Title
Identification of Shared Molecular Pathways and Networks between Alzheimer's Disease and Type 2 Diabetes

Author
Wei, Katherine Wei

Publication Date
2016

UNIVERSITY OF CALIFORNIA

Los Angeles

Identification of Shared Molecular Pathways and Networks between Alzheimer's Disease and Type 2 Diabetes

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Science

in Physiological Science

by

Katherine Jung Wei

2016

ABSTRACT OF THE THESIS

Identification of Shared Molecular Pathways and Networks between Alzheimer's Disease and Type 2 Diabetes

by

Katherine Jung Wei

Master of Science in Physiological Science

University of California, Los Angeles, 2016

Professor Xia Yang, Chair

We hypothesize that Alzheimer's disease (AD) and Type 2 Diabetes (T2D) share genetically perturbed molecular pathways, with T2D also inducing biological processes upstream of AD that promote AD development. Our study employs a systems biology approach integrating human genetic association studies, gene expression profiling studies, biological pathways, and tissue-specific gene networks to investigate the mechanistic links between the two diseases. Our approach has identified tissue-specific gene subnetworks from adipose, brain, liver, and skeletal muscle tissues enriched for both AD and T2D genetic signals. These subnetworks are involved in immune regulation, cell cycle processes, oxidative phosphorylation, extracellular matrix

function, and keratan sulfate degradation and biosynthesis. These biological annotations are of particular interest as they are consistently identified across multiple tissues. The identification of known and novel pathways and genes testifies to the power of our systems-wide approach to identifying causal mechanisms and genes shared between metabolic and neurodegenerative diseases.

The thesis of Katherine Jung Wei is approved.

Fernando Gomez-Pinilla

Aldons J. Lusis

Xia Yang, Committee Chair

University of California, Los Angeles

2016

Table of Contents

Introduction

Overview of Alzheimer's disease and Type 2 Diabetes

Type 2 diabetes mellitus (T2D) and Alzheimer's disease (AD) are two of the most prevalent and debilitating diseases among aging individuals. T2D is a common metabolic disease featuring hyperglycemia and insulin resistance, whereas AD is a neurodegenerative disorder mainly characterized by cognitive decline and amyloid plaque accumulation. In the United States, approximately 5.4 million individuals are affected by AD and 25.8 million individuals are affected by T2D[1]. Among individuals 65 and over, the incidence of diabetes is 26.9%, compared to an incidence of 8.3% in the overall population[1]. Both AD and T2D are among the top leading causes of death in the United States.

Epidemiological, biochemical, molecular, and clinical supporting evidence connecting AD and T2D

An increasing volume of epidemiological evidence suggests that a link exists between T2D and AD. Individuals with diabetes, particularly T2D, have lower cognitive function and are at twice the risk for developing AD compared to those without diabetes[2]. A study employing a neuropsychological evaluation known as the Minimum Cognitive Examination revealed that having diabetes contributed significantly to the negative impact of age on cognitive ability, particularly in areas of global cognition and executive function[3]. In another meta-analysis involving over 67,000 participants, onset of T2D was associated with an increased risk of dementia[4].

At the biochemical and molecular levels, AD and T2D also share multiple pathological characteristics. One of these shared hallmarks between the two diseases is insulin resistance. In the case of AD, this characteristic insulin resistance occurs in the neurons. Insulin is also

speculated to regulate neuronal connectivity and activity throughout the brain, and deficiency of or resistance to insulin could potentially impair transmission of information between neurons, leading to decline in cognitive ability[5; 6]. A second feature that the two pathologies have in common is the presence of amyloid β plaques and tau protein deposits. Pancreatic islet amyloidosis, or the presence of amyloid β plaque and tau deposits, is a key feature of T2D, and shares key structural features with the neurofibrillary tangles in amyloid plaque found in the neurons of AD patients.[7] Another recent study also found that doubling the blood glucose levels of young mice without amyloid deposits in their brains raised the amount of amyloid deposits by 20%. In older mice that had already developed amyloid plaques in the brain, doubling blood glucose levels caused a 40% increase in amyloid deposits[8]. These findings suggest that having chronically high blood glucose levels could potentially promote amyloid β production, which eventually aggregates into plaques and contribute to the development of AD. A third shared molecular feature between the two diseases is the sustained activation of inflammatory processes and presence of cellular oxidative stress[9; 10].

In addition to the above epidemiological and molecular studies that clearly demonstrate the co-occurrence and shared pathological features between the two diseases, experimental and clinical studies also support shared therapeutic strategies. Several medications commonly prescribed for the treatment of diabetes, such as metformin, intranasal insulin administration, PPARγ inhibitors, and GLP-1 activators, have been shown to be useful in the attenuation of neurodegenerative symptoms characteristic of AD[11; 12]. A study of metformin, one of the most common drugs used to treat T2D, demonstrated the neuroprotective effects of metformin on lessening the progression of AD-like changes in the brains of obese, leptin-resistant mice. Metformin was shown to attenuate the increase of total accumulated tau protein, phosphorylated

tau, and activated JNK (N-terminal kinase) that is responsible for tau phosphorylation, polymerization and neuronal dysfunction and death[13]. In a meta-analysis involving over 67,000 participants, the overall dementia risk decreased when participants consistently took antidiabetic medications sulfonylureas and metformin[4]. Furthermore, a study including 60 adults who were diagnosed with either mild cognitive impairment or mild to moderate Alzheimer's disease showed that participants receiving 40 IU (International Units) of intranasal insulin spray demonstrated significant improvements in working memory compared to those receiving 20 IU or a placebo[12]. These findings strongly substantiate the causal and mechanistic conjectures surrounding the pathological relatedness of AD and T2D.

To tackle the underlying molecular connections between T2D and AD, candidate gene and pathway studies have revealed the potential role of insulin signaling, growth factor, protein misfolding, inflammation, and apoptosis triggering pathways[14]. However, whether these pathways play causal roles in both diseases is unclear and a comprehensive understanding of the shared causal mechanisms has not been achieved.

Integrative genomics study overview

To systematically investigate the causal molecular connections between the two diseases, in this study we conduct a data-driven, systems genomics analysis by integrating genetic data from human genome-wide association studies (GWAS) of AD and T2D, functional genomics data from tissue-specific expression quantitative trait loci (eQTLs) and Encyclopedia of DNA Elements (ENCODE) studies, tissue-specific transcriptome profiling, canonical pathways, and molecular networks to provide a systems-level understanding of AD and T2D as well as their connections. Specifically, we address whether T2D and AD share genetically driven pathogenic

pathways (i.e., common genetic factors drive both T2D and AD), occur in a sequential manner (i.e., T2D onset drives molecular changes that in turn induce AD), or both scenarios occur to some extent. A better understanding of the mechanistic connections between T2D and AD will unravel novel therapeutic targets and open new avenues for the treatment of these prevalent and debilitating diseases.

Materials and Methods

Overall Study Design

As illustrated in **Figure 1**, we first delineated the molecular pathways and gene networks that are perturbed by genetic risks of AD and T2D separately. We then compared the genetically perturbed pathways and networks between T2D and AD to derive shared genetic mechanisms that may be causally linked to both diseases. To test whether molecular pathways downstream of T2D (i.e., reactive to and non-causal for T2D) also pose predisposition to AD development, we extracted tissue-specific transcriptomic profiles of T2D and compared the T2D gene signatures with those tested causal for AD from the genetic analysis.

Computational Pipeline for Genomic Data Integration

We have previously developed a computational pipeline, Mergeomics, to integrate a diverse array of disease-related datasets, including disease GWAS, eQTL studies, ENCODE studies, knowledge-based canonical pathways, data-driven co-expression networks, and gene regulatory networks[15]. Systematic data integration enables identification of disease key drivers and biological pathways that are perturbed in disease pathogenesis. The overall framework for our study can be divided into four key steps: 1) SNP (single nucleotide polymorphism) Set

Enrichment Analysis (SSEA), 2) merging of overlapping pathways or gene sets into independent supersets, 3) Key Driver Analysis (KDA) to identify potential regulators of the disease-associated supersets, and 4) gene network visualization, as depicted in **Figure 2**.

AD and T2D GWAS datasets

The AD GWAS dataset was from the International Genomics of Alzheimer's Disease (IGAP) consortium. The IGAP GWAS dataset includes over 7 million genotyped and imputed SNPs from 17,008 Alzheimer's disease cases and 37,154 controls of European ancestry[16]. The dataset excluded any SNPs with call rates of less than 95%, and the meta-analysis included only SNPs genotyped or successfully imputed based on 1000G in at least 40% of AD cases and 40% of the controls.

The T2D GWAS dataset was from the Diabetes Genetics Replication and Meta-analysis (DIAGRAM) consortium. The DIAGRAM GWAS dataset was a meta-analysis of 12,171 cases and 56,862 controls of primarily European descent[17]. The DIAGRAM consortium comprises datasets from approximately 40 groups investigating T2D genetics, including studies from groups such as WTCCC (Wellcome Trust Case Control Consortium), DGI (Diabetes Genetics Initiative), FUSION (Finland-United States Investigation of NIDDM Genetics), DGDG (Diabetes Gene Discovery Group), KORA, Rotterdam, DeCODE (Diabetes epidemiology: collaborative analysis of diagnostic criteria in Europe), EUROSPAN, Framingham, ARIC (Atherosclerosis Risk in Communities), and NHS (Nurses' Health Study). Large-scale genotyping was performed using the Metabochip, a custom array consisting of 196,725 variants[17] and was imputed to up to 2.5 million autosomal SNPs[17].

Biological pathways

A total of 1825 metabolic, biochemical, and signaling pathways were curated from Reactome[18], BioCarta (www.biocarta.com), and KEGG (Kyoto Encyclopedia of Genes and Genomes)[19]. databases. Reactome is an open-source publicly available, manually curated biological database which includes pathway annotations from UCSC and HapMap Genome Browsers, KEGG Compound and ChEBI small molecule databases, PubMed, and Gene Ontology[18]. The main entities, such as genes, RNAs, proteins, complexes, and small molecules, that are involved in the same network of biological interactions are grouped together and given the same pathway annotation. BioCarta is a community-fed database that catalogues the biological interactions of signal transduction pathways, and involves over 120,000 genes from multiple species and proteomic and genomic information, as well as canonical biological pathways and suggestions for new pathways. KEGG is an aggregation of databases involved with genomes, biological pathways, diseases, drugs, and chemical substances often used in bioinformatics studies, various omics studies, and translational drug research[19].

