PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Bayesian integrative analysis of epigenomic and transcriptomic data identifies Alzheimer's disease candidate genes and networks

Hans-Ulrich Klein[1,2☯]*, **Martin Schäfer**[3☯], **David A. Bennett**[4], **Holger Schwender**[3], **Philip L. De Jager**[1,2]

1 Center for Translational & Computational Neuroimmunology, Department of Neurology, Columbia University Irving Medical Center, New York, New York, United States of America, **2** Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Irving Medical Center, New York, New York, United States of America, **3** Mathematical Institute, Heinrich Heine University, Düsseldorf, Germany, **4** Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois, United States of America

☯ These authors contributed equally to this work.
* hk2948@cumc.columbia.edu

Citation: Klein H-U, Schäfer M, Bennett DA, Schwender H, De Jager PL (2020) Bayesian integrative analysis of epigenomic and transcriptomic data identifies Alzheimer's disease candidate genes and networks. PLoS Comput Biol 16(4): e1007771. https://doi.org/10.1371/journal.pcbi.1007771

Editor: Donna K. Slonim, Tufts University, UNITED STATES

Received: July 1, 2019

Accepted: March 3, 2020

Published: April 7, 2020

Data Availability Statement: All relevant data are available from the AMP-AD Knowledge Portal at Synapse (https://adknowledgeportal.synapse.org).

Funding: Research reported in this publication was supported by the National Institute on Aging under award numbers P30AG010161 (DAB), R01AG015819 (DAB), R01AG017917 (DAB), R01AG036042 (DAB), R01AG036836 (PLDJ), U01AG046152 (PLDJ, DAB), U01AG061356 (PLDJ, DAB). MS and HS were supported by the

Abstract

Biomedical research studies have generated large multi-omic datasets to study complex diseases like Alzheimer's disease (AD). An important aim of these studies is the identification of candidate genes that demonstrate congruent disease-related alterations across the different data types measured by the study. We developed a new method to detect such candidate genes in large multi-omic case-control studies that measure multiple data types in the same set of samples. The method is based on a gene-centric integrative coefficient quantifying to what degree consistent differences are observed in the different data types. For statistical inference, a Bayesian hierarchical model is used to study the distribution of the integrative coefficient. The model employs a conditional autoregressive prior to integrate a functional gene network and to share information between genes known to be functionally related. We applied the method to an AD dataset consisting of histone acetylation, DNA methylation, and RNA transcription data from human cortical tissue samples of 233 subjects, and we detected 816 genes with consistent differences between persons with AD and controls. The findings were validated in protein data and in RNA transcription data from two independent AD studies. Finally, we found three subnetworks of jointly dysregulated genes within the functional gene network which capture three distinct biological processes: *myeloid cell differentiation*, *protein phosphorylation* and *synaptic signaling*. Further investigation of the myeloid network indicated an upregulation of this network in early stages of AD prior to accumulation of hyperphosphorylated tau and suggested that increased *CSF1* transcription in astrocytes may contribute to microglial activation in AD. Thus, we developed a method that integrates multiple data types and external knowledge of gene function to detect candidate genes, applied the method to an AD dataset, and identified several disease-related genes and processes demonstrating the usefulness of the integrative approach.

Deutsche Forschungsgemeinschaft grant SCHW 1508/3-1. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

Author summary

Recent technological advances have led to a new generation of studies that interrogate multiple molecular levels in the same target tissue of a set of subjects, generating complex multi-omic datasets with which to study disease mechanism. These datasets of genetic, epigenomic, transcriptomic, and other data have the potential to reveal novel biological insights; however, integrative analyses remain challenging and require new computational methods. We developed an integrative Bayesian approach to detect genes with consistent differences between case and control samples across multiple data types. The method further integrates prior knowledge about gene function in the form of a gene functional similarity network to improve statistical inference by sharing information between related genes. We applied our method to an Alzheimer's disease dataset of epigenomic and transcriptomic data and detected and then validated several novel and known candidate genes as well as three major disease-related biological processes. One of these processes reflected microglial activation and included the cytokine *CSF1*. Single-nucleus data revealed that *CSF1* was primarily upregulated in astrocytes, implicating the involvement of this cell type in microglial activation. Hence, we demonstrated that integrative analysis approaches to multi-omic datasets can improve candidate gene detection and thereby generate new insights into complex diseases.

Introduction

Alzheimer's disease (AD) is a complex progressive neurodegenerative disease characterized clinically by impaired episodic memory and other impaired cognitive abilities [1]. To better understand disease mechanisms and to identify novel therapeutic targets, several studies of aging and AD have generated large molecular datasets from blood and/or post-mortem human brain samples. Some of these studies targeted multiple molecular levels and measured, for example, genetic variants, epigenetic modifications, mRNAs, or proteins, in the same set of samples [2–4]. However, jointly analyzing genome-wide multi-omic datasets remains challenging and requires novel computational methods to fully utilize these datasets [5].

Integration of multiple data types from the same set of samples has been referred to as vertical data integration [6]. Methods for vertical data integration can be further characterized by the primary goal of the integrative analysis as outlined in recent reviews [5–11]. While the main objective of the method presented in this work is the detection of genes with consistent differences between cases and controls in multiple data types, we review a wider range of vertical data integration methods in the introduction with a focus on those that were either successfully applied to data from AD or related complex diseases or share methodological similarity with our approach. Among the most frequently used are methods for integrating genetic and transcriptomic data. These have successfully identified genetic variants that affect gene transcription thereby improving our understanding of how the transcriptome mediates the effect of risk variants for various diseases including AD [12–14]. Most of these methods regress gene transcription on one or more genetic variants. A simple approach for integrating epigenomic and transcriptomic data is to replace the genetic with epigenetic measurements in the regression model [15–17]. Stepwise regression procedures were proposed to study interaction effects between different histone modifications that may not act independently on gene transcription [18]. Moreover, various machine learning approaches were applied to predict gene transcription levels based on transcription factor binding, histone modification or other epigenomic

data [19, 20]. If the primary goal is to predict the effect of an intervention experiment, the often unknown causal structure between different variables has to be learned [21]. Recently, Bayesian networks were applied to infer directed gene-wise graphs that model the relationships between epigenomic, transcriptomic, pathologic and clinical variables in the AD brain [22]. In the classic gene prioritization setting, the primary goal of an integrative analysis is to detect genes with consistent differences between case and control samples across different data types. The motivation for integrating data in this setting is twofold. First, gene prioritization can be improved assuming that a gene with differences in more than one data type is more likely to be a true positive finding. Second, if a specific epigenetic mechanism of a drug is known, target genes for this drug should ideally not only demonstrate differences in the epigenetic data, but also consistent differences at a functionally more relevant level such as gene transcription or protein expression [23]. Methods for detecting differential genes in a joint analysis of multiple data types were developed for experiments with only a few or no replicates [24, 25]. These methods provide probabilistic frameworks for studying the relationship between data types and for classifying genes, but since heterogeneity among replicates is not modelled, statistical inference is challenging and can only be carried out by relying heavily on prior information.

Further need for data integration stems from our constantly increasing knowledge about gene functions and pathways which can be represented as a network where functionally related genes are connected by edges [26]. This prior knowledge can be integrated into an analysis to share information between related genes and thus improve statistical inference and interpretability of the results. For example, gene networks have been used to improve parameter estimation and gene selection in penalized regression models [27–29]. In Bayesian models, conditionally autoregressive (CAR) Markov random field priors were frequently used to incorporate gene networks into genome-wide data analyses [25, 30–32].

Here, we propose a new integrative method to detect genes with consistent differences between case and control groups across multiple data types. The method is based on an integrative coefficient that summarizes information from multiple data types in a case-control study design. To improve statistical inference, a CAR prior is used in a hierarchical Bayesian model to share information between functionally similar genes defined by a gene network. We applied our method to a large AD case-control study consisting of histone ChIP-seq, DNA methylation, and RNA-seq profiles from 233 subjects. Identified genes were validated using protein data and two independent RNA-seq studies of AD. Finally, in a post hoc analysis, we identified differential networks reflecting AD-related processes. Our new method is outlined in the flow chart S1 Fig and described in detail in the Methods section. The validation of our method and findings from the analysis of the AD data are presented in the Results section. We note that our approach can be adapted and applied to other similar multi-omic case-control studies.

Methods

Coefficient for integration of multiple genomic variates

We propose an integrative coefficient Z that summarizes observations made in different genomic data types for the same subjects and genes in a case-control study. Let $X_{ij}^{(k)}$ denote the value observed for gene i in individual j of the patient group in data type k, and $Y_{ij}^{(k)}$ denotes the respective value in the matched control subject. We define Z_{ij} as the sum of standardized

differences

$$Z_{ij} = \sum_{k=1}^{K} S^{(k)} \frac{X_{ij}^{(k)} - Y_{ij}^{(k)}}{\sigma_{X^{(k)}Y^{(k)}}}, \quad i = 1, \ldots, n;\ j = 1, \ldots, m. \tag{1}$$

The variances of the differences $\sigma^2_{X^{(k)}Y^{(k)}} = \frac{1}{nm}\sum_{i=1,\ldots,n;\ j=1,\ldots,m}(X_{ij}^{(k)} - Y_{ij}^{(k)})^2$ are calculated across all genes and used to standardize the differences which may have a different range of values depending on the data type and technical platform used to generate the data. Whether a positive or negative association is expected between the genomic data types is modelled by the factor $S^{(k)} \in \{-1,1\}$.

If the differences $X_{ij}^{(k)} - Y_{ij}^{(k)}$ for a gene i and individual j show consistent directions as modelled by $S^{(k)}$ across all K data types, the absolute value of Z_{ij} is large. In contrast, a difference in one data type might be cancelled out by a difference in another data type if the directions of the differences do not meet the assumption defined by $S^{(k)}$. Z_{ij} is also expected to be close to zero if a gene does not have any differences in any data type. Previous work suggested multiplicative instead of additive coefficients for data integration [24, 25, 33], however, replacing the sum by a product in Eq (1) is a very conservative approach when multiple different data types are modelled. A small difference in a single data type would result in a small multiplicative coefficient even if distinct differences are observed in the other data types. In this work, we jointly analyzed gene transcription, histone modification and CG methylation both at promoters and at exons resulting in $K = 4$ different data types. To model the negative association between promoter methylation and transcription [34], we set $S^{(k)} = -1$ for promoter methylation and $S^{(k)} = 1$ for the other three data types.

Bayesian hierarchical model

Model. The Z_{ij} are assumed to be normally distributed. The normal distributions' means are regressed on two gene-specific effects, H_i and U_i. The former is simply assigned a normal distribution, while the latter represents a spatial effect sharing information between functionally similar genes. Similarity information is extracted from an external gene network where two functionally related genes i and g are connected by an edge and the weight ω_{ig} of the edge represents the strength or confidence of the relation,

$$Z_{ij} \sim N(\mu_i, \sigma_i^2),$$

$$\mu_i = \beta_0 + U_i + H_i, \tag{2}$$

$$1/\sigma_i^2 \sim Gamma(a_\sigma, b_\sigma), \tag{3}$$

$$U_i | U_r, r \neq i \sim N(m_i, v_i),$$

$$H_i \sim N(0, v_H).$$

The spatially structured effect U_i is given an intrinsic Gaussian CAR prior, and the network's similarity values ω_{ig} are employed as weights,

$$m_i = \frac{\sum_{g\in\delta_i}\omega_{ig}u_g}{\sum_{g\in\delta_i}\omega_{ig}} \text{ and } v_i = \frac{\tilde{v}}{\sum_{g\in\delta_i}\omega_{ig}},$$

where δ_i denotes a set of $\tilde{n}_i$ genes neighboring gene i in the gene network. Further, β_0 is

assigned an improper flat prior on R, and the variances $\tilde{v}$ and v_H are assigned inverse Gamma distributions,

$$\beta_0 \sim R(-\infty, \infty),$$

$$1/\tilde{v} \sim Gamma(a_{\tilde{v}}, b_{\tilde{v}}), \quad (4)$$

$$1/v_H \sim Gamma(a_{v_H}, b_{v_H}). \quad (5)$$

The posterior distributions of the quantities $E_i := U_i + H_i, i = 1,\ldots,n$, are used to classify genes as consistently differential, i.e. as presenting congruent differences between cases and controls across different data types. Specifically, gene i is assumed to be consistently differential if the 99% credible interval for E_i lies either above or below zero. An implementation of the model in the BUGS language is given in S1 File.

Prior elicitation. The hyperparameters of the distributions for $1/\sigma_i^2$, $1/\tilde{v}$ and $1/v_H$ are important as they regulate the degree of confidence in the gene-level data and, additionally, in the functional similarities between transcripts reported by the gene network. We suggest an empirical Bayesian approach where the prior in the model is chosen based on the empirical variance observed in the data. To obtain hyperparameters for $\tilde{v}$ and v_H, we decompose the variability of the gene-wise mean coefficients $Z_{i\bullet} = \frac{1}{m}\sum_{j=1}^{m} Z_{ij}$ into a non-structural part and a structural part explained by the neighborhood relationship. The structural part of the variance is used to derive a prior for $\tilde{v}$, and the remaining variance is used to derive the prior for v_H. Specifically, we assume

$$Var(H_i) = v_H \approx Var(Z_{i\bullet} - \tilde{m}_i) \text{ and } Var(U_i | U_r, r \neq i) = \tilde{v}/\omega_{\delta_i} \approx Var(\tilde{m}_i),$$

with $\tilde{m}_i = \sum_{g\in\delta_i} \omega_{ig} Z_{g\bullet} / \sum_{g\in\delta_i} \omega_{ig}$ and $\omega_{\delta_i} = \sum_{g\in\delta_i} \omega_{ig}$. The hyperparameters in Eqs (4) and (5) are calculated using that $E(X) = \alpha/\beta$ and $Var(X) = \alpha/\beta^2$ for $X \sim Gamma(\alpha,\beta)$. To solve the equations, ω_{δ_i} is replaced by the average number of neighbors and the variance of the priors is set to 10^4.

The parameters σ_i^2, $i = 1,\ldots,n$, model the variability of Z_{ij} within a gene across subjects. To derive the hyperparameters in Eq (3), we assume $\sigma_i^2 \approx \text{median}_{i\in\{1,\ldots n\}} \frac{1}{m}\sum_{j=1}^{m} (Z_{ij} - Z_{i\bullet})^2$ and use a prior variance of 10^4. The hyperparameters obtained for the presented analysis are given in S1 Table.

