costessi, A Iaconcig, E Tongiorgi, and AF Muro. β-adducin (add2) ko mice show synaptic plasticity, motor coordination and behavioral deficits accompanied by changes in the expression and phosphorylation levels of the α- and γ-adducin subunits. *Genes, Brain and Behavior*, 9:84–96, 2010.

R Postina, A Schroeder, I Dewachter, J Bohl, U Schmitt, E Kojro, C Prinzen, K Endres, C Hiemke, M Blessing, P Flamez, A Dequenne, E Godaux, F van Leuven, and F Fahrenholz. A disintegrin-metalloproteinase prevents amyloid plaque formation and hippocampal defects in an alzheimer disease mouse model. *The Journal of Clinical Investigation*, 113:1456–1464, 2004.

C Prinzen, D Trümbach, W Wurst, K Endres, R Postina, and F Fahrenholz. Differential gene expression in adam10 and mutant adam10 transgenic mice. *BMC Genomics*, 10:66, 2009.

D Puzzo, O Vitolo, F Trinchese, JP Jacob, A Palmeri, and O Arancio. Amyloid-beta peptide inhibits activation of the nitric oxide/cgmp/camp-responsive element-binding protein pathway during hippocampal synaptic plasticity. *The Journal of Neuroscience*, 25:6887–6897, 2005.

L Qin, X Wu, ML Block, Y Liu, GR Breese, JS Hong, DJ Knapp, and FT Crews. Systemic lps causes chronic neuroinflammation and progressive neurodegeneration. *Glia*, 55:453–62, 2007.

RL Rabenstein, NA Addy, BJ Caldarone, Y Asaka, LM Gruenbaum, LL Peters, DM Gilligan, RM Fitzsimonds, and MR Picciotto. Impaired synaptic plasticity and learning in mice lacking beta-adducin, an actin-regulating protein. *The Journal of Neuroscience*, 25:2138–2145, 2005.

JS Rao and SI Rapoport. Mood-stabilizers target the brain arachidonic acid cascade. *Current Molecular Pharmacology*, 2:207–214, 2009.

MA Rapp, M Schnaider-Beeri, HT Grossman, M Sano, DP Perl, DP Purohit, JM Gorman, and V Haroutunian. Increased hippocampal plaques and tangles in patients with alzheimer disease with a lifetime history of major depression. *Archives of general Psychiatry*, 63:161–7, 2006.

M Rehmsmeier, P Steffen, M Hochsmann, and R Giegerich. Fast and effective prediction of microrna/target duplexes. *RNA*, 10:1507–1517, 2004.

A Reiner, D Yekutieli, and Y Benjamini. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics*, 19:368–75, 2003.

Y Ren, HW Xu, F Davey, M Taylor, J Aiton, P Coote, F Fang, J Yao, D Chen, JX Chen, SD Yan, and FJ Gunn-Moore. Endophilin i expression is increased in the brains of alzheimer disease patients. *The Journal of Biological Chemistry*, 283:5685–91, 2008.

P Renbaum, R Beeri, E Gabai, M Amiel, M Gal, MU Ehrengruber, and E Levy-Lahad. Egr-1 upregulates the alzheimer's disease presenilin-2 gene in neuronal cells. *Gene*, 318:113–124, 2003.

EH Reynolds. Folic acid, ageing, depression, and dementia. *BMJ*, 324:1512–5, 2002.

GV Ronnett, S Ramamurthy, AM Kleman, LE Landree, and S Aja. Ampk in the brain: its roles in energy balance and neuroprotection. *Journal of Neurochemistry*, 109:17–23, 2009.

C Roodveldt, A Labrador-Garrido, E Gonzalez-Rey, R Fernandez-Montesinos, M Caro, CC Lachaud, CA Waudby, M Delgado, CM Dobson, and D Pozo. Glial innate immunity generated by non-aggregated alpha-synuclein in mouse: differences between wild-type and parkinson's disease-linked mutants. *PLoS One*, 5:e13481, 2010.

KM Rosen, CE Moussa, HK Lee, P Kumar, T Kitada, G Qin, Q Fu, and HW Querfurth. Parkin reverses intracellular beta-amyloid accumulation and its negative effects on proteasome function. *Journal of Neuroscience Research*, 88:167–78, 2010.

S Roßner, M Sastre, K Bourne, and SF Lichtenthaler. Transcriptional and translational regulation of bace1 expression–implications for alzheimer's disease. *Progress in Neurobiology*, 79:95–111, 2006.

M Rylski, R Amborska, K Zybura, FA Konopacki, GM Wilczynski, and L Kaczmarek. Yin yang 1 expression in the adult rodent brain. *Neurochemical Research*, 33:2556–2564, 2008.

H Sahai and A Khurshid. On analysis of epidemiological data involving a 2x2 contingency table: an overview of fisher's exact test and yates' correction for continuity. *Journal of Biopharmaceutical Statistics*, 5:43–70, 1995.

S Saita, M Shirane, T Natume, S Iemura, and KI Nakayama. Promotion of neurite extension by protrudin requires its interaction with vesicle-associated membrane protein-associated protein. *Journal of Biological Chemistry*, 284:13766–77, 2009.

T Sakata, M Kang, M Kurokawa, and H Yoshimatsu. Hypothalamic neuronal histamine modulates adaptive behavior and thermogenesis in response to endogenous pyrogen. *Obesity research*, 3: 707S–712S, 1995.

A Salminen, A Kauppinen, T Suuronen, K Kaarniranta, and J Ojala. Er stress in alzheimer's disease: a novel neuronal trigger for inflammation and alzheimer's pathology. *Journal of Neuroinflammation*, 6:41, 2009.

G Santpere, M Nieto, B Puig, and I Ferrer. Abnormal sp1 transcription factor expression in alzheimer disease and tauopathies. *Neuroscience Letters*, 397:30–34, 2006.

RM Sapolsky. The possibility of neurotoxicity in the hippocampus in major depression: a primer on neuron death. *Biological Psychiatry*, 48:755–765, 2000.

W Satake, Y Nakabayashi, I Mizuta, Y Hirota, C Ito, M Kubo, T Kawaguchi, T Tsunoda, M Watanabe, A Takeda, H Tomiyama, K Nakashima, K Hasegawa, F Obata, T Yoshikawa, H Kawakami, S Sakoda, M Yamamoto, N Hattori, M Murata, Y Nakamura, and T Toda. Genome-wide association study identifies common variants at four loci as genetic risk factors for parkinson's disease. *Nature Genetics*, 41:1303–1307, 2009.

S Sato, J Xu, S Okuyama, LB Martinez, SM Walsh, MT Jacobsen, RJ Swan, JD Schlautman, P Ciborowski, and T Ikezu. Spatial learning impairment, enhanced cdk5/p35 activity, and downregulation of nmda receptor expression in transgenic mice expressing tau-tubulin kinase 1. *The Journal of Neuroscience*, 28:14511–21, 2008.

J Satoh. Micrornas and their therapeutic potential for human diseases: aberrant microrna expression in alzheimer's disease brains. *Journal of Pharmacological Sciences*, 114:269–275, 2010.

JM Savitt, VL Dawson, and TM Dawson. Diagnosis and treatment of parkinson disease: molecules to medicine. *The journal of clinical investigations*, 116:1744–54, 2006.

DE Schmechel, AM Saunders, WJ Strittmatter, BJ Crain, CM Hulette, SH Joo, MA Pericak-Vance, D Goldgaber, and AD Roses. Increased amyloid β-peptide deposition in cerebral cortex as a consequence of apolipoprotein e genotype in late-onset alzheimer disease. *Proceedings of the National Academy of Sciences USA*, 90:9649–53, 1993.

DJ Selkoe. Alzheimer's disease: Genes, proteins, therapy. *Physiological Reviews*, 81:741–766, 2001.

DJ Selkoe and D Schenk. Alzheimer's disease: molecular understanding predicts amyloid-based therapeutics. *Annual Review of Pharmacology and Toxicology*, 43:545–584, 2003.

P Sethupathy, M Megraw, and AG Hatzigeorgiou. A guide through present computational approaches for the identification of mammalian microrna targets. *Nature Methods*, 3:881–886, 2006.

L Shao, MV Martin, SJ Watson, A Schatzberg, H Akil, RM Myers, EG Jones, WE Bunney, and MP Vawter. Mitochondrial involvement in psychiatric disorders. *Annals of medicine*, 40:281–95, 2008.

S Shimohama. Apoptosis in alzheimer's disease–an update. *Apoptosis*, 5:9–16, 2000.

JH Shin, HS Ko, H Kang, Y Lee, YI Lee, O Pletinkova, JC Troconso, VL Dawson, and TM Dawson. Paris (znf746) repression of pgc-1α contributes to neurodegeneration in parkinson's disease. *Cell*, 144:689–702, 2011.

V Shukla, SK Mishra, and HC Pant. Oxidative stress in neurodegeneration. *Advances in Pharmacological Sciences*, 2011:572634, 2011.

M Sjöbeck and E Englund. Alzheimer's disease and the cerebellum: a morphologic study on neuronal and glial changes. *Dementia and Geriatric Cognitive Disorders*, 12:211–8, 2001.

DK Slonim and I Yanai. Getting started in gene expression microarray analysis. *PLoS Computational Biology*, 5:e1000543, 2009.

M Smriga, M Kameishi, H Uneyama, and K Torii. Dietary l-lysine deficiency increases stress-induced anxiety and fecal excretion in rats. *The Journal of Nutrition*, 132:3744–3746, 2002.

M Smriga, S Ghosh, Y Mouneimne, PL Pellett, and NS Scrimshaw. Lysine fortification reduces anxiety and lessens stress in family members in economically weak communities in northwest syria. *Proceedings of the National Academy of Sciences of the USA*, 101:8285–8288, 2004.

M Solas, B Aisa, MC Mugueta, J Del Río, RM Tordera, and MJ Ramíirez. Interactions between age, stress and insulin on cognition: implications for alzheimer's disease. *Neuropsychopharmacology*, 35:1664–73, 2010.

H Steiner, E Winkler, and C Haass. Chemical cross-linking provides a model of the gamma-secretase complex subunit architecture and evidence for close proximity of the c-terminal fragment of presenilin with aph-1. *Journal of Biological Chemistry*, 283:34677–34686, 2008.