Gene coexpression networks

Tissue-specific coexpression networks of adipose, blood, brain, heart, islet cells, kidney, liver, and muscle tissues were constructed from data obtained from independent, publically available transcriptomics studies from human and mouse (Table 1). We also included human brain region-specific coexpression networks using the data from the Kang et al study[20]. This dataset included transcriptome data for 16 brain regions: the orbital prefrontal cortex (OFC), dorsolateral prefrontal cortex (DFC), ventrolateral prefrontal cortex (VFC), medial prefrontal cortex (MFC), primary motor cortex (M1C), primary somatosensory cortex (S1C), posterior

inferior parietal cortex (IPC), primary auditory cortex (A1C), posterior superior temporal cortex (STC), inferior temporal cortex (ITC), primary visual cortex (V1C), hippocampus (HIP), amygdala (AMY), striatum (STR), mediodorsal nucleus of the thalamus (MD), and cerebellar cortex (CBC). The brain samples were taken from clinically unremarkable males and females of multiple ethnicities from a range of 15 different developmental periods, with period 1 being the developing embryonic brain and period 15 being the late adulthood brain (older than 60 years of age)[20].

We used the Weighted Gene Co-expression Network Analysis (WGCNA) R package to construct coexpression networks from each transcriptome dataset [21]. The overall method for clustering genes into coexpression modules involves performance of a pairwise correlation to determine the correlation between different genes. Genes that not only show high pair-wise correlation (either positively or negatively) but also share similar correlation patterns with the rest of the transcriptome were clustered into the same module in the co-expression network. Gene co-expression networks provide useful biological information as co-regulated genes are typically either functionally related, regulated by the same transcriptional mechanism, or belong to members of the same biological pathway[22]. The dynamic method for branch cutting of the clustering dendrogram was used to define individual modules in the gene co-expression networks. A minimum module size of 30 was used in the hierarchical clustering process. The lowest power at which the scale-free topology fit index reaches a height of 0.9 was used to generate the gene modules and the clustering dendrogram of genes, with each module assigned a unique module color[23]. A total of 3020 coexpression modules, including 339 brain coexpression modules, were constructed using WGCNA. These data-driven co-expression modules, along

with knowledge-based canonical pathways described in the previous section, were then used in downstream analysis.

Functional Genomics Data

As GWAS SNPs frequently fall in non-coding regions, many disease-associated gene variants may be responsible for gene regulation rather than altering protein functions. Thus, tissue-specific eQTL mapping as well as regulatory element information may provide functional support to connect genetic variants with altered gene expression and ultimately with disease susceptibility. Human expression SNPs (eSNPs) under eQTLs from adipose, brain, liver, lymphoblastoid, blood, artery, pituitary, lymphocytes, sigmoid colon, transverse colon, esophagus mucosa, esophagus Muscularis, small intestine, stomach, spleen, and skeletal muscle tissue were curated from GTEx (Genotype-Tissue Expression project)[24] and additional eQTL studies . Additionally, functional information in the RegulomeDB database[25] based on the ENCODE (Encyclopedia of DNA Elements) project[26] were collected. These various types of functional information were used to perform the gene to GWAS SNP mapping.

SNP Set Enrichment Analysis (SSEA)

The GWAS, network, pathway, and functional genomics datasets described above were used in SSEA to determine if any of the modules from our data-drive co-expression networks or knowledge-driven biological pathways demonstrate significant genetic association with AD and T2D (**Figure 2**). SSEA comprises of 4 general steps. First, gene sets from knowledge-driven pathways and data-driven co-expression modules are collected. Second, the gene sets are converted to SNP sets with functional support according to tissue-specific eQTL and ENCODE

studies, or without functional support based on chromosomal distance mapping, or both. Third, P-values from T2D and AD GWAS are extracted for each mapped SNP to provide associations between the SNPs and the respective diseases. Finally, the GWAS P-values within the SNP sets representing the pathways or modules are compared against permutated sets of random genes in an enrichment analysis to derive pathways and network modules enriched for T2D and AD genetic signals.

Before performing the SSEA, we preprocessed the GWAS datasets for AD and T2D to prevent linkage disequilibrium (LD) structure in the SNPs from producing spurious results. To adjust for LD structure, we performed SNP pruning using a LD cutoff of $r^2 < 0.7$. The LD information was obtained from HapMap for the CEU population, as the majority of study participants in the AD and T2D GWAS are of European descent.

After LD pruning of the DIAGRAM and IGAP GWAS datasets for T2D and AD, we tested for enrichment of T2D and AD associated SNPs in the tissue-specific co-expression networks and knowledge-driven canonical pathways using SSEA. When using eSNPs from the tissue-specific eQTL studies to map the GWAS SNPs to genes, we focused on T2D and AD-relevant tissues including adipose, blood (including lymphoblastoid), brain, liver, islet, and skeletal muscle. We also used regulome-based and chromosomal distance-based mapping methods as well as a pooled set of eSNPs from all tissues and cell types curated ("All eSNPs"), and a combined set including all eQTL-based, distance-based, and RegulomeDB-based mapping methods ("Combined"). Each set of SNPs mapped to a coexpression module or pathway based on a mapping method was then compared to a background of SNPs from sets of random genes to determine whether a particular pathway or module was enriched for disease-associated SNPs. After performing the SSEA using both GWAS datasets, we adjusted the enrichment p values by

estimating false discovery rate (FDR) using the Benjamini-Hochberg method. Significant modules at FDR<25% and suggestive modules at $p<0.05$ were selected for T2D and AD, and then compared between the two diseases.

Gene Module Merging, Construction of Supersets, and Second Round of SSEA

There are redundancies among the pathways and coexpression modules in terms of gene and functional overlaps. To reduce the redundancy, we merged the pathways and modules with significant sharing (>20%) of member genes into supersets. To prevent very large supersets that may lose functional specificity, we restricted the gene number in each superset to include only the most consistent core genes shared between pathways or modules. After merging the shared pathways or modules between T2D and AD into supersets, a second round of SSEA was performed on these supersets to confirm that they still retained the significant association to both diseases as demonstrated by their constituent pathways or modules. The supersets were functionally annotated using KEGG and Reactome pathway databases.

Key Driver Analysis (KDA) of the Shared Supersets

To identify potential key regulatory genes that contribute to both T2D and AD, we performed KDA on the significant supersets shared between diseases. Bayesian networks, which illustrate detailed gene-gene interactions, for adipose, brain, liver, and skeletal muscle tissue were used for the KDA as these tissues are biologically relevant to both T2D and AD pathologies and they were found to be informative for both diseases in our SSEA analysis. To perform the KDA, the member genes of the shared supersets are overlaid over the Bayesian gene networks to determine the key driver genes whose network neighborhoods are enriched for AD-T2D superset

genes based on the network topology. Enrichment was assessed with Fisher's Exact Test and Bonferroni-corrected $p<0.05$ was used to determine the significance of key drivers. After identifying the key drivers from the supersets in this manner, we focused on the top 5 statistically significant key drivers identified from each tissue. The regulatory subnetworks for each key driver were retrieved from each tissue-specific Bayesian network and visualized using Cytoscape[27].

Identification of Differentially Expressed Genes in T2D to Define Tissue-specific T2D signatures

To test whether biological pathways downstream of T2D may drive AD development, we analyzed tissue-specific T2D transcriptome profiles to identify genes that are differentially expressed between T2D cases and controls. We extracted transcriptome profiles from T2D-relevant tissues including adipose, brain, islet, liver, and muscle tissues from the Gene Expression Omnibus (GEO). We included only datasets generated using common microarray platforms (Illumina, Affymetrix, and Agilent) that examined both T2D and non-T2D subjects with $n>= 3$ samples in each group, and excluded datasets in which other comorbidities were involved (such as heart failure and diabetic kidney disease). Studies from human, mouse, and rat were used. Of the datasets we analyzed, 10, 2, 5, 2, and 11 were from adipose tissue, brain, islet, liver, and muscle, respectively.

To identify differentially expressed genes between T2D and non-T2D subjects, we selected the genes that exhibited at least a 2 fold change in expression levels between T2D cases and non-T2D controls in at least 1/3 of the studies for each tissue, or both studies for liver and brain since only two datasets were included for each of these tissues. The differentially expressed

genes in each tissue were defined as tissue-specific T2D transcriptome signatures. These signatures were compared with the pathways and supersets associated with AD, T2D, or both, derived from the GWAS (genetics)-based SSEA analysis. To determine the significance of overlap between T2D signatures and the genetically associated AD/T2D gene sets, we used a Fisher's exact test and Bonferroni-corrected p-value cutoff of 0.05.

Results

Pathways and coexpression modules enriched for AD GWAS signals

At FDR<25%, a total of 53 unique pathways or coexpression modules out of the 4,845 sets tested were found to be enriched for AD GWAS signals. A majority (48) of the significant signals were derived when chromosomal distance was used as the SNP-gene mapping method, whereas the tissue eQTLs and ENCODE information from RegulomeDB did not appear to be informative for AD GWAS interpretation. A likely explanation is that the genetic risks of AD affect gene functions that are poorly captured in the functional genome information that is currently available.

As summarized in **Table 1**, the AD-associated pathways or modules are involved in metabolic processes (ascorbate and aldarate metabolism, aminoacyl tRNA biosynthesis, metabolism and transport of lipids and lipoproteins, metabolism of vitamins and cofactors), immune system (chemokine signaling pathways, cytokine-cytokine receptor interactions, interferon signaling, intestinal immune network for IGA production, antigen processing and cross presentation, complement pathways, primary immunodeficiency, FC gamma R mediated phagocytosis), cell communication (extracellular matrix (ECM) glycoproteins, adherens

junctions, and cell-cell junction organization), signaling pathways (NGF signaling, calcineurin signaling pathway), hematopoietic cell lineage, and development.