Data

Study cohort and case/control definition. The dataset was taken from the longitudinal Religious Orders Study and Rush Memory and Aging Project (ROS/MAP) [35]. Participants of these two studies were without known dementia at time of enrollment and underwent annual cognitive and clinical tests. The studies were approved by an Institutional Review Board of Rush University Medical Center. All participants signed an informed consent, an Anatomic Gift Act for brain donation, and a repository consent to allow their data and biospecimens to be shared. More information on the study and resources can be found on our Resource Sharing Hub at www.radc.rush.edu. Post-mortem neuropathologic evaluation was performed to assess AD and other brain pathologies common in aging and dementia. Gray matter was dissected from biopsies of the dorsolateral prefrontal cortex (DLPFC) and used to generate profiles of the histone acetylome, DNA methylome, and transcriptome [3]. For this study, we defined AD cases based on the NIA Reagan diagnosis (*high* or *intermediate* likelihood of AD) [36] and on the clinical diagnosis of dementia status at time of death (*AD and no*

other cause of cognitive impairment) [37]. Subjects with a NIA Reagan diagnosis of *low* likelihood of AD or *no AD* and a clinical diagnosis of *no cognitive impairment* were considered as controls (persons with mild cognitive impairment were excluded from these analyses). This definition of AD and control cases resulted in a total of 141 AD cases (40 males, Ø 90.1 years; 101 females, Ø 91.7 years) and 92 control cases (34 males, Ø 83.0 years; 58 females, Ø 85.9 years) who had complete genome-wide molecular data after quality control. For matching AD cases to controls, age of death was stratified into five-year intervals between 70 and 95 years and an additional >95 years stratum. Then, each AD case was randomly matched to a control of the same gender and age stratum.

Matching of data types. Different genomic data types were matched at the transcript (isoform) level following a previously suggested approach [38]. Transcript abundances were estimated from RNA-seq data using RSEM [39]. Only active transcripts with an fpkm value $\geq$2 in at least 25% of the samples were considered. The targeted histone mark histone 3 lysine 9 acetylation (H3K9ac) is primarily located at active transcriptional start sites (TSS) [40]. To quantify the H3K9ac level for a given transcript *i*, we counted the number of ChIP-seq reads aligned within a genomic region R_i^H of 5,000 bp centered at the transcript's TSS. Often, multiple transcripts of a gene share the same TSS or have TSSs in close genomic vicinity. These transcripts cannot be distinguished at the H3K9ac level, and thus, we merged genomic regions R_i^H of two or more transcripts if they overlapped and summed their respective fpkm values (Fig 1A). In total, we observed 23,674 active transcripts which were merged to 14,796 groups of transcripts with disjoint promoter regions R_i^H, and hence, distinct H3K9ac values. DNA methylation levels were measured by the Illumina HumanMethylation450 BeadChip limiting the methylation data to CG dinucleotides included in the chip design. DNA methylation in promoter regions is generally assumed to be negatively correlated with transcription, whereas the correlation with transcription has been reported to be positive when looking at exon methylation [34]. We calculated the promoter methylation for a transcript by averaging the methylation values of all probes located within the promoter region $R_i^{M_1}$ defined as 2,000 bp upstream of the transcript's TSS. Similarly, exon methylation was calculated by averaging the methylation values of all probes located within the transcript's exonic region $R_i^{M_2}$. As illustrated in Fig 1A, for groups of transcripts, we combined the regions $R_i^{M_1}$ and $R_i^{M_2}$ of the individual transcripts. Overall, we obtained 10,857 transcripts or groups of transcripts with non-missing values for all four data types. For simplicity, we will use the term gene hereafter ignoring the detail that our features actually represent either a single transcript or a group of transcripts that share a promoter, and thus, that a gene with two or more active promoters is represented by two or more features in our dataset.

Data normalization. Large genomic datasets are inevitably affected by technical confounders. To reduce the effect of these covariates, we used the preprocessed datasets after quality control as described in the respective original publications [41–43] and subsequently regressed out technical covariates, biological covariates, and the estimated proportion of neurons in the cell type composition of the neocortical tissue. Pearson residuals obtained from the regression models were then plugged into Eq (1) as our normalized observations $X_{ij}^{(k)}$ and $Y_{ij}^{(k)}$. Fig 1B depicts the correlation of the residuals of the same gene between different data types. Correlation between the residuals from gene transcription and H3K9ac data were shifted towards positive values, whereas correlation coefficients were almost centered for promoter and exon methylation. A strong correlation is not expected for the majority of genes, since we regressed out known factors like gender that impose a correlation structure between the data types. The remaining correlation structure is caused by AD or by unknown environmental

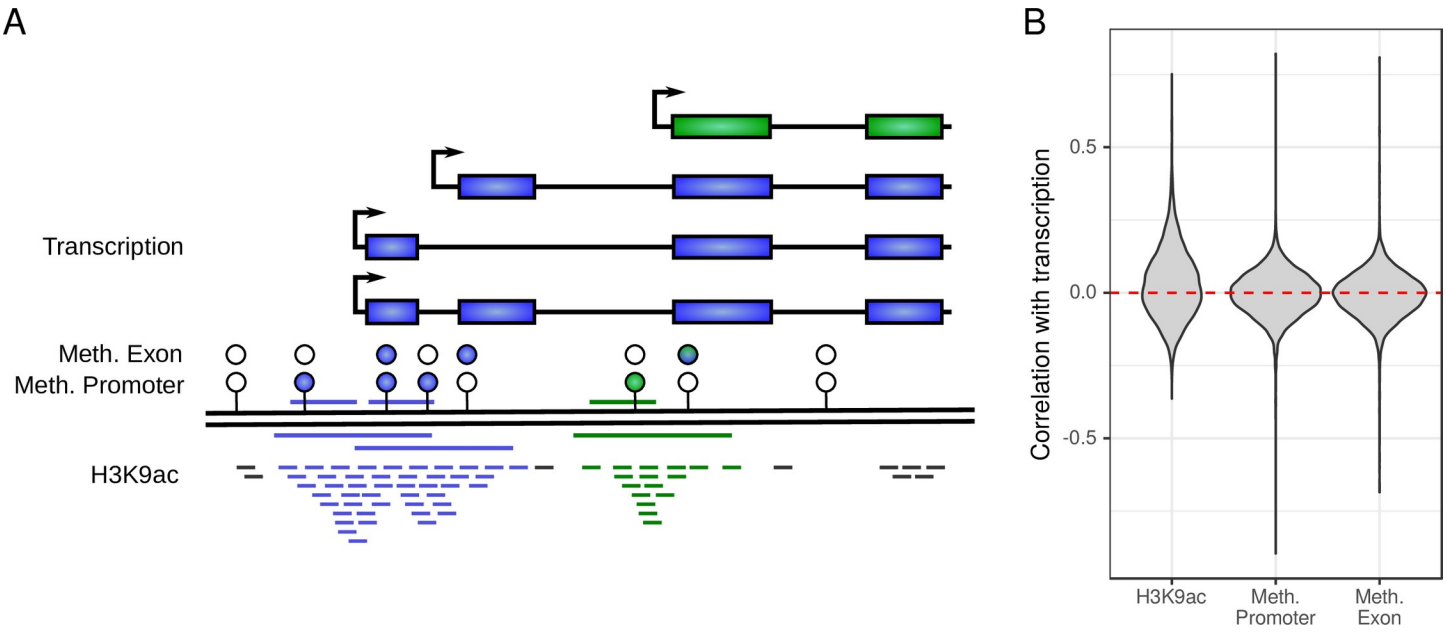


Fig 1. Matching different data types to genes. (A) The figure shows an exemplary gene with four transcripts and their TSSs (small arrows), CG methylation probes (circles), and H3K9ac ChIP-seq reads (small dashes at the bottom) aligned to the genome (black double line). H3K9ac data is matched to transcripts by counting the number of reads in the promoter region (long blue and green lines below the genome). Since the promoter regions (±2.5 kbp around TSS) of the three blue transcripts overlap, the blue transcripts are merged and all ChIP-seq reads are added together. Transcipt-level expression values from RNA-seq data for the blue transcripts are summed accordingly, whereas the green transcript constitutes a separate feature in the final dataset. Methylation levels are calculated separately for promoter and exon methylation. Promoter methylation is calculated as the average methylation level of all probes in the 2 kbp upstream promoter regions of the transcripts (blue and green lines above the genome). Selected probes are indicated by blue and green circles (lower row). Similarly, exon methylation is calculated as the average methylation level of all probes in the respective transcripts' exons (blue and green circles in the upper row). (B) Violin plots show the correlation between transcription data and H3K9ac, promoter methylation, and exon methylation respectively. Pearson correlation was calculated for each gene across the n = 233 subjects after removing the effects of technical variables, proportion of neurons, age and gender.

https://doi.org/10.1371/journal.pcbi.1007771.g001

and genetic factors. Environmental and genetic factors likely have a small effect due to the homogeneity of the ROS/MAP cohort.

In more detail, the following regression models were used. For the RNA-seq data (SynapseID: syn3388564), we log-transformed the transcript-level fpkm values and fitted a linear regression model for each transcript with the covariates RNA integrity score, log-transformed total number of sequence reads, batch, postmortem interval, age of death, gender and proportion of neurons. For the ChIP-seq data (Synapse ID: syn4896408), we used the number of reads observed in the genomic region R_i^H as outcome and fitted a negative binomial regression model for each transcript with the log transformed total number of reads as offset and the covariates cross correlation, postmortem interval, age at death, gender and proportion of neurons. DNA methylation data (Synapse ID: syn3157275) contained methylation values between 0 and 1 for each CG dinucleotide. We applied beta regression models with the covariates bisulfite conversion rate, batch, postmortem interval, age at death, gender and proportion of neurons.

The proportion of neurons used in the regression models to adjust for changes in cell type composition during the course of AD were estimated from the RNA-seq data. We applied the Digital Sorting Algorithm (DSA) [44] to the expression values of the five neuronal marker genes *GABBR2*, *MYT1L*, *ARL4C*, *CADPS* and *NRXN3* that were previously identified using an external human brain RNA-seq reference dataset of purified cells [42, 45]. We observed a decrease of the proportion of neurons from 66.9% to 65.5% in AD subjects (p = 0.01, Wilcoxon rank-sum test, n = 141 AD subjects and 92 controls) indicating the need to adjust for neuronal

proportion. RNA-seq-derived estimations were also used to adjust the H3K9ac and DNA methylation data since these data were generated from adjacent specimens of the same tissue block.

Functional gene network

Information about the functional similarity of genes was obtained from the HumanNet [46]. HumanNet is a functional gene network consisting of 16,243 genes connected by 476,399 weighted edges. Weights ω_{ig} range between 0.41 and 4.26 (Ø 1.14) and reflect the likelihood of a functional linkage between the two connected genes. The HumanNet was not developed specifically for the human brain and many genes and functional relationships are not observed in the human neocortex. Therefore, we first removed all genes that were not detected in our data. Then, edges of the induced subgraph were removed if the connected genes did not show a correlation coefficient larger or equal to 0.35 (85^{th} percentile) in an external gene transcription dataset from the Mount Sinai Brain Bank (MSBB) AD study [47]. The MSBB dataset consisted of 753 RNA-seq profiles of human aged and Alzheimer's brain samples generated from four different brain regions (S1 File). The modified network consisted of 6,470 genes connected by 41,093 edges with a mean weight of 1.24. The remaining 4,387 genes in our dataset that were either not represented in the original network or not connected by any edges after pruning remained in the analysis, but for these genes the structural component U_i is ignored in Eq (2).

Detection of differential subnetworks

We applied the prize-collecting Steiner tree (PCST) algorithm implemented in the R package PCSF [48] to detect subnetworks that were enriched with consistently differential genes identified by our integrative Bayesian analysis. To detect multiple trees in the network, the algorithm introduces an extra root node connected to each node in the network with cost ω_0 [49]. After the PCST problem has been solved, the artificial root node is removed from the tree and the remaining forest structure F is returned. Specifically, the algorithm maximized the objective function

$$f(F) = \beta \sum_{i \in V_F} |\hat{E}_i| - \sum_{(i,j) \in L_F} c(\omega_{ij}) - \omega_0 \kappa_F$$

where V_F denotes the set of all vertices (genes) in the forest and L_F the set of all edges (functional links between genes) in the forest. Variable κ_F denotes the number of trees in the forest. The cost for an edge (i, j) was defined as $c(\omega_{ij}) = \omega_c - \omega_{ij}$ with constant ω_c set to the sum of the maximum and minimum weight observed in the functional similarity gene network. The optimal solution depends on the two tuning parameters β and ω_0 that need to be specified. β balances the prizes associated with genes, i.e., the absolute values of the integrative statistics $|\hat{E}_i|$, and the costs assigned to the edges. A larger value of β results in larger trees. We set $\beta = 22$, which corresponded to the 75^{th} percentile of the costs $c(\omega_{ij})$ divided by the 95^{th} percentile of the prizes $|\hat{E}_i|$ in our data. The parameter ω_0 defines the costs for adding a tree to the forest. A larger value of ω_0 results in fewer trees in the optimal solution F. We set ω_0 to the 99^{th} percentile of $\beta|\hat{E}_i|$. Our choices for β and ω_0 preferred a solution with a few small trees, which is often better for biological interpretability. The optimal forest F with these settings consisted of nine trees. Three trees consisted of ten or more genes and were studied in more detail. These three trees were extended into the subnetworks by adding all edges to a tree that existed between any two genes of the tree in the initial functional gene similarity network. Thus, each of the three

subnetworks corresponds to one single tree found by the PCSF algorithm with the identical set of genes but additional edges. Differential subnetworks were tested for an enrichment of gene ontology (GO) terms from the biological process ontology using the R package topGO [50, 51]. Fisher's test was applied to compare genes within a subnetwork to the background set of all genes included in the initial gene network. GO terms with less than 10 genes were excluded from the analysis.

Model fitting

The hierarchical Bayesian model was implemented in the BUGS language. The code is available in S1 File. The Gibbs sampler implemented by WinBUGS was used to carry out 400,000 iterations after an initial number of 40,000 burn-in iterations. A thinning of 200 was applied resulting in 2,000 samples from the posterior distributions. Median values and 99% credible intervals were obtained from these 2,000 samples to perform inference. Estimates for the parameters β_0, v_H, and $\tilde{v}$, and their respective trace plots are given in S2 Table and S2 Fig.