X Sun, Y Wang, H Qing, MA Christensen, Y Liu, W Zhou, Y Tong, C Xiao, Y Huang, S Zhang, X Liu, and W Song. Distinct transcriptional regulation and function of the human bace2 and bace1 genes. *The FASEB Journal*, 19:739–749, 2005.

T Taira, Y Saito, T Niki, SM Iguchi-Ariga, K Takahashi, and H Ariga. Dj-1 has a role in antioxidative stress to prevent cell death. *EMBO reports*, 5:213–8, 2004.

M Takagi, F Suto, T Suga, and J Yamada. Sterol regulatory element-binding protein-2 modulates human brain acyl-coa hydrolase gene transcription. *Molecular and cell biochemistry*, 275:199–206, 2005.

A Takashima, H Yamaguchi, K Noguchi, G Michel, K Ishiguro, K Sato, T Hoshino, M Hoshi, and K Imahori. Amyloid beta peptide induces cytoplasmic accumulation of amyloid protein precursor via tau protein kinase i/glycogen synthase kinase-3 beta in rat hippocampal neurons. *Neuroscience Letters*, 198:83–86, 1995.

PN Tan, M Steinbach, and V Kumar. *Introduction to Data Mining; Chapter: 8 Cluster Analysis*. Addison-Wesley, 2005.

H Taru and T Suzuki. Facilitation of stress-induced phosphorylation of beta-amyloid precursor protein family members by x11-like/mint2 protein. *Journal of Biological Chemistry*, 279:21628–21636, 2004.

SJ Teipel, W Bayer, GE Alexander, Y Zebuhr, D Teichberg, L Kulic, MB Schapiro, HJ Moller, SI Rapoport, and H Hampel. Progression of corpus callosum atrophy in alzheimer disease. *Archives of Neurology*, 59:243–248, 2002.

J Tenenbaum, M Walker, P Utz, and A Butte. Expression-based pathway signature analysis (epsa): Mining publicly available microarray data for insight into human disease. *BMC Medical Genomics*, 1:51, 2008.

E Teuling, S Ahmed, E Haasdijk, J Demmers, MO Steinmetz, A Akhmanova, D Jaarsma, and CC Hoogenraad. Motor neuron disease-associated mutant vesicle-associated membrane protein-associated protein (vap) b recruits wild-type vaps into endoplasmic reticulum-derived tubular aggregates. *The Journal of Neuroscience*, 27:9801–9815, 2007.

M Thambisetty, A Simmons, L Velayudhan, A Hye, J Campbell, Y Zhang, LO Wahlund, E Westman, A Kinsey, A Güntert, P Proitsi, J Powell, M Causevic, R Killick, K Lunnon, S Lynham, M Broadstock, F Choudhry, DR Howlett, RJ Williams, SI Sharp, C Mitchelmore, C Tunnard, R Leung, C Foy, D O'Brien, G Breen, SJ Furney, M Ward, I Kloszewska, P Mecocci, H Soininen, M Tsolaki, B Vellas, A Hodges, DG Murphy, S Parkins, JC Richardson, SM Resnick, L Ferrucci, DF Wong, Y Zhou, S Muehlboeck, A Evans, PT Francis, C Spenger, and S Lovestone. Association of plasma clusterin concentration with severity, pathology, and progression in alzheimer disease. *Archives of general psychiatry*, 67:739–48, 2010.

RK Thimmulappa, KH Mai, S Srisuma, TW Kensler, M Yamamoto, and S Biswal. Identification of nrf2-regulated genes induced by the chemopreventive agent sulforaphane by oligonucleotide microarray. *Cancer research*, 62:5196–203, 2002.

JE Tillman. *DJ-1, a novel androgen receptor binding protein, activates receptor signaling in prostate cancer and correlates with the development of androgen-independent disease*. PhD thesis, Vanderbilt University, 2007.

F Tippmann, J Hundt, A Schneider, K Endres, and F Fahrenholz. Up-regulation of the alpha-secretase adam10 by retinoic acid receptors and acitretin. *FASEB Journal*, 23:1643–54, 2009.

C Touma, M Bunck, L Glasl, M Nussbaumer, R Palme, H Stein, M Wolferstätter, R Zeh, M Zimbelmann, F Holsboer, and R Landgraf. Mice selected for high versus low stress reactivity: A new animal model for affective disorders. *Psychoneuroendocrinology*, 33:839–862, 2008.

V Trevino and F Falciani. Galgo: an r package for multivariate variable selection using genetic algorithms. *Bioinformatics*, 22:1154–6, 2006.

D Trümbach, C Graf, B Pütz, C Kühne, M Panhuysen, P Weber, F Holsboer, W Wurst, G Welzl, and JM Deussing. Deducing corticotropin-releasing hormone receptor type 1 signaling networks from gene expression data by usage of genetic algorithms and graphical gaussian models. *BMC Systems Biology*, 4:159, 2010.

A Tsolakidou, L Czibere, B Pütz, D Trümbach, M Panhuysen, JM Deussing, W Wurst, I Sillaber, R Landgraf, F Holsboer, and T Rein. Gene expression profiling in the stress control brain region hypothalamic paraventricular nucleus reveals a novel gene network including amyloid beta precursor protein. *BMC Genomics*, 11:546, 2010.

LE Tune. Depression and alzheimer's disease. *Depression and anxiety*, 8 Suppl1:91–95, 1998.

H Turner, T Bailey, and W Krzanowski. Improved biclustering of microarray data demonstrated through systematic performance tests. *Computational Statistics & Data Analysis*, 48:235–254, 2005.

J Turner, H Nicholas, D Bishop, JM Matthews, and M Crossley. The lim protein fhl3 binds basic krüppel-like factor/krüppel-like factor 3 and its co-repressor c-terminal-binding protein 2. *Journal of Biological Chemistry*, 278:12786–12795, 2003.

A Vagnoni, L Rodriguez, C Manser, KJ De Vos, and CC Miller. Phosphorylation of kinesin light chain 1 at serine 460 modulates binding and trafficking of calsyntenin-1. *Journal of cell science*, 124:1032–42, 2011.

S Vasudevan and JA Steitz. Au-rich-element-mediated upregulation of translation by fxr1 and argonaute 2. *Cell*, 128:1105–1118, 2007.

JS Verducci, VF Melfi, S Lin, Z Wang, S Roy, and CK Sen. Microarray analysis of gene expression: considerations in data mining and statistical treatment. *Physiological Genomics*, 25:355–63, 2006.

PM Visscher, MA Brown, MI McCarthy, and J Yang. Five years of gwas discovery. *American journal of human genetics*, 90:7–24, 2012.

GK Voeltz, WA Prinz, Y Shibata, JM Rist, and TA Rapoport. A class of membrane proteins shaping the tubular endoplasmic reticulum. *Cell*, 124:573–586, 2006.

AA Vostrov and WW Quitschke. The zinc finger protein ctcf binds to the apbbeta domain of the amyloid beta-protein precursor promoter. *Journal of Biological Chemistry*, 272:33353–33359, 1997.

L Wang, H Rajan, JL Pitman, M McKeown, and CC Tsai. Histone deacetylase-associating atrophin proteins are nuclear receptor corepressors. *Genes & Development*, 20:525–30, 2006.

S Wang and Q Cheng. Microarray analysis in drug discovery and clinical applications. *Methods in Molecular Biology*, 316:49–65, 2006.

WX Wang, BW Rajeev, AJ Stromberg, N Ren, G Tang, Q Huang, I Rigoutsos, and PT Nelson. The expression of microrna mir-107 decreases early in alzheimer's disease and may accelerate disease progression through regulation of beta-site amyloid precursor protein-cleaving enzyme 1. *The Journal of Neuroscience*, 28:1213–1223, 2008a.

WX Wang, BR Wilfred, SK Madathil, G Tang, Y Hu, J Dimayuga, AJ Stromberg, Q Huang, KE Saatman, and PT Nelson. mir-107 regulates granulin/progranulin with implications for traumatic brain injury and neurodegenerative disease. *The American Journal of Pathology*, 177: 334–345, 2010.

WX Wang, Q Huang, Y Hu, AJ Stromberg, and PT Nelson. Patterns of microrna expression in normal and early alzheimer's disease human temporal cortex: white matter versus gray matter. *Acta Neuropathologica*, 121:193–205, 2011.

X Wang, B Su, SL Siedlak, PI Moreira, H Fujioka, Y Wang, G Casadesus, and X Zhu. Amyloid-beta overproduction causes abnormal mitochondrial dynamics via differential modulation of mitochondrial fission/fusion proteins. *Proceedings of the National of Academy Sciences U S A*, 105: 19318–23, 2008b.

Y Wang, R Medvid, C Melton, R Jaenisch, and R Blelloch. Dgcr8 is essential for microrna biogenesis and silencing of embryonic stem cell self-renewal. *Nature Genetics*, 39:380–385, 2007.

YT Wang. Probing the role of ampar endocytosis and long-term depression in behavioural sensitization: relevance to treatment of brain disorders, including drug addiction. *British Journal of Pharmacology*, 153:S389–95, 2008.

Y Watanabe, M Tomita, and A Kanai. Computational methods for microrna target prediction. *Methods in Enzymology*, 427:65–86, 2007.

JA Webster, JR Gibbs, J Clarke, M Ray, W Zhang, P Holmans, K Rohrer, A Zhao, L Marlowe, M Kaleem, DS 3rd McCorquodale, C Cuello, D Leung, L Bryden, P Nath, VL Zismann, K Joshipura, MJ Huentelman, D Hu-Lince, KD Coon, DW Craig, JV Pearson, NACC-Neuropathology Group, CB Heward, EM Reiman, D Stephan, J Hardy, and AJ Myers. Genetic control of human brain transcript expression in alzheimer disease. *The American Journal of Human Genetics*, 84: 445–458, 2009.

B Wefers, C Hitz, SM Hölter, D Trümbach, J Hansen, P Weber, B Pütz, JM Deussing, MH de Angelis, T Roenneberg, F Zheng, C Alzheimer, A Silva, W Wurst, and R Kühn. Mapk signaling determines anxiety in the juvenile mouse brain but depression-like behavior in adults. *PLoS One*, 7:e35035, 2012.

T Werner. Models for prediction and recognition of eukaryotic promoters. *Mammalian Genome*, 10:168–175, 1999.