Pathways and coexpression modules that are significantly enriched for T2D GWAS signals

At FDR<25%, a total of 176 unique pathways and coexpression modules out of the 4,845 sets used in our analysis showed significant enrichment for T2D GWAS signals. In contrast to AD, tissue-specific eQTLs were found to be highly informative for T2D GWAS signals: 34, 45, 68, and 27 modules were detected when using eQTLs from blood, liver, skeletal muscle, and all eSNPs, respectively. In addition, 103 modules were identified when functional information from the ENCODE-based RegulomeDB was used as the SNP-gene mapping method. Distance-based mapping method and the combined mapping method revealed 34 and 42 modules, respectively.

As summarized in **Table 2**, these significant biological pathways and coexpression modules are related to signal transduction (G alpha signaling, JNK-signaling, platelet derived growth factor (PDGF) receptor beta signaling, PI3K cascade, MAPK signaling, olfactory transduction signaling, calcium signaling pathways), metabolic pathways (type 2 diabetes, steroid hormone biosynthesis, amino sugar and nucleotide sugar metabolism, cholesterol biosynthesis, PPAR signaling, triglyceride biosynthesis, lipid and lipoprotein metabolism, integration of energy metabolism, respiratory electron transport and ATP synthesis, mitochondrial protein import, oxidative phosphorylation, drug metabolism involving cytochrome P450), immune system (cytokine receptor interactions, complement pathways, chemokine signaling pathways, interferon signaling), neuronal processes (axon guidance, serotonin receptors), cell cycle regulation (apoptosis pathways, p53 downstream pathways), translation (deadenylation dependent mRNA decay, ribosome function), coagulation (fibrin clotting

cascades, platelet activation and aggregation), and cell-cell communication (ECM glycoproteins, focal adhesion). Based on these broad functional categories, metabolic pathways, immune system, and cell communication are related to both AD and T2D. Interestingly, a pathway related to Alzheimer's disease was also found among the T2D-associated modules detected by our SSEA.

Significant pathways and coexpression modules shared between AD and T2D

A total of 16 coexpression modules directly overlapped between the AD and T2D analyses. These shared modules are involved in metabolic pathways (metabolism of lipids and lipoproteins, metabolism of xenobiotics by cytochrome P450, immune system (cytokine-cytokine receptor interaction, chemokine signaling pathway, IL12 family signaling pathways, chemokine receptors, interferon signaling, antigen processing and cross-presentation, complement pathway), cell communication (core matrisome function, ECM glycoproteins, cell adhesion molecules), and ribosome function.

Suggestive pathways and coexpression modules shared between AD and T2D

We noted that many of the previously implicated processes such as apoptotic pathways and insulin signaling pathways were not among the top significant pathways for AD and T2D at FDR<25%. However, they appeared to show enrichment signals at a nominal p-value of $p < 0.05$ for both AD and T2D. We reason that these weaker pathways may still be of biological significance given the ample supporting evidence in the literature. Therefore, we further extracted the shared pathways and modules at a nominal p value cutoff of $p<0.05$, and defined 72 suggestive gene sets as shared between T2D and AD. These gene sets are characterized by

annotation such as apoptosis pathways, regulation of insulin-like growth factors, WNT signaling, glycosphingolipid metabolism, actin cytoskeleton regulation, respiratory electron transport, glycan degradation.

Merging of the overlapping T2D and AD-associated gene sets into non-overlapping supersets

Since the pathway and gene modules come from a variety of sources, it is possible that some of these gene sets overlap in member genes as well as functional annotation. To account for this redundancy, we merge the shared AD-T2D gene sets (including both significant and suggestive sets discovered in the above analyses) with >30% member gene overlap into 42 relatively independent supersets. We confirmed that 39 of these supersets retained significant enrichment for both AD and T2D GWAS signals at a nominal p-value of $p<0.05$ and 1 superset retained significance at a Bonferroni-corrected $p<0.05$ in the round of SSEA.

As summarized in **Table 3**, the shared supersets were associated with processes such as metabolism (insulin like growth factor activity regulation, cholesterol biosynthesis, PPAR signaling), immune regulation, cell cycle regulation and apoptosis, transport, ECM-related functions, autophagy, and cell communication or cell signaling (MAPK signaling, JAK STAT signaling).

Identification of key network drivers for the shared supersets between AD and T2D

To pinpoint the most influential regulatory genes in the AD-T2D shared supersets, we used tissue-specific Bayesian network models from adipose, brain, liver, and skeletal muscle tissues. We mapped our T2D and AD-associated supersets to determine key driver genes based on the Bayesian network topology. A key driver was defined as a gene that has an over-

representation of genes in the T2D and AD-associated supersets in its network neighborhood. We identified 51, 18, 20, and 32 key drivers from the adipose, brain, liver, and muscle networks. The subnetworks for the top 5 key drivers for each tissue were visualized using Cytoscape (**shown in Figures 3, 4, 5, and 6** for adipose, brain, liver, and muscle subnetworks, respectively).

The adipose subnetwork (**Figure 3**) connected immune system functions including interferon and cytokine signaling with FC gamma R mediated phagocytosis and keratan sulfate biosynthesis and degradation via key drivers *AIF1*, *GPM6B*, *RTP4*, *OAS2*, and *IFI44*. The brain subnetwork (**Figure 4**), orchestrated by key drivers *SERPING1*, *CYP1B1*, *OMD*, *CTGF*, and *ADORA2A*, mainly regulates core matrisome and proteoglycan production, G alpha S signaling events, nucleotide-like purinergic receptor function, and keratan sulfate degradation and biosynthesis. The liver subnetwork (**Figure 5**) centered at key drivers *C1QC*, *CSF1R*, *HLA-F*, *IFIT1*, and *LILRB3* primarily represents immune system regulation including the complement pathway, endosomal vacuolar pathways, antigen processing and presentation, FC gamma R mediated phagocytosis, natural killer cell-mediated cytotoxicity, interferon signaling, and interactions between lymphoid and non-lymphoid cells. The muscle subnetwork (**Figure 6**) containing key drivers *HLA-DRB1*, *MMT00074772*, *PCOLCE2*, *PLAC8*, and *POSTN* is enriched for genes involved in antigen processing and presentation, allograft rejection, ECM, and chemokine signaling pathways.

Identification of pathways downstream of T2D and upstream of AD

To test the hypothesis that some downstream pathways perturbed after the onset of T2D can be causal for AD, we collected T2D transcriptome signatures from different tissues and obtained 2956, 5471, 2793, 596, and 1547 differentially expressed genes as T2D transcriptome signatures in adipose tissue, brain, islet, liver, and muscle, respectively. Of the supersets shared between T2D and AD in the genetic analysis, 25 supersets significantly overlapped with the T2D transcriptome signatures at a Bonferroni-corrected p-value < 0.05 (**Table 5**). In addition, 27 (42%) AD modules and 70 (45%) T2D modules from the genetic analysis significantly overlapped with the T2D transcriptome signatures. These results support that transcriptome profiling can capture causal pathways detected from genetic studies. The biological pathways and coexpression modules that are genetically associated with AD and significantly overlap with T2D transcriptome signatures are listed in **Table 6**. As shown in **Figure 7**, specific overlaps exist between gene sets that are genetically associated with AD or T2D, and between these genes sets and T2D transcriptome signatures. In particular, the pathways that are downstream of T2D, as evidenced by significant perturbation at the transcriptome level but lack of evidence for genetic perturbation in T2D GWAS, but are putatively causal for AD based on AD GWAS, include annotations such as toll endogenous pathway, T-cell receptor (TCR) pathway, hematopoietic cell lineage, phagocytosis, NGF signaling, and calcineurin signaling.

Discussion

To better understand the shared the molecular mechanisms between AD and T2D, we conducted a highly integrative analyses leveraging a multitude of genetic and genomic data. The aggregation of diverse types of data allowed us to identify diverse biological processes common to both diseases and to support a causal role of T2D in AD development. Our systems-level

investigation revealed both previously implicated pathways such as antigen processing and presentation, interferon signaling, overall immune regulation and adaptive immune function, PPAR signaling, and IGF activity regulation, as well as novel processes such as oxidative phosphorylation and respiratory chain function, cell cycle processes, MAPK signaling, autophagy regulation, cell adhesion molecules, and extracellular matrix-related functions in the pathogenesis of both diseases. Our network analysis revealed tissue-specific key regulators of these shared pathways, which may serve as effective targets to mitigate these common debilitating diseases.

Known Biological Pathways Involved in AD and T2D

Among the shared pathways identified through our integrative study, immune and inflammatory signaling are one of the most well characterized similarities between AD and T2D pathologies. Both AD and T2D are diseases associated with a chronic inflammatory state. Obesity and metabolic dysregulation are strongly linked to insulin resistance, hypertension, dyslipidemia, and increased levels of pro-inflammatory adipokines released by adipose tissues. Type 2 diabetes is known to be strongly associated with obesity, as well as highly correlated with high circulating levels of inflammatory compounds, particularly CRP (C-reactive protein) and IL-6 (interleukin-6)[28]. AD development is also characterized by a state of systemic inflammation, leading to an increase of pro-inflammatory factors as well. This chronic inflammatory state can reduce the ability of the body to clear beta amyloid plaque buildup, leading to neuronal damage and further increasing the levels of inflammatory mediators in circulation. Ultimately, when inflammatory signaling is dysregulated in various organs or tissues, this can lead to progression of a variety of diseases, from T2D and AD to cancer and cardiovascular disease. Because

inflammatory factors and other immune regulators are released into the bloodstream, immune dysregulation can potentially have a systems-wide effect and can manifest as the nexus of a host of complex diseases and comorbidities. As multiple shared gene sets were found to be enriched with immune signals in our data-driven genetic analyses, our findings not only provide further support for the previously discovered immune connection between AD and T2D, but poses immune signaling as causal mechanisms for both diseases based on genetic evidence.