Results

Detection of genes with consistent differences across data types in AD

The primary goal of our integrative analysis was to identify genes with consistent alterations of the epigenome and transcriptome in AD. Evidence for differential gene regulation across transcription, H3K9ac, promoter methylation and exon methylation data was summarized by the integrative coefficient Z. In Eq (1), we set $S^{(k)} = -1$ for promoter methylation and $S^{(k)} = 1$ for the other three data types to model the negative association between promoter methylation and transcription [34]. Genes were classified as consistently differential if the 99% credible interval for the integrative statistic E_i excluded 0. In total, 393 genes were significantly upregulated and 423 genes were significantly downregulated in AD. S3 Table contains the statistics for all genes included in the analysis. The first ten genes sorted by $|E_i|$ are shown in Table 1.

Among the differential genes in Table 1 are the neurotransmitter transporters *SLC6A9* and *SLC6A12*, which were associated with cognition in human AD patients and AD model systems [52–54]. A recently developed *SLC6A9* inhibitor is currently tested in a clinical trial [55]. The downregulated phosphatase *DUSP9* and the upregulated kinase *CDK18* have been suggested to modulate pathological tau phosphorylation in AD [56, 57]. *KIF5A* is a motor protein that has been reported to be upregulated in AD and may contribute to AD-related mitochondrial defects [58, 59]. Another candidate gene is *APOD*, which has a neuroprotective role and is upregulated in the aging and AD brain [60, 61]. Overall, the differential genes identified by our analysis are involved in various biological processes in different cell types reflecting the complexity of AD. Before we highlight some of these processes and discuss novel insights, we first study and validate our integrative model in more detail.

The integrative statistic E_i consists of a non-structural component H_i and a structural component U_i defined by the functional similarity gene network. If functionally related genes demonstrate congruent up- or downregulation in AD, we expect to observe large absolute values for U_i. To assess the importance of the network in our analysis, we approximated the fraction of the variance of E_i that is contributed by U_i, $Var(U_i)/(Var(U_i)+Var(H_i)) = 0.06$. A fraction of 6% indicates that while H_i accounted for most of the observed differences between AD and controls, some functionally related genes were jointly dysregulated in AD. The structural component U_i alone was significant for a total of 6 genes.

Table 1. Top 10 differential genes ranked by $|\hat{E}_i|$.

Gene	$\hat{E}_i$ [99% CI]	$\tilde{n}_i$	Ø ω_{ig}	+\-	$Z_{i\bullet}^{(exprs)}$	$Z_{i\bullet}^{(H3K9ac)}$	$Z_{i\bullet}^{(methP)}$	$Z_{i\bullet}^{(methE)}$
SLC6A9	0.93 [0.49, 1.34]	1	1.05	69.5%	0.72	0.32	-0.07	0.06
KIF5A	0.77 [0.43, 1.12]	6	1.41	75.2%	0.69	0.21	-0.28	0.04
CRB2	0.76 [0.38, 1.14]	0	-	74.5%	0.53	0.28	-0.17	0.29
SLC6A12	0.74 [0.37, 1.13]	0	-	73.8%	0.98	0.33	0.10	0.08
PRELP	0.74 [0.40, 1.07]	2	1.13	72.3%	0.67	0.24	-0.07	0.10
DUSP6	-0.71 [-1.11, -0.30]	1	0.54	38.3%	-0.18	-0.46	0.09	-0.03
MAP4K3-DT	-0.70 [-1.04, -0.36]	0	-	24.8%	-0.56	-0.28	0.07	-0.20
SLC14A1	-0.70 [-1.19, -0.16]	1	1.00	28.4%	-1.01	-0.44	0.02	-0.09
APOD	0.69 [0.31, 1.06]	4	0.91	67.4%	0.34	0.24	-0.41	0.06
CDK18	0.68 [0.32, 1.06]	28	0.78	69.5%	0.68	0.32	-0.06	0.01

Columns from left to right show the gene symbol, estimated integrative statistic $\hat{E}_i$ and its 99% credible interval, number of neighbors in the gene network, mean weights of neighbors, percentage of subjects j with positive coefficient Z_{ij}. The last four columns show the mean coefficient $Z_{i\bullet}$ calculated using only genes expression, H3K9ac, promoter methylation, and exon methylation data.

https://doi.org/10.1371/journal.pcbi.1007771.t001

Assessment of specificity and sensitivity using simulated data

To validate our model and prior development, we simulated a dataset based on the 92 control subjects, which were randomly split into 46 cases and 46 controls. A total of 10,000 genes were randomly selected and assigned to 500 pathways with 20 genes each. Differences between cases and controls were simulated for half of the pathways by adding or subtracting a value Δ to the observed variables $X_{ij}^{(k)}, k = 1, \ldots, 4$, of all genes in the pathway in the case samples. Then, a noisy gene network was simulated consisting of 47,500 correct edges between genes of the same pathway and 99,800 incorrect edges between genes of different pathways.

Only 19 out of the 5,000 non-differential genes were falsely classified as differential based on a 99% credible interval. The method correctly identified 1,173 of the 5,000 differential genes resulting in an FDR of 0.016 and a sensitivity ranging from 0.027 to 0.535 depending on the magnitude of the simulated difference Δ as depicted in Fig 2A. Notably, even the largest simulated difference Δ = 0.15 corresponded to a small standardized effect size of 0.083 (Cohen's d). Next, we compared the performance of our Bayesian model to alternative approaches using our simulated dataset. For our Bayesian approach, we observed an area under the receiver operating characteristic curve (AUC) of 0.84 (Fig 2B), which is a modest improvement over the gene-wise one-sample t-tests applied to the Z values (AUC of 0.80). A moderated t-test on the Z values as implemented in the limma software performed equally well as the regular t-test (AUC of 0.80) [62]. All three approaches operate on the integrative coefficient Z and therefore leverage information from all four data types and consider the directionality of the effects observed in the different data types. Alternatively, an integrative analysis can be performed as a meta-analysis on significance levels or rankings obtained from separate analyses of the different data types. To implement this strategy, we first ran paired two-tailed t-tests on the simulated $X_{ij}^{(k)}$ and $Y_{ij}^{(k)}$ values separately for each of the k = 1,. . .,4 data types to obtain a p-value per gene and data type. Subsequently, the results from the different data types were combined by either calculating meta-p-values using the z-score method (AUC of 0.76), or by calculating the genes' mean ranks (AUC of 0.75), or by applying the Robust Rank Aggregation method (AUC of 0.75). The latter method was specifically developed for integrating multi-omic data [63]. However, when combining p-values or gene ranks, information about the directionality of the effect sizes observed in the different datasets is lost resulting

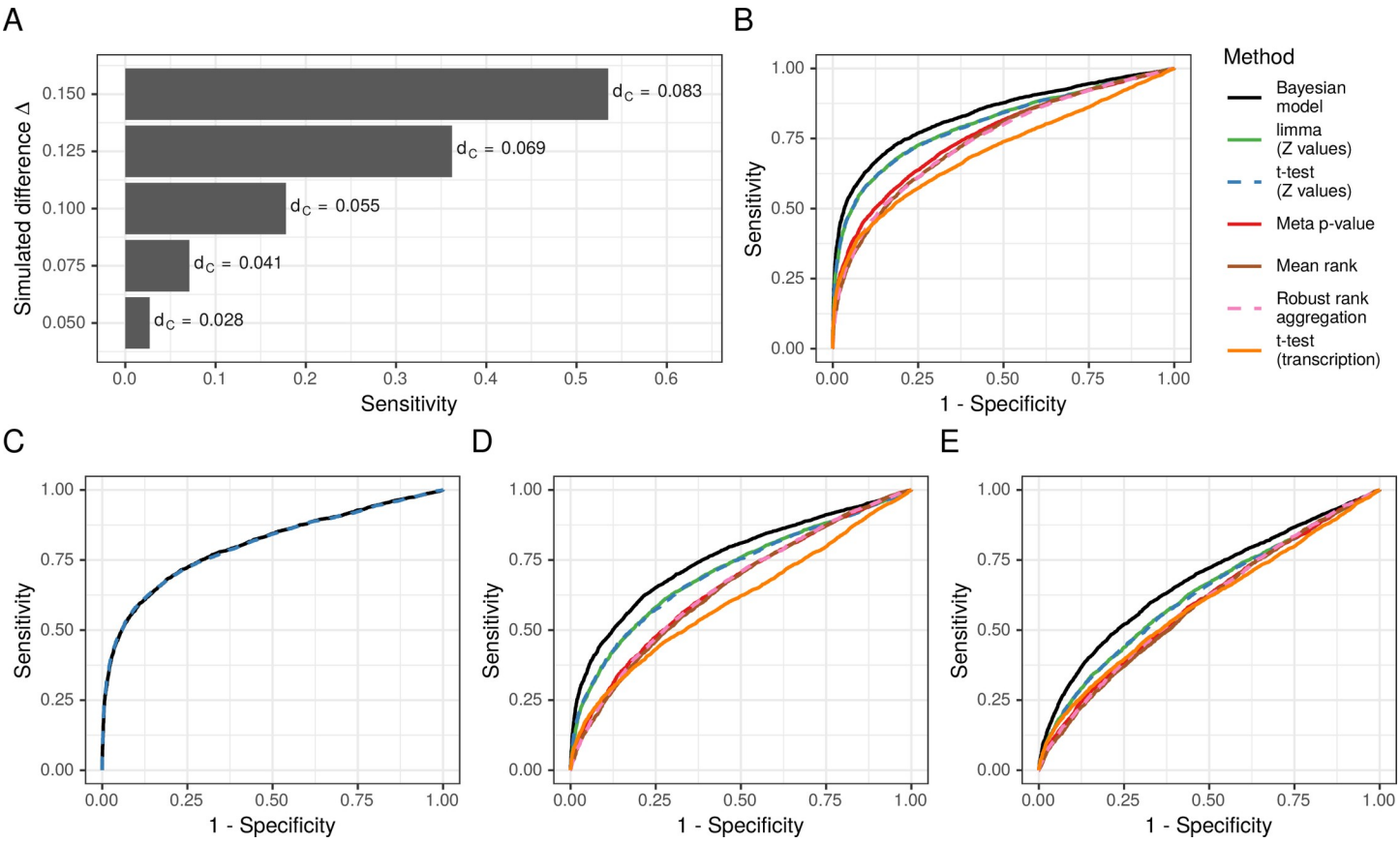


Fig 2. Sensitivity and specificity analysis. (A) The sensitivity achieved by the Bayesian model on the simulated dataset (n = 92) is shown on the x-axis for various simulated differences Δ on the y-axis. The standardized effect size d_C (Cohen's d) is depicted next to the bars. (B) Sensitivity is plotted against 1—specificity as observed in the simulated data for the Bayesian model and six alternative approaches. (C) Sensitivity is plotted against 1—specificity observed when using a random gene network. For better comparison, the curve observed for the t-test identical as in (B) was added to the plot. (D, E) Sensitivity is plotted versus 1—specificity as in (B) using a smaller sample size of n = 46 (D) and n = 20 (E).

https://doi.org/10.1371/journal.pcbi.1007771.g002

in lower AUC values. Finally, we added the results obtained from applying a t-test to only one data type (transcription data as example) to our comparison in Fig 2B. As expected, the AUC of 0.71 was smaller compared to the integrative methods, since a large fraction of the data was not used.

To study how the Bayesian method performs if a non-informative network is given, we randomly permuted the edges of the simulated gene network. Fig 2C shows the results from the Bayesian method with the random network (AUC of 0.80) next to the unchanged results from the t-test for a better comparison. These results indicate that the improvement of the Bayesian method depicted in Fig 2B stems from the information provided by the gene network, and that if a random network is given, the Bayesian methods performs equally well as the t-test. Finally, we studied the effect of different sample sizes on the methods by repeating the simulation study on subsets of n = 46 (Fig 2D) and n = 20 (Fig 2E) samples. While the performance of all methods declined with smaller sample sizes, the Bayesian method maintained an advantage as more relative weight was given to the prior. In summary, the simulation study demonstrated that the Bayesian model with the selected priors results in a small false positive rate, and, if an appropriate network is given, performs better than simple gene-wise t-tests on the integrative coefficients or methods that integrate p-values or ranks.

Validation in independent datasets

We first studied whether genes with consistent differences in epigenomic and transcriptomic data identified by our analysis also presented differences at the protein level. To do this, we utilized a targeted proteomics dataset generated from the same sample type and sample collection as our multi-omic data: DLPFC samples of ROS/MAP participants (Synapse ID: syn10468856). The targeted proteins were candidate genes from previous AD studies [64] and measured by liquid chromatography-selected reaction monitoring [65]. We applied the same case/control definition as for the main analysis resulting in 393 AD cases and 214 control subjects. Each protein was tested for difference in abundance in AD versus control subjects, adjusting for gender, age and postmortem interval (S1 File). Overall, we observed a positive correlation (Pearson's r = 0.54) between the integrative statistic $\hat{E}_i$ and the observed differences in the protein data based on 98 proteins encoded by genes considered in our integrative analysis (Fig 3A). Out of 18 differential genes from our integrative analysis, 9 genes demonstrated significantly altered protein levels in AD (family-wise error rate ≤ 0.05); the direction of effect was consistent between the two sets of results.

Since the protein data is limited to selected AD candidate genes and was not generated from an independent cohort, we additionally validated our findings in two RNA-seq datasets from other AD sample collections. The first dataset consisted of 79 AD samples and 37 control samples from the inferior frontal gyrus included in the MSBB study [47]. We identified 747 genes that were classified as differential in our integrative analysis and passed the detection threshold in the inferior frontal gyrus samples of the MSBB dataset (S1 File). When comparing AD to control samples, a majority of 601 out of the 747 genes showed a change in transcription consistent with the results from the integrative analysis (Fig 3B). These changes were significant at an unadjusted p-value of 0.05 for 102 out of 354 upregulated genes and for 97 out of the 393 downregulated genes. Similar results were observed for the second dataset of temporal cortex samples from the Mayo LOAD study (n = 71 control samples, n = 80 AD samples) [4]. We detected 759 of our differential genes in the temporal cortex (S1 File), and 553 of these genes showed a consistent increase or decrease in AD (Fig 3C). At an unadjusted p-value of 0.05, 154 out of 359 upregulated and 200 out of 400 downregulated genes were validated in the Mayo LOAD study.