T Werner. Promoters can contribute to the elucidation of protein function. *Trends in Biotechnology*, 21:9–13, 2003.

T Werner, S Fessele, H Maier, and PJ Nelson. Computer modeling of promoter organization as a tool to study transcriptional coregulation. *The FASEB Journal*, 17:1228–1237, 2003.

JL Whitwell, SD Weigand, MM Shiung, BF Boeve, TJ Ferman, GE Smith, DS Knopman, RC Petersen, EE Benarroch, KA Josephs, and CR Jr Jack. Focal atrophy in dementia with lewy bodies on mri: a distinct pattern from alzheimer's disease. *Brain*, 130:708–19, 2007.

C Wider and ZK Wszolek. Clinical genetics of parkinson's disease and related disorders. *Parkinsonism & Related Disorders*, 13:S229–32, 2007.

RS Wilson, JA Schneider, JL Bienias, DA Evans, and DA Bennett. Parkinsonianlike signs and risk of incident alzheimer disease in older persons. *Archives of Neurology*, 60:539–544, 2003.

M Wines-Samuelson, M Handler, and J Shen. Role of presenilin-1 in cortical lamination and survival of cajal-retzius neurons. *Developmental Biology*, 277:332–346, 2005.

KF Winklhofer and C Haass. Mitochondrial dysfunction in parkinson's disease. *Biochimica et Biophysica Acta*, 1802:29–44, 2010.

TM Witkos, E Koscianska, and WJ Krzyzosiak. Practical aspects of microrna target prediction. *Current Molecular Medicine*, 11:93–109, 2011.

GE Wood, LT Young, LP Reagan, B Chen, and BS McEwen. Stress-induced structural remodeling in hippocampus: prevention by lithium treatment. *Proceedings of the National Academy of Sciences USA*, 101:3973–8, 2004.

RE Wragg and DV Jeste. Overview of depression and psychosis in alzheimer's disease. *The American Journal of Psychiatry*, 146:577–87, 1989.

YR Wu, CM Chen, JC Hwang, ST Chen, IH Feng, HC Hsu, CN Liu, YT Liu, YY Lai, HJ Huang, and GJ Lee-Chen. Interleukin-1α polymorphism has influence on late-onset sporadic parkinson's disease in taiwan. *Journal of Neural Transmission*, 114:1173–1177, 2007.

S Wuchty, W Fontana, IL Hofacker, and P Schuster. Complete suboptimal folding of rna and the stability of secondary structures. *Biopolymers*, 49:145–165, 1999.

JP Wyles, CR McMaster, and ND Ridgway. Vesicle-associated membrane protein-associated protein-a (vap-a) interacts with the oxysterol-binding protein to modify export from the endoplasmic reticulum. *The Journal of Biological Chemistry*, 277:29908–18, 2002.

H Xie, B Lim, and HF Lodish. Micrornas induced during adipogenesis that accelerate fat cell development are downregulated in obesity. *Diabetes*, 58:1050–1057, 2009.

Y Yamaguchi-Kabata, MK Shimada, Y Hayakawa, S Minoshima, R Chakraborty, T Gojobori, and T Imanishi. Distribution and effects of nonsense polymorphisms in human genes. *PLoS One*, 3: e3393, 2008.

M Yamakuchi, CD Lotterman, C Bao, RH Hruban, B Karim, JT Mendell, D Huso, and CJ Lowenstein. P53-induced microrna-107 inhibits hif-1 and tumor angiogenesis. *Proceedings of the National Academy of Sciences*, 107:6334–6339, 2010.

JO Yang, WY Kim, SY Jeong, JH Oh, S Jho, J Bhak, and NS Kim. Pdbase: a database of parkinson's disease-related genes and genetic variation using substantia nigra ests. *BMC Genomics*, 10:S32, 2009.

J Yao, T Hennessey, A Flynt, E Lai, MF Beal, and MT Lin. Microrna-related cofilin abnormality in alzheimer's disease. *PLoS ONE*, 5:e15546, 2010.

M Yao, RH Zhou, M Petreaca, L Zheng, J Shyy, and M Martins-Green. Activation of sterol regulatory element-binding proteins (srebps) is critical in il-8-induced angiogenesis. *Journal of leukocyte biology*, 80:608–20, 2006.

NH Zawia, DK Lahiri, and F Cardozo-Pelaez. Epigenetics, oxidative stress, and alzheimer disease. *Free Radical Biology and Medicine*, 46:1241–1249, 2009.

L Zhang, J Gough, D Christmas, DL Mattey, SC M Richards, J Main, D Enlander, D Honeybourne, JG Ayres, DJ Nutt, and JR Kerr. Microbial infections in eight genomic subtypes of chronic fatigue syndrome/myalgic encephalomyelitis. *Journal of Clinical Pathology*, 63:156–164, 2010.

S Zhang and J Cao. A close examination of double filtering with fold change and t test in microarray analysis. *BMC Bioinformatics*, 10:402, 2009.

Y Zhang, VL Dawson, and TM Dawson. Oxidative stress and genetics in the pathogenesis of parkinson's disease. *Neurobiology of Disease*, 7:240–250, 2000.

Z Zhang and M Gerstein. Of mice and men: phylogenetic footprinting aids the discovery of regulatory elements. *Journal of Biology*, 2:11, 2003.

N Zhong and J Xu. Synergistic activation of the human mnsod promoter by dj-1 and pgc-1alpha: regulation by sumoylation and oxidation. *Human Molecular Genetics*, 17:3357–67, 2008.

X Zhou and DP Tuck. Msvm-rfe: extensions of svm-rfe for multiclass gene selection on dna microarray data. *Bioinformatics*, 23:1106–1114, 2007.

X Zhu, HG Lee, AK Raina, G Perry, and MA Smith. The role of mitogen-activated protein kinase pathways in alzheimer's disease. *Neurosignals*, 11:270–81, 2002.

M Zuker and P Stiegler. Optimal computer folding of large rna sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9:133–148, 1981.

PA Zunszain, C Anacker, A Cattaneo, LA Carvalho, and CM Pariante. Glucocorticoids, cytokines and brain abnormalities in depression. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 35:722–729, 2011.

Appendix

List of abbreviations

Aβ	Amyloid-beta
ACOT7	Acyl-CoA thioesterase 7
ACTH	Adrenocorticotropic hormone
AD	Alzheimer's disease
ADAM10	A disintegrin and metalloproteinase 10
ADD3	Adducin 3
Aldh1l1	Aldehyde dehydrogenase 1 family, member L1
ALS	Amyotrophic lateral sclerosis
AMPK	AMP-activated protein kinase
APH1A	Anterior pharynx defective 1 homolog
APOE	Apolipoprotein E
APP	Amyloid-beta precursor protein
Ass1	Argininosuccinate synthase 1
ATF	Activating transcription factor
BACE1	Beta-site APP-cleaving enzyme 1
BACE2	Beta-site APP-cleaving enzyme 2
BCL2	B-cell CLL/lymphoma 2
bp	Basepairs
BS	Binding site
Ca^{2+}	Calcium
Cacna2d1	Calcium channel, voltage-dependent, alpha2/delta subunit 1
cAMP	Cyclic adenosine monophosphate
Ccnd2	Cyclin D2
Cdc42	Cell division cycle 42
CDK2	Cyclin-dependent kinase 2
CEBP	Ccaat/Enhancer Binding Protein
CLSTN1	Calsyntenin 1
CLU	Clusterin
CREB	cAMP-responsive element binding proteins
CRH	Corticotropin-releasing hormone
CTBP2	C-terminal binding protein 2
CTCF	CTCF and BORIS gene family, transcriptional regulators with 11 highly conserved zinc finger domains
DEAF	Homolog to deformed epidermal autoregulatory factor-1 from D. melanogaster
DGCR8	DiGeorge syndrome critical region gene 8
DHRS3	Dehydrogenase/reductase (SDR family) member 3
DJ-1/PARK7	Parkinson protein 7

dnADAM10 .... Dominant-negative ADAM10
EBOX ..... E-box binding factors
EGR ..... Early growth response
EGRF ..... EGR/nerve growth factor-induced protein C and related factors
EIF ..... Eukaryotic translation initiation factor
EIF2AK2 ..... Eukaryotic translation initiation factor 2-alpha kinase 2
EIF2alpha ..... Eukaryotic translation initiation factor-2 alpha
ER ..... Endoplasmic reticulum
Erk ..... Extracellular signal-regulated kinase
ETSF ..... Human and murine ETS1 factors
FABP4 ..... Fatty acid binding protein 4
FDR ..... False discovery rate
FN ..... False Negative
Fn1 ..... Fibronectin 1
Fos ..... FBJ osteosarcoma oncogene
FP ..... False Positive
GNAS ..... Guanine nucleotide binding protein (G protein), alpha stimulating activity polypeptide 1
GO ..... Gene Ontology
GR ..... Glucocorticoid receptor
GRK5 ..... G-protein-coupled receptor kinase 5
GRN ..... Granulin
Gsk3b ..... Glycogen synthase kinase 3 beta
GWAS ..... Genome-wide association studies
HNMT ..... Histamine N-methyltransferase
HPA ..... Hypothalamic-pituitary-adrenocortical
HR ..... High reactivity
IP_3 ..... Inositol trisphosphate
JNK ..... c-Jun NH_2-terminal kinase
KEGG ..... Kyoto Encyclopedia of Genes and Genomes
KLF ..... Kruppel-like factor
KLFS ..... Krueppel-like transcription factors
LOAD ..... Late onset Alzheimer's disease
LPS ..... Lipopolysaccharide
LR ..... Low reactivity
Lrrk2 ..... Leucine-rich repeat kinase 2
Map2k4 ..... Mitogen-activated protein kinase kinase 4
MAPK ..... Mitogen-activated protein kinase
MAPT ..... Microtubule-associated protein tau
MGI ..... Mouse Genome Informatics
Miip ..... Migration and invasion inhibitory protein
miRNAs ..... MicroRNAs
mSVM-RFE ... Multiclass support vector machine recursive feature elimination
MTHFR ..... Methylenetetrahydrofolate reductase (NAD(P)H)
NCSTN ..... Nicastrin
NFIA ..... Nuclear factor I/A
NFKB ..... Nuclear Factor-KappaB
NLP ..... Natural language processing