An increasing volume of evidence demonstrates that insulin signaling abnormalities and insulin resistance, the hallmark of T2D, can play a role in AD as well. Insulin receptors are selectively expressed in certain brain regions, including areas that are involved in memory formation[29]. Studies suggest that insulin may be involved in the processes of memory function, as well as play roles in regulation of amyloid precursor protein and beta-amyloid levels in the brain[29]. It has also been conjectured that insulin plays a role in the transport of beta-amyloid proteins and obstructs its degradation and eventual clearance from the brain[29]. Insulin-sensitizing actions throughout the body seem to be modulated by PPAR-gamma (peroxisome proliferator-activated receptor gamma), a nuclear receptor that plays essential roles in adipocyte differentiation, as well as other developmental and metabolic processes throughout the body. PPAR agonists, a common medication used to treat T2D, have also been shown to exert neuroprotective effects as well and slow both the onset and the progression of AD[30]. It is thought that PPAR gamma alleviates the symptoms and slows the progression of AD by downregulating inflammatory processes that occur in the AD brain[31]. With multiple studies pointing to the neuroprotective effects and insulin-sensitizing actions of PPAR agonists, our study further point to a causal role of PPAR signaling in driving both T2D and AD pathologies.

Impaired IGF (insulin like growth factor) signaling and impaired insulin receptor function are pathophysiological conditions that characterize both T2D and AD. As discussed earlier, the impairment of IGF-1 and proper insulin signaling prevents the degradation of beta-amyloid protein aggregations in the brain[32]. In the case of insulin resistance, insulin competes with beta-amyloid for the insulin-degrading enzyme as well, leading to reduced beta-amyloid clearance and sustained hyperinsulinemia conditions in the brain as well[32]. Insulin signaling appears to be central to the development of both diseases, both according to preexisting literature and to our data-driven analysis.

Novel Biological Pathways Involved in AD and T2D

Our data-driven analysis also uncovered novel processes shared between AD and T2D such as oxidative phosphorylation and respiratory chain function, cell cycle processes, MAPK signaling, autophagy regulation, cell adhesion molecules, and extracellular matrix-related functions. While these biological pathways have not yet been identified or well-characterized in the literature as processes connected to both AD and T2D, they are biologically plausible and warrant future study.

A potential novel connection between AD and T2D as identified by our study is the involvement of mitochondrial respiratory dysfunction in both diseases, as well as many other complex aging-related diseases[33]. Mitochondrial dysfunction plays a largely mechanistic role in aging and the development of complex diseases, mainly in the generation of reactive oxygen species and subsequent oxidative stress and metabolic dysregulation that ensues[33]. Mitochondrial malfunction has also been shown to impair autophagy, a process in which damaged or aged cellular components are degraded, which further contributes to the aging process and the

development of age-related diseases[33]. Furthermore, the frequency of mutations in mitochondrial DNA may be as high as 1/5000 in the adult population[34], pointing to possible genetic factors at play, particularly at the mitochondrial function loci, in the process of complex aging-related disease development. As T2D and AD both affect considerably large proportions of the aging population, our integrative genomics pipeline seems to accurately capture this aspect shared by both diseases, and may provide a causal explanation for this similarity between AD and T2D.

Cell cycle processes were also found to be pathways associated with both T2D and AD in our analysis. Pathogenic aggregation of amyloidogenic proteins as a result of protein misfolding characterize both AD and T2D[35]. Beta-amyloid peptide aggregations lead to eventual degradation of neural tissue and loss of synaptic connections in AD, while islet amyloid polypeptide (IAPP) is one of the major products secreted by pancreatic β cells and major contributing factors to the destruction of islet cells in T2D[35]. The tumor suppressor p53 is involved with cell cycle and apoptosis regulation, and is a known amyloidogenic protein that promotes the formation of amyloid oligomers and fibrils in the event of misfolding[35]. While p53 dysregulation is mostly implicated in the pathophysiology of cancers, studies have also shown that p53 is involved in both AD and T2D as well, specifically in the upregulation of its apoptotic activity in the brain and pancreatic β cells respectively[36; 37]. According to our results, apoptotic pathways were also shared between T2D and AD. These convergent findings of upregulation of p53 activity in both T2D and AD pathologies as well as the amyloidogenicity of p53 in the event of protein misfolding poses an intriguing causal mechanism in both diseases. While AD and T2D disease conditions are characterized by increased p53 activity as a result of amyloid protein

aggregation in brain and pancreatic tissue, it seems that p53 itself could be involved in further contributing to the pathogenesis due to its amyloidogenicity.

The insulin signaling involves a complex protein cascade, which includes MAPK (mitogen activation protein kinase)[38]. MAPK is associated with most of the metabolic effects of insulin, including uptake of glucose and amino acids, inhibition of fat breakdown, and the synthesis of fatty acids in the liver. Interestingly, the physiological effects of insulin in the brain are different from its metabolic effects in the periphery. Instead of regulating glucose uptake, the central effects of insulin include altered neurotransmitter release, neuronal growth, tubulin activity, and synaptic plasticity[39]. In this way, insulin activity can adversely affect neuronal function, primarily by inducing susceptibility to various stress effects[39]. Thus, chronic hyperinsulinemia, a hallmark of T2D, appears to play pathogenic roles both in T2D and AD, with insulin exerting different central and peripheral effects that are consistent with both disease conditions. MAPK signaling pathways that were found in our analysis could be related to the dysregulated insulin signaling that occurs in both disorders and suggest an involvement of specific aspects of the insulin signaling cascade as causal mechanisms of both diseases as supported by genetic evidence.

Autophagy, the catabolic process responsible for the breakdown of damaged or aged cellular components, has been shown to exhibit impaired efficiency in affected tissues of both T2D and AD[40]. Autophagy regulation is dependent on age, as well as AMPK (AMP activated protein kinase) and mTOR (mammalian target of rapamycin) signaling pathways[40]. Medications such as metformin, the most commonly prescribed T2D treatment, are also AMPK- activators and are known to diminish the symptoms of T2D and other age-related metabolic diseases[40]. Furthermore, additional studies also demonstrate that alterations in the autophagy pathway also

cause synaptic damage in both AD and T2D[41]. These findings from existing literature are consistent with our data-driven results, and again suggest that autophagy pathway genes could potentially act as susceptibility loci for both T2D and AD.

As one of the most consistent findings across different analyses carried out in our study, ECM-related pathways were captured as multiple gene sets and tissue-specific networks shared between AD and T2D. Although little literature exists to highlight ECM as a shared mechanism between AD and T2D, the importance of ECM in both metabolic diseases and brain disorders have been noted very recently. For instance, our previous data-driven research indicates that ECM plays a central role in driving and regulating the metabolically connected cardiovascular disease and type 2 diabetes, as well as functions as key regulators of cognitive functions[42-45]. Structural components of the ECM have been shown to be important orchestrators of synaptic plasticity in the brain, as well as play an important role in regulation of neurite outgrowth, supporting the connection between ECM perturbation and neurodegenerative diseases[46]. Disruption of ECM-involved gene networks and pathways have also been found to be important in promoting metabolic disease such as T2D and cardiovascular disease, likely through perturbation of cell integrity and cell communication pathways that are vital for vascular functions[43]. The cross-disease pathophysiological role of ECM has also been supported by the systemic perturbation of lipid metabolism, glucose homeostasis, and cognitive functions in knockout mouse models lacking key ECM proteins[45]. The role of the ECM and cell adhesion proteins in overall structural integrity and biological scaffolding could provide a systems-wide link between complex metabolic and brain diseases.

Tissue-Specificity of Significant Coexpression Networks from SSEA

From the AD genetic analysis, disease gene-enriched modules appear to have moderate overlap among different tissues. Immune function related coexpression networks with functional annotations such as “immune system”, “intestinal immune network for IGA production”, and “IL12-2 pathway” seem to have especially high overlap among various tissues. “Immune system” annotations represent 16 distinct modules found across four different tissues, specifically adipose tissue, liver, muscle, and the kidney medulla. “Intestinal immune network for IGA production” annotations represent 19 distinct modules found across five different tissues, including the liver, brain, gonadal white adipose tissue, kidney cortex, and kidney medulla. The “IL12-2 pathway” is found in six distinct modules and represented across five different tissues—the kidney medulla, kidney cortex, adipose, liver, and gonadal white adipose tissue. All other functional annotations of the coexpression networks found to be enriched with AD-associated genes do not exhibit as much cross-tissue overlap, and represent at most two different tissues. Given that AD is a disease primarily localized to the brain, the degree of tissue overlap seen in the biological pathways found to be perturbed in AD supports the possibility of a peripheral origin of AD, and suggests that multiple peripheral tissues relevant to metabolism may contribute to the development of AD, thus raising the likelihood of AD as a metabolic disease in addition to a neurodegenerative disorder.

T2D-associated coexpression networks as identified by our study show more cross-tissue overlap compared to the AD-associated gene modules. Annotation such as “immune system”, “antigen processing and presentation”, “respiratory electron transport and oxidative phosphorylation”, “fatty acid and cholesterol biosynthesis”, and “ribosome/translation” show the highest degree of cross-tissue representation. Modules associated with “antigen processing and presentation” are found in the brain, liver, muscle, adipose, and kidney medulla, modules

associated with “fatty acid and cholesterol biosynthesis” are found in the liver, adipose, muscle, and kidney medulla, modules with the annotation “immune system” are found in the brain, liver, kidney cortex, kidney medulla, and adipose, “respiratory electron transport and oxidative phosphorylation”-associated modules are found in the brain, adipose, muscle, omental, and liver, and “ribosome/translation”-associated modules are found in the adipose, brain, liver, omental, and gonadal white adipose tissue. Overall there appears to be tighter cross-tissue correlations among the different biological pathways in T2D, with most pathways present in multiple tissues. The exceptions are the “Alzheimer’s disease”-associated modules which are found only in the muscle, the “platelet activation, signaling, and aggregation” modules which are found only in the brain, “MAPK signaling”-associated modules which are found only in the kidney cortex, and “PPAR signaling pathway”-associated modules which are found exclusively in the muscle. The implication of multiple tissues engaging shared or different biological pathways to affect T2D further confirms the systems-level origin of this metabolic disease.

