

Research Article

Bioinformatics Identification of Modules of Transcription Factor Binding Sites in Alzheimer's Disease-Related Genes by In Silico Promoter Analysis and Microarrays

Regina Augustin,¹ Stefan F. Lichtenthaler,² Michael Greeff,³ Jens Hansen,¹ Wolfgang Wurst,^{1,2,4} and Dietrich Trümbach^{1,4}

¹ Institute of Developmental Genetics, Helmholtz Centre Munich, German Research Centre for Environmental Health (GmbH), Technical University Munich, Ingolstädter Landstraße 1, Munich 85764, Neuherberg, Germany

² DZNE, German Center for Neurodegenerative Diseases, Schillerstraße 44, 80336 Munich, Germany

³ Institute of Bioinformatics and Systems Biology, Helmholtz Centre Munich, German Research Centre for Environmental Health (GmbH), Ingolstädter Landstraße 1, Munich 85764, Neuherberg, Germany

⁴ Clinical Cooperation Group Molecular Neurogenetics, Max Planck Institute of Psychiatry, Kraepelinstraße, 2-10, 80804 Munich, Germany

Correspondence should be addressed to Wolfgang Wurst, wurst@helmholtz-muenchen.de and Dietrich Trümbach, dietrich.truembach@helmholtz-muenchen.de

Received 21 December 2010; Accepted 15 February 2011

Academic Editor: Jeff Kuret

Copyright © 2011 Regina Augustin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The molecular mechanisms and genetic risk factors underlying Alzheimer's disease (AD) pathogenesis are only partly understood. To identify new factors, which may contribute to AD, different approaches are taken including proteomics, genetics, and functional genomics. Here, we used a bioinformatics approach and found that distinct AD-related genes share modules of transcription factor binding sites, suggesting a transcriptional coregulation. To detect additional coregulated genes, which may potentially contribute to AD, we established a new bioinformatics workflow with known multivariate methods like support vector machines, biclustering, and predicted transcription factor binding site modules by using in silico analysis and over 400 expression arrays from human and mouse. Two significant modules are composed of three transcription factor families: CTCF, SP1F, and EGRF/ZBPF, which are conserved between human and mouse APP promoter sequences. The specific combination of in silico promoter and multivariate analysis can identify regulation mechanisms of genes involved in multifactorial diseases.

1. Introduction

Alzheimer's disease (AD) is the most common form of dementia, which slowly destroys neurons and causes serious cognitive disability [1]. 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 [2]. AD exists in a sporadic and familial (heritable) form. Mutations in amyloid-beta precursor protein (APP), presenilin1, and presenilin2 are associated with early-onset forms of familial AD, whereas sporadic AD occurs in people over the age of 65 years [3].

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. These proteases control the generation of the amyloid- β peptide ($A\beta$), which is considered the culprit in AD. β - and γ -secretase cleavage leads to $A\beta$ formation. β -secretase is the aspartyl protease BACE1 [4, 5]. A homolog of BACE1, BACE2, cleaves within the $A\beta$ domain and does not contribute to $A\beta$ generation. γ -secretase is a heterotetramer consisting of the four subunits presenilin 1 or 2 (PS1, PS2), Aph1, Nicastrin, and Pen-2 [6]. Aggregates of $A\beta$ are neurotoxic and start the so-called amyloid cascade, which describes the molecular mechanisms

leading to AD, including formation of plaques and tangles [1]. The third protease, the alpha-secretase ADAM10 [7, 8], avoids formation of A β , because it cleaves APP inside the A β sequence [9]. Additionally, α -secretase generates the soluble sAPP α , which enhances memory in normal and amnesic mice [10]. In AD patients, the amount of sAPP α in the cerebrospinal fluid is reduced [11].

Microarray studies measure the changes in expression level for thousands of genes simultaneously, detect single nucleotide polymorphisms, and therefore are an unbiased approach to identify genes with an altered expression, for example, in diseases such as AD [12, 13]. Aims of the analysis of microarray datasets include the discovery of gene function [14], getting insights into human disease progression [15], and prediction of gene regulatory elements like transcription factor binding sites (TFBSs) of coregulated genes [16].

TFs bind predominantly to the upstream region of genes, and each TF has its own specific binding motif. 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 [17]. 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 [18]. 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 AD.

Our aim was to identify modules of transcription factor binding sites in the promoters of AD-linked genes, which may contribute to AD pathogenesis. In our study, we developed a workflow and included 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) [19]. We hypothesized that beta- and gamma-secretase, which are responsible for A β formation, are coregulated. Thus, we started by analyzing TF binding modules in the genes for β - and γ -secretase, that is, BACE1, PS1, and PS2. We also included the BACE1 homolog BACE2.

2. Materials and Methods

2.1. 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 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).

2.2. Microarray Datasets. We used three microarray datasets downloaded from the Gene Expression Omnibus in this study. The dataset of Blalock et al. [12] 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 monotransgenic APP control) mice (GSE10908) from Prinzen et al. [20]. Gene expression 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. [21] 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.

2.3. Multivariate Analysis. Statistical analysis was performed with *R* statistical software (*R* version 2.8.0, <http://www.r-project.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. We applied multiclass support vector machines with recursive feature elimination (mSVM-RFE). A SVM considers a set of objects (i.e., biological replicates) as classes, so that around the class boundaries the broadest possible range remains, which is free of data points. We 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 [22] and Guyon et al. [23]. We 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 into stratified four folds, which are

four equally sized folds, 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 [24]. 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, we did three groupings into four folds with four times mSVM-RFE per grouping, because of four combinations of three folds per grouping). We 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 dataset GSE15222 was not filtered by mSVM-RFE but by the illumina detection score [25]. Transcripts that have a detection score ≥ 0.99 in less than 90% of cases or 90% of controls are excluded, and 8650 probesets remain [21]. 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 correction [26] 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, we 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. [27] 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, we 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* (colour of lines in plot) = *c*(1), *type* (type of plot) = "l", and *lty* (line type) = "solid". The coloured 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 coloured lines is enlarged by the parameter *lwd* = 4.

2.4. 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). 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 [26] was applied by the *R* function *p.adjust* with method = "fdr". For the pathway analysis, we report the *P* value and FDR value. Additionally, we 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 [26] 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.

2.5. 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. 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/>) [28] 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/>) [29] was searched for expression tissue of genes, and gene ontology was used for the functional annotation and classification of the target genes [30].