NR2F ..... Nuclear receptor subfamily 2 factors
NRF1 ..... Nuclear respiratory factor 1
NUCKS1 ..... Nuclear casein kinase and cyclin-dependent kinase substrate 1
OPA1 ..... Optic atrophy 1
PARK7/DJ-1 .. Parkinson protein 7
Parkin ..... Parkinson protein 2, E3 ubiquitin protein ligase
PCBE ..... PREB core-binding element
PCR ..... Polymerase Chain Reaction
PD ..... Parkinson's disease
PEN-2 ..... Presenilin enhancer 2
Pink1 ..... PTEN induced putative kinase 1
PLOD1 ..... Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1
Ppm1a ..... Protein phosphatase 1A, magnesium dependent, alpha isoform
PRKAA1 ..... Protein kinase, AMP-activated, alpha 1 catalytic subunit
PRKAA2 ..... Protein kinase, AMP-activated, alpha 2 catalytic subunit
PRKAG1 ..... Protein kinase, AMP-activated, gamma 1 non-catalytic subunit
PS1 ..... Presenilin 1
PS2 ..... Presenilin 2
qPCR ..... Quantitative real-time PCR
REEP5 ..... Receptor accessory protein 5
RERE ..... Arginine-glutamic acid dipeptide (RE) repeats
RFE ..... Recursive feature elimination
ROC ..... Receiver operator characteristic
RXRA ..... Retinoid X receptor, alpha
RXRF ..... RXR heterodimer binding sites
S100a4 ..... S100 calcium binding protein A4
SAGE ..... Serial analysis of gene expression
sAPPα ..... α-secretase-cleaved fragment of the amyloid precursor protein
Serpine1 ..... Serpin peptidase inhibitor, clade E
Sh3gl2 ..... SH3-domain GRB2-like 2
SNCA ..... Synuclein alpha
SNP ..... Single Nucleotide Polymorphism
SP1F ..... GC-Box factors SP1/GC
SVM ..... Support vector machine
SYP ..... Synaptophysin
TFBSs ..... Transcription Factor Binding Sites
TFs ..... Transcription Factors
TN ..... True Negative
TNF ..... Tumor necrosis factor
TP ..... True Positive
TSS ..... Transcription start site
Ttbk1 ..... Tau-tubulin kinase-1
UBIAD1 ..... UbiA prenyltransferase domain containing 1
UTR ..... Untranslated region
VAPA ..... Vesicle-associated membrane protein-associated protein A
VAPB ..... Vesicle-associated membrane protein-associated protein B
VEGFA ..... Vascular endothelial growth factor A
YY1 ..... Yin-Yang-1

YY1F ..... Activator/repressor binding to transcription initiation site
ZBPF ..... Zinc binding protein factors
ZF02 ..... C2H2 zinc finger transcription factors 2
ZNF ..... Zinc finger protein

List of Figures

List of Tables

Target genes of TFBS modules

Gene symbol	EntrezID	Molecular function
AP3S1	1176	protein binding, protein transporter activity, transporter activity
BAI2	576	G-protein coupled receptor activity, brain-specific angiogenesis inhibitor activity
C1orf95	375057	-
C20orf24	55969	molecular function, protein binding
CD164	8763	protein binding
EFNB2	1948	ephrin receptor binding
EIF5	1983	GTP binding, GTPase activity, nucleotide binding, translation factor activity, nucleic acid binding, translation initiation factor activity
EPAS1	2034	DNA binding, RNA polymerase II transcription factor activity, enhancer binding, histone acetyltransferase binding, protein binding, protein heterodimerization activity, contributes to sequence-specific DNA binding, signal transducer activity, specific RNA polymerase II transcription factor activity, transcription coactivator activity, transcription factor binding
GLUD1	2746	ADP, ATP, GTP, NAD binding, glutamate dehydrogenase activity, glutamate dehydrogenase activity, identical protein binding, leucine binding, nucleotide binding, oxidoreductase activity, protein binding
GNAS	2778	GTP binding, GTPase activity, guanyl nucleotide binding, identical protein binding, molecular function, nucleotide binding, protein binding, signal transducer activity
MCTP1	79772	calcium ion binding, NOT calcium-dependent phospholipid binding
NFIB	4781	transcription factor activity
NPTX2	4885	metal ion binding, molecular function, sugar binding
RB1CC1	9821	protein binding
RHEB	6009	GTP binding, GTPase activity, metal ion binding, nucleotide binding, protein binding

Gene symbol	EntrezID	Molecular function
VAPA	9218	protein binding, protein heterodimerization activity, signal transducer activity, structural molecule activity

Table A.1: AD patients target genes of the module CTCF-EGRF-SP1F derived from biclustering analysis with eight clusters. The columns describe the Gene symbol, GeneID and the molecular function as described in GO.

Gene symbol	EntrezID	Molecular function
Abi2	329165	-
Brunol5	319586	molecular function
Cacna2d1	12293	calcium channel activity, ion channel activity, metal ion binding, protein binding, voltage-gated calcium channel activity, voltage-gated ion channel activity
Ccdc47	67163	calcium ion binding
Dennd5a	19347	Rab GTPase binding, protein binding
Eid1	58521	protein binding, specific transcriptional repressor activity, transcription corepressor activity
Eif5	217869	GTP binding, binding, nucleotide binding, translation initiation factor activity
Gsk3b*)	56637	ATP binding, beta-catenin binding, glycogen synthase kinase 3 activity, integrin binding, ionotropic glutamate receptor binding, kinase activity, nucleotide binding, p53 binding, protein binding, protein kinase activity, protein serine/threonine kinase activity, tau-protein kinase activity
Hnrnpk	15387	DNA binding, RNA binding, nucleic acid binding, protein binding, single-stranded DNA binding
Jmjd1c	108829	metal ion binding, molecular function, oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen
Klc1	16593	binding, microtubule motor activity, motor activity, protein binding
Myh10	77579	ADP, ATP binding, actin binding, actin filament binding, actin-dependent ATPase activity, calmodulin binding, microfilament motor activity, motor activity, nucleotide binding, protein binding
Nfix	18032	DNA binding, protein binding, transcription activator activity, transcription factor activity
Nptxr	73340	pentraxin receptor activity, protein complex binding

Gene symbol	**EntrezID**	**Molecular function**
Pafah1b1	18472	dynein intermediate chain binding, hydrolase activity, microtubule binding, phosphoprotein binding, protein complex binding, protein homodimerization activity
Pde10a	23984	3',5'-cyclic-GMP phosphodiesterase activity, 3',5'-cyclic-nucleotide phosphodiesterase activity, cAMP binding, catalytic activity, cyclic-nucleotide phosphodiesterase activity, drug binding, hydrolase activity, metal ion binding, nucleotide binding
Pebp1	23980	ATP binding, lipid binding, kinase binding, mitogen-activated protein kinase binding, nucleotide binding, peptidase inhibitor activity, protein kinase binding, receptor binding, serine-type endopeptidase inhibitor activity
Ppp3cb	19056	calcium-dependent protein serine/threonine phosphatase activity, calmodulin binding, hydrolase activity, metal ion binding, phosphoprotein phosphatase activity, protein heterodimerization activity, protein serine/threonine phosphatase activity
Prei4	74182	carbohydrate binding, catalytic activity, glycerophosphodiester phosphodiesterase activity, hydrolase activity, phosphoric diester hydrolase activity
Ptbp2	56195	RNA binding, mRNA binding, nucleic acid binding, nucleotide binding, protein binding
Rab6	19346	ATPase activator activity, GTP binding, nucleotide binding, protein N-terminus binding, protein binding
Reep5	13476	protein binding, receptor activity
Sgtb	218544	Binding, protein heterodimerization activity, protein homodimerization activity
Syp	20977	SH2 domain binding, identical protein binding, protein binding, protein complex binding, syntaxin-1 binding, transporter activity
Tmed7	66676	molecular function
Tsc22d1	21807	transcription factor activity
Tspan2	70747	molecular function
Ttc7b	104718	molecular function
Ubr3	68795	ligase activity, metal ion binding, protein binding, ubiquitin-protein ligase activity, zinc ion binding
Vamp2	22318	SNARE binding, calmodulin binding, myosin binding, phospholipid binding, protein binding, protein complex binding, syntaxin binding, syntaxin-1 binding
Vapa	30960	protein heterodimerization activity, structural molecule activity
Wdr6	83669	molecular function

Gene symbol	EntrezID	Molecular function
Ywhae	22627	enzyme binding, monooxygenase activity, protein binding, protein complex binding, protein domain specific binding
Ywhaq	22630	monooxygenase activity, protein domain specific binding

Table A.2: Double transgenic mice target genes of the module CTCF-EGRF-SP1F derived from biclustering analysis with 13 clusters. The columns describe the Gene symbol, GeneID and the molecular function as described in GO. *) incorporated in AlzGene database

Gene symbol	EntrezID	Molecular function
ATP1A3	478	ATP binding, ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism, hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances, metal ion binding, monovalent inorganic cation transmembrane transporter activity, nucleotide binding, sodium:potassium-exchanging ATPase activity
ATP6V1A	523	ATP binding, hydrogen ion transporting ATP synthase activity, rotational mechanism, hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances, nucleotide binding, proton-transporting ATPase activity, rotational mechanism
CALY	50632	clathrin light chain binding, dopamine receptor binding
CCK	885	hormone activity, neuropeptide hormone activity, protein binding
EEF1A2	1917	GTP binding, GTPase activity, nucleotide binding, protein binding, translation elongation factor activity, translation factor activity, nucleic acid binding
GARS	2617	ATP binding, glycine-tRNA ligase activity, nucleotide binding, protein dimerization activity
GOT1*)	2805	L-aspartate: 2-oxoglutarate aminotransferase activity, carboxylic acid binding, phosphatidylserine decarboxylase activity, pyridoxal phosphate binding
NAPB	63908	binding
NELL2	4753	calcium ion binding, protein binding, structural molecule activity
PNMA2	10687	protein binding
REEP5	7905	molecular function, protein binding

Gene symbol	EntrezID	Molecular function
SYP	6855	SH2 domain binding, calcium ion binding, cholesterol binding, identical protein binding, protein complex binding, syntaxin-1 binding, transporter activity