When examining the shared AD-T2D processes after merging overlapping pathways, we again observed both cross-tissue and tissue-specific properties. A few functional pathways appear to be highly conserved across tissues, including the “interferon signaling”, “Type I Diabetes Mellitus”, and “allograft rejection”, reflecting the systemic nature of immune function signals. Interestingly, perturbation of insulin like growth factor (IGF) pathways were identified to be specific to the brain in both AD and T2D, suggesting that brain IGF signaling, when genetically perturbed, may contribute to the two diseases. Again the involvement of a wide variety of tissues supports systems-wide perturbations that are genetic in origin in both diseases.

Top Key Drivers of AD-T2D Subnetworks

We identified numerous potential regulatory genes, termed key drivers, in tissue-specific gene subnetworks governing the genes and pathways shared between AD and T2D. Key drivers found in the adipose network include: *AIF1* (involved in the immune system/fc gamma r mediated phagocytosis), *GPM6B* (involved in keratan sulfate degradation and biosynthesis), *IFI44* (involved in interferon signaling and cytokine signaling in the immune system), *OAS2* (involved in interferon signaling), and *RTP4* (involved in interferon signaling). Overall, the top key drivers and their respective subnetwork genes seem to be predominantly involved in immune regulation, although the keratan sulfate synthesis and degradation subnetwork suggests potential association of adipose key driver genes with cellular processes such as protein ligand recognition, axonal guidance, cell motility, and an array of other nervous system functions[47]. The *GPM6B* subnetwork in particular deals primarily with keratan sulfate biosynthesis and degradation. This particular functional annotation found in the adipose gene subnetwork is, to our knowledge, a novel connection between AD and T2D. Keratan sulfate is a structural carbohydrate synthesized in the CNS typically associated with cell-surface or extracellular matrix proteins. It participates in neuronal development and glial scar formation in the event of CNS injury. *GPM6B* and *OMD*—another key driver found in brain tissue—are both found to be associated with keratan sulfate synthesis and degradation. Extracellular protein misfolding and aggregation contributes significantly to both AD and T2D pathologies. Extracellular chaperones are likely to play an important role in maintaining "proteostasis", or protein homeostasis and preventing amyloidosis from occurring[48]. Mainly extracellular proteins maintain proteostasis by facilitating clearance of protein aggregates via endocytic receptors and monitoring the extracellular fluid for the presence of misfolded proteins[48]. In addition to its potential role in facilitating amyloidogenic protein clearance, extracellular matrix pathways have also been

identified by another systems biology study as a pathway shared in T2D among diverse ethnic groups[43]. Although initially thought to function predominantly in the cornea, keratan sulfate is actually synthesized in many tissues throughout the body, including cartilage, bone, oocytes, the epidermis, synaptic vesicles, and the brain[47]. In the brain, keratan sulfate in microglia is reduced during inflammation and also reduced in the cerebrum due to Alzheimer's disease[47]. This suggests a potential link between production of inflammatory cytokines and the synthesis of keratan synthesis. The upregulation of certain glycoproteins such as lumican, vasorin, and retinol binding protein-4 were identified in a clinical study in T2D patients with diabetic nephropathy, a common and serious complication associated with T2D[49]. An altered glycoproteome was determined by the study to be a hallmark of this condition, and glycoprotein biomarkers were shown to be a potential clinical method for predicting onset of diabetic nephropathy. As keratan sulfate seems to present a link between inflammatory aspect and the altered blood glycoproteome aspect of AD and T2D respectively, it seems worthwhile to investigate this biological molecule further and unravel its potential causal impact on the development of both diseases.

The liver key driver subnetworks constructed for both AD and T2D are heavily involved in immune function, just like the adipose key driver networks. Key drivers found in the liver subnetwork include: *C1QC* (involved in the triggering of the complement system and complement pathway), *HLA-F* (involved in the endosomal vacuolar pathway/antigen processing and presentation), *CSF1R* (involved in fc gamma r mediated phagocytosis and natural killer cell-mediated cytotoxicity), *IFIT1* (involved in interferon signaling), and LILRB3 (involved in interactions between lymphoid and non-lymphoid cells, as well as the adaptive immune system). Genetic variants in genes *HLA-DRB1* and *HLA-DQB1*, part of the *HLA-F* and *C1QC* subnetworks respectively, are significantly associated with both AD and T2D, according to AD

and T2D GWAS data. *HLA-DRB1* was found to be a member gene belonging to five significant supersets identified to be enriched with T2D and AD genes as well, while *HLA-DQB1* was found to belong to two significant supersets. Different alleles of the *HLA-DQ* and *HLA-DR* loci have been found to be associated with an increased risk of developing type 1 diabetes (T1D), specifically with the autoimmunity aspect of the disease[50-52]. Interestingly, since these genetic loci typically associated with T1D were captured in our analysis as significantly associated with T2D and AD as well, our results suggest a potential connection of autoimmunity with T2D and AD. Islet cell autoimmunity, which typically characterizes T1D, has been found in 10-15% of T2D patients as well[53], suggesting the possible existence of an autoimmune form of T2D. Additionally, neurons exhibiting apoptotic features have been found to contain abnormal vascular-derived immunoglobulins, supporting the conjecture that neuronal cell death in AD may be induced by autoimmunity[54]. Coupled with our finding on autoimmunity genes as key regulators of both diseases through the data-driven systems biology study, it is highly likely that this pathway plays a causal role in both diseases.

The muscle key driver subnetworks appear to connect ECM related processes with autoimmunity. Key drivers found in the muscle subnetworks are: *HLA-DRB1* (involved in antigen processing and presentation, as well as allograft rejection), *PLAC8* (involved in the matrisome and the chemokine signaling pathway), *PCOLCE2* (involved in the matrisome and collagen formation), *MMT00074772* (involved in the matrisome and collagen formation), and *POSTN* (involved with the core matrisome and glycoproteins). The immune and inflammatory response pathways are still represented in the muscle subnetworks, as shown by the antigen processing and chemokine signaling pathway annotations for the HLA-DRB1 network, but most of these key driver subnetworks detected appear to be involved with the ECM. Interestingly,

HLA-DRB1 was detected to be a key driver in the muscle subnetworks, as well as determined as an important part of the liver subnetworks associated with both T2D and AD. *HLA-DQB1* (also identified as a significant AD and T2D-associated node in the liver subnetworks) was also found to be part of the *HLA-DRB1* subnetwork in the muscle. The connection between these immunity genes with ECM is intriguing.

Finally, the brain subnetwork annotations revealed that many key regulators were heavily involved in the ECM as well. The key drivers found in the brain subnetwork include: *SERPING1*, *CYP1B1*, *CTGF* (all related to core matrisome and proteoglycan function), OMD (related to keratan sulfate degradation and biosynthesis), and *ADORA2A* (g alpha s signaling events and nucleotide like purinergic receptors). *OMD* is a key driver involved in keratan sulfate degradation and biosynthesis, which is a pathway represented in both adipose and brain tissues, again supporting the novel connection between this pathway and ECM-related processes in general and the onset of AD and T2D. Further experimental validation of the key regulator genes identified in this study could yield promising results into our understanding of both diseases.

Disease Association Strengths in the Shared Subnetworks

Despite the sharing of a number of gene subnetworks between AD and T2D, we found that the member genes in these subnetworks do not necessarily demonstrate similar genetic association strengths for the two diseases. For instance, in adipose tissue, genetic variants of *OASL* and *AIF1* appear to show stronger association with AD than with T2D, as can be seen by the much larger node sizes in the AD subnetwork; the key driver *OAS2* appears to be much more significantly associated with T2D than with AD; *PSMB9*, on the other hand, appear to be highly associated with both AD and T2D. Similar results can be observed in the other tissue

subnetworks. This observation may indicate that perturbations of these subnetworks by genetic variants in different genes have non-equivalent effects on the two diseases.

Similarities and Differences of Different Tissue Analyses

The results of our analyses of the different tissues reveal interesting commonalities as well as tissue-specific implications. In terms of our key driver analysis, the muscle and brain key driver networks show more extracellular matrix (ECM), core matrisome (ECM constituents and ECM-modifying enzymes), and proteoglycan (found in a lot of connective tissues) related process involvement, while the liver and adipose key driver networks are heavily involved in immune functions, such as cytokine and interferon signaling. In terms of the subnetwork structure for both T2D and AD, differences exist in strength of disease association at the tissue-specific level. In adipose tissue, the subnetwork member genes appear to show relatively similar association strength (according to disease GWAS p-values) in regards to both AD and T2D, implying that molecular pathways perturbed in adipose tissue appear to be significantly affected in both diseases. The brain subnetworks reveal a number of genetic nodes that appear to be highly associated with both AD and T2D, including *HLA-DRB1*, which is seen in five of the supersets shared between AD and T2D. The immune aspect of AD pathology again appears to be heavily represented in the brain subnetworks as several immune genes show high GWAS significance, while in terms of T2D, potential more widespread effects involving *TMEM40* and *SLEC22A18* appear to be involved. The liver subnetworks, which regulate largely immune functions, show more distinctly different subnetwork features between T2D and AD. Since genes that are involved in the immune regulatory process appear to be differentially associated with the two diseases in the liver subnetworks, it appears that different aspects of immune function may

be involved in the two disease pathologies, although immune signals are strongly captured in genetic analyses of both AD and T2D. The muscle subnetworks show better agreement between genes associated with both diseases, in terms of immune function related genes. The significant genetic nodes seen in the AD and T2D subnetworks appear to be involved in mostly different biological functions and include biological annotations such as signal sequence receptors, in lipid metabolism, Ras oncogene family, and calcium/calmodulin-dependent protein kinases, although most of the genes seem to be involved in general signal transduction cascades. Considering our KDA results and the visualization of each key driver's subnetwork structures, the results from muscle and adipose tissue align better with each other while those from the brain and liver tissues show more differences in terms of disease associations. These tissue-specific similarities and differences at the subnetwork structure level could suggest that biological networks in brain and liver tissues are more similarly perturbed in AD and T2D than those in muscle and adipose tissues. In terms of AD pathology, these results may highlight the strong association previously found between cognitive decline and metabolic dysregulation, as the liver is the primary organ responsible for maintaining glucose homeostasis. Perturbations in the body's ability to metabolize glucose could in turn be driving aberrations in brain as well, as evidenced by the crucial role of IDE (insulin degrading enzyme) in preventing beta amyloid formation in the brain as well as promoting glucose tolerance.