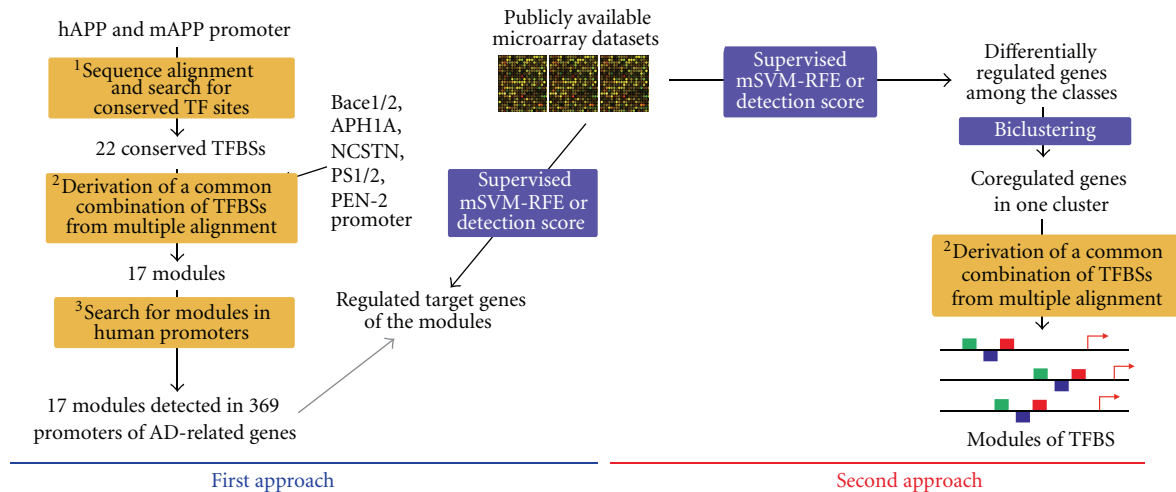


FIGURE 1: Workflow of bioinformatics analysis of promoter sequences and gene expression data to identify modules of TFBSs in AD-related genes. The workflow is divided into two parts, the first and second approach. The coloured boxes describe the methods which were used. The yellow boxes represent tools of the Genomatrix Software (¹DiAlignTF, ²FrameWorker, ³ModelInspector), 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 grey arrow denotes the comparison of the target genes of the modules with genes differentially regulated in microarray analyses. The scheme at the end of the second approach 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.

3. Results and Discussion

3.1. Workflow. The workflow (Figure 1) consists of two different approaches starting at different points, but resulting in similar modules in the end.

The first approach starts with an alignment between human and mouse APP promoter sequences to search for conserved TFBSs. The result yielded 22 conserved TFBSs (data not shown). Some of them have previously been verified. For example, the transcription factors CTCF and AP1F (including JUN) bind to APP [31, 32], and some are expressed in the hippocampus according to MGI database. In the next step of the approach, the conserved TFBSs and the promoter sequences of genes involved in A β 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 a subset of AD key genes. With this analysis, we got 17 modules composed of two or more TFBSs families (Table 1). Modules regulating key genes of AD are thought to play a role in AD. The three microarray datasets should verify these modules by searching for significantly regulated genes on the microarray. We applied mSVM-RFE, a frequently used tool, because it is very accurate and fast in classification and has a low error rate [22]. 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 2). After searching for these 17 modules in all human promoters, we limited the result to genes, which have been tested for genetic association with AD according to the AlzGene database (<http://www.alzgene.org/>; Version: 12.05.2010) [33] 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 (genome-wide association studies) performed on AD phenotypes. It is an important resource for AD and contains all considerable genetic association studies and key genes of AD. We detected these 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. We obtained for the AD patients dataset, which is composed of data from AD patients at different stages of severity and control, in the end 948 genes after mSVM-RFE and an overlap with the putative AD-risk genes of 31 genes. Applying a chi-squared test, we got a *P* value of 0.0001182, which shows that 31 is a significant high number of overlapping genes between the two gene sets compared to the number of genes in the whole dataset. Additionally, we 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 expressed in the brain, and the amount of plaques in the mice is controlled by the active as well as the dominant-negative 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 A β peptide levels are increased, and neuroprotective sAPP α 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

TABLE 1: Modules identified by the first approach: 17 modules composed of two or more TFBSs families. The TFBSs families consist of several TFs (Table 2). The second column specifies the key genes of AD the TFs of the module 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.

Module	AD key genes—targets of 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 2: Composition of the TF families of the modules. Each TF family consists of several transcription factors (TFs). Additional information about description and binding domains of the families is given in the second and fourth column, respectively.

TF family	Description	TFs	Binding domains
CTCF	CTCF and BORIS gene family, transcriptional regulators with 11 highly conserved zinc finger domains	CTCF, CTCFL	C2H2 zinc finger domain
E2FF	E2F-myc activator/cell cycle regulator	E2F1, E2F2, E2F3, E2F4, E2F5, E2F6, E2F7, E2F8, TFDPI, TFDPI2, TFDPI3	E2F winged helix
EGRF	EGR/nerve growth factor-induced protein C and related factors	EGR1, EGR2, EGR3, EGR4, WT1, ZBTB7A, ZBTB7B	C2H2 zinc finger domain
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	bHLH
KLFS	Krueppel-like transcription factors	KLF1, KLF2, KLF3, KLF4, KLF6, KLF7, KLF8, KLF9, KLF12, KLF13, KLF15	—
NRF1	Nuclear respiratory factor 1	NRF1	bZIP
SP1F	GC-Box factors SP1/GC	KLF10, KLF11, KLF16, KLF5, SP1, SP2, SP3, SP4, SP5, SP6, SP7, SP8	C2H2 zinc finger domain
ZBPF	Zinc binding protein factors	ZKSCAN3, ZNF148, ZNF202, ZNF219, ZNF281, ZNF300	C2H2 zinc finger domain

like AD patients at different stages of severity [34]. Thus, this AD mouse model and its microarray dataset are appropriate to be included in this analysis to verify modules [20]. The double-transgenic mice dataset was reduced by mSVM-RFE to 878 genes with an overlap of 26 to the 369 AD-related genes. A chi-squared test with P value = 4.21×10^{-12} shows that 26 is a significant high number of overlapping genes compared to the whole number of genes on the array. The

third dataset 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, we got an overlap of 199 genes between these two genesets. The P value = 8.99×10^{-15} of the chi-squared test indicates a significant number of overlapping genes.