Differential subnetworks

When we analyzed the posterior distributions we found that about 6% of the variance of E_i is contributed by U_i, suggesting that some parts of the gene network are collectively dysregulated in AD. Such a subnetwork of jointly differential genes often represents a disease-related biological process and is easier to interpret than single genes. To identify differential subnetworks post hoc, we applied a prize-collecting Steiner tree (PCST) algorithm [48, 66]. The objective of the PCST algorithm was to find a subnetwork that maximizes the sum of $|\hat{E}_i|$ of the genes in the subnetwork minus the costs for the edges $c(\omega_{ij})$ needed to construct the subnetwork.

Three differential subnetworks with at least 10 genes were identified. The first subnetwork (Fig 4A) was enriched with genes involved in *myeloid cell differentiation* (S4 Table) and reflected the immune component of AD [67]. The network included myeloid transcription factors such as *NFIC* [68], and cytokines such as *CSF1* and the corresponding receptor *CSF1R*, which have recently been studied in the context of microglia activation [69–71]. Cellular functions of the upregulated genes *RHOQ* and *TRIP10* include endocytosis and regulation of cell shape and motility [72, 73]. To verify that this gene network is transcribed by myeloid cells, we compared the genes' transcription levels in an external RNA-seq dataset of purified human brain cells (Fig 4B) [45]. Further, all five significant genes in the network were also

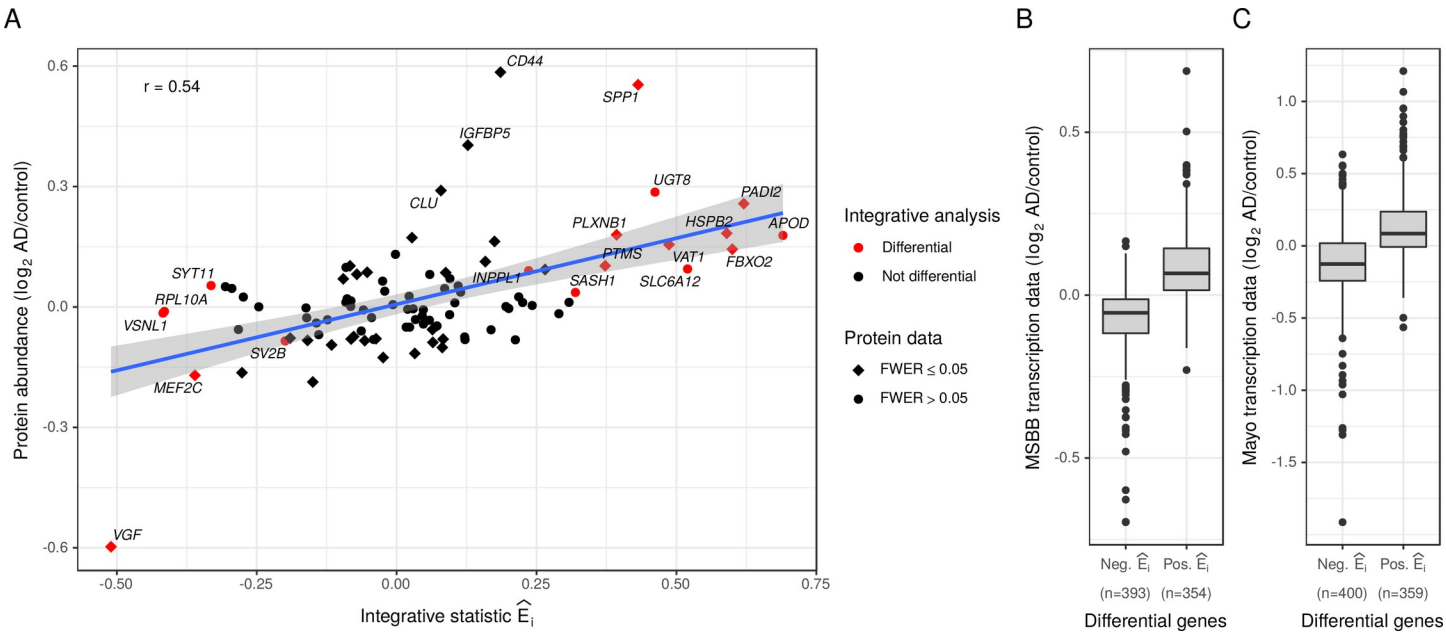
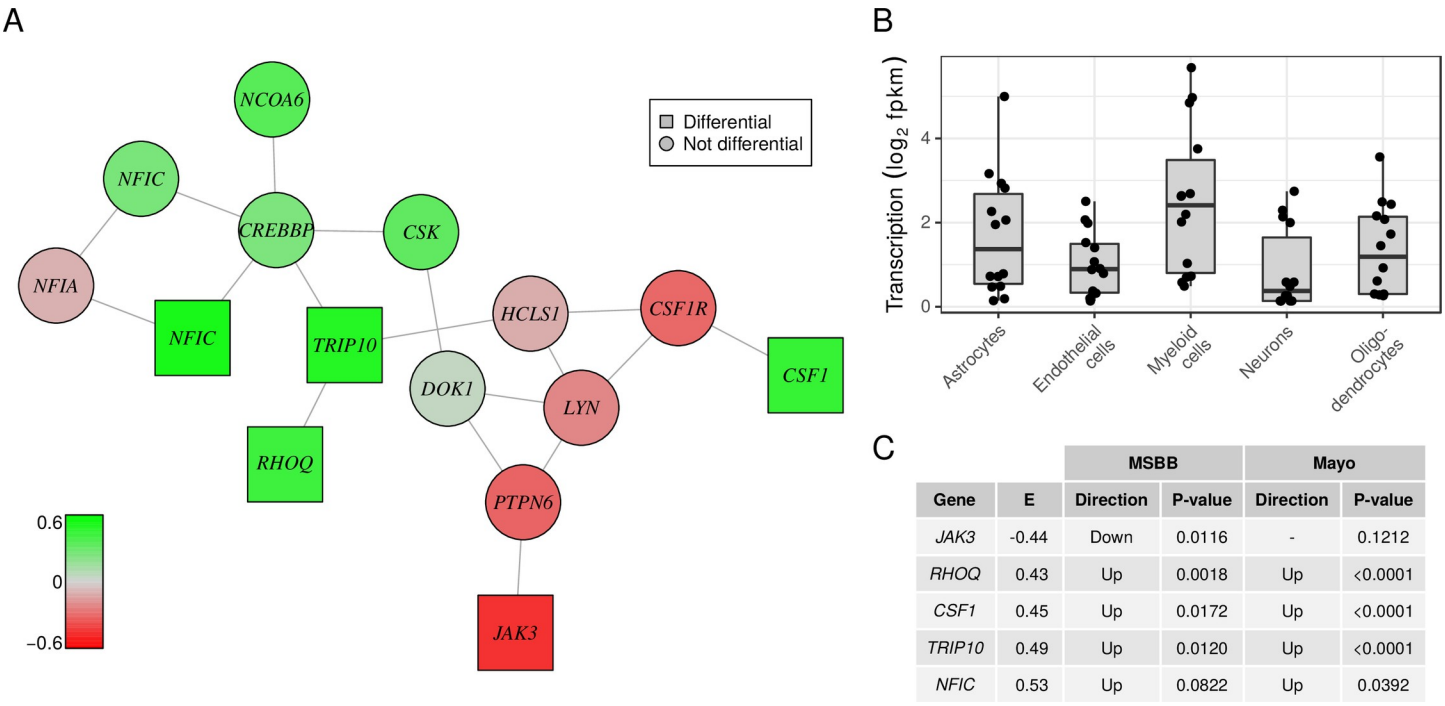


Fig 3. Validation of differential genes identified by the integrative analysis. (A) The integrative statistic for 98 genes that were included in a targeted proteomic dataset is plotted on the x-axis versus the observed differences between AD and control cases in the protein data on the y-axis. Red color indicates genes that were detected as differential in the integrative analysis (n = 233 samples). Squares indicate significant differences in the protein data (n = 607 samples) at a family-wise error rate of 0.05. (B) Differences in gene transcription between AD and controls observed in the MSBB RNA-seq study (inferior frontal gyrus, n = 116 samples) are shown separately for genes identified as up- or downregulated in the integrative analysis. (C) Similarly, differences in gene transcription between AD and controls observed in the Mayo LOAD RNA-seq study (temporal cortex, n = 151 samples) are shown separately for genes identified as up- or downregulated.

https://doi.org/10.1371/journal.pcbi.1007771.g003

differentially transcribed in the MSBB or the Mayo LOAD dataset (Fig 4C). We note that the myeloid genes that we prioritize are different from the well-validated myeloid AD susceptibility genes that have emerged from genome-wide association studies.

The second differential network (S3 Fig) was enriched for the Gene Ontology term *protein phosphorylation* (S5 Table). Protein phosphorylation regulates various cellular processes by altering protein activity, localization and stability, and this mechanism has been implicated in AD [74]. For example, the gene *PRKAA2* (alias *AMPK*) encodes a kinase that regulates cellular energy homeostasis, is activated by amyloid-β, and phosphorylates tau at multiple sites [75–77]. Another kinase directly involved in the phosphorylation and accumulation of tau is TTBK1 [78, 79]. Interestingly, TTBK1 also phosphorylates TDP-43, a protein which forms pathologic aggregates in aged and AD brains [80, 81]. MAP2K4 (alias MKK4) has been suggested to phosphorylate tau [82] and to modulate amyloid-β toxicity [83]. Most kinases and phosphatases were depicted in the right half of the network (S3 Fig). The lower left part of the network included two genes, *TUBA1B* and *TUBB2A*, that encode major constituents of microtubules, which are disrupted by hyperphosphorylated tau in AD [84]. The tubulin genes were connected to the mitochondrial fission gene *DNM1L* (alias *DRP1*) in the network. The protein DNM1L interacts with amyloid-β and hyperphosphorylated tau, causing mitochondria fragmentation and thereby affecting mitochondrial health and axonal transport in AD neurons [85–87]. Thus, the lower left part of the network reflected impaired energy metabolism in AD synapses. The upper left part of the network consisted of dysregulated genes of the ubiquitin proteasomal system, such as *PSMD2*, *BTRC*, *CUL9* and *UBQLN1* [88]. UBQLN1 is involved in the degradation of PSEN1 and APP, two proteins which are essential for the generation of amyloid-β peptides (*APP* is the gene that encodes the amyloid-β peptide) [89, 90]. Overexpression of *UBQLN1* alleviates symptoms in some AD mouse models [91]. Thus, while this



C

		MSBB		Mayo	
Gene	E	Direction	P-value	Direction	P-value
JAK3	-0.44	Down	0.0116	-	0.1212
RHOQ	0.43	Up	0.0018	Up	<0.0001
CSF1	0.45	Up	0.0172	Up	<0.0001
TRIP10	0.49	Up	0.0120	Up	<0.0001
NFIC	0.53	Up	0.0822	Up	0.0392

Fig 4. Myeloid cell differentiation network. (A) Graph shows the subnetwork of differential genes largely involved in myeloid cell differentiation. Color encodes the value of the integrative statistic from green (upregulated in AD) to red (downregulated in AD). Squares indicate significantly differential genes (99% credible interval). The gene *NFIC* is represented twice reflecting two alternative active promoters. (B) Boxplots depict the transcription levels of the subnetwork's genes in each of five major brain cell types obtained from an external RNA-seq dataset of purified cell types. (C) Table shows the value of the integrative statistic $\hat{E}_i$ and the unadjusted p-value from the two external validation datasets for each significant gene in the subnetwork. The directionality in the validation studies (up- or downregulated in AD) is given if the p-value was less than 0.1.

https://doi.org/10.1371/journal.pcbi.1007771.g004

complex network captures several different processes, we refer to the network as protein phosphorylation network because of the enrichment with kinases and phosphatases (S5 Table). Many genes of this network were transcribed by neurons (S3B Fig).

The third differential network was characterized by the GO term *synaptic signaling* (S6 Table) and mainly consisted of downregulated synaptic genes (S4 Fig). For example, *RGS7* regulates synaptic plasticity by modulating the signaling pathway downstream of the $GABA_B$ receptor [92]. *RPH3A*, another downregulated gene involved in synaptic signaling, correlates with cognitive decline and is specifically downregulated by amyloid-β [93]. Similarly, the glutamate receptors GRIA2 (alias GluR2) and GRIN2A (alias GluN2A) have been shown to be reduced in the postsynaptic density in AD and are associated with memory deficit [94, 95]. Overall, this network, which is mainly transcribed by neurons (S4B Fig), reflects abnormalities in synaptic signaling and a reduction of synaptic density, which is a hallmark of AD and occurs before neuronal cell death [96, 97].

Upregulation of the myeloid network is associated with amyloid-β pathology and promoted by *CSF1* expressing astrocytes

To further investigate the role of the biological processes underlying the three differential networks in neurodegeneration, we leveraged the RNA-seq data from the Mayo LOAD study. In addition to 80 AD and 71 control samples, the dataset from the Mayo LOAD study also included 30 samples with a post mortem diagnosis of pathologic aging [4]. Individuals with pathologic aging have widespread cortical amyloid-β plaque deposits but demonstrate no or

only minimal neurofibrillary tau pathology and are cognitively non-impaired [98]. It is unclear whether pathologic aging is an early stage of AD or whether this condition develops in individuals who have protective factors that block processes downstream of amyloid-β pathology [99]. We summarized the transcriptional activity of a network in the Mayo cohort by calculating the first principal component of the normalized RNA-seq transcription profiles of the network's genes (S1 File). First, we verified that the three networks were differentially transcribed between AD and controls in this independent cohort (Fig 5A–5C). As expected, the myeloid cell differentiation network was significantly upregulated in AD compared to controls ($p = 0.001$, Wilcoxon rank-sum test), and the protein phosphorylation ($p = 0.002$, Wilcoxon rank-sum test) and the synaptic signaling ($p<0.001$, Wilcoxon rank-sum test) networks were significantly downregulated in AD. Interestingly, the individuals diagnosed with pathologic aging demonstrated an upregulation of the myeloid cell differentiation network ($p = 0.047$, Wilcoxon test) to a level similar to that seen in AD subjects (Fig 5A), whereas the protein phosphorylation and synaptic signaling networks were not dysregulated in this group of samples (Fig 5B and 5C). These findings indicate that the upregulation of the myeloid cell differentiation network does not require tau pathology and is probably an early event in the pathogenesis of AD preceding tau pathology and neuronal dysfunction that manifest as impairment in cognitive function. These results are consistent with a recent study suggesting that microglia interact with amyloid-β pathology to contribute to tau proteinopathy and downstream cognitive decline [100].