Pathways and Networks Downstream of T2D but Causal for AD

Our GWAS-based analysis strongly support that genetic perturbation of certain pathways and gene networks contribute to both AD and T2D. To explore whether genes perturbed as a result of T2D onset were also responsible for driving downstream development of AD as an

additional mechanistic connection between the two diseases, we used T2D transcriptome signatures which can capture both causal (upstream) and reactive (downstream) processes relative to T2D status.

We found that genes in the toll endogenous pathway, T cell receptor (TCR) pathway, hematopoietic cell lineage, phagocytosis, NGF signaling, and calcineurin signaling pathways that were likely downstream of T2D due to the lack of causal inference based on our T2D GWAS analysis, but were among the AD genetic pathways based on the AD GWAS analysis. These processes likely contribute to the sequential development of AD after development of T2D.

T-cell receptor signaling, or TCR, pathways were identified through our analysis to be processes downstream of T2D development and genetically causal for AD. This finding reflects the robust immune signals captured for the AD analysis through both the SSEA and KDA portions of our pipeline, suggesting that these immune-related modules are gene networks strongly perturbed in both AD genetic level and in T2D after disease onset. Chronic low-grade inflammation, or widespread activation of the innate immune system involving upregulated T-cell signaling processes, is a characteristic of T2D pathogenesis and is also closely associated with complications such as dyslipidemia and atherosclerosis[55]. Inflammatory markers such as C-reactive peptide (CRP) and interleukin-6 (IL-6) levels have also been shown to be predictive of T2D development, and drugs with anti-inflammatory effects have been shown to reduce both serum levels of inflammatory compounds as well as glycemic index, potentially decreasing the risk of T2D development[55]. These findings suggest that inflammation and increased activation of innate immune signaling and T-cell response plays a pathogenic role in T2D, and the results of our analysis further implicate T2D-driven inflammatory changes may be involved in sequential development of AD. Convergence of immune pathway perturbations in the event of T2D

development on the same biological pathways causally involved in AD could be subsequently driving AD following T2D development. Furthermore, defects in T-lymphocyte function and signaling result in global immune response changes as well, including abnormal cytokine profiles, signal transduction, ECM interaction, and adhesion molecule expression[56]. The result of these disturbances in immune function is increased susceptibility to a host of metabolic and neurodegenerative diseases. Ultimately, this increased risk, conferred following development of T2D, could converge on preexisting genetic risk to drive AD onset.

Phagocytosis pathways were also detected as biological processes perturbed in T2D and causal for AD development. This pathway captured in our analysis could be reflecting the dysregulation of autophagy and mitochondrial function seen in both diseases. Normal autophagy processes, which involves mTOR and AMPK signaling pathways, is dependent on cell metabolic status[40]. Drug agonists of AMPK, particularly metformin, are known to attenuate the symptoms of T2D and restore insulin sensitivity and glucose homeostatic balance to some extent. Although the mechanism of action of AMPK activators in improving insulin resistance symptoms in T2D patients has yet to be fully understood, the efficacy of metformin and other AMPK activators that mimic the effects of caloric restriction and improve metabolic dysregulation characteristic of T2D suggests a central role of AMPK pathways in T2D pathophysiology. Perturbations of AMPK pathways in T2D could subsequently be affecting autophagy pathways and lead to alterations of proteome homeostasis and aberrant protein aggregation[57]. Additionally, autophagy was found to be protective in neurons by promoting degradation of damaged proteins and organelles[58]. The phagocytotic pathways detected as causally involved in AD and downstream of T2D development suggests that perturbation of AMPK signaling in T2D could be affecting

autophagy pathways and subsequently obstructing clearance of harmful protein aggregates that could be driving AD onset.

Calcineurin signaling pathways were also detected as an overlapping pathway between AD causal processes and T2D transcriptome profile gene signatures. Dysregulation of calpain, an important cellular calcium sensor, and calcineurin signaling pathways has been implicated in the pathogenesis of numerous metabolic disorders such as cardiovascular disease, hypertension, diabetes, and Alzheimer's disease. Although the role of calpain-calcineurin signaling cascades in T2D pathology has yet to be fully elucidated, there is considerable evidence pointing to greatly increased risk of new onset diabetes after transplantation as a result of immunosuppressive medications[59]. Calcineurin inhibitors given to treat insulin resistance developed as a result of new onset diabetes mellitus have been shown to acutely improve insulin sensitivity, implicating calcineurin signaling pathways as interacting with insulin signaling processes. In addition to the potential convergence of calcineurin and insulin signaling pathways in the event of immunosuppression, other lines of evidence also demonstrate involvement of calcineurin pathways in neurons as well. Recent studies have shown that calpain is directly involved in the regulation of calcineurin activity through proteolysis events in glutamate-stimulated neurons[60]. Calpain-mediated proteolytic cleavage of calcineurin upregulates phosphatase activity, and ultimately promotes neuronal cell death, a defining hallmark of AD and neurodegenerative disease in general [60]. As networks significantly enriched for calcineurin signaling pathways were identified as causal for AD as well as enriched for T2D signatures, this provides support for calcineurin signaling processes as a potential mediator between T2D pathogenesis and d development of AD.

Interestingly, NGF or nerve growth factor signaling was also found to be the functional annotation of an AD-associated module significantly enriched for T2D differentially expressed genes found across multiple tissues, specifically the islet and the muscle. NGF signaling, or neutrophic factor signaling, which is known to be impaired in both AD and T2D[61]. As this particular pathway was identified through performing an enrichment analysis of AD-associated modules for tissue-specific genes differentially expressed in T2D, it seems likely that this particular module may be genetically perturbed as a result of T2D, thus driving AD development downstream. Thus, perturbation of this module during T2D can be interpreted as a risk factor for AD. As growth factors such as BDNF (brain-derived neurotrophic factor), IGF-1 (insulin-like growth factor 1), and NGF have been shown in clinical studies to be neuroprotective, perturbation of these cellular pathways in T2D could trigger a cascade of neurodegenerative events eventually leading to AD[62].

Cancer-related pathways, such as apoptosis, VEGF signaling, and MAPK signaling, as well as overall immune function and regulatory pathways seem to be pathways that are strongly perturbed in both AD and T2D, as well as perturbed after T2D onset and prior to AD development. These biological processes may also be interpreted as pathways that may become dysregulated as a result of T2D development, as they are represented in the T2D signatures identified across multiple studies, as well as pathways that could be causal for both T2D and AD, as they are represented in both AD and T2D supersets enriched for T2D and AD GWAS genes. As these pathways and coexpression networks were identified to be enriched for both signatures downstream of T2D development as well as those genetically causal for AD and T2D, there is the possibility that some of these coexpression networks and biological pathways contain susceptibility loci as well as genetic loci that are altered as a result of T2D onset. These

perturbations can then subsequently contribute to driving AD onset downstream as well. Since many of these pathways and gene networks identified to contain T2D signatures as well as causal AD and T2D genes converge on immune processes and cancer-related pathways, it seems that genetic associations with other disease such as cancer may be captured by these modules and networks as well. This makes intuitive sense, as development of cancers have been linked to metabolic disease and dysregulation as well, such as that seen in T2D and AD pathologies[63]. T2D, AD, cancer, and chronic metabolic diseases, share many similar characteristics including excessive generation of free radicals, oxidative DNA damage, apoptotic pathway dysregulation, and mitochondrial DNA abnormalities. Oftentimes, individuals affected by one of these conditions also experience comorbidities with other cancers, metabolic diseases, or neurodegenerative diseases. Since the disease burden of these conditions is enormous, and presents both greatly diminished quality of life to patients who suffer from these conditions as well as tremendous economic strain to our healthcare system, a systematic analysis of genetic loci linked to these conditions is necessary to better understand the intricate relationships between these comorbidities.

Advantages and Limitations of the Study

This study integrates a large amount of genomic and genetic data, and provides a data-driven approach to understanding the genetic architecture of both AD and T2D. Our approach involving the identification of coexpression networks and biological pathways potentially involved in the two diseases and the subsequent network modeling to identify disease key drivers provides a holistic, systems-wide outlook on AD and T2D. Furthermore, as both diseases can be classified as metabolic disorders and have wide-ranging effects throughout the body, an

approach such as ours is needed to fully elucidate how the two diseases may be interconnected and provide novel insights into genes and biological processes in multiple tissues involved in both diseases. However, given our data-driven methods, it is difficult to substantiate the involvement of our detected novel biological pathways and key driver genes in AD and T2D using purely computational techniques. Thus, validation in a laboratory setting is needed to fully understand the biological implications of our findings. Some key genes identified as important for driving both AD and T2D pathologies that would be interesting to validate in a laboratory setting to are *GPM6B* and *OMD* – the keratan sulfate metabolism genes found to be key drivers in the adipose and brain tissue respectively. As the brain and adipose tissue are tissues centrally affected in neurodegenerative and metabolic diseases respectively, it is interesting that key drivers that to orchestrate similar biological processes were found to be in involved in these tissues. Extracellular matrix related processes were also robustly indicated in both the SSEA and KDA portions of the analysis for both T2D and AD, and has been suggested to play a critical role in maintaining protein homeostasis and clearing away misfolded proteins and protein aggregates. It would be interesting to validate some shared key driver genes involved in ECM and proteosome processes such as PCOLCE2 and POSTN (found in the muscle), as well as SERPING1, CYP1B1, and CTGF (found in the brain). Knockdown and overexpression of these genes in mouse models of AD and T2D with matched controls could provide experimental insight into the roles of these key driver genes in terms of driving the two disease pathologies.