Table A.3: LOAD patients target genes of the module CTCF-EGRF-SP1F derived from biclustering analysis with 18 clusters. The columns describe the Gene symbol, GeneID and the molecular function as described in GO. *) incorporated in AlzGene database

Gene symbol	EntrezID	Molecular function
ADD3	120	actin binding, calmodulin binding, metal ion binding, protein kinase C binding, structural constituent of cytoskeleton
CDC42EP4	23580	GTP-Rho binding, protein binding
CTBP2	1488	NAD or NADH binding, cofactor binding, oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor, protein binding, transcription repressor activity
FXN	2395	2 iron, 2 sulfur cluster binding, ferric iron binding, ferrous iron binding, iron chaperone activity, iron-sulfur cluster binding, protein binding
MAFF	23764	sequence-specific DNA binding, transcription factor activity
NBPF14	25832	-
PAX6	5080	DNA binding, protein binding, sequence-specific DNA binding, transcription factor activity
RAB31	11031	GTP binding, GTPase activity, nucleotide binding
RHOQ	23433	GBD domain binding, GTP binding, GTPase activity, nucleotide binding, profilin binding, protein binding
SEPT9	10801	GTP binding, GTPase activity, nucleotide binding, protein binding
SOX10	6663	RNA polymerase II transcription factor activity, enhancer binding, chromatin binding, identical protein binding, promoter binding, protein binding, transcription coactivator activity
SUV420H1	51111	histone methyltransferase activity (H4-K20 specific), methyltransferase activity, protein binding, transferase activity
TMEM184B	25829	-

Gene symbol	EntrezID	Molecular function
ZBTB16	7704	DNA binding, double-stranded DNA binding, identical protein binding, metal ion binding, protein C-terminus binding, protein binding, protein domain specific binding, protein homodimerization activity, specific transcriptional repressor activity, transcription factor/repressor activity, zinc ion binding

Table A.4: AD patients target genes of the module CTCF-SP1F-ZBPF derived from biclustering analysis with five clusters. The columns describe the Gene symbol, GeneID and the molecular function as described in GO.

Gene symbol	EntrezID	Molecular function
ADD3	120	actin binding, calmodulin binding, metal ion binding, protein kinase C binding, structural constituent of cytoskeleton
ATG10	83734	Atg12 ligase activity, ligase activity, protein binding
CXorf41	139212	-
EEF1D	1936	protein binding, signal transducer activity, translation elongation factor activity, translation factor activity, nucleic acid binding
GLUL	2752	ATP binding, glutamate decarboxylase activity, glutamate-ammonia ligase activity, identical protein binding, ligase activity, lyase activity, nucleotide binding
MT3	4504	antioxidant activity, copper ion binding, metal ion binding, zinc ion binding
NUCKS1	64710	-
PIP4K2B	8396	1-phosphatidylinositol-4-phosphate 5-kinase activity, 1-phosphatidylinositol-5-phosphate 4-kinase activity, ATP binding, kinase activity, nucleotide binding, protein binding, receptor signaling protein activity, transferase activity
RBX1	9978	NEDD8 ligase activity, ligase activity, metal ion binding, protein binding, contributes to ubiquitin-protein ligase activity, zinc ion binding
RPL35	11224	mRNA binding, protein binding, structural constituent of ribosome
SHCBP1	79801	SH2 domain binding, protein binding

Table A.5: LOAD patients target genes of the module CTCF-SP1F-ZBPF derived from biclustering analysis with 18 clusters. The columns describe the Gene symbol, GeneID and the molecular function as described in GO.

Gene symbol	EntrezID	Molecular function
ADD3	120	actin binding, calmodulin binding, metal ion binding, protein kinase C binding, structural constituent of cytoskeleton
CDC42EP4	23580	GTP-Rho binding, protein binding
CLU*)	1191	misfolded protein binding, protein binding
CTBP2	1488	NAD or NADH binding, cofactor binding, oxidoreductase activity, oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor, protein binding, transcription repressor activity
FXN	2395	2 iron, 2 sulfur cluster binding, ferric iron binding, ferrous iron binding, iron chaperone activity, iron-sulfur cluster binding, protein binding
NFE2L1	4779	protein dimerization activity, sequence-specific DNA binding, transcription cofactor activity, transcription factor activity
NUCKS1	64710	-
PAX6	5080	DNA binding, protein binding, sequence-specific DNA binding, transcription factor activity
RAB31	11031	GTP binding, GTPase activity, nucleotide binding
RHOQ	23433	GBD domain binding, GTP binding, GTPase activity, nucleotide binding, profilin binding, protein binding
SEPT9	10801	GTP binding, GTPase activity, nucleotide binding, protein binding
STOM	2040	protein binding
SUV420H1	51111	histone methyltransferase activity (H4-K20 specific), methyltransferase activity, protein binding, transferase activity
TRAM1	23471	protein binding, receptor activity
WNK1	65125	ATP binding, molecular function, nucleotide binding, protein binding, protein kinase inhibitor activity, protein serine/threonine kinase activity, transferase activity
ZBTB16	7704	DNA binding, double-stranded DNA binding, identical protein binding, metal ion binding, protein C-terminus binding, protein binding, protein domain specific binding, protein homodimerization activity, specific transcriptional repressor activity, transcription factor/repressor activity, zinc ion binding

Table A.6: AD patients target genes of the module KLFS-SP1F-ZBPF derived from biclustering analysis with five clusters. The columns describe the Gene symbol, GeneID and the molecular function as described in GO. *) incorporated in AlzGene database

Gene symbol	EntrezID	Molecular function
ADD3	120	actin binding, calmodulin binding, metal ion binding, protein kinase C binding, structural constituent of cytoskeleton
ATG10	83734	Atg12 ligase activity, ligase activity, protein binding
CKB	1152	ATP binding, creatine kinase activity, nucleotide binding, protein binding
CLU*)	1191	misfolded protein binding, protein binding
CXorf41	139212	-
EEF1D	1936	protein binding, signal transducer activity, translation elongation factor activity, translation factor activity, nucleic acid binding
GLUL	2752	ATP binding, glutamate decarboxylase activity, glutamate-ammonia ligase activity, identical protein binding, ligase activity, lyase activity, nucleotide binding
NUCKS1	64710	-
PIP4K2B	8396	1-phosphatidylinositol-4-phosphate 5-kinase activity, 1-phosphatidylinositol-5-phosphate 4-kinase activity, ATP binding, kinase activity, nucleotide binding, protein binding, receptor signaling protein activity, transferase activity
RPL35	11224	mRNA binding, protein binding, structural constituent of ribosome
SHCBP1	79801	SH2 domain binding, protein binding

Table A.7: LOAD patients target genes of the module KLFS-SP1F-ZBPF derived from biclustering analysis with 18 clusters. The columns describe the Gene symbol, GeneID and the molecular function as described in GO. *) incorporated in AlzGene database

Gene symbol	EntrezID	Molecular function
1500012F01Rik	68949	-
Acsl1	14081	ATP binding, acetate-CoA ligase (ADP-forming) activity, catalytic activity, ligase activity, long-chain fatty acid-CoA ligase activity, nucleotide binding
Actr10	56444	protein binding
Add1	11518	T cell receptor binding, actin binding, calmodulin binding, metal ion binding, structural molecule activity
Adss	11566	GTP binding, adenylosuccinate synthase activity, ligase activity, magnesium ion binding, metal ion binding, nucleotide binding
Atp11b	76295	ATP binding, hydrolase activity, molecular function, nucleotide binding

Gene symbol	**EntrezID**	**Molecular function**
Atp8b2	54667	ATP binding, ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism, hydrolase activity, hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances, magnesium ion binding, metal ion binding, nucleotide binding, phospholipid-translocating ATPase activity, protein binding
BC018507	218333	molecular function
Brsk2	75770	ATP binding, kinase activity, magnesium ion binding, metal ion binding, nucleotide binding, protein kinase activity, protein serine/threonine kinase activity, transferase activity
Bzw1	66882	binding, molecular function
Ccl27a	20301	Cytokine activity, protein binding
Cdc42	12540	GTP-dependent protein binding, GTPase activity, mitogen-activated protein kinase kinase kinase binding, nucleotide binding, protein binding
Cdc42se2	72729	molecular function
Col25a1*)	77018	molecular function
Ctnna2	12386	Cadherin binding, structural molecule activity
D10Ertd610e	52666	Rho guanyl-nucleotide exchange factor activity, guanyl-nucleotide exchange factor activity, protein binding
D1Ertd622e	52392	molecular function
Dlat	235339	acyltransferase activity, dihydrolipoyllysine-residue acetyltransferase activity, lipoic acid binding, protein binding, transferase activity
Dnpep	13437	aminopeptidase activity, hydrolase activity, metal ion binding, metallopeptidase activity, peptidase activity, zinc ion binding
Dpysl4	26757	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, protein binding
Eif1ay	66235	RNA binding, molecular function, translation initiation factor activity
Epb4.1l1	13821	actin binding, binding, cytoskeletal protein binding, protein binding, structural molecule activity
Fam135a	68187	molecular function
Fam169a	320557	molecular function
Fam49b	223601	molecular function
Fbxo3	57443	-
Fkrp	243853	Transferase activity
Fnbp1	14269	Lipid binding, protein binding