The myeloid cell differentiation network consisted of 14 genes of which 5 were classified as differential (Fig 4A). The cytokine *CSF1*, one of the 5 differential genes, is an interesting candidate gene, because of its role as a regulator of myeloid cell frequency and function during homeostasis and inflammation [69, 101]. Previous studies of the corresponding receptor *CSF1R* in AD mouse models showed that blocking *CSF1R* reduced microglia density and attenuated the burden of AD pathology in the animals [70, 102]. To investigate which cell types trigger *CSF1* signaling, we employed single-nucleus RNA-sequencing (snRNA-seq) data from n = 48 subjects from the ROS/MAP study (S1 File) [103]. As shown in Fig 5D, *CSF1* was primarily transcribed in astrocytes, oligodendrocytes and oligodendrocyte progenitor cells. Next, we confirmed that *CSF1R* was exclusively transcribed by myeloid cells in the human prefrontal cortex (Fig 5E). Finally, we tested which of the three cell types that transcribed *CSF1* contributed to the differences between AD and controls observed at the tissue level. Interestingly, an upregulation of *CSF1* in AD was only observed in astrocytes ($p = 0.016$, Wilcoxon rank-sum test), but not in oligodendrocytes or oligodendrocyte progenitor cells (Fig 5F). Although the sample size of the snRNA-seq data is limited, these findings suggest that astrocytes activate microglia cells via *CSF1* signaling. The alternative *CSF1R* ligand *IL34* was not detected as differential in our integrative analysis (S3 Table).

Discussion

Large multi-omic datasets are becoming more common in biomedical research and require novel integrative bioinformatics approaches to fully harness their potential. We developed an integrative method to detect genes consistently altered in multiple data types in case-control studies. In addition to integrating information from different data types, our method also utilizes functional gene similarity to share information across genes and thereby improve statistical inference.

Information from different data types is aggregated by the integrative coefficient given in Eq (1) at the gene level. Data was matched to genes based on genome annotation in our AD study, however, other data types, e.g. some enhancer marks, may require a more complex

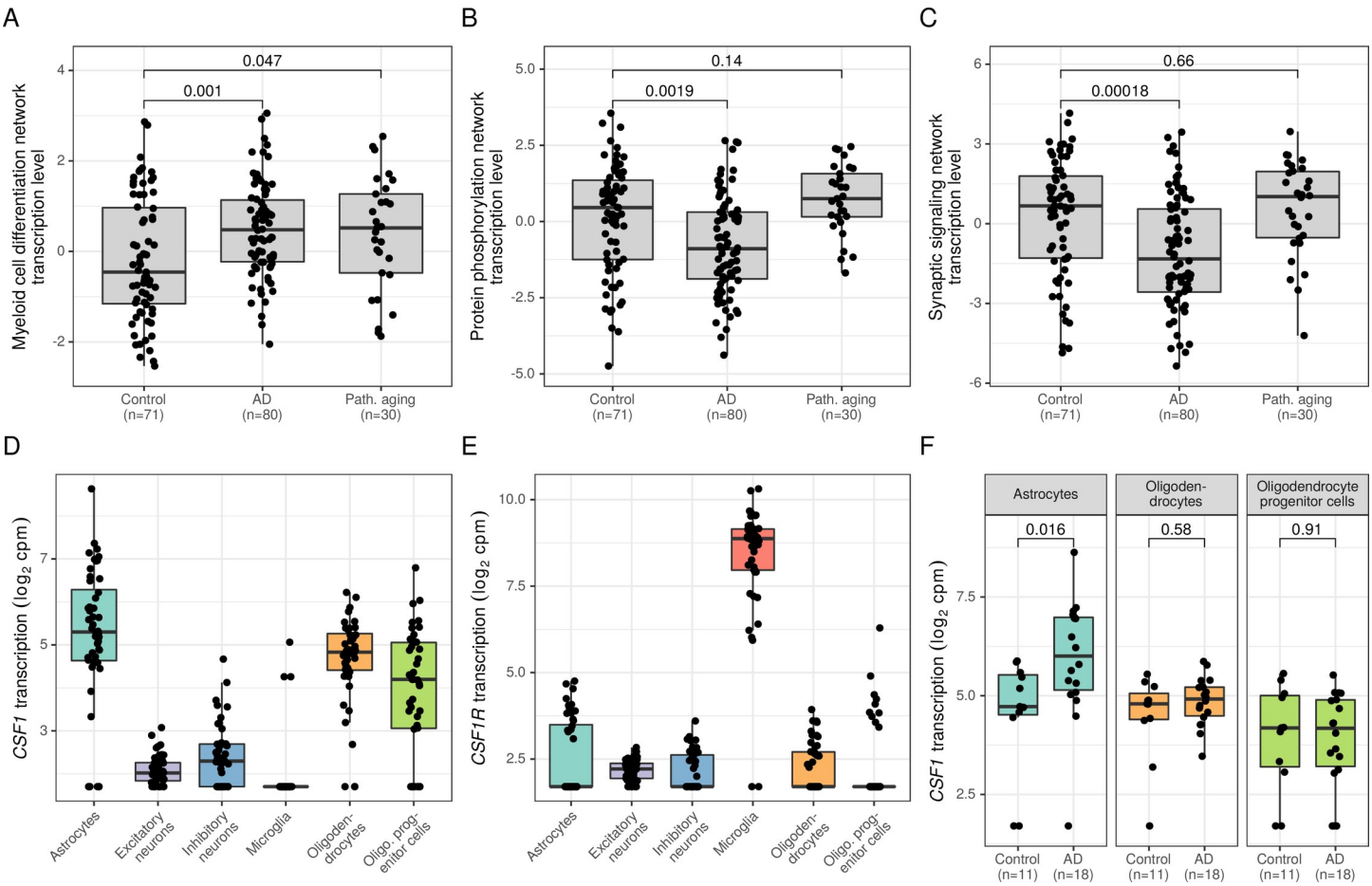


Fig 5. Increased *CSF1* transcription in astrocytes contributes to amyloid-β-related activation of the myeloid cell differentiation network. (A) Boxplots show transcription levels of the myeloid cell differentiation network (first principal component) in control, AD, and pathological aging samples from the Mayo LOAD study (Wilcoxon rank-sum tests, unadjusted p-values). (B, C) Similarly, network transcription levels are shown for the protein phosphorylation network (B), and for the synaptic signaling network (C). (D, E) Boxplots depict transcription levels of *CSF1* (D) and *CSF1R* (E) in six major human brain cell types measured in the prefrontal cortex from 48 individuals. (F) *CSF1* transcription levels are shown separately for controls and AD cases in astrocytes, oligodendrocytes and oligodendrocyte progenitor cells (Wilcoxon rank-sum tests, unadjusted p-values).

https://doi.org/10.1371/journal.pcbi.1007771.g005

matching strategy as outlined elsewhere [38, 104]. After matching, we observed primarily positive gene-wise correlations between transcription and H3K9ac, whereas no clear trend was observed for promoter or exon methylation (Fig 1B). This may reflect the complex relation between DNA methylation and transcriptional activity in the brain, including the role of hydroxymethylation in neurons [105], but we also note that a large correlation between the residuals of different data types should generally not be expected, since we regressed out the effects of major factors such as age, gender and proportion of neurons that impose a correlation structure on the data. The remaining correlation structure was likely caused by unknown genetic and environmental factors as well as the AD status, which may not affect many genes in all data types. To limit the effect of data types that are not associated with the outcome, we modelled an additive instead of a multiplicative coefficient suggested by previous studies [24, 25]. Further, multiplicative coefficients follow a more complex product distribution and the sign of the coefficient is difficult to interpret if more than two data types are involved. In Eq (1), the factors $S^{(k)}$ model the relationship between data types so that the sign of the coefficient corresponds to an up- or downregulation, respectively. The factors $S^{(k)}$ can often be chosen

based on prior knowledge derived from studies like Encode or Roadmap Epigenomics [106, 107].

A hierarchical Bayesian model is used to study the distribution of the integrative coefficient. An innovation of the model is the representation of the differences between AD cases and controls by a non-structural component H_i and a structural component U_i which shares information between functionally related genes. Based on the results in a recent comparative review, we selected the HumanNet to define functional similarity [108]. Functional similarity networks are constructed from various datasets from different tissues and organisms and are not brain specific. Thus, we had to customize the network and prune edges which were not supported by an external brain RNA-seq dataset in order to better utilize the information contained in the network for our specific analysis. After pruning the network, we observed that overall approximately 6% of the AD effects were contributed by U_i indicating that parts of the gene network were jointly dysregulated in AD. The fraction of variation attributable to the structural component depends on the gene network and the structure of the studied data and may vary between diseases and tissues. In other domains like spatial epidemiology, a wide range of values has been observed that can be as large as 71% in extreme cases [109]. Future studies will have to show whether a fraction of 6% as observed in this study is a common value for genome-wide molecular data.

We validated our model using simulated and independent data from other studies. The simulation study was important to demonstrate that our prior choices result in a reasonable small FDR of 0.016 when using 99% credible intervals to classify genes. Further, the simulation study showed that our method outperforms a one-sample t-test on the integrative coefficients, if an appropriate network is given. The one-sample t-test on the integrative coefficient resembles a paired t-test as the coefficient is the sum of the differences between the matched samples across data types, and thus, can be expected to be powerful in the setting of a matched case-control study. Consequently, when a random network was given, both methods performed equally well indicating that the advantage of the Bayesian method stemmed from the information provided by the network. Methods that integrate results from separate analyses were inferior as these methods ignore the directionality of the observed differences in the different data types. Further, these methods do not provide a statistical framework for assessing significance and controlling error rates. For example, the z-score approach as included in the comparison will likely result in inflated p-values since the data sets are not independent.

Comparing our results from the integrative analysis with the protein data and the external transcription data revealed that a majority of our findings can be reproduced at the protein level and at the mRNA level in an independent cohort, even though the aim of our method was not to predict differences at the protein or transcription level, but to identify genes with consistent differences across the given data types. In line with the validation results, we found that many of the differential genes given in Table 1 have been studied as candidate genes for AD. Thus, we anticipate that the complete result from the gene-wise analysis (S3 Table) is a useful resource for AD candidate genes. However, we take these results further, prioritizing a subset of genes that may be of greater interest: based on the gene-wise results, we studied which parts of the gene similarity network were collectively dysregulated in AD. Three different dysregulated AD subnetworks were identified: *myeloid cell differentiation*, *protein phosphorylation*, and *synaptic signaling*. Similar network-based approaches have been suggested to reveal disease related pathways which may not become obvious in a gene-wise analyses [110]. In contrast to single genes, a network signature can usually be replicated more robustly in model systems or independent datasets, and thus, can be helpful in follow-up studies to address questions such as the temporal progression of these three processes during the course of AD.

In this study, we further investigated the status of the three differential networks in pathologic aging, which is characterized by high amyloid-β loads similar as in AD but a lack of distinct tau pathology [98]. Consistent with the normal cognition of individuals with pathologic aging, the synaptic signaling network was not altered compared to controls. We also found no evidence for altered transcription of the protein phosphorylation network; however, the myeloid cell differentiation network was upregulated to a similar level as observed in AD. Although it is unclear whether the amyloid-β aggregation in pathologic aging reflects an early stage of AD, these findings support the hypothesis that microglia are already activated at the preclinical stage of AD before accumulation of hyperphosphorylated tau. An interesting member of the myeloid cell differentiation network is *CSF1* because of its role as a regulator of myeloid cell numbers and functions [101]. A few studies of AD focused on the corresponding receptor *CSF1R* as a potential therapeutic target. Microglia cells depend on *CSF1R* signaling [111, 112] and treatment of AD mice with *CSF1R* inhibitors results in reduced microglia activation and improved memory function [70, 113], but little is known about the cells that contribute to *CSF1R* triggering in AD. Using snRNA-seq data, we showed that the upregulation of *CSF1* observed at the tissue level is primarily caused by astrocytes in human AD brains. Altogether, our findings suggest that astrocytes contribute to microglial activation by expressing *CSF1* at an early stage of AD preceding tau accumulation. Whether *CSF1* overexpression by astrocytes is directly provoked by amyloid-β cannot be concluded from our data. Interestingly, activated microglia in return secrete signals that induce reactive astrocytes illustrating the complex relationship between these two cell types during the pathogenesis of AD [114].

In summary, we proposed a novel method for the joint analysis of multiple genome-wide datasets that utilizes external information about functional gene similarity. We applied the method to transcription, histone acetylation and DNA methylation data from a large AD study and discovered multiple well-known and new target genes as well as AD processes. Further study of one of these processes indicated that astrocytes may contribute to microglia activation by *CSF1* expression at early stages of AD. Our approach can be adapted to analyze other multi-omic case-control datasets and thereby promotes integrative analyses to fully utilize these complex datasets.