Starting point of the second approach is the three microarray datasets established from AD patients at three different stages of severity [12], double-transgenic ADAM10/APP, dominant negative ADAM10/APP as well as APP control mice [20], and AD patients with late-onset AD (LOAD) [21]. 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, we 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, we 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 from 45101 to 5198 probesets, and after biclustering, we obtained 13 clusters of coregulated genes. The third dataset of LOAD patients is already reduced by illumina detection score to 8650 probesets, and therefore, we did not apply mSVM-RFE. By biclustering, we 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, we 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, we obtained several modules for each cluster of coregulated genes. The target genes of three selected modules, which are described in more detail in section 3.3, 3.4, and 3.5, are listed in the Supplementary Tables 1–8 (see Tables 1–8 in Supplementary Material available online at doi: 10.4061/2011/154325). 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, we got four different sets of TFBSs modules. We 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. We compared these four sets with regard to similar modules and found two modules in common: CTCF-EGRF-SP1F as well as CTCF-SP1F-ZBPF (Figure 2). 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 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.

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, and therefore these modules could not be found by the first approach focussing on conserved TFBSs. Moreover, the

starting points of the two approaches are different, because in the first approach we include conservation and in the second approach not. This leads to slightly different modules between the two approaches, but in the end, we have two modules in common.

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. The arrangement of TFBSs of a promoter module seems to be much more restricted than the variety and distribution of TFBSs in the promoter sequence [35]. However, TFBSs modules found in the promoter of genes suggest functional connection but do not prove it [36]. 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 nonconserved between the promoters [37, 38].

This workflow with the combination of multivariate analysis methods and in silico promoter analysis based on the hypothesis that the key genes of AD are coregulated is a new approach and different to already existing ones to find possible transcriptional regulations addressing AD pathogenesis. The confirmation of this hypothesis would offer better possibilities to develop new strategies for the therapeutic treatment of AD. 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.

3.2. 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. 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 and supplementary Figure 1). Additionally, a third significant module was detected established from a set of coregulated genes of the AD (Figure 3(b)) and LOAD

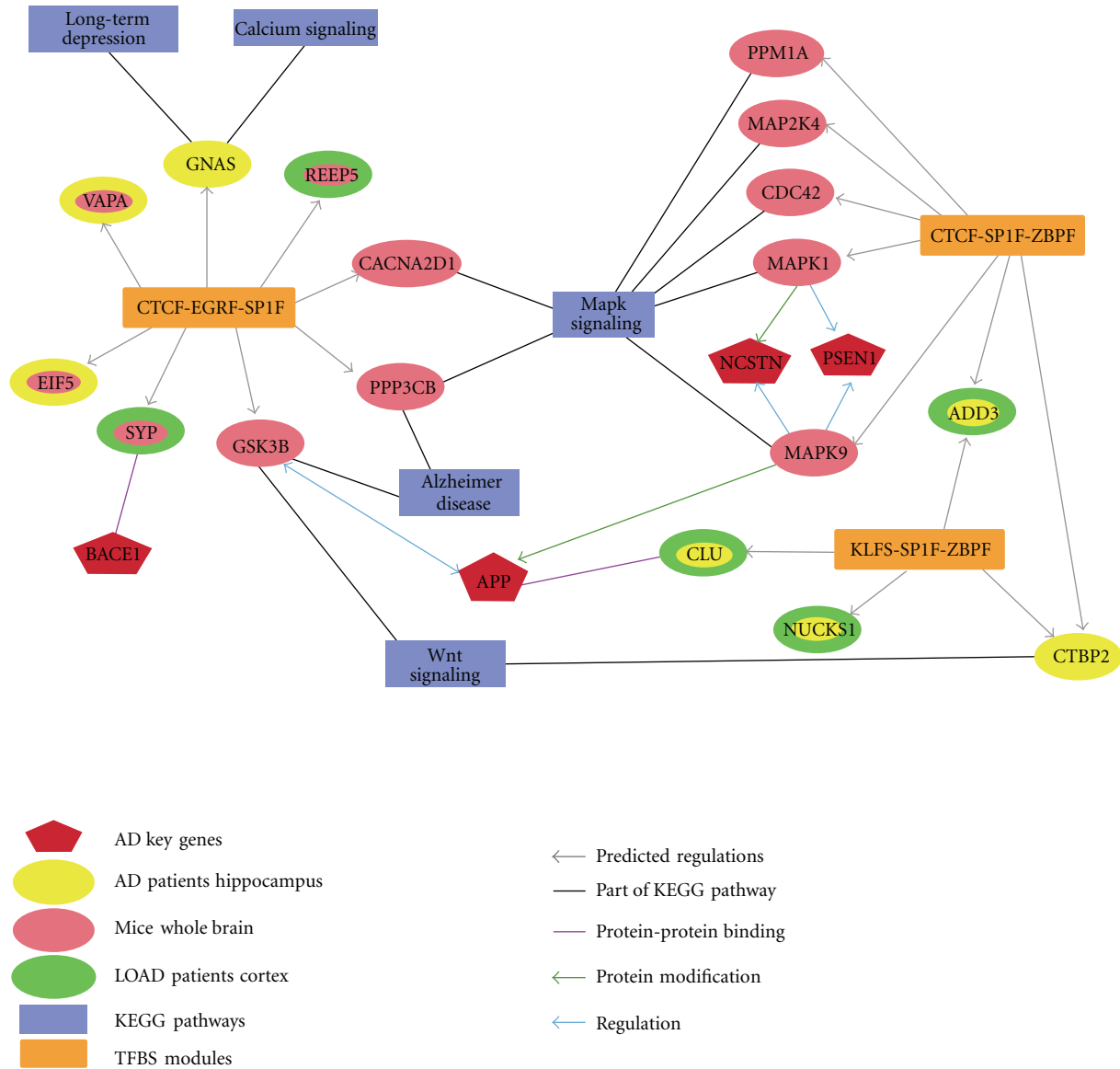


FIGURE 2: Relations of predicted target genes of three TFBSs modules. This picture summarizes important target genes of the modules, the relation of target genes to KEGG pathways playing a role in AD (blue rectangle), and the relation of the target genes to some AD key genes (red pentagon). The target genes are coloured according to their membership to microarray studies, and some target genes with two colours are derived from analysis of two different microarray studies. The grey 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.

patient's dataset (Figure 3(c)). 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 [39]. 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.