Gene symbol	**EntrezID**	**Molecular function**
Foxj3	230700	DNA binding, sequence-specific DNA binding, transcription factor activity
Fry	320365	molecular function
Gabra4	14397	GABA-A receptor activity, chloride channel activity, extracellular ligand-gated ion channel activity, ion channel activity, receptor activity
Gabrb3	14402	GABA-A receptor activity, chloride channel activity, extracellular ligand-gated ion channel activity, ion channel activity, receptor activity
Gria2	14800	PDZ domain binding, extracellular-glutamate-gated ion channel activity, ion channel activity, ionotropic glutamate receptor activity, protein binding, protein kinase binding, receptor activity
Hdgfrp3	29877	chromatin binding, growth factor activity
Hnrnpu	51810	ATP, DNA, RNA binding, nucleic acid binding, nucleotide binding, protein binding
Hspa12a	73442	ATP binding, molecular function, nucleotide binding
Il33*)	77125	Cytokine activity
Khdrbs1	20218	RNA binding, SH3 domain binding, SH3/SH2 adaptor activity, protein binding, transcription repressor activity
Kndc1	76484	Ras guanyl-nucleotide exchange factor activity, guanyl-nucleotide exchange factor activity, kinase activity, protein binding, protein serine/threonine kinase activity
Larp5	217980	RNA binding, molecular function, nucleic acid binding
Map2k4	26398	ATP binding, JUN kinase kinase activity, MAP kinase kinase activity, kinase activity, mitogen-activated protein kinase kinase kinase binding, nucleotide binding, protein kinase activity, protein serine/threonine kinase activity, protein tyrosine kinase activity, transferase activity
Mapk1	26413	ATP binding, MAP kinase 2 activity, RNA polymerase II carboxy-terminal domain kinase activity, kinase activity, mitogen-activated protein kinase kinase kinase binding, nucleotide binding, phosphotyrosine binding, protein binding, protein kinase activity, protein serine/threonine kinase activity, transcription factor binding, transcription regulator activity, transferase activity
Mapk9	26420	ATP binding, JUN kinase activity, MAP kinase activity, caspase activator activity, kinase activity, mitogen-activated protein kinase kinase kinase binding, nucleotide binding, protein binding, protein kinase activity, protein serine/threonine kinase activity, transferase activity

Gene symbol	**EntrezID**	**Molecular function**
March7	57438	ligase activity, metal ion binding, molecular function, zinc ion binding
Mll5	69188	histone-lysine N-methyltransferase activity, metal ion binding, methyltransferase activity, protein binding, transferase activity, zinc ion binding
Mtap4	17758	-
Nap1l1	53605	protein binding
Nr2f1	13865	DNA binding, ligand-dependent nuclear receptor activity, metal ion binding, receptor activity, sequence-specific DNA binding, steroid hormone receptor activity, transcription activator activity, transcription factor activity, zinc ion binding
Ola1	67059	ATP binding, GTP binding, hydrolase activity, molecular function, nucleotide binding
Pcmt1	18537	S-adenosylmethionine-dependent methyltransferase activity, methyltransferase activity, protein binding, protein-L-isoaspartate (D-aspartate) O-methyltransferase activity, transferase activity
Pcsk2	18549	endopeptidase activity, hydrolase activity, peptidase activity, protein complex binding, serine-type endopeptidase activity
Pfn1	18643	Rho GTPase binding, actin binding, phosphatidylinositol-4,5-bisphosphate binding, protein binding, receptor binding
Polr1d	20018	Protein binding
Ppm1a	19042	catalytic activity, hydrolase activity, magnesium ion binding, manganese ion binding, metal ion binding, phosphoprotein phosphatase activity, protein C-terminus binding, protein serine/threonine phosphatase activity
Prpf39	328110	Binding, molecular function
Psma7	26444	endopeptidase activity, hydrolase activity, peptidase activity, threonine-type endopeptidase activity
Ptpra	19262	hydrolase activity, phosphatase activity, phosphoprotein phosphatase activity, protein complex binding, protein tyrosine phosphatase activity, receptor activity
Ptprz1	19283	hydrolase activity, phosphoprotein phosphatase activity, protein binding, receptor activity
Rasgef1b	320292	guanyl-nucleotide exchange factor activity, molecular function
Rnf115	67845	ligase activity, metal ion binding, protein binding, zinc ion binding
Rundc3b	242819	molecular function

Gene symbol	**EntrezID**	**Molecular function**
Sesn3	75747	molecular function
Slc23a2	54338	L-ascorbate:sodium symporter activity, L-ascorbic acid transporter activity, symporter activity, transporter activity
Srrm1	51796	DNA binding, RNA binding, NOT RNA binding
Ssx2ip	99167	actinin binding, protein binding
Tanc2*)	77097	binding, molecular function
Tlk1	228012	ATP binding, kinase activity, molecular function, nucleotide binding, protein kinase activity, protein serine/threonine kinase activity, transferase activity
Tm7sf3	67623	molecular function
Trim3	55992	Metal ion binding, protein binding, zinc ion binding
Ttll1	319953	ligase activity, molecular function, tubulin-tyrosine ligase activity
Txndc14	66958	molecular function
Ube2e2	218793	ATP binding, ligase activity, molecular function, nucleotide binding, small conjugating protein ligase activity, ubiquitin-protein ligase activity
Zbtb4	75580	metal ion binding, molecular function
Zc3h14	75553	molecular function

Table A.8: Double transgenic mice target genes of the module CTCF-SP1F-ZBPF derived from biclustering analysis with 13 clusters. The columns describe the Gene symbol, GeneID and the molecular function as described in GO. *) incorporated in AlzGene database

Enrichment analysis

Dataset	TFBS module	r	n	R	N	p-value
AD patients	CTCF-EGRF-SP1F	17	8885	68	97259	0.0001
	KLFS-SP1F-ZBPF	17	9649	66	97259	0.0002
	CTCF-SP1F-ZBPF	16	8742	66	97259	0.0002
double transgenic mice	CTCF-SP1F-ZBPF	76	13296	232	101113	1.79×10^{-14}
	CTCF-EGRF-SP1F	35	11123	119	101113	3.5×10^{-8}
LOAD patients	CTCF-SP1F-ZBPF	11	16118	26	97259	0.0017
	KLFS-SP1F-ZBPF	11	17417	26	97259	0.0033
	CTCF-EGRF-SP1F	14	13910	45	97259	0.0041

Table A.9: TFBS module enrichment analysis. The first and second column describe the dataset and the TFBS module, respectively. N is the overall number of promoter regions in the corresponding genome from the Genomatix database ElDorado (NCBI build 37) and R corresponds to the amount of promoters in the appropriate cluster. n and r are the number of promoters in the genome and cluster, respectively, maintaining the TFBS module (column 2). The four parameters are used for p-value computation (last column).

TFBS module	KEGG pathway	r	n	R	N	p-value
CTCF-EGRF-SP1F	Long-term depression	2	69	25	5368	0.0404
	Calcium signaling pathway	3	177	25	5368	0.0476
	MAPK signaling pathway	7	265	43	6060	0.0024
	Alzheimer's disease	6	191	43	6060	0.0021
	Wnt signaling pathway	4	149	43	6060	0.0206
CTCF-SP1F-ZBPF	MAPK signaling pathway	8	265	48	6060	0.0010
KLFS-SP1F-ZBPF	Wnt signaling pathway	3	150	19	5368	0.0149

Table A.10: TFBS modules: KEGG pathway enrichment analysis. The first and second column describe the TFBS module and the KEGG pathway, respectively. N is the amount of genes of all human (5368) and mouse (6060) KEGG pathways and R corresponds to the number of genes in the appropriate cluster used for TFBS module prediction and also contained in the KEGG pathways. n is the number of genes in the KEGG pathway (column 2) and r are all genes in the cluster being involved in the KEGG pathway (column 2). The four parameters are used for p-value computation (last column).

Dataset	KEGG pathway	r	n	R	N	p-value
mSVM-RFE	One carbon pool by folate	2	18	53	6662	0.0088
	Arachidonic acid metabolism	3	90	53	6662	0.0344
	Lysine degradation	2	45	53	6662	0.0495
Cluster	Neuroactive ligand-receptor interaction	2	277	8	6662	0.0409
	Endocytosis	2	220	10	6662	0.0410

Table A.11: HR versus LR microarray: KEGG pathway enrichment analysis. The first column describes the dataset for KEGG enrichment analysis either genes occurring in all gene selections of mSVM-RFE or coregulated genes of a cluster. The second column shows the KEGG pathway. N is the amount of genes of all mouse KEGG pathways and R corresponds to the number of genes in the appropriate dataset, which are also contained in the KEGG pathways. n is the number of genes in the KEGG pathway (column 2) and r are all genes in the dataset being involved in the KEGG pathway (column 2). The four parameters are used for p-value computation (last column).

KEGG pathway	r	n	R	N	p-value
Huntington's disease	5	190	16	6662	6.05×10^{-5}
Citrate cycle (TCA cycle)	2	31	16	6662	0.0024
Oxidative phosphorylation	3	140	16	6662	0.0042
Parkinson's disease	3	140	16	6662	0.0042
Alzheimer's disease	3	180	16	6662	0.0084
Cell cycle	2	125	16	6662	0.0353

Table A.12: DJ-1 knockout microarray: KEGG pathway enrichment analysis. Coregulated genes of one cluster were analysed for enrichment in KEGG pathways (column 1). N is the amount of genes of all mouse KEGG pathways and R corresponds to the number of genes in the cluster, which are also contained in the KEGG pathways. n is the number of genes in the KEGG pathway (column 1) and r are all genes in the cluster being involved in the KEGG pathway (column 1). The four parameters are used for p-value computation (last column).

DJ-1 knockout microarray: genes adjusted p-value < 0.35

Gene information			rawp	adj.P.Val	foldchange			
Probeset	**Gene symbol**	**Entrez ID**	**LPS ko vs. LPS wt**	**LPS ko vs. LPS wt**	**LPS ko vs. LPS wt**	**Ctrl ko vs. Ctrl wt**	**LPS ko vs. Ctrl ko**	**LPS wt vs. Ctrl wt**
10518774	Park7	57320	7.63×10^{-9}	0.00022	0.06	0.05	0.83	0.76
10510129	Dhrs3	20148	3.89×10^{-7}	0.00562	2.02	1.6	0.57	0.45
10518520	Ubiad1	71707	2.03×10^{-6}	0.01948	1.89	1.73	0.97	0.89
10510687	Acot7	70025	2.97×10^{-6}	0.02145	0.49	0.63	1.11	1.45
10510270	Mthfr	17769	8.36×10^{-6}	0.02412	0.42	0.92	1.34	2.91
10510552	Rere	68703	3.45×10^{-5}	0.09047	1.39	1.23	0.74	0.66
10522744	Mthfr	17769	4.47×10^{-5}	0.10750	0.67	1	1.07	1.59
10428004	Ankrd33b	67434	6.08×10^{-5}	0.13493	0.78	1.02	1.5	1.95
10510482	Clstn1	65945	7.61×10^{-5}	0.15678	1.59	1.34	1.02	0.86
10480459	Hnmt	140483	0.00016	0.27257	0.73	0.43	0.79	0.46
10518408	Plod1	18822	0.00022	0.31861	0.61	0.68	0.63	0.71
10510061	Pramef8	242736	0.00025	0.34772	0.73	0.88	0.83	1.01
10588283	Rab6b	270192	0.00027	0.34772	0.76	0.34	1.17	0.52

Table A.13: Genes with adjusted p-value < 0.35 of DJ-1 knockout microarray analysis. The first three columns describe the gene by probeset ID, Gene symbol and Entrez ID. Column 4 and 5 show p-value and adjusted p-value of the classes LPS knockout versus LPS wildtype. The last four columns specify the foldchange (linear ratio) between the different classes.