Conclusions

In summary, through a comprehensive integrative genomics study incorporating genetic, transcriptome, functional genome, pathways, and molecular networks, we examined the shared

molecular mechanisms between two interconnected diseases, T2D and AD. We provide compelling evidence supporting that genetic risks of these two diseases drive perturbations in a large number of pathways such as immune signaling, cell adhesion, ECM, PPAR signaling, cell cycle regulation, autophagy, and oxidative phosphorylation. Moreover, molecular processes downstream of T2D, such as NGF signaling, NK dynamin pathway, core matrisome function, calcineurin signaling pathways, and various immune pathways, were found to be putatively causal for AD, supporting a sequential role of T2D in driving AD onset. Finally, the ECM component appears to be a novel recurring pathway found in our analysis that is captured across multiple tissues, and could open up a new direction for researching the connection between different complex human diseases. With the emergence of systems biology, studying diseases in a more integrated and comprehensive way has yielded more possibilities for identifying causal genes underlying diseases, discovering drug targets for therapeutic interventions, and bringing us closer to the era of personalized medicine. With compelling evidence from our data-driven systems study supporting multiple shared pathways between T2D and AD, developing pharmacological therapies that can simultaneously treat both diseases serves as an important future direction in achieving effective reduction of the health burden imposed by these common complex diseases.

Tables

Table 1. Network resources. Datasets from several studies were used to construct coexpression networks used in the analysis. The references and descriptions for the datasets are listed in the chart below.

Tissue	Species	Dataset descriptions	References
Adipose tissue	Human	1,675 individuals from two Icelandic cohorts	Emillson, 2008
	Mouse	C57BL/6J x A/J mouse cross	Derry, 2010
	Mouse	C57BL/6J x C3H ApoE -/- mouse cross	Wang, 2007
	Mouse	C57BL/6J x C3H wildtype mouse cross	Schadt, 2008
	Mouse	C57BL/6J x BTBR Lepob mouse cross	Tu, 2012
Blood	Human	1,675 individuals from two Icelandic cohorts	Emilsson, 2008
Brain	Mouse	C57BL/6J x A/J mouse cross	Derry, 2010
	Mouse	C57BL/6J x C3H ApoE -/- mouse cross	Wang, 2007 Yang, 2006
	Mouse	C57BL/6J x BTBR Lepob mouse cross	Tu, 2012
	Human	57 subjects ranging from 5.7 weeks post-conception to 82 years; samples taken from 16 regions and from both hemispheres; 1,340 samples total	Kang, 2011
Heart	Mouse	C57BL/6J x A/J mouse cross	Derry, 2010
Islet cells	Mouse	C57BL/6J x BTBR Lepob mouse cross	Tu, 2012
Kidney	Mouse	C57BL/6J x A/J mouse cross	Derry, 2010
Liver	Human	427 individuals	Schadt, 2008
	Mouse	C57BL/6J x A/J mouse cross	Derry, 2010
	Mouse	C57BL/6J x C3H ApoE -/- mouse cross	Wang, 2007 Yang, 2006
	Mouse	C57BL/6J x C3H wildtype mouse cross	Schadt, 2008
	Mouse	C57BL/6J x BTBR Lepob mouse cross	Tu, 2012
Muscle	Mouse	C57BL/6J x A/J mouse cross	Derry, 2010
	Mouse	C57BL/6J x C3H ApoE -/- mouse cross	Wang, 2007 Yang, 2012
	Mouse	C57BL/6J x C3H wildtype mouse cross	Tu, 2012
	Mouse	C57BL/6J x BTBR Lepob mouse cross	Tu, 2012

Table 2. Top 20 representative functional categories showing significant enrichment for AD GWAS signals. Biological pathways or coexpression modules meeting a FDR cutoff <25% in our AD SSEA were taken and a gene ontology enrichment analysis was performed to identify the biological pathways represented by the modules. A total of 53 modules met the FDR cutoff and were determined to be significantly associated with AD disease genes as identified by GWAS. Only the top 20 most represented module annotations are shown. The ones shared with T2D in Table 3 are highlighted in bold.

Annotations	Modules Represented	Tissues Represented
antigen processing cross presentation	4386	muscle
calcineurin pathway	M5940	N/A
cell-cell junction organization	rctm0763	N/A
chemokine signaling pathway	5605, 4083, 26, 27	liver, kidney medulla
comp pathway	4386	muscle
cxcr4 pathway	AD Positive	N/A
cytokine signaling in immune system	4351, 4139, 4351, 4139, 5605, 7123	liver, adipose
DNAPK pathway	4121	brain
ECM glycoproteins	4186	muscle
fc gamma r mediated phagocytosis	4936	liver
HDL mediated li transport	rctm0647, rctm0521	N/A
hematopoietic cell lineage	4568, AD Positive	muscle
il12 2pathway	4416, 26, 5136, 87, 5405, 4998	kidney medulla, liver, gonadal white adipose tissue, kidney cortex, adipose
immune system	4080, 4344, 5054, 4911, 4483, 4416, 4080, 4083, 5266, 4351, 4139, 4479, 5447, 4936, 4911, 4289	adipose, liver, muscle, kidney medulla
interferon signaling	7123, 4804	muscle, adipose
intestinal immune network for IGA production	5266, 4479, 5532, 5136, 87, 5447, 5405, 5315, 4998, 4833, 5366, 63, 5532, 5216, 5366, 5656, 5315, 5659, 5330	liver, brain, gonadal white adipose tissue, kidney cortex, kidney medulla, kidney
metabolism of lipids and lipoproteins	rctm0647, 5354, rctm0239, rctm0521	adipose
metabolism of xenobiotics by cytochrome p450	4148	islet
ribosome	rctm0493	N/A
TCR pathway	4483, 4568, 4344	muscle, adipose

Table 3. Top 20 representative functional categories showing significant enrichment for T2D GWAS signals. Biological pathways or coexpression modules meeting a FDR cutoff <25% in our T2D SSEA were taken and a gene ontology enrichment analysis was performed to identify the biological pathways represented by the modules. A total of 176 modules met the FDR cutoff and were determined to be significantly associated with T2D disease genes as identified by GWAS. Only the top 20 most represented functional categories are shown. The ones shared with AD in Table 2 are highlighted in bold.

Annotation	**Pathways or Modules Represented**	**Tissues Represented**
allograft rejection	5677, 7020, 7107, 183, 5348	adipose, liver
Alzheimer's disease	4334, 7088	muscle
antigen processing and presentation	5016, 5768, 6919, 7040, 5459, 4386, 5607	brain, kidney medulla, liver, muscle, adipose
cell cycle	VFCturquoise, STCturquoise, ITCturquoise, 4091, 4230	brain, liver
complement and coagulation cascades	4919, 4195, 4118, 4320, 4735, 5014, 4750, CBCwhite, 4744	brain, liver
diabetes (maturity onset diabetes of the young, Type 1, Type 2)	7136, 5768, 6919, 7040, 37, 5459, 4820, rctm1011, M18312, rctm1014	brain, kidney medulla, muscle
drug metabolism cytochrome p450	4469, 4212, 4735, 4465, 5018, 5014, 4750, 4744	muscle, liver, brain
extracellular matrix	4074, 5545	adipose, kidney
fatty acid and cholesterol biosynthesis	4336, 4566, 4087, 7222, rctm0646, 4308, 6675, 5428	liver, adipose, muscle, kidney medulla
GPCR signaling	4141, 4038, CBCyellow	brain, adipose, islet
HIV infection	4766, 4821, 4421, IPCorange, 4357	brain, liver
immune system	HIPlightyellow, STRdarkorange, 5532, 5216, 6782, 5065, 4351, 4139, 4083, 5266, 4139, 5447, 4234, 6630, 110, 5266, S1Cblack, 5656, 5136, 87, 5447	brain, liver, kidney cortex, kidney medulla, adipose
lectin pathway	5016, 4302, 37, 5212	brain, kidney medulla, adipose
MAPK signaling	M13191, M14631, M19888, M9670, 52	kidney cortex
metabolism of lipids and lipoproteins	5725, 4336, 4566, 4308, 4791, 6675, rctm1380, rctm1381	adipose, muscle, liver
olfactory signaling pathway	5503, 5658	adipose, brain
platelet activation signaling and aggregation	4368, 4295	brain
PPAR signaling pathways	5725, 7032, 4212, 4791	muscle
respiratory electron transport/oxidative phosphorylation	6649, IPCdarkgreen, 5498, 4683, 4651, 4486, 4078, 7088, 4334	brain, adipose, muscle, liver
ribosome/translation	5671, 4124, 4457, 4452, 4026, 5512, 4678, 4393, 4720	adipose, brain, liver

Table 4. Top 20 representative functional categories of the merged supersets showing significant enrichment for both AD and T2D. The suggestive pathways or coexpression modules shared between AD and T2D were assessed for overlap, and those sharing at least 30% of genes were merged into relatively independent supersets. These supersets were then used to perform a second round of SSEA to confirm AD/T2D genetic enrichment. The top 20 functional categories over-represented in the 39 significant supersets shared between AD and T2D are listed.

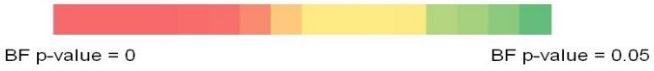
Annotation	Modules Represented	Tissues Represented
allograft rejection	7136, 7040,.., M12618	muscle, brain, adipose, kidney
antigen processing cross presentation	M16005, 4386	muscle, adipose
apoptosis	M9670,..	N/A
cell adhesion molecules	rctm0567	N/A
cell cycle	5033, STCturquoise,..	brain
cholesterol biosynthesis	4932,..	liver
complement and coagulation cascades	4932,..	liver
core matrisome	STCbrown	brain
cytokine signaling in immune system	M6910,.., 4634	blood
fatty acyl CoA biosynthesis	4212,..	muscle
GPCR downstream signaling	4487,.., 4770, 4866	adipose, brain
HIV life cycle	4723	liver
immune system	HIPdarkgrey,.., S1Cblack, rctm0568	brain
interferon signaling	26,.., 4634	kidney, liver, brain, adipose, blood
MAPK signaling pathway	rctm0090, M9670,..	N/A
oxidative phosphorylation/respiratory transport chain	IPCdarkgreen, 4683	brain, adipose
PPAR signaling pathway	4212,.., 5376	muscle, adipose
regulation of autophagy	M16004	N/A
regulation of insulin like growth factor IGF activity by insulin like growth factor binding proteins IGFBPs	STCbrown	brain
type I diabetes mellitus	7040,.., M12617, 7136	muscle, adipose, kidney, brain

Table 5. Shared genetic supersets between AD and T2D that overlap with T2D transcriptome signatures. Of the merged supersets shared between T2D and AD, 25 supersets were found to be significantly overlap with T2D transcriptome signatures at a Bonferroni (BF)-corrected p-value < 0.05 based on Fisher's exact test. Numbers in the T2D signature columns are BF p values. The colors correspond to the significance level of overlap, with red to green representing more significant to less significant.