Supporting information

S1 Fig. Schematic overview of the integrative multi-omic analysis. Blue boxes indicate input datasets, red boxes indicate our novel integrative analysis, and green boxes indicate the explorative post hoc analysis of differential subnetworks. The genomic data from our AD case-control study consisted of four different data types, which were matched to genes and summarized by the integrative coefficient Z_{ij}. Z_{ij} modeled the differences observed across data types for gene i when comparing sample j to its matched control sample. The distribution of Z_{ij} was modeled by a hierarchical Bayesian model to identify consistently differential genes. The Bayesian model incorporated a gene network (HumanNet) to share information between functionally related genes. Genes with consistent differences in the epigenomic and transcriptomic data between AD and control samples were the primary result of our integrative analysis (S3 Table). To further analyze and interpret the results, we subsequently employed an explorative network-based approach to detect AD-related subnetworks (green boxes). Therefore, we annotated the genes in the HumanNet with the integrative statistic $\hat{E}_i$ derived from our Bayesian model and used a prize-collecting Steiner tree (PCST) algorithm to identify subnetworks enriched with consistently differential genes.
(TIF)

S2 Fig. Trace plots of MCMC draws. (A) Trace plot for parameter β_0 after removing the burn-in period. A thinning of 200 was applied. (B) Trace plot for parameter v_H after removing the burn-in period. A thinning of 200 was applied. (C) Trace plot for parameter $\tilde{v}$ after removing the burn-in period. A thinning of 200 was applied.
(TIF)

S3 Fig. Protein phosphorylation network. (A) Graph shows the subnetwork of differential genes largely involved in protein phosphorylation. Color encodes the value of the integrative statistic from green (upregulated in AD) to red (downregulated in AD). Squares indicate significantly differential genes (99% credible interval). (B) Boxplots depict the transcription levels of the subnetwork's genes in each of five major brain cell types obtained from an external RNA-seq dataset of purified cell types. (C) Table shows the value of the integrative statistic $\hat{E}_i$ and the unadjusted p-value from the two external validation datasets for each significant gene in the subnetwork. The directionality in the validation studies (up- or downregulated in AD) is given if the p-value was less than 0.1.
(TIF)

S4 Fig. Synaptic signaling network. (A) Graph shows the subnetwork of differential genes largely involved in synaptic signaling. Color encodes the value of the integrative statistic from green (upregulated in AD) to red (downregulated in AD). Squares indicate significantly differential genes (99% credible interval). The gene *ANK2* is represented twice reflecting two alternative active promoters. (B) Boxplots depict the transcription levels of the subnetwork's genes in each of five major brain cell types obtained from an external RNA-seq dataset of purified cell types. (C) Table shows the value of the integrative statistic $\hat{E}_i$ and the unadjusted p-value from the two external validation datasets for each significant gene in the subnetwork. The directionality in the validation studies (up- or downregulated in AD) is given if the p-value was less than 0.1.
(TIF)

S1 Table. Hyperparameters of the hierarchical Bayesian model.
(DOCX)

S2 Table. Parameter estimates of the hierarchical Bayesian model.
(DOCX)

S3 Table. Analysis results for all 10,857 genes.
(XLSX)

S4 Table. GO analysis of the myeloid cell differentiation subnetwork.
(DOCX)

S5 Table. GO analysis of the protein phosphorylation network.
(DOCX)

S6 Table. GO analysis of the synaptic signaling network.
(DOCX)

S1 File. Detailed description of the datasets used in this study including data preprocessing. BUGS code for the hierarchical Bayesian model.
(PDF)

Acknowledgments

We thank AMP-AD's RNA-seq reprocessing work group for sharing their normalized versions of the MSBB and Mayo LOAD RNA-seq data.

Author Contributions

Conceptualization: Hans-Ulrich Klein, Martin Schäfer, David A. Bennett, Holger Schwender, Philip L. De Jager.

Data curation: Hans-Ulrich Klein, David A. Bennett, Philip L. De Jager.

Formal analysis: Hans-Ulrich Klein, Martin Schäfer.

Funding acquisition: David A. Bennett, Holger Schwender, Philip L. De Jager.

Investigation: Hans-Ulrich Klein, Martin Schäfer, David A. Bennett, Holger Schwender, Philip L. De Jager.

Methodology: Hans-Ulrich Klein, Martin Schäfer.

Project administration: Hans-Ulrich Klein, Philip L. De Jager.

Resources: David A. Bennett, Holger Schwender, Philip L. De Jager.

Software: Hans-Ulrich Klein, Martin Schäfer.

Supervision: David A. Bennett, Holger Schwender, Philip L. De Jager.

Validation: Hans-Ulrich Klein, Martin Schäfer, Philip L. De Jager.

Visualization: Hans-Ulrich Klein, Martin Schäfer.

Writing – original draft: Hans-Ulrich Klein, Martin Schäfer.

Writing – review & editing: Hans-Ulrich Klein, Martin Schäfer, David A. Bennett, Holger Schwender, Philip L. De Jager.

References

1. Jack CR Jr., Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimers Dement. 2018; 14 (4):535–62. Epub 2018/04/15. https://doi.org/10.1016/j.jalz.2018.02.018 PMID: 29653606; PubMed Central PMCID: PMCPMC5958625.
2. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome Biol. 2017; 18(1):83. Epub 2017/05/10. https://doi.org/10.1186/s13059-017-1215-1 PMID: 28476144; PubMed Central PMCID: PMCPMC5418815.
3. De Jager PL, Ma Y, McCabe C, Xu J, Vardarajan BN, Felsky D, et al. A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. Sci Data. 2018; 5:180142. Epub 2018/08/08. https://doi.org/10.1038/sdata.2018.142 PMID: 30084846; PubMed Central PMCID: PMCPMC6080491.
4. Allen M, Carrasquillo MM, Funk C, Heavner BD, Zou F, Younkin CS, et al. Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. Sci Data. 2016; 3:160089. Epub 2016/10/12. https://doi.org/10.1038/sdata.2016.89 PMID: 27727239; PubMed Central PMCID: PMCPMC5058336.
5. Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. Methods of integrating data to uncover genotype-phenotype interactions. Nat Rev Genet. 2015; 16(2):85–97. Epub 2015/01/15. https://doi.org/10.1038/nrg3868 PMID: 25582081.
6. Richardson S, Tseng GC, Sun W. Statistical Methods in Integrative Genomics. Annu Rev Stat Appl. 2016; 3:181–209. Epub 2016/08/03. https://doi.org/10.1146/annurev-statistics-041715-033506 PMID: 27482531; PubMed Central PMCID: PMCPMC4963036.
7. Angelini C, Costa V. Understanding gene regulatory mechanisms by integrating ChIP-seq and RNA-seq data: statistical solutions to biological problems. Front Cell Dev Biol. 2014; 2:51. Epub 2014/11/05.

https://doi.org/10.3389/fcell.2014.00051 PMID: 25364758; PubMed Central PMCID: PMCPMC4207007.

8. Ickstadt K, Schäfer M, Zucknick M. Toward Integrative Bayesian Analysis in Molecular Biology. Annual Review of Statistics and Its Application, Vol 5. 2018; 5:141–67. https://doi.org/10.1146/annurev-statistics-031017-100438 WOS:000429191800007.
9. Huang S, Chaudhary K, Garmire LX. More Is Better: Recent Progress in Multi-Omics Data Integration Methods. Front Genet. 2017; 8:84. Epub 2017/07/04. https://doi.org/10.3389/fgene.2017.00084 PMID: 28670325; PubMed Central PMCID: PMCPMC5472696.
10. Moreau Y, Tranchevent LC. Computational tools for prioritizing candidate genes: boosting disease gene discovery. Nat Rev Genet. 2012; 13(8):523–36. Epub 2012/07/04. https://doi.org/10.1038/nrg3253 PMID: 22751426.
11. Bersanelli M, Mosca E, Remondini D, Giampieri E, Sala C, Castellani G, et al. Methods for the integration of multi-omics data: mathematical aspects. BMC Bioinformatics. 2016; 17 Suppl 2:15. Epub 2016/01/30. https://doi.org/10.1186/s12859-015-0857-9 PMID: 26821531; PubMed Central PMCID: PMCPMC4959355.
12. Gamazon ER, Segre AV, van de Bunt M, Wen X, Xi HS, Hormozdiari F, et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. Nat Genet. 2018; 50(7):956–67. Epub 2018/06/30. https://doi.org/10.1038/s41588-018-0154-4 PMID: 29955180; PubMed Central PMCID: PMCPMC6248311.
13. Raj T, Li YI, Wong G, Humphrey J, Wang M, Ramdhani S, et al. Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. Nat Genet. 2018; 50(11):1584–92. Epub 2018/10/10. https://doi.org/10.1038/s41588-018-0238-1 PMID: 30297968; PubMed Central PMCID: PMCPMC6354244.
14. Ng B, White CC, Klein HU, Sieberts SK, McCabe C, Patrick E, et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. Nat Neurosci. 2017; 20(10):1418–26. Epub 2017/09/05. https://doi.org/10.1038/nn.4632 PMID: 28869584; PubMed Central PMCID: PMCPMC5785926.
15. Karlic R, Chung HR, Lasserre J, Vlahovicek K, Vingron M. Histone modification levels are predictive for gene expression. Proc Natl Acad Sci U S A. 2010; 107(7):2926–31. Epub 2010/02/06. https://doi.org/10.1073/pnas.0909344107 PMID: 20133639; PubMed Central PMCID: PMCPMC2814872.
16. Park SJ, Nakai K. A regression analysis of gene expression in ES cells reveals two gene classes that are significantly different in epigenetic patterns. BMC Bioinformatics. 2011; 12 Suppl 1:S50. Epub 2011/03/05. https://doi.org/10.1186/1471-2105-12-S1-S50 PMID: 21342583; PubMed Central PMCID: PMCPMC3044308.
17. Ouyang Z, Zhou Q, Wong WH. ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. Proc Natl Acad Sci U S A. 2009; 106(51):21521–6. Epub 2009/12/10. https://doi.org/10.1073/pnas.0904863106 PMID: 19995984; PubMed Central PMCID: PMCPMC2789751.
18. Xu X, Hoang S, Mayo MW, Bekiranov S. Application of machine learning methods to histone methylation ChIP-Seq data reveals H4R3me2 globally represses gene expression. BMC Bioinformatics. 2010; 11:396. Epub 2010/07/27. https://doi.org/10.1186/1471-2105-11-396 PMID: 20653935; PubMed Central PMCID: PMCPMC2928206.
19. Cheng C, Alexander R, Min R, Leng J, Yip KY, Rozowsky J, et al. Understanding transcriptional regulation by integrative analysis of transcription factor binding data. Genome Res. 2012; 22(9):1658–67. Epub 2012/09/08. https://doi.org/10.1101/gr.136838.111 PMID: 22955978; PubMed Central PMCID: PMCPMC3431483.
20. Dong X, Greven MC, Kundaje A, Djebali S, Brown JB, Cheng C, et al. Modeling gene expression using chromatin features in various cellular contexts. Genome Biol. 2012; 13(9):R53. Epub 2012/09/07. https://doi.org/10.1186/gb-2012-13-9-r53 PMID: 22950368; PubMed Central PMCID: PMCPMC3491397.
21. Hu P, Jiao R, Jin L, Xiong M. Application of Causal Inference to Genomic Analysis: Advances in Methodology. Front Genet. 2018; 9:238. Epub 2018/07/26. https://doi.org/10.3389/fgene.2018.00238 PMID: 30042787; PubMed Central PMCID: PMCPMC6048229.
22. Tasaki S, Gaiteri C, Mostafavi S, Yu L, Wang Y, De Jager PL, et al. Multi-omic Directed Networks Describe Features of Gene Regulation in Aged Brains and Expand the Set of Genes Driving Cognitive Decline. Front Genet. 2018; 9:294. Epub 2018/08/25. https://doi.org/10.3389/fgene.2018.00294 PMID: 30140277; PubMed Central PMCID: PMCPMC6095043.
23. Schenk T, Chen WC, Gollner S, Howell L, Jin L, Hebestreit K, et al. Inhibition of the LSD1 (KDM1A) demethylase reactivates the all-trans-retinoic acid differentiation pathway in acute myeloid leukemia.

Nat Med. 2012; 18(4):605–11. Epub 2012/03/13. https://doi.org/10.1038/nm.2661 PMID: 22406747; PubMed Central PMCID: PMCPMC3539284.

24. Klein HU, Schäfer M, Porse BT, Hasemann MS, Ickstadt K, Dugas M. Integrative analysis of histone ChIP-seq and transcription data using Bayesian mixture models. Bioinformatics. 2014; 30(8):1154–62. Epub 2014/04/15. https://doi.org/10.1093/bioinformatics/btu003 PMID: 24403540.

25. Schäfer M, Klein HU, Schwender H. Integrative analysis of multiple genomic variables using a hierarchical Bayesian model. Bioinformatics. 2017; 33(20):3220–7. Epub 2017/06/06. https://doi.org/10.1093/bioinformatics/btx356 PMID: 28582573.

26. Fuxman Bass JI, Diallo A, Nelson J, Soto JM, Myers CL, Walhout AJ. Using networks to measure similarity between genes: association index selection. Nat Methods. 2013; 10(12):1169–76. Epub 2013/12/04. https://doi.org/10.1038/nmeth.2728 PMID: 24296474; PubMed Central PMCID: PMCPMC3959882.

27. Huang J, Ma S, Li H, Zhang CH. The Sparse Laplacian Shrinkage Estimator for High-Dimensional Regression. Ann Stat. 2011; 39(4):2021–46. Epub 2011/11/22. https://doi.org/10.1214/11-aos897 PMID: 22102764; PubMed Central PMCID: PMCPMC3217586.

28. Pan W, Xie B, Shen X. Incorporating predictor network in penalized regression with application to microarray data. Biometrics. 2010; 66(2):474–84. Epub 2009/08/04. https://doi.org/10.1111/j.1541-0420.2009.01296.x PMID: 19645699; PubMed Central PMCID: PMCPMC3338337.

29. Kim S, Pan W, Shen X. Network-based penalized regression with application to genomic data. Biometrics. 2013; 69(3):582–93. Epub 2013/07/05. https://doi.org/10.1111/biom.12035 PMID: 23822182; PubMed Central PMCID: PMCPMC4007772.

30. Chen M, Cho J, Zhao H. Incorporating biological pathways via a Markov random field model in genome-wide association studies. PLoS Genet. 2011; 7(4):e1001353. Epub 2011/04/15. https://doi.org/10.1371/journal.pgen.1001353 PMID: 21490723; PubMed Central PMCID: PMCPMC3072362.

31. Stingo FC, Vannucci M. Variable selection for discriminant analysis with Markov random field priors for the analysis of microarray data. Bioinformatics. 2011; 27(4):495–501. Epub 2010/12/17. https://doi.org/10.1093/bioinformatics/btq690 PMID: 21159623; PubMed Central PMCID: PMCPMC3105481.

32. Robinson S, Nevalainen J, Pinna G, Campalans A, Radicella JP, Guyon L. Incorporating interaction networks into the determination of functionally related hit genes in genomic experiments with Markov random fields. Bioinformatics. 2017; 33(14):i170–i9. Epub 2017/09/09. https://doi.org/10.1093/bioinformatics/btx244 PMID: 28881978; PubMed Central PMCID: PMCPMC5870666.