Already known regulations of AD key genes by the transcription factors CTCF, EGR1, SP1, and ZNF202 confirm these modules. EGR1 is known to upregulate Presenilin2 [40], CTCF binds to APB β domain a nuclear factor binding site in proximal APP promoter as well as acts as a transcriptional activator in the APP gene promoter [31], and APOE is repressed by ZNF202 (ZBPF) according to TRANSPATH database. SP1 can regulate both BACE1 and BACE2 genes [41], activates the transcription of PS1 [42], and upregulates APP gene expression [43]. In addition, we

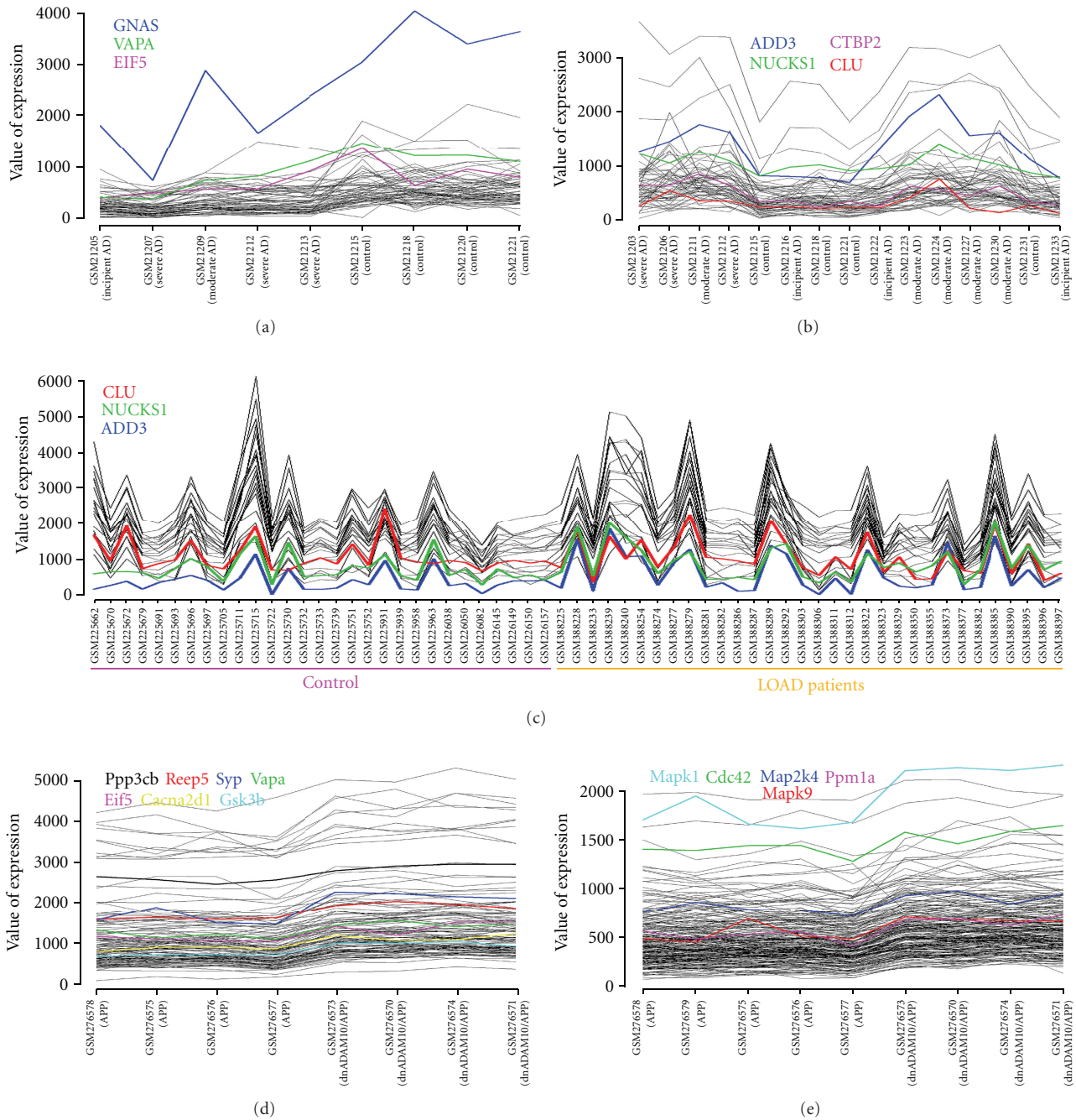


FIGURE 3: Expression profiles of five different clusters of coregulated genes. On the x-axis, the sample IDs (specified by accession numbers of GEO/NCBI) incorporated in the cluster are given, and y-axis indicates values of expression. One gene corresponds to a single line in the profile, and the target genes of the modules as mentioned in the text are coloured. The two upper profiles (a, b) are clusters of coregulated genes of the AD patients dataset, the profile in the middle (c) corresponds to coregulated genes of the LOAD patients dataset, and the two profiles below (d, e) correspond to coregulated genes of the double-transgenic mice dataset. The target genes of the profiles (a) and (d) were used for the establishment of the module CTCF-EGRF-SP1F and the profiles (b) and (e) for the module CTCF-SP1F-ZBPF, at which (b) was also used for the module KLFS-SP1F-ZBPF. The five lines in profile (e) correspond to genes, which are involved in the MAPK signaling pathway. The target genes of the profile (c) were used for the establishment of the modules CTCF-SP1F-ZBPF and KLFS-SP1F-ZBPF.

found SP1 significantly regulated on the LOAD patients microarrays ($FDR = 3.4 \times 10^{-9}$), which is in agreement with the results of a dysregulation of this transcription factor in AD [44, 45]. 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 (KLFs) 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, we incorporated information of the AlzGene database. Some target genes of the modules are already mentioned in AlzGene database to be associated to AD like Gsk3b, GOT1 (CTCF-EGRF-SP1F), Col25a1, Il33, and Tanc2 (CTCF-SP1F-ZBPF), and CLU (KLFs-SP1F-ZBPF).

3.3. First Module CTCF-EGRF-SP1F. Literature mining revealed known relations between the target genes of the first module and AD (supplementary Tables 1, 2, and 3) confirming this module. Long-term depression and calcium signaling are pathways which are connected to AD by the upregulation of calcium signaling leading to amyloid metabolism with neuronal cell apoptosis and an enhancement of long-term depression, which leads to loss of memory initiated by long-term potentiation [46]. GNAS is incorporated in both pathways, which occur significantly often in the cluster of the coreregulated 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)). The study from Zhang et al. shows that the gene expression of APP and GNAS is significantly upregulated in patients with endogenous depression [47]. Interestingly, clinically significant depression develops in at least 40% of all demented patients, and depressive symptoms like significant loss of appetite, insomnia, and fatigue occur commonly in the course of Alzheimer's disease [48].