MAPT and SNCA - SNPs in TFBSs

Poly name	Study name	Chr position	TFBS family
rs3744456	Camuzat, 2008 PMID: 18785640	43972176	V$PRDM
rs3744457	Mizuta, 2006 PMID: 16500997	43972915	V$EGRF, V$EBOX, V$PAX9, V$PAX5, V$DMTF
rs1560310	Oliveira, 2004 PMID: 15459824	43978534	V$GFI1, V$CREB
rs1467966	Tobin, 2008 PMID: 18509094	43984399	V$YY1F, V$CEBP, V$HMTB, V$ETSF, V$IKRS, V$XBBF
rs1467967	Das, 2009 PMID: 19450659	43986179	V$RUSH, V$ABDB
rs17563965	Maraganore, 2005 PMID: 16252231	43990919	V$PAX6, V$KLFS, V$LEFF, V$BTBF
rs17563986	Saad, 2010 PMID: 21084426	43991272	-
rs3785880	Camuzat, 2008 PMID: 18785640	43993376	V$PARF, V$HNFP, V$ETSF, V$EGRF, V$KLFS, V$SP1F
rs35908989	Camuzat, 2008 PMID: 18785640	43994021	V$KLFS, V$MAZF, V$RXRF
rs17649641	Maraganore, 2005 PMID: 16252231	43997372	V$RUSH, V$MYBL, V$MOKF
rs767057	Oliveira, 2004 PMID: 15459824	43998822	-
rs4792894	Camuzat, 2008 PMID: 18785640	43999203	V$CREB, V$ZF35, V$GLIF
rs242556	Camuzat, 2008 PMID: 18785640	44002250	V$HEAT, V$NFAT, V$E2FF
rs17564829	Maraganore, 2005 PMID: 16252231	44006601	V$TALE, V$SRFF
rs17650417	Maraganore, 2005 PMID: 16252231	44013103	V$PCBE, V$YY1F, V$EVI1, V$TALE
rs11867549	Simon-Sanchez, 2009 PMID: 19915575	44013235	-
rs2435205	Fung, 2006 PMID: 17052657	44018764	V$KLFS

Poly name	**Study name**	**Chr position**	**TFBS family**
rs242557	Camuzat, 2008 PMID: 18785640	44019712	V$KLFS, V$CP2F, V$STAT
rs242559	Tobin, 2008 PMID: 18509094	44025888	V$NR2F
rs242562	Fidani, 2006 PMID: 16552760	44026739	V$DEAF, V$NFKB, V$ETSF, V$ZF01, V$NOLF
rs17571857	Maraganore, 2005 PMID: 16252231	44035706	V$NF1F, V$SRFF, V$PERO, V$RXRF, V$LEFF

Table A.14: All SNPs of MAPT lying in the first intron. The study, which revealed the polymorphism (column 1), with corresponding PubMedID as well as the position of the SNP in the human genome is shown in column 2 and 3, respectively. All predicted TF-families with binding site located at the SNP are listed in the last column.

Poly name	**Study name**	**Chr position**	**Gene region**	**TFBS family**
rs3756059	Kobayashi, 2006 PMID: 17078049	90757272	1. Intron	V$MYBL, V$GREF, V$OVOL
rs1372519	Fung, 2006 PMID: 17052657	90757309	1. Intron	V$HAND, V$ZICF, V$CTCF, V$FKHD
rs2870027	Kobayashi, 2006 PMID: 17078049	90757312	1. Intron	V$HAND, V$ZICF, V$CTCF
rs3756063	Kobayashi, 2006 PMID: 17078049	90757394	1. Intron	V$RXRF
rs1372520	Kobayashi, 2006 PMID: 17078049	90757505	1. Intron	V$BRN5, V$IRFF, V$ETSF, V$RXRF, V$AP1R
rs2245801	Kobayashi, 2006 PMID: 17078049	90757840	1. Intron	V$DEAF, V$DMTF, V$ETSF, V$PAX3, V$HEAT, V$YY1F
rs2301136	Kobayashi, 2006 PMID: 17078049	90757843	1. Intron	V$DEAF, V$DMTF, V$ETSF, V$PAX3
rs2301135	Farrer, 2001 PMID: 11532993	90758389	Promoter	V$KLFS, V$SP1F, V$ZF02, V$AP1R
rs3216775	Kobayashi, 2006 PMID: 17078049	90758437: 90758438	Promoter	V$GATA

Table A.15: All SNPs of SNCA lying in the promoter and first intron. The study, which revealed the polymorphism (column 1), with corresponding PubMedID, the position of the SNP in the human genome as well as the SNCA gene region is shown in column 2, 3 and 4, respectively. All predicted TF-families with binding site located at the SNP are listed in the last column.

Predicted microRNA target genes

Gene symbol	Entrez geneID	miR target gene / AlzGene DB gene
ABCA1	19	miR-1306, AlzGeneDB
ABHD2	11057	miR-103, miR-107
AC007546.6-201	400073	miR-103, miR-107
ADAM10	102	miR-1306, AlzGeneDB
ADAM7	8756	miR-103, miR-107
AMOT	154796	miR-103, miR-107
APH1A	51107	miR-1306, AlzGeneDB
AQP11	282679	miR-103, miR-107
ARHGAP5	394	miR-103, miR-107
Ari2	10425	miR-103, miR-107
ARL6IP2	64225	miR-103, miR-107
ARNT	405	miR-103, miR-107
ARPC2	10109	miR-103, miR-107
BACH2	60468	miR-103, miR-107
BDNF	627	miR-103, miR-107, AlzGeneDB
BMX	660	miR-103, miR-107
BSDC1	55108	miR-103, miR-107
BTG2	7832	miR-103, miR-107
C20orf39	79953	miR-103, miR-107
C21orf55	54943	miR-1306, AlzGeneDB
C2orf42	54980	miR-103, miR-107
C5orf41	153222	miR-103, miR-107
CACNA2D1	781	miR-103, miR-107
CACNB2	783	miR-1306, AlzGeneDB
CAPZA2	830	miR-103, miR-107
CCNE1	898	miR-103, miR-107
CDK5R1	8851	miR-103, miR-107, AlzGeneDB
CDK6	1021	miR-103, miR-107
CELSR2	1952	miR-103, miR-107
CHD1	1105	miR-103, miR-107
CHEK1	1111	miR-103, miR-107
CHRNA3	1136	miR-1306, AlzGeneDB
CHRNA4	1137	miR-1306, AlzGeneDB
CLCN5	1184	miR-103, miR-107
COBLL1	22837	miR-103, miR-107
COX15	1355	miR-1306, AlzGeneDB

Gene symbol	Entrez geneID	miR target gene / AlzGene DB gene
CPEB3	22849	miR-103, miR-107
CPNE6	9362	miR-103, miR-107
CXCL12	6387	miR-1306, AlzGeneDB
CXorf23	256643	miR-103, miR-107
CYP19A1	1588	miR-1306, AlzGeneDB
DCBLD2	131566	miR-103, miR-107
DCUN1D4	23142	miR-103, miR-107
DICER1	23405	miR-103, miR-107
DKFZp564E0482	79012	miR-103, miR-107
DLL1	28514	miR-103, miR-107
DSC1	1823	miR-1306, AlzGeneDB
DVL1	1855	miR-103, miR-107, AlzGeneDB
EIF4B	1975	miR-103, miR-107
EIF5	1983	miR-103, miR-107
ENSA	2029	miR-103, miR-107
ESR1	2099	miR-103, miR-107, AlzGeneDB
ESRRA	2101	miR-103, miR-107
EXOC5	10640	miR-103, miR-107
FAM81A	145773	miR-103, miR-107
FBXW7	55294	miR-103, miR-107
FOXJ2	55810	miR-103, miR-107
FRYL	285527	miR-103, miR-107
FSTL4	23105	miR-103, miR-107
GALNTL6	442117	miR-103, miR-107
GNA11	2767	miR-1306, AlzGeneDB
GPATCH8	23131	miR-103, miR-107
HPCAL1	3241	miR-1306, AlzGeneDB
IGF1	3479	miR-1306, AlzGeneDB
IGF1R	3480	miR-1306, AlzGeneDB
IHH	3549	miR-103, miR-107
JAKMIP2	9832	miR-103, miR-107
JUB	84962	miR-103, miR-107
KIAA0774	23281	miR-103, miR-107
KIAA1005	23322	miR-103, miR-107
KIAA1033	23325	miR-103, miR-107
KIAA1726	85463	miR-103, miR-107
KIAA2018	205717	miR-103, miR-107
KIF21A	55605	miR-103, miR-107
KIF23	9493	miR-103, miR-107
KLHL6	89857	miR-103, miR-107
LDLR	3949	miR-1306, AlzGeneDB
LRP1	4035	miR-103, miR-107, AlzGeneDB