33. Schäfer M, Lkhagvasuren O, Klein HU, Elling C, Wustefeld T, Muller-Tidow C, et al. Integrative analyses for omics data: a Bayesian mixture model to assess the concordance of ChIP-chip and ChIP-seq measurements. J Toxicol Environ Health A. 2012; 75(8–10):461–70. Epub 2012/06/13. https://doi.org/10.1080/15287394.2012.674914 PMID: 22686305.

34. Wagner JR, Busche S, Ge B, Kwan T, Pastinen T, Blanchette M. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. Genome Biol. 2014; 15(2):R37. Epub 2014/02/22. https://doi.org/10.1186/gb-2014-15-2-r37 PMID: 24555846; PubMed Central PMCID: PMCPMC4053980.

35. Bennett DA, Buchman AS, Boyle PA, Barnes LL, Wilson RS, Schneider JA. Religious Orders Study and Rush Memory and Aging Project. J Alzheimers Dis. 2018; 64(s1):S161–S89. Epub 2018/06/06. https://doi.org/10.3233/JAD-179939 PMID: 29865057; PubMed Central PMCID: PMCPMC6380522.

36. Bennett DA, Schneider JA, Arvanitakis Z, Kelly JF, Aggarwal NT, Shah RC, et al. Neuropathology of older persons without cognitive impairment from two community-based studies. Neurology. 2006; 66(12):1837–44. Epub 2006/06/28. https://doi.org/10.1212/01.wnl.0000219668.47116.e6 PMID: 16801647.

37. Schneider JA, Arvanitakis Z, Bang W, Bennett DA. Mixed brain pathologies account for most dementia cases in community-dwelling older persons. Neurology. 2007; 69(24):2197–204. Epub 2007/06/15. https://doi.org/10.1212/01.wnl.0000271090.28148.24 PMID: 17568013.

38. Klein HU, Schäfer M. Integrative Analysis of Histone ChIP-seq and RNA-seq Data. Curr Protoc Hum Genet. 2016; 90:20 3 1–3 16. Epub 2016/07/02. https://doi.org/10.1002/cphg.17 PMID: 27367165.

39. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12:323. Epub 2011/08/06. https://doi.org/10.1186/1471-2105-12-323 PMID: 21816040; PubMed Central PMCID: PMCPMC3163565.

40. Kratz A, Arner E, Saito R, Kubosaki A, Kawai J, Suzuki H, et al. Core promoter structure and genomic context reflect histone 3 lysine 9 acetylation patterns. BMC Genomics. 2010; 11:257. Epub 2010/04/23. https://doi.org/10.1186/1471-2164-11-257 PMID: 20409305; PubMed Central PMCID: PMCPMC2867832.

41. Mostafavi S, Gaiteri C, Sullivan SE, White CC, Tasaki S, Xu J, et al. A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. Nat

Neurosci. 2018; 21(6):811–9. Epub 2018/05/29. https://doi.org/10.1038/s41593-018-0154-9 PMID: 29802388.

42. Klein HU, McCabe C, Gjoneska E, Sullivan SE, Kaskow BJ, Tang A, et al. Epigenome-wide study uncovers large-scale changes in histone acetylation driven by tau pathology in aging and Alzheimer's human brains. Nat Neurosci. 2019; 22(1):37–46. Epub 2018/12/19. https://doi.org/10.1038/s41593-018-0291-1 PMID: 30559478

43. De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nat Neurosci. 2014; 17(9):1156–63. Epub 2014/08/19. https://doi.org/10.1038/nn.3786 PMID: 25129075; PubMed Central PMCID: PMCPMC4292795.

44. Zhong Y, Wan YW, Pang K, Chow LM, Liu Z. Digital sorting of complex tissues for cell type-specific gene expression profiles. BMC Bioinformatics. 2013; 14:89. Epub 2013/03/19. https://doi.org/10.1186/1471-2105-14-89 PMID: 23497278; PubMed Central PMCID: PMCPMC3626856.

45. Zhang Y, Sloan SA, Clarke LE, Caneda C, Plaza CA, Blumenthal PD, et al. Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. Neuron. 2016; 89(1):37–53. Epub 2015/12/22. https://doi.org/10.1016/j.neuron.2015.11.013 PMID: 26687838; PubMed Central PMCID: PMCPMC4707064.

46. Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. Genome Res. 2011; 21(7):1109–21. Epub 2011/05/04. https://doi.org/10.1101/gr.118992.110 PMID: 21536720; PubMed Central PMCID: PMCPMC3129253.

47. Wang M, Beckmann ND, Roussos P, Wang E, Zhou X, Wang Q, et al. The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. Sci Data. 2018; 5:180185. Epub 2018/09/12. https://doi.org/10.1038/sdata.2018.185 PMID: 30204156; PubMed Central PMCID: PMCPMC6132187.

48. Akhmedov M, Kedaigle A, Chong RE, Montemanni R, Bertoni F, Fraenkel E, et al. PCSF: An R-package for network-based interpretation of high-throughput data. PLoS Comput Biol. 2017; 13(7):e1005694. Epub 2017/08/02. https://doi.org/10.1371/journal.pcbi.1005694 PMID: 28759592; PubMed Central PMCID: PMCPMC5552342.

49. Tuncbag N, Braunstein A, Pagnani A, Huang SS, Chayes J, Borgs C, et al. Simultaneous reconstruction of multiple signaling pathways via the prize-collecting steiner forest problem. J Comput Biol. 2013; 20(2):124–36. Epub 2013/02/07. https://doi.org/10.1089/cmb.2012.0092 PMID: 23383998; PubMed Central PMCID: PMCPMC3576906.

50. Alexa A, Rahnenfuhrer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics. 2006; 22(13):1600–7. Epub 2006/04/12. https://doi.org/10.1093/bioinformatics/btl140 PMID: 16606683.

51. The Gene Ontology C. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 2019; 47(D1):D330–D8. Epub 2018/11/06. https://doi.org/10.1093/nar/gky1055 PMID: 30395331; PubMed Central PMCID: PMCPMC6323945.

52. Barbash S, Garfinkel BP, Maoz R, Simchovitz A, Nadorp B, Guffanti A, et al. Alzheimer's brains show inter-related changes in RNA and lipid metabolism. Neurobiol Dis. 2017; 106:1–13. Epub 2017/06/21. https://doi.org/10.1016/j.nbd.2017.06.008 PMID: 28630030; PubMed Central PMCID: PMCPMC5560656.

53. Harada K, Nakato K, Yarimizu J, Yamazaki M, Morita M, Takahashi S, et al. A novel glycine transporter-1 (GlyT1) inhibitor, ASP2535 (4-[3-isopropyl-5-(6-phenyl-3-pyridyl)-4H-1,2,4-triazol-4-yl]-2,1,3-benzoxadiazol e), improves cognition in animal models of cognitive impairment in schizophrenia and Alzheimer's disease. Eur J Pharmacol. 2012; 685(1–3):59–69. Epub 2012/05/01. https://doi.org/10.1016/j.ejphar.2012.04.013 PMID: 22542656.

54. Ibi D, Tsuchihashi A, Nomura T, Hiramatsu M. Involvement of GAT2/BGT-1 in the preventive effects of betaine on cognitive impairment and brain oxidative stress in amyloid beta peptide-injected mice. Eur J Pharmacol. 2019; 842:57–63. Epub 2018/11/06. https://doi.org/10.1016/j.ejphar.2018.10.037 PMID: 30393201.

55. Rosenbrock H, Desch M, Kleiner O, Dorner-Ciossek C, Schmid B, Keller S, et al. Evaluation of Pharmacokinetics and Pharmacodynamics of BI 425809, a Novel GlyT1 Inhibitor: Translational Studies. Clin Transl Sci. 2018; 11(6):616–23. Epub 2018/08/24. https://doi.org/10.1111/cts.12578 PMID: 30136756; PubMed Central PMCID: PMCPMC6226115.

56. Banzhaf-Strathmann J, Benito E, May S, Arzberger T, Tahirovic S, Kretzschmar H, et al. MicroRNA-125b induces tau hyperphosphorylation and cognitive deficits in Alzheimer's disease. EMBO J. 2014; 33(15):1667–80. Epub 2014/07/09. https://doi.org/10.15252/embj.201387576 PMID: 25001178; PubMed Central PMCID: PMCPMC4194100.

57. Herskovits AZ, Davies P. The regulation of tau phosphorylation by PCTAIRE 3: implications for the pathogenesis of Alzheimer's disease. Neurobiol Dis. 2006; 23(2):398–408. Epub 2006/06/13. https://doi.org/10.1016/j.nbd.2006.04.004 PMID: 16766195.

58. Hares K, Miners JS, Cook AJ, Rice C, Scolding N, Love S, et al. Overexpression of Kinesin Superfamily Motor Proteins in Alzheimer's Disease. J Alzheimers Dis. 2017; 60(4):1511–24. Epub 2017/10/25. https://doi.org/10.3233/JAD-170094 PMID: 29060936.

59. Wang Q, Tian J, Chen H, Du H, Guo L. Amyloid beta-mediated KIF5A deficiency disrupts anterograde axonal mitochondrial movement. Neurobiol Dis. 2019; 127:410–8. Epub 2019/03/30. https://doi.org/10.1016/j.nbd.2019.03.021 PMID: 30923004.

60. Dassati S, Waldner A, Schweigreiter R. Apolipoprotein D takes center stage in the stress response of the aging and degenerative brain. Neurobiol Aging. 2014; 35(7):1632–42. Epub 2014/03/13. https://doi.org/10.1016/j.neurobiolaging.2014.01.148 PMID: 24612673; PubMed Central PMCID: PMCPMC3988949.

61. Li H, Ruberu K, Munoz SS, Jenner AM, Spiro A, Zhao H, et al. Apolipoprotein D modulates amyloid pathology in APP/PS1 Alzheimer's disease mice. Neurobiol Aging. 2015; 36(5):1820–33. Epub 2015/03/19. https://doi.org/10.1016/j.neurobiolaging.2015.02.010 PMID: 25784209.

62. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004; 3:Article3. Epub 2006/05/02. https://doi.org/10.2202/1544-6115.1027 PMID: 16646809.

63. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics. 2012; 28(4):573–80. Epub 2012/01/17. https://doi.org/10.1093/bioinformatics/btr709 PMID: 22247279; PubMed Central PMCID: PMCPMC3278763.

64. Yu L, Petyuk VA, Gaiteri C, Mostafavi S, Young-Pearse T, Shah RC, et al. Targeted brain proteomics uncover multiple pathways to Alzheimer's dementia. Ann Neurol. 2018; 84(1):78–88. Epub 2018/06/17. https://doi.org/10.1002/ana.25266 PMID: 29908079; PubMed Central PMCID: PMCPMC6119500.

65. Andreev VP, Petyuk VA, Brewer HM, Karpievitch YV, Xie F, Clarke J, et al. Label-free quantitative LC-MS proteomics of Alzheimer's disease and normally aged human brains. J Proteome Res. 2012; 11(6):3053–67. Epub 2012/05/09. https://doi.org/10.1021/pr3001546 PMID: 22559202; PubMed Central PMCID: PMCPMC3445701.

66. Akhmedov M, Kwee I, Montemanni R. A divide and conquer matheuristic algorithm for the Prize-collecting Steiner Tree Problem. Comput Oper Res. 2016; 70:18–25. https://doi.org/10.1016/j.cor.2015.12.015 WOS:000372380400003.

67. Sarlus H, Heneka MT. Microglia in Alzheimer's disease. J Clin Invest. 2017; 127(9):3240–9. Epub 2017/09/02. https://doi.org/10.1172/JCI90606 PMID: 28862638; PubMed Central PMCID: PMCPMC5669553.

68. Wahlestedt M, Ladopoulos V, Hidalgo I, Sanchez Castillo M, Hannah R, Sawen P, et al. Critical Modulation of Hematopoietic Lineage Fate by Hepatic Leukemia Factor. Cell Rep. 2017; 21(8):2251–63. Epub 2017/11/23. https://doi.org/10.1016/j.celrep.2017.10.112 PMID: 29166614; PubMed Central PMCID: PMCPMC5714592.

69. De I, Nikodemova M, Steffen MD, Sokn E, Maklakova VI, Watters JJ, et al. CSF1 overexpression has pleiotropic effects on microglia in vivo. Glia. 2014; 62(12):1955–67. Epub 2014/07/22. https://doi.org/10.1002/glia.22717 PMID: 25042473; PubMed Central PMCID: PMCPMC4205273.

70. Olmos-Alonso A, Schetters ST, Sri S, Askew K, Mancuso R, Vargas-Caballero M, et al. Pharmacological targeting of CSF1R inhibits microglial proliferation and prevents the progression of Alzheimer's-like pathology. Brain. 2016; 139(Pt 3):891–907. Epub 2016/01/10. https://doi.org/10.1093/brain/awv379 PMID: 26747862; PubMed Central PMCID: PMCPMC4766375.

71. Oosterhof N, Kuil LE, van der Linde HC, Burm SM, Berdowski W, van Ijcken WFJ, et al. Colony-Stimulating Factor 1 Receptor (CSF1R) Regulates Microglia Density and Distribution, but Not Microglia Differentiation In Vivo. Cell Rep. 2018; 24(5):1203–17 e6. Epub 2018/08/02. https://doi.org/10.1016/j.celrep.2018.06.113 PMID: 30067976.

72. Holmes WR, Edelstein-Keshet L. Analysis of a minimal Rho-GTPase circuit regulating cell shape. Phys Biol. 2016; 13(4):046001. Epub 2016/07/21. https://doi.org/10.1088/1478-3975/13/4/046001 PMID: 27434017.

73. Shimada A, Niwa H, Tsujita K, Suetsugu S, Nitta K, Hanawa-Suetsugu K, et al. Curved EFC/F-BAR-domain dimers are joined end to end into a filament for membrane invagination in endocytosis. Cell. 2007; 129(4):761–72. Epub 2007/05/22. https://doi.org/10.1016/j.cell.2007.03.040 PMID: 17512409.

74. Oliveira J, Costa M, de Almeida MSC, da Cruz ESOAB, Henriques AG. Protein Phosphorylation is a Key Mechanism in Alzheimer's Disease. J Alzheimers Dis. 2017; 58(4):953–78. Epub 2017/05/21. https://doi.org/10.3233/JAD-170176 PMID: 28527217.

75. Cai Z, Yan LJ, Li K, Quazi SH, Zhao B. Roles of AMP-activated protein kinase in Alzheimer's disease. Neuromolecular Med. 2012; 14(1):1–14. Epub 2012/03/01. https://doi.org/10.1007/s12017-012-8173-2 PMID: 22367557.