Another pathway related to AD is the mitogen-activated protein kinase (MAPK) signaling pathway, which is involved in the production of proinflammatory cytokines in the hippocampus induced by A β and is a potential target for future therapeutics in AD [49]. Cacna2d1, a target gene of the double-transgenic mice dataset, is involved in MAPK signaling in the mouse, and this pathway is significantly overrepresented among the coreregulated genes of the double-transgenic mice dataset (P value = 0.0024, FDR = 0.0121). 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)) and regulates the APP accumulation after A β formation [50]. Another target gene Ppp3cb is also involved in AD and MAPK signaling in the mouse according to KEGG pathways, which are significantly enriched 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, which provide more evidence for the functionality of the module. Additional literature search and text mining reveal relations to AD. The target gene EIF5 (eukaryotic translation initiation factor 5) is not mentioned in AlzGene database, but in the EIF2 regulation pathway, it is downstream from its family member

EIF2AK2, which is listed in AlzGene database. 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 [51]. 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#).

The second overlapping target gene VAPA, a vesicle-associated membrane protein, interacts with its family member VAPB through the transmembrane domain [52], and both are reduced in human amyotrophic lateral sclerosis (ALS) patients, another neurodegenerative disease. Additionally, both genes interact with lipid binding proteins; especially VAPA is involved in lipid export and neurite outgrowth. 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 [53]. Approximately 30% of ALS patients with dementia have AD [54].

Furthermore, two overlapping target genes from the double transgenic mice and LOAD patients dataset were found: SYP, REEP5. SYP, a synaptic vesicle marker, 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 [55]. Other diseases associated with SYP are schizophrenia, ALS, and dementia, and it is expressed in hippocampus and cortex according to TRANSPATH database. The promoter region of SYP contains four SP1 binding sites located within 100 bp from the transcription start point [56]. 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 A β , and the loss of the associated presynaptic protein SYP is a key pathological feature of AD [57].

The second overlapping gene REEP5 (receptor accessory protein 5) is expressed in the brain and central nervous system according to MGI database and induces apoptosis according to gene ontology. 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 [58]. 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. IP₃ (inositol trisphosphate) receptor was a candidate to be involved in ER network formation as well as rapid Ca²⁺ efflux correlating with ER network formation [59]. Moreover, there exist indications that ER stress is involved

in AD pathogenesis. 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 [60].

Concluding, literature search revealed several target genes, which show a relation to AD, like SYP and Gsk3b strengthening the assumption that this module may be part of neurodegenerative processes as observed in AD.

3.4. Second Module CTCF-SP1F-ZBPF. The AD and LOAD patients target genes of the second module (supplementary Tables 4, 5) partially overlap with the AD and LOAD patients target genes of the third module KLFS-SP1F-ZBPF (supplementary Tables 6, 7), respectively. 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 (supplementary Table 8). For example, the C-terminal binding protein CTBP2 (target gene of AD patients dataset) binds ZNF219, a member of the ZBPF TF family, in vitro [61].

Verifications of the second module by double transgenic mice target genes revealed Cdc42 to regulate synapse formation in neurons and an increased Cdc42-GTPase activity in neurons stimulated with $\text{A}\beta_{1-42}$. Additionally, Cdc42 is upregulated in neuronal populations of AD brains in comparison to controls [62] and involved in the mouse MAPK signaling pathway (enrichment analysis: P -value = 0.0010, FDR = 0.0242). The MAPK pathway, which is significantly enriched in the coregulated genes of the double transgenic mice, incorporates the target genes Map2k4, Mapk1, Mapk9, and Ppm1a of the second module (Figure 3(e)). Further, PS1, a AD key gene, inhibits the Map2k4 activity [63]. The phosphorylation of Mapk1, an extracellular signal-regulated kinase (Erk), is activated by $\text{A}\beta$, and this Erk signal is involved in repression of L-glutamate uptake in astrocytes possibly defending neurodegeneration in the pathogenesis of AD [64]. Additionally, the activation of Erk cascades through EGF yields an increased CTCF expression showing that CTCF is a downstream target of Erk cascades [65]. 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 [66]. Some studies indicate that Mapk9 possibly interacts with SP1 and the JNK pathway targets SP1 [67].

The gene ADD3 (adducin 3 (gamma)) overlaps between the target genes of the AD and LOAD patients dataset. According to TRANSPATH database, it is known to be upregulated in ALS and to play a role in apoptosis, which leads to neuronal death in AD. ADD3 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 [68, 69]. Memory loss is the cardinal and one of the earliest clinical manifestations of AD, and studies suggest that aging

promotes the formation of soluble $\text{A}\beta$ assemblies mediating negative effects on memory [70].

In summary, five target genes (Cdc42, Map2k4, Mapk1, Mapk9, and Ppm1a) of the module are involved in MAPK signaling confirming the relation of the module to AD, and a possible new gene involved in AD, ADD3, was identified.

3.5. Additional Information of the Third Module KLFS-SP1F-ZBPF. One TF of the third module, KLF3 (BKLF), a member of the KLFS TF family, binds the AD patients target gene CTBP2 in vivo [71], which is involved in the Wnt signaling pathway. In our study, the human Wnt pathway genes are significantly enriched in the cluster of coregulated genes (P -value = 0.0149, FDR = 0.0464), and Wnt pathway has been found to prevent neurodegenerative diseases like AD by inhibiting $\text{A}\beta$ -dependent cytotoxic effects [72].

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, and a relation to AD by learning and memory is already known in literature. Verifications by literature exist also for the target gene CLU, which is induced by KLF4 overexpression [73] and contains binding sites for SP1 [74]. 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. Biologically, CLU seems to be involved in the pathogenesis of AD by interacting with different molecules like lipids or amyloid proteins, but also in brains of AD patients the level of CLU mRNA is significantly higher than in control brains [39]. Furthermore, APP/PS1 transgenic mice showed increased plasma CLU, age-dependent increase in brain CLU, and amyloid and CLU colocalization in plaques [75]. Another disease associated with CLU is ALS and according to gene ontology, CLU, which has increased protein levels in frontal cortex and hippocampus in AD [76], is involved in lipid transporter activity, apoptosis, and neuron development.

The third overlapping target gene NUCKS1 (nuclear casein kinase and cyclin-dependent kinase substrate 1) is known to be a strongly associated gene to Parkinson's disease (PD) [77], another neurodegenerative disease. A study in 2003 by Wilson et al. 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 [78]. Furthermore, NUCKS1 may play a role in cell proliferation [79]. 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 $\text{A}\beta$ can affect the proliferation of neural progenitor cells [80].