Gene symbol	Entrez geneID	miR target gene / AlzGene DB gene
LRP2	4036	miR-103, miR-107, AlzGeneDB
LRP8	7804	miR-1306, AlzGeneDB
LRRN3	54674	miR-103, miR-107
MECP2	4204	miR-103, miR-107
MED26	9441	miR-103, miR-107
MFN2	9927	miR-103, miR-107
MTMR4	9110	miR-103, miR-107
MYB	4602	miR-103, miR-107
MYH9	4627	miR-103, miR-107
N4BP1	9683	miR-103, miR-107
NAV2	89797	miR-103, miR-107
NEDD9	4739	miR-107, AlzGeneDB
NF1	4763	miR-103, miR-107
NOTCH2	4853	miR-103, miR-107
NOVA1	4857	miR-103, miR-107
NR4A3	8013	miR-103, miR-107
NTRK2	4915	miR-103, miR-107, miR-1306, AlzGeneDB
OTUD4	54726	miR-103, miR-107
PCGF5	84333	miR-103, miR-107, AlzGeneDB
PDE3B	5140	miR-103, miR-107, AlzGeneDB
PELI2	57161	miR-103, miR-107
PER3	8863	miR-103, miR-107
PIK3R1	5295	miR-103, miR-107, AlzGeneDB
PLAG1	5324	miR-103, miR-107
PLCB1	23236	miR-103, miR-107
PLS3	5358	miR-103, miR-107
PPARD	5467	miR-1306, AlzGeneDB
PPARGC1A	10891	miR-1306, AlzGeneDB
PRKAB2	5565	miR-107, AlzGeneDB
PRKG1	5592	miR-103, miR-107
PTH	5741	miR-103, miR-107
RAB11FIP2	22841	miR-103, miR-107
RAI14	26064	miR-103, miR-107
RANGRF	29098	miR-103, miR-107
RASSF5	83593	miR-103, miR-107
RBM24	221662	miR-103, miR-107
RGS4	5999	miR-103, miR-107, AlzGeneDB
RNF38	152006	miR-103, miR-107
RUNX1T1	862	miR-103, miR-107
RXRA	6256	miR-1306, AlzGeneDB
SAP130	79595	miR-103, miR-107
SCN1A	6323	miR-103, miR-107

Gene symbol	Entrez geneID	miR target gene / AlzGene DB gene
SCN2A2	6326	miR-103, miR-107
SEPT3	55964	miR-1306, AlzGeneDB
SEPT8	23176	miR-103, miR-107
SH2D2A	9047	miR-103, miR-107
SH3GL2	6456	miR-103, miR-107
SIPA1L2	57568	miR-103, miR-107
SLN	6588	miR-103, miR-107
SNRK	54861	miR-103, miR-107
SNX3	8724	miR-103, miR-107
SPINK5	11005	miR-103, miR-107
SPRED1	161742	miR-103, miR-107
ST8SIA3	51046	miR-103, miR-107
SYNJ1	8867	miR-103, miR-107
TANC2	26115	miR-1306, AlzGeneDB
TARBP2	6895	miR-103, miR-107
TBC1D19	55296	miR-103, miR-107
TDG	6996	miR-103, miR-107
TGFBR3	7049	miR-103, miR-107
TLK1	9874	miR-103, miR-107
TNFRSF1B	7133	miR-1306, AlzGeneDB
TSPAN8	7103	miR-103, miR-107
UBE4A	9354	miR-103, miR-107
UMOD	7369	miR-103, miR-107
UPF2	26019	miR-103, miR-107
VAMP1	6843	miR-103, miR-107
VCP	7415	miR-103, miR-107, AlzGeneDB
VDP	8615	miR-103, miR-107
WDR22	8816	miR-103, miR-107
WNT3A	89780	miR-103, miR-107
YTHDC1	91746	miR-103, miR-107
YWHAH	7533	miR-103, miR-107
ZBTB10	65986	miR-103, miR-107
ZC3H12B	340554	miR-103, miR-107
ZKSCAN1	7586	miR-103, miR-107
ZNF449	203523	miR-103, miR-107

Table A.16: The table shows a list of predicted target genes of miR-103, miR-107, miR-1306 common in 4 out of 6 DBs. The target genes are either listed in AlzGene DB and target gene of at least one miRNA or target gene of at least two miRNAs (column 3). Beside the gene symbol the Entrez GeneID is given.

Danksagung

Zuerst möchte ich mich bei Prof. Dr. Wolfgang Wurst bedanken, dass ich in seinem Institut meine Promotion vollziehen durfte. Durch seine konstruktive Kritik und anregenden Diskussionen gab er mir wertvolle Unterstützung für diese Arbeit.

Ganz herzlich Danke sagen möchte ich meinem Betreuer Dr. Dietrich Trümbach für seine vielfältige Unterstützung sowohl wissenschaftlich als auch administrativ. Er setzte sich immer für mich und meine Arbeit ein und durch sein fundiertes Wissen in zahlreichen Bereichen habe ich viel in meiner Zeit als Doktorandin gelernt.

Ich möchte mich bei Prof. Dr. Erwin Grill, Prof. Dr. Wolfgang Wurst und Prof. Dr. Dr. Fabian Theis für die Durchführung meiner Prüfung und Bewertung meiner Dissertation bedanken.

Außerdem danke ich meinem Thesis Committee Dr. Stefan Lichtenthaler und Prof. Dr. Dr. Fabian Theis für die großartigen wissenschaftlichen Diskussionen in unseren Treffen und ihren Ideenreichtum.

Ich möchte meinem Teamleiter der Bioinformatik Gruppe Dr. Jens Hansen danken. Er hat mit seinen Fragestellungen und seiner konstruktiven Kritik einen wesentlichen Beitrag zur Arbeit geleistet und mich dazu gebracht meinen Blickwinkel auf die Arbeit zu erweitern.

Bedanken möchte ich mich auch bei den zahlreichen Kooperationspartnern:

- Dr. Stefan Lichtenthaler für seinen Beitrag bei der Planung der Projekte Vorhersage von TFBS Modulen und microRNAs.
- Dr. Daniela Vogt-Weisenhorn für die DJ-1 knockout microarray Daten.
- Dr. Konstanze Winklhofer für die erfolgreiche Zusammenarbeit in der Validierung von TFBSs in Parkinson relevanten Genen.
- Dr. Regina Widner für den HR LR Maus Microarray Datensatz, die SAGE Daten und die Durchführung der qPCR Experimente.
- Dr. Michael Greeff für Pathway Analysen im TFBSs Module Projekt.

- Dr. Peer-Hendrik Kuhn für seinen Beitrag zur Validierung der microRNAs.
- Dr. Kristina Endres für das Entwerfen des experimentiellen Systems für die microRNA Validierung in ADAM10.
- Sven Reinhardt für die Durchführung der Transfektion und des Luciferase Reporter Assay Experiments.

Ein großer Dank geht auch an Dr. Gerhard Welzl und Theresia Faus-Keßler, die mich bei statistischen Fragestellungen immer unterstützten.
Danken möchte ich auch dem Sekretariat des IDG und Dr. Daniela Vogt-Weisenhorn für ihre Hilfsbereitschaft und stetige Unterstützung.

Sehr dankbar bin ich dem Bioinformatik Team des IDG, in dem ich bei einer einmaligen Arbeitsatmosphäre arbeiten durfte: Bernd, der mich bei Computerproblemen stets unterstützte und immer für viel Erheiterung sorgte sowie meinen beiden Zimmerkollegen Anh-Thu und Rainer, mit denen ich zwei neue Freunde gewonnen habe.

Ein besonders lieber Dank geht an meine Eltern, die mich bedingungslos in allen Lebenslagen unterstützen, und an Florian für seine liebevolle Unterstützung.

Lebenslauf

Regina Peis, geb. Augustin, Dipl. Bioinf. (Uni)
geboren am 16.09.1983 in Landshut

Berufserfahrung

12/2008 - 09/2012	Helmholtz Zentrum München, Institut für Entwicklungsgenetik, Abteilung Bioinformatik Erstellung einer Disseration mit dem Thema 'Prediction of new regulatory molecular networks in neurodegenerative diseases with the help of bioinformatical methods'

Studium

10/2003 - 09/2008	Ludwig-Maximilians-Universität und Technische Universität München, Studium der Bioinformatik mit Abschluss Diplom Schwerpunkte: Datenbanken, Data Mining, Biochemie und Genomik Diplomarbeit: 'Data Mining in HIV-Resistenzmutationen' Praktika: 'Programmierpraktikum', 'Genomorientiertes Programmierpraktikum'

Schulbildung

09/1994 - 06/2003	Gymnasium Seligenthal, Landshut Abschluss der Allgemeinen Hochschulreife

Publikationen

Müller-Rischart, Anne Kathrin and Pilsl, Anna and Beaudette, Patrick and Patra, Maria and Hadian, Kamyar and Funke, Maria and **Peis, Regina** and Deinlein, Alexandra and Schweimer, Carolin and Kuhn, Peer-Hendrik and Lichtenthaler, Stefan F. and Motori, Elisa and Hrelia, Silvana and Wurst, Wolfgang and Trümbach, Dietrich and Langer, Thomas and Krappmann, Daniel and Dittmar, Gunnar and Tatzelt, Jörg and Winklhofer, Konstanze F.: **The E3 Ligase Parkin Maintains Mitochondrial Integrity by Increasing Linear Ubiquitination of NEMO.** *Molecular Cell*; 2013 Mar 7;49(5):908-21

Augustin, Regina and Endres, Kristina and Reinhardt, Sven and Kuhn, Peer-Hendrik and Lichtenthaler, Stefan F. and Hansen, Jens and Wurst, Wolfgang and Trümbach, Dietrich: **Computational identification and experimental validation of microRNAs binding to the Alzheimer-related gene ADAM10.** *BMC Medical Genetics*; 2012 May 17;13(1):35

Augustin, Regina and Lichtenthaler, Stefan F. and Greeff, Michael and Hansen, Jens and Wurst, Wolfgang and Trümbach, Dietrich: **Bioinformatics identification of modules of transcription factor binding sites in Alzheimer's Disease related genes by in silico promoter analysis and microarrays.** *International Journal of Alzheimer's Disease*, 2011 Apr 26; 2011: 154325

Bouman, Lena and Schlierf, Anita and Lutz, A. Kathrin and Shan, Jixiu and Deinlein, Alexandra and Kast, Jessica and Galehdar, Zohreh and Palmisano, Vincenza and Patenge, Nadja and Berg, Daniela and Gasser, Thomas and **Augustin, Regina** and Trümbach, Dietrich and Irrcher, Isabella and Park, David S. and Wurst, Wolfgang and Kilberg, Michael S. and Tatzelt, Jörg and Winklhofer, Konstanze F.: **Parkin is transcriptionally regulated by ATF4: evidence for an interconnection between mitochondrial stress and ER stress.** *Cell Death and Differentiation.* 2011 May; 18(5): 769-82. Epub 2010 Nov 26.

Richter, Lothar and **Augustin, Regina** and Kramer, Stefan: **Finding Relational Associations in HIV Resistance Mutation Data.** In *Inductive Logic Programming. Volume* 5989. Edited by De Raedt L: Springer Berlin / Heidelberg; 2010: 202-208: *Lecture Notes in Computer Science.*