76. Domise M, Didier S, Marinangeli C, Zhao H, Chandakkar P, Buee L, et al. AMP-activated protein kinase modulates tau phosphorylation and tau pathology in vivo. Sci Rep. 2016; 6:26758. Epub 2016/05/28. https://doi.org/10.1038/srep26758 PMID: 27230293; PubMed Central PMCID: PMCPMC4882625.

77. Thornton C, Bright NJ, Sastre M, Muckett PJ, Carling D. AMP-activated protein kinase (AMPK) is a tau kinase, activated in response to amyloid beta-peptide exposure. Biochem J. 2011; 434(3):503–12. Epub 2011/01/06. https://doi.org/10.1042/BJ20101485 PMID: 21204788.

78. Sato S, Cerny RL, Buescher JL, Ikezu T. Tau-tubulin kinase 1 (TTBK1), a neuron-specific tau kinase candidate, is involved in tau phosphorylation and aggregation. J Neurochem. 2006; 98(5):1573–84. Epub 2006/08/23. https://doi.org/10.1111/j.1471-4159.2006.04059.x PMID: 16923168.

79. Lund H, Cowburn RF, Gustafsson E, Stromberg K, Svensson A, Dahllund L, et al. Tau-tubulin kinase 1 expression, phosphorylation and co-localization with phospho-Ser422 tau in the Alzheimer's disease brain. Brain Pathol. 2013; 23(4):378–89. Epub 2012/10/24. https://doi.org/10.1111/bpa.12001 PMID: 23088643.

80. Liachko NF, McMillan PJ, Strovas TJ, Loomis E, Greenup L, Murrell JR, et al. The tau tubulin kinases TTBK1/2 promote accumulation of pathological TDP-43. PLoS Genet. 2014; 10(12):e1004803. Epub 2014/12/05. https://doi.org/10.1371/journal.pgen.1004803 PMID: 25473830; PubMed Central PMCID: PMCPMC4256087.

81. Nag S, Yu L, Boyle PA, Leurgans SE, Bennett DA, Schneider JA. TDP-43 pathology in anterior temporal pole cortex in aging and Alzheimer's disease. Acta Neuropathol Commun. 2018; 6(1):33. Epub 2018/05/03. https://doi.org/10.1186/s40478-018-0531-3 PMID: 29716643; PubMed Central PMCID: PMCPMC5928580.

82. Grueninger F, Bohrmann B, Christensen K, Graf M, Roth D, Czech C. Novel screening cascade identifies MKK4 as key kinase regulating Tau phosphorylation at Ser422. Mol Cell Biochem. 2011; 357(1–2):199–207. Epub 2011/06/04. https://doi.org/10.1007/s11010-011-0890-6 PMID: 21638028.

83. Mazzitelli S, Xu P, Ferrer I, Davis RJ, Tournier C. The loss of c-Jun N-terminal protein kinase activity prevents the amyloidogenic cleavage of amyloid precursor protein and the formation of amyloid plaques in vivo. J Neurosci. 2011; 31(47):16969–76. Epub 2011/11/25. https://doi.org/10.1523/JNEUROSCI.4491-11.2011 PMID: 22114267; PubMed Central PMCID: PMCPMC6623849.

84. Li B, Chohan MO, Grundke-Iqbal I, Iqbal K. Disruption of microtubule network by Alzheimer abnormally hyperphosphorylated tau. Acta Neuropathol. 2007; 113(5):501–11. Epub 2007/03/21. https://doi.org/10.1007/s00401-007-0207-8 PMID: 17372746; PubMed Central PMCID: PMCPMC3191942.

85. Manczak M, Kandimalla R, Fry D, Sesaki H, Reddy PH. Protective effects of reduced dynamin-related protein 1 against amyloid beta-induced mitochondrial dysfunction and synaptic damage in Alzheimer's disease. Hum Mol Genet. 2016; 25(23):5148–66. Epub 2016/09/30. https://doi.org/10.1093/hmg/ddw330 PMID: 27677309; PubMed Central PMCID: PMCPMC6078633.

86. Kandimalla R, Reddy PH. Multiple faces of dynamin-related protein 1 and its role in Alzheimer's disease pathogenesis. Biochim Biophys Acta. 2016; 1862(4):814–28. Epub 2015/12/29. https://doi.org/10.1016/j.bbadis.2015.12.018 PMID: 26708942; PubMed Central PMCID: PMCPMC5343673.

87. Devine MJ, Kittler JT. Mitochondria at the neuronal presynapse in health and disease. Nat Rev Neurosci. 2018; 19(2):63–80. Epub 2018/01/20. https://doi.org/10.1038/nrn.2017.170 PMID: 29348666.

88. Gadhave K, Bolshette N, Ahire A, Pardeshi R, Thakur K, Trandafir C, et al. The ubiquitin proteasomal system: a potential target for the management of Alzheimer's disease. J Cell Mol Med. 2016; 20(7):1392–407. Epub 2016/03/31. https://doi.org/10.1111/jcmm.12817 PMID: 27028664; PubMed Central PMCID: PMCPMC4929298.

89. El Ayadi A, Stieren ES, Barral JM, Boehning D. Ubiquilin-1 regulates amyloid precursor protein maturation and degradation by stimulating K63-linked polyubiquitination of lysine 688. Proc Natl Acad Sci U S A. 2012; 109(33):13416–21. Epub 2012/08/01. https://doi.org/10.1073/pnas.1206786109 PMID: 22847417; PubMed Central PMCID: PMCPMC3421158.

90. Viswanathan J, Haapasalo A, Bottcher C, Miettinen R, Kurkinen KM, Lu A, et al. Alzheimer's disease-associated ubiquilin-1 regulates presenilin-1 accumulation and aggresome formation. Traffic. 2011; 12(3):330–48. Epub 2010/12/15. https://doi.org/10.1111/j.1600-0854.2010.01149.x PMID: 21143716; PubMed Central PMCID: PMCPMC3050036.

91. Adegoke OO, Qiao F, Liu Y, Longley K, Feng S, Wang H. Overexpression of Ubiquilin-1 Alleviates Alzheimer's Disease-Caused Cognitive and Motor Deficits and Reduces Amyloid-beta Accumulation in Mice. J Alzheimers Dis. 2017; 59(2):575–90. Epub 2017/06/10. https://doi.org/10.3233/JAD-170173 PMID: 28598849; PubMed Central PMCID: PMCPMC5791527.

92. Ostrovskaya O, Xie K, Masuho I, Fajardo-Serrano A, Lujan R, Wickman K, et al. RGS7/Gbeta5/R7BP complex regulates synaptic plasticity and memory by modulating hippocampal GABABR-GIRK signaling. Elife. 2014; 3:e02053. Epub 2014/04/24. https://doi.org/10.7554/eLife.02053 PMID: 24755289; PubMed Central PMCID: PMCPMC3988575.

93. Tan MG, Lee C, Lee JH, Francis PT, Williams RJ, Ramirez MJ, et al. Decreased rabphilin 3A immunoreactivity in Alzheimer's disease is associated with Abeta burden. Neurochem Int. 2014; 64:29–36. Epub 2013/11/10. https://doi.org/10.1016/j.neuint.2013.10.013 PMID: 24200817.

94. Gong Y, Lippa CF, Zhu J, Lin Q, Rosso AL. Disruption of glutamate receptors at Shank-postsynaptic platform in Alzheimer's disease. Brain Res. 2009; 1292:191–8. Epub 2009/07/29. https://doi.org/10.1016/j.brainres.2009.07.056 PMID: 19635471; PubMed Central PMCID: PMCPMC2745956.

95. Leuba G, Vernay A, Kraftsik R, Tardif E, Riederer BM, Savioz A. Pathological reorganization of NMDA receptors subunits and postsynaptic protein PSD-95 distribution in Alzheimer's disease. Curr Alzheimer Res. 2014; 11(1):86–96. Epub 2013/10/26. https://doi.org/10.2174/1567205011310660170 PMID: 24156266.

96. Briggs CA, Chakroborty S, Stutzmann GE. Emerging pathways driving early synaptic pathology in Alzheimer's disease. Biochem Biophys Res Commun. 2017; 483(4):988–97. Epub 2016/09/24. https://doi.org/10.1016/j.bbrc.2016.09.088 PMID: 27659710; PubMed Central PMCID: PMCPMC5303639.

97. Scheff SW, Price DA. Alzheimer's disease-related alterations in synaptic density: neocortex and hippocampus. J Alzheimers Dis. 2006; 9(3 Suppl):101–15. Epub 2006/08/18. https://doi.org/10.3233/jad-2006-9s312 PMID: 16914849.

98. Dickson DW, Crystal HA, Mattiace LA, Masur DM, Blau AD, Davies P, et al. Identification of normal and pathological aging in prospectively studied nondemented elderly humans. Neurobiol Aging. 1992; 13(1):179–89. Epub 1992/01/01. https://doi.org/10.1016/0197-4580(92)90027-u PMID: 1311804.

99. Murray ME, Dickson DW. Is pathological aging a successful resistance against amyloid-beta or preclinical Alzheimer's disease? Alzheimers Res Ther. 2014; 6(3):24. Epub 2014/07/18. https://doi.org/10.1186/alzrt254 PMID: 25031637; PubMed Central PMCID: PMCPMC4055017.

100. Felsky D, Roostaei T, Nho K, Risacher SL, Bradshaw EM, Petyuk V, et al. Neuropathological correlates and genetic architecture of microglial activation in elderly human brain. Nat Commun. 2019; 10 (1):409. Epub 2019/01/27. https://doi.org/10.1038/s41467-018-08279-3 PMID: 30679421; PubMed Central PMCID: PMCPMC6345810.

101. Hamilton JA, Achuthan A. Colony stimulating factors and myeloid cell biology in health and disease. Trends Immunol. 2013; 34(2):81–9. Epub 2012/09/25. https://doi.org/10.1016/j.it.2012.08.006 PMID: 23000011.

102. Spangenberg E, Severson PL, Hohsfield LA, Crapser J, Zhang J, Burton EA, et al. Sustained microglial depletion with CSF1R inhibitor impairs parenchymal plaque development in an Alzheimer's disease model. Nat Commun. 2019; 10(1):3758. Epub 2019/08/23. https://doi.org/10.1038/s41467-019-11674-z PMID: 31434879; PubMed Central PMCID: PMCPMC6704256.

103. Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, et al. Single-cell transcriptomic analysis of Alzheimer's disease. Nature. 2019; 570(7761):332–7. Epub 2019/05/03. https://doi.org/10.1038/s41586-019-1195-2 PMID: 31042697; PubMed Central PMCID: PMCPMC6865822.

104. Whalen S, Truty RM, Pollard KS. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. Nat Genet. 2016; 48(5):488–96. Epub 2016/04/12. https://doi.org/10.1038/ng.3539 PMID: 27064255; PubMed Central PMCID: PMCPMC4910881.

105. Klein HU, De Jager PL. Uncovering the Role of the Methylome in Dementia and Neurodegeneration. Trends Mol Med. 2016; 22(8):687–700. Epub 2016/07/18. https://doi.org/10.1016/j.molmed.2016.06.008 PMID: 27423266.

106. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489(7414):57–74. Epub 2012/09/08. https://doi.org/10.1038/nature11247 PMID: 22955616; PubMed Central PMCID: PMCPMC3439153.

107. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015; 518(7539):317–30. Epub 2015/02/20. https://doi.org/10.1038/nature14248 PMID: 25693563; PubMed Central PMCID: PMCPMC4530010.

108. Huang JK, Carlin DE, Yu MK, Zhang W, Kreisberg JF, Tamayo P, et al. Systematic Evaluation of Molecular Networks for Discovery of Disease Genes. Cell Syst. 2018; 6(4):484–95 e5. Epub 2018/04/02. https://doi.org/10.1016/j.cels.2018.03.001 PubMed Central PMCID: PMCPMC5920724. PMID: 29605183

109. Ibanez-Beroiz B, Librero-Lopez J, Peiro-Moreno S, Bernal-Delgado E. Shared component modelling as an alternative to assess geographical variations in medical practice: gender inequalities in hospital admissions for chronic diseases. BMC Med Res Methodol. 2011; 11:172. Epub 2011/12/23. https://doi.org/10.1186/1471-2288-11-172 PMID: 22188979; PubMed Central PMCID: PMCPMC3273448.

110. Tuncbag N, Gosline SJ, Kedaigle A, Soltis AR, Gitter A, Fraenkel E. Network-Based Interpretation of Diverse High-Throughput Datasets through the Omics Integrator Software Package. PLoS Comput Biol. 2016; 12(4):e1004879. Epub 2016/04/21. https://doi.org/10.1371/journal.pcbi.1004879 PMID: 27096930; PubMed Central PMCID: PMCPMC4838263.

111. Ginhoux F, Greter M, Leboeuf M, Nandi S, See P, Gokhan S, et al. Fate mapping analysis reveals that adult microglia derive from primitive macrophages. Science. 2010; 330(6005):841–5. Epub 2010/10/23. https://doi.org/10.1126/science.1194637 PMID: 20966214; PubMed Central PMCID: PMCPMC3719181.

112. Elmore MR, Najafi AR, Koike MA, Dagher NN, Spangenberg EE, Rice RA, et al. Colony-stimulating factor 1 receptor signaling is necessary for microglia viability, unmasking a microglia progenitor cell in the adult brain. Neuron. 2014; 82(2):380–97. Epub 2014/04/20. https://doi.org/10.1016/j.neuron.2014.02.040 PMID: 24742461; PubMed Central PMCID: PMCPMC4161285.

113. Mancuso R, Fryatt G, Cleal M, Obst J, Pipi E, Monzon-Sandoval J, et al. CSF1R inhibitor JNJ-40346527 attenuates microglial proliferation and neurodegeneration in P301S mice. Brain. 2019; 142(10):3243–64. Epub 2019/09/11. https://doi.org/10.1093/brain/awz241 PMID: 31504240.

114. Liddelow SA, Guttenplan KA, Clarke LE, Bennett FC, Bohlen CJ, Schirmer L, et al. Neurotoxic reactive astrocytes are induced by activated microglia. Nature. 2017; 541(7638):481–7. Epub 2017/01/19. https://doi.org/10.1038/nature21029 PMID: 28099414; PubMed Central PMCID: PMCPMC5404890.