Taken together, all relations described here between the target genes of the modules to AD key genes and

KEGG pathways are shown in Figure 2. The three modules composed of TFs, which are mainly expressed in brain and partially associated with AD, 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, we 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.

4. Conclusion

As described above, many of the genes that we identified in our bioinformatics analysis have links to the molecular mechanisms of AD. In particular, the binding sites of the TF families: CTCF, EGRE, 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 regulations of target genes by transcription factors 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. The most likely candidate genes are *ADD3*, *CLU*, *EIF5*, *NUCKS1*, *REEP5*, *SYP*, and *VAPA*, which are derived from analysis of two different microarray studies. A next step is the experimental validation of TFBSs target interactions to verify the modules. At first, the differential regulation of candidate genes from AD mouse models can be validated by qRT-PCR. We also suggest a knockout (or knockdown) of single or combinations of TFs in mice and then the measurement of the expression compared to wildtype mice by qRT-PCR 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.

Acknowledgments

The Project underlying this report "Helmholtz Alliance for Mental Health in an Ageing Society" (HelMA, HA-215) was supported in the frame of the "Helmholtz Impuls- und Vernetzungsfonds" funded by the German "Bundesministerium für Bildung und Forschung" (BMBF), by the BMBF Project KNDD (FKZ 01G1704) to W. Wurst and S. F. Lichtenthaler, by the NGFN Project "Alzheimer" (FKZ 01GS08133) funded by the BMBF, and by the Deutsche Forschungsgemeinschaft (DFG) Project SFB596 to W. Wurst and S. F. Lichtenthaler. The authors are responsible for the content of this paper.

References

- [1] D. J. Selkoe and D. Schenk, "Alzheimer's disease: molecular understanding predicts amyloid-based therapeutics," *Annual Review of Pharmacology and Toxicology*, vol. 43, no. 1, pp. 545–584, 2003.
- [2] L. Crews, E. Rockenstein, and E. Masliah, "APP transgenic modeling of alzheimer's disease: mechanisms of neurodegeneration and aberrant neurogenesis," *Brain Structure and Function*, vol. 214, no. 2-3, pp. 111–126, 2010.
- [3] D. J. Selkoe, "Alzheimer's disease: genes, proteins, and therapy," *Physiological Reviews*, vol. 8, no. 2, pp. 741–766, 2001.
- [4] S. L. Cole and R. Vassar, "The role of amyloid precursor protein processing by BACE1, the β -secretase, in alzheimer disease pathophysiology," *Journal of Biological Chemistry*, vol. 283, no. 44, pp. 29621–29625, 2008.
- [5] S. Roßner, M. Sastre, K. Bourne, and S. F. Lichtenthaler, "Transcriptional and translational regulation of BACE1 expression—Implications for alzheimer's disease," *Progress in Neurobiology*, vol. 79, no. 2, pp. 95–111, 2006.
- [6] H. Steiner, E. Winkler, and C. Haass, "Chemical cross-linking provides a model of the γ -secretase complex subunit architecture and evidence for close proximity of the C-terminal fragment of presenilin with APH-1," *Journal of Biological Chemistry*, vol. 283, no. 50, pp. 34677–34686, 2008.
- [7] S. Lammich, E. Kojro, R. Postina et al., "Constitutive and regulated α -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*, vol. 96, no. 7, pp. 3922–3927, 1999.
- [8] P. H. Kuhn, H. Wang, B. Dislich et al., "ADAM10 is the physiologically relevant, constitutive alpha-secretase of the amyloid precursor protein in primary neurons," *The EMBO Journal*, vol. 29, no. 17, pp. 3020–3032, 2010.
- [9] F. Fahrenholz, "Alpha-secretase as a therapeutic target," *Current Alzheimer Research*, vol. 4, no. 4, pp. 412–417, 2007.
- [10] H. Mezziane, J.-C. Dodart, C. Mathis et al., "Memory-enhancing effects of secreted forms of the β -amyloid precursor protein in normal and amnesic mice," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 21, pp. 12683–12688, 1998.
- [11] L. Lannfelt, H. Basun, L. O. Wahlund, B. A. Rowe, and S. L. Wagner, "Decreased alpha-secretase-cleaved amyloid precursor protein as a diagnostic marker for alzheimer's disease," *Nature Medicine*, vol. 1, no. 8, pp. 829–832, 1995.
- [12] E. M. Blalock, J. W. Geddes, K. C. Chen, N. M. Porter, W. R. Markesbery, and P. W. 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*, vol. 101, no. 7, pp. 2173–2178, 2004.
- [13] J. A. Miller, S. Horvath, and D. H. Geschwind, "Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 28, pp. 12698–12703, 2010.
- [14] W. Pan, "Incorporating gene functions as priors in model-based clustering of microarray gene expression data," *Bioinformatics*, vol. 22, no. 7, pp. 795–801, 2006.
- [15] 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*, vol. 1, no. 1, p. 51, 2008.

- [16] P. J. Park, A. J. Butte, and I. S. Kohane, "Comparing expression profiles of genes with similar promoter regions," *Bioinformatics*, vol. 18, no. 12, pp. 1576–1584, 2002.
- [17] S. Döhr, A. Klingenhoff, H. Maier, M. H. de Angelis, T. Werner, and R. Schneider, "Linking disease-associated genes to regulatory networks via promoter organization," *Nucleic Acids Research*, vol. 33, no. 3, pp. 864–872, 2005.
- [18] C. D. Cohen, A. Klingenhoff, A. Boucherot et al., "Comparative promoter analysis allows de novo identification of specialized cell junction-associated proteins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 15, pp. 5682–5687, 2006.
- [19] D. Trümbach, C. Graf, B. Pütz et al., "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*, vol. 4, no. 1, p. 159, 2010.
- [20] 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*, vol. 10, no. 1, p. 66, 2009.
- [21] J. A. Webster, J. R. Gibbs, J. Clarke et al., "Genetic control of human brain transcript expression in alzheimer disease," *American Journal of Human Genetics*, vol. 84, no. 4, pp. 445–458, 2009.
- [22] X. Zhou and D. P. Tuck, "MSVM-RFE: extensions of SVM-RFE for multiclass gene selection on DNA microarray data," *Bioinformatics*, vol. 23, no. 9, pp. 1106–1114, 2007.
- [23] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1–3, pp. 389–422, 2002.
- [24] 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*, vol. 2, Morgan Kaufmann, Montreal, Canada, 1995.
- [25] Illumina, "Beadstudio gene expression module v3.4 user guide (11317265 Rev. A)," July 2008.
- [26] 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*, vol. 57, no. 1, pp. 289–300, 1995.
- [27] H. Turner, T. Bailey, and W. Krzanowski, "Improved biclustering of microarray data demonstrated through systematic performance tests," *Computational Statistics and Data Analysis*, vol. 48, no. 2, pp. 235–254, 2005.
- [28] T. Barnickel, J. Weston, R. Collobert, H. W. Mewes, and V. Stümpflen, "Large scale application of neural network based semantic role labeling for automated relation extraction from biomedical texts," *Plos One*, vol. 4, no. 7, Article ID e6393, 2009.
- [29] C. J. Bult, J. T. Eppig, J. A. Kadin, J. E. Richardson, and J. A. Blake, "The mouse genome database (MGD): mouse biology and model systems," *Nucleic Acids Research*, no. 36, pp. D724–D728, 2008.
- [30] M. Ashburner, C. A. Ball, J. A. Blake et al., "Gene ontology: tool for the unification of biology," *Nature Genetics*, vol. 25, no. 1, pp. 25–29, 2000.
- [31] A. A. Vostrov and W. W. Quitschke, "The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter," *Journal of Biological Chemistry*, vol. 272, no. 52, pp. 33353–33359, 1997.
- [32] W. J. Lukiw, E. I. Rogaev, L. Wong, G. Vaula, D. R. C. McLachlan, and P. St George Hyslop, "Protein-DNA interactions in the promoter region of the amyloid precursor protein (APP) gene in human neocortex," *Molecular Brain Research*, vol. 22, no. 1–4, pp. 121–131, 1994.
- [33] L. Bertram, M. B. McQueen, K. Mullin, D. Blacker, and R. E. Tanzi, "Systematic meta-analyses of alzheimer disease genetic association studies: the alzgene database," *Nature Genetics*, vol. 39, no. 1, pp. 17–23, 2007.
- [34] R. Postina, A. Schroeder, I. Dewachter et al., "A disintegrin-metalloproteinase prevents amyloid plaque formation and hippocampal defects in an alzheimer disease mouse model," *Journal of Clinical Investigation*, vol. 113, no. 10, pp. 1456–1464, 2004.
- [35] T. Werner, "Models for prediction and recognition of eukaryotic promoters," *Mammalian Genome*, vol. 10, no. 2, pp. 168–175, 1999.
- [36] 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*, vol. 12, no. 1, pp. 67–72, 2001.
- [37] 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*, vol. 2, no. 1, pp. S17–S26, 2002.
- [38] T. Werner, "Promoters can contribute to the elucidation of protein function," *Trends in Biotechnology*, vol. 21, no. 1, pp. 9–13, 2003.
- [39] R. J. Guerreiro, J. Beck, J. R. Gibbs et al., "Genetic variability in CLU and its association with alzheimer's disease," *PLoS ONE*, vol. 5, no. 3, p. e9510, 2010.
- [40] R. Wang, Y. W. Zhang, P. Sun et al., "Transcriptional regulation of PEN-2, a key component of the γ -secretase complex, by CREB," *Molecular and Cellular Biology*, vol. 26, no. 4, pp. 1347–1354, 2006.
- [41] X. Sun, Y. Wang, H. Qing et al., "Distinct transcriptional regulation and function of the human BACE2 and BACE1 genes," *FASEB Journal*, vol. 19, no. 7, pp. 739–749, 2005.
- [42] M. Pastorcic and H. K. Das, "An upstream element containing an ETS binding site is crucial for transcription of the human presenilin-1 gene," *European Journal of Biochemistry*, vol. 274, no. 34, pp. 24297–24307, 1999.
- [43] J. Wu, M. R. Basha, B. Brock et al., "Alzheimer's disease (AD)-like pathology in aged monkeys after infantile exposure to environmental metal lead (Pb): evidence for a developmental origin and environmental link for AD," *Journal of Neuroscience*, vol. 28, no. 1, pp. 3–9, 2008.
- [44] B. A. Citron, J. S. Dennis, R. S. Zeitlin, and V. Echeverria, "Transcription factor Sp1 dysregulation in alzheimer's disease," *Journal of Neuroscience Research*, vol. 86, no. 11, pp. 2499–2504, 2008.
- [45] G. Santpere, M. Nieto, B. Puig, and I. Ferrer, "Abnormal Sp1 transcription factor expression in Alzheimer disease and tauopathies," *Neuroscience Letters*, vol. 397, no. 1–2, pp. 30–34, 2006.
- [46] M. J. Berridge, "Calcium hypothesis of Alzheimer's disease," *Pflügers Archive European Journal of Physiology*, vol. 459, no. 3, pp. 441–449, 2010.
- [47] L. Zhang, J. Gough, D. Christmas et al., "Microbial infections in eight genomic subtypes of chronic fatigue syndrome/myalgic encephalomyelitis," *Journal of Clinical Pathology*, vol. 63, no. 2, pp. 156–164, 2010.
- [48] L. E. Tune, "Depression and Alzheimer's disease," *Depression and Anxiety*, vol. 8, pp. 91–95, 1998.

- [49] L. Munoz, H. Ranaivo, S. M. Roy et al., "A novel p38 α 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*, vol. 4, no. 1, p. 21, 2007.
- [50] A. Takashima, H. Yamaguchi, K. Noguchi et al., "Amyloid, β peptide induces cytoplasmic accumulation of amyloid protein precursor via tau protein kinase I/glycogen synthase kinase-3 β in rat hippocampal neurons," *Neuroscience Letters*, vol. 198, no. 2, pp. 83–86, 1995.
- [51] R. C. Chang, A. K. Wong, H. K. Ng, and J. Hugon, "Phosphorylation of eukaryotic initiation factor-2 α (eIF2 α) is associated with neuronal degeneration in alzheimer's disease," *Neuroreport*, vol. 13, no. 18, pp. 2429–2432, 2002.
- [52] I. Hamamoto, Y. Nishimura, T. Okamoto et al., "Human VAP-B is involved in hepatitis C virus replication through interaction with NS5A and NS5B," *Journal of Virology*, vol. 79, no. 21, pp. 13473–13482, 2005.
- [53] E. Teuling, S. Ahmed, E. Haasdijk et al., "Motor neuron disease-associated mutant vesicle-associated membrane protein-associated protein (VAP) B recruits wild-type VAPs into endoplasmic reticulum-derived tubular aggregates," *Journal of Neuroscience*, vol. 27, no. 36, pp. 9801–9815, 2007.
- [54] R. L. Hamilton and R. Bowser, "Alzheimer disease pathology in amyotrophic lateral sclerosis," *Acta Neuropathologica*, vol. 107, no. 6, pp. 515–522, 2004.
- [55] D. Beher, C. Elle, J. Underwood et al., "Proteolytic fragments of alzheimer's disease-associated presenilin 1 are present in synaptic organelles and growth cone membranes of rat brain," *Journal of Neurochemistry*, vol. 72, no. 4, pp. 1564–1573, 1999.
- [56] R. C. Bargou and R. E. Leube, "The synaptophysin-encoding gene in rat and man is specifically transcribed in neuroendocrine cells," *Gene*, vol. 99, no. 2, pp. 197–204, 1991.
- [57] L. McConlogue, M. Buttini, J. P. Anderson et al., "Partial reduction of BACE1 has dramatic effects on alzheimer plaque and synaptic pathology in APP transgenic mice," *Journal of Biological Chemistry*, vol. 282, no. 36, pp. 26326–26334, 2007.
- [58] S. Shimohama, "Apoptosis in alzheimer's disease-an update," *Apoptosis*, vol. 5, no. 1, pp. 9–16, 2000.
- [59] G. K. Voeltz, W. A. Prinz, Y. Shibata, J. M. Rist, and T. A. Rapoport, "A class of membrane proteins shaping the tubular endoplasmic reticulum," *Cell*, vol. 124, no. 3, pp. 573–586, 2006.
- [60] M. P. Mattson, "ER calcium and Alzheimer's disease: in a state of flux," *Science Signaling*, vol. 3, no. 114, p. pe10, 2010.
- [61] J. D. Hildebrand and P. Soriano, "Overlapping and unique roles for C-terminal binding protein 1 (CtBP1) and CtBP2 during mouse development," *Molecular and Cellular Biology*, vol. 22, no. 15, pp. 5296–5307, 2002.
- [62] A. Mendoza-Naranjo, C. Gonzalez-Billault, and R. B. MacCioni, "A β 1-42 stimulates actin polymerization in hippocampal neurons through Rac1 and Cdc42 Rho GTPases," *Journal of Cell Science*, vol. 120, no. 2, pp. 279–288, 2007.
- [63] J. W. Kim, T. S. Chang, J. E. Lee et al., "Negative regulation of the SAPK/JNK signaling pathway by presenilin 1," *Journal of Cell Biology*, vol. 153, no. 3, pp. 457–464, 2001.
- [64] K. Abe and M. Misawa, "The extracellular signal-regulated kinase cascade suppresses amyloid β protein-induced promotion of glutamate clearance in cultured rat cortical astrocytes," *Brain Research*, vol. 979, no. 1-2, pp. 179–187, 2003.
- [65] T. Li and L. Lu, "Epidermal growth factor-induced proliferation requires down-regulation of Pax6 in corneal epithelial cells," *Journal of Biological Chemistry*, vol. 280, no. 13, pp. 12988–12995, 2005.
- [66] 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*, vol. 279, no. 20, pp. 21628–21636, 2004.
- [67] J. Y. Chuang, Y. T. Wang, S. H. Yeh, Y. W. Liu, W. C. Chang, and J. J. Hung, "Phosphorylation by c-Jun NH2-terminal kinase 1 regulates the stability of transcription factor Sp1 during mitosis," *Molecular Biology of the Cell*, vol. 19, no. 3, pp. 1139–1151, 2008.
- [68] R. L. Rabenstein, N. A. Addy, B. J. Caldarone et al., "Impaired synaptic plasticity and learning in mice lacking β -adducin, an actin-regulating protein," *Journal of Neuroscience*, vol. 25, no. 8, pp. 2138–2145, 2005.
- [69] F. Porro, M. Rosato-Siri, E. Leone et al., " 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*, vol. 9, no. 1, pp. 84–96, 2010.
- [70] K. H. Ashe, "Learning and memory in transgenic mice modeling alzheimer's disease," *Learning and Memory*, vol. 8, no. 6, pp. 301–308, 2001.
- [71] J. Turner, H. Nicholas, D. Bishop, J. M. 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*, vol. 278, no. 15, pp. 12786–12795, 2003.
- [72] W. Cerpa, E. M. Toledo, L. Varela-Nallar, and N. C. Inestrosa, "The role of WNT signaling in neuroprotection," *Drug News and Perspectives*, vol. 22, no. 10, pp. 579–591, 2009.
- [73] D. T. Dang, X. Chen, J. Feng, M. Torbenson, L. H. Dang, and V. W. Yang, "Overexpression of Krüppel-like factor 4 in the human colon cancer cell line RKO leads to reduced tumorigenicity," *Oncogene*, vol. 22, no. 22, pp. 3424–3430, 2003.
- [74] S. Lee, E. Nakamura, H. Yang et al., "Neuronal apoptosis linked to EglN3 prolyl hydroxylase and familial pheochromocytoma genes: developmental culling and cancer," *Cancer Cell*, vol. 8, no. 2, pp. 155–167, 2005.
- [75] M. Thambisetty, A. Simmons, L. Velayudhan et al., "Association of plasma clusterin concentration with severity, pathology, and progression in alzheimer disease," *Archives of General Psychiatry*, vol. 67, no. 7, pp. 739–748, 2010.
- [76] A. M. 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*, vol. 154, no. 2, pp. 511–521, 1998.
- [77] W. Satake, Y. Nakabayashi, I. Mizuta et al., "Genome-wide association study identifies common variants at four loci as genetic risk factors for parkinson's disease," *Nature Genetics*, vol. 41, no. 12, pp. 1303–1307, 2009.
- [78] R. S. Wilson, J. A. Schneider, J. L. Bienias, D. A. Evans, and D. A. Bennett, "Parkinsonianlike signs and risk of incident alzheimer disease in older persons," *Archives of Neurology*, vol. 60, no. 4, pp. 539–544, 2003.
- [79] K. Grundt, I. V. Haga, H. S. Huitfeldt, and A. C. Østfold, "Identification and characterization of two putative nuclear localization signals (NLS) in the DNA-binding protein NUCKS," *Biochimica et Biophysica Acta*, vol. 1773, no. 9, pp. 1398–1406, 2007.
- [80] N. J. Haughey, A. Nath, S. L. Chan, A. C. Borchard, M. S. Rao, and M. P. Mattson, "Disruption of neurogenesis by amyloid β -peptide, and perturbed neural progenitor cell homeostasis, in models of alzheimer's disease," *Journal of Neurochemistry*, vol. 83, no. 6, pp. 1509–1524, 2002.