Superset	Annotation	Adipose T2D signature	Brain T2D signature	Islet T2D signature	Liver T2D signature	Muscle T2D signature
4723	HIV life cycle	3.17E-46	---	1.23E-126	---	4.74E-14
rctm0089	MAPK signaling	5.44E-11	---	8.01E-35	---	1.07E-17
5033	Cell cycle	5.83E-09	---	2.20E-19	---	3.49E-02
IPCred	VEGF ligand receptor	1.85E-07	---	1.87E-23	---	3.72E-02
rctm0116	Immune system	1.81E-06	---	1.61E-05	---	4.32E-03
S1Cblack	Interferon signaling	2.59E-06	---	1.96E-07	---	1.06E-02
4457	Peptide ligand receptors	1.30E-05	---	9.32E-03	1.86E-05	1.53E-08
STCturquoise,..	Cell cycle	3.66E-05	---	9.54E-05	---	7.73E-20
4678	ribosome	6.15E-05	---	8.65E-19	---	2.68E-04
4683	Oxidative phosphorylation	1.46E-04	---	1.34E-08	---	---
4212,..	PPAR signaling	7.13E-04	---	5.01E-05	---	7.90E-06
4770	GPCR signaling	1.03E-03	---	8.15E-05	---	2.48E-08
M16004	Autophagy	1.9E-02	---	03.18E-02	---	2.42E-03
M9670,..	MAPK signaling	2.28E-02	---	---	---	3.16E-02
26,..	Interferon signaling	2.95E-02	---	---	---	2.08E-07
4393	ribosome	3.96E-02	---	5.66E-03	---	---
4386	Endosomal vacuolar pathway	---	---	3.08E-02	---	1.41E-05
5194	Integrin interactions	---	---	---	---	3.33E-04
STCbrown	Core matrisome	---	---	4.33E-02	---	8.78E-04
4932,..	Cholesterol biosynthesis	---	---	5.13E-10	---	9.52E-04
4634	Interferon signaling	---	---	---	---	6.44E-03
HIPdarkgrey,..	Immune system	---	---	---	---	1.59E-02
4715	Bile synthesis	---	---	4.21E-02	---	---
IPCdarkgreen	Electron transport chain	---	---	1.53E-03	---	---
ITCgreenyellow	Transcription	---	---	1.05E-03	---	---

Table 6. AD genetic pathways or coexpression modules that overlap with T2D transcriptome signatures. Using a Bonferroni-corrected (BF) p-value cutoff of 0.05, we found that 6 AD pathways or coexpression modules overlapped with transcriptome signatures defined as downstream to T2D. These AD-associated modules that are significantly enriched for T2D downstream signatures can be interpreted as molecular processes causal for AD and downstream of T2D that could be contributing to sequential development of AD after development of T2D. In other words, these causal AD genes, as identified by GWAS studies, could be perturbed by T2D and subsequently drive onset of AD. The colored label corresponds to how significantly enriched for T2D tissue-specific DEGs the supersets were (i.e. high or low Bonferroni-corrected p-value).

Module	Annotation	Adipose T2D signature	Brain T2D signature	Islet T2D signature	Liver T2D signature	Muscle T2D signature
rctm0493	tRNA biosynthesis	2.05E-27	---	5.12E-46	---	1.63E-20
4568	TCR pathway	---	---	---	---	2.33E-04
4483	TCR pathway	---	---	---	2.74E-03	5.18E-03
4344	TCR pathway	---	---	---	1.04E-02	5.76E-03
rctm1111	NGF signaling	---	---	1.97E-04	---	1.12E-02
4936	phagocytosis	---	---	---	---	2.34E-02
27	Toll endogenous pathway	---	---	---	5.05E-03	---
M5940	Calcineurin pathway	---	---	3.96E-02	---	---



Figures

Figure 1. Overall study design. We first delineate the molecular pathways and gene networks that are perturbed by genetic risks of AD and T2D separately. We then compare the genetically perturbed pathways and networks between T2D and AD to derive shared genetic mechanisms that may be causally linked to both diseases. To test whether molecular pathways downstream of T2D (i.e., non-causal for T2D) also pose predisposition to AD development, we will also extract tissue-specific transcriptomic profiles of T2D and compare their signatures with the AD causal processes derived from the genetic analysis of AD.

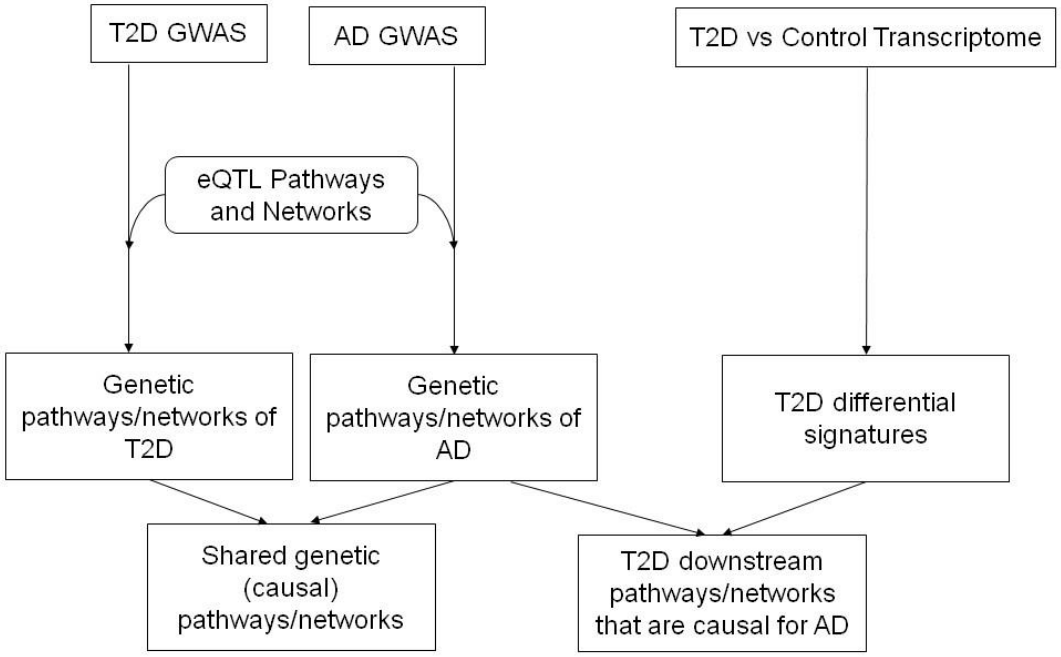


Figure 2. Conceptual framework for the Mergeomics pipeline. The first step of the Mergeomics pipeline is the SSEA, or SNP set enrichment analysis, which integrates disease GWAS, functional genomics from eQTL and ENCODE studies, biological pathways, gene regulatory networks, and returns gene sets that are significantly enriched for disease genes according to GWAS data. The next step of the pipeline is the module merging and trimming step, in which modules that share a high proportion of member genes are merged into non-overlapping gene sets. The final step of the pipeline is the key driver analysis, in which Bayesian gene regulatory networks are used as frameworks to identify key driver genes in the gene sets. The subnetworks for these key driver genes are subsequently retrieved and visualized using a software such as Cytoscape.

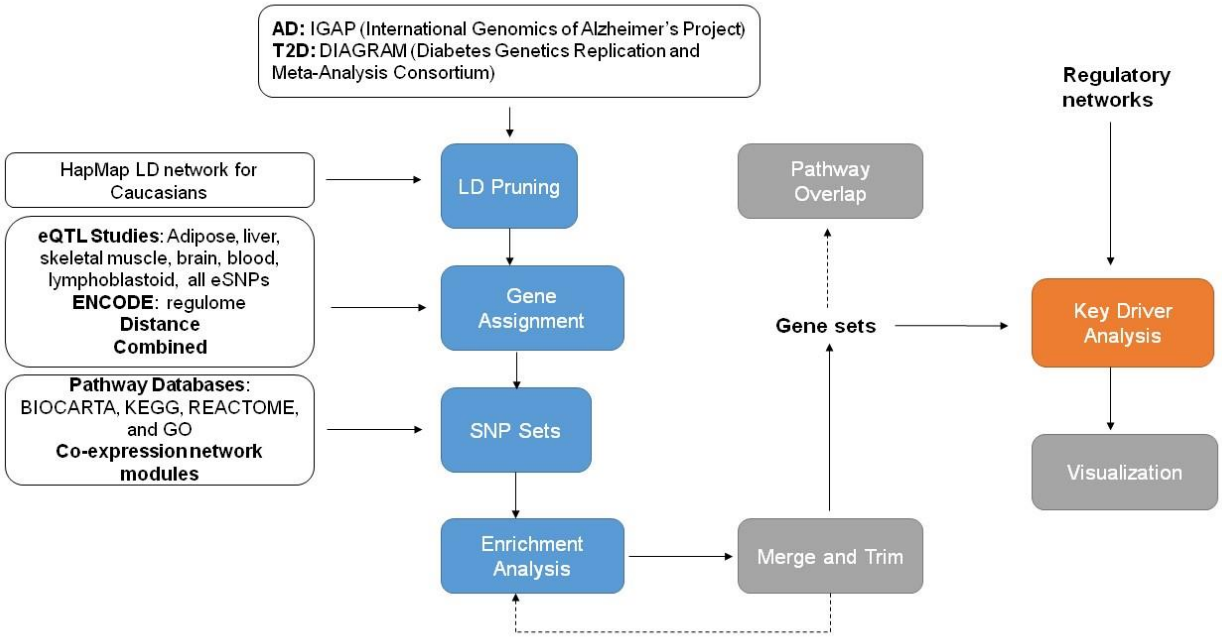


Figure 3. AD and T2D key drivers in adipose tissue. **Figure 3. AD and T2D key drivers in adipose tissue.** **A.)** This panel depicts the adipose tissue subnetworks in AD. The adipose subnetworks depict the following key drivers—AIF1, GPM6B, RTP4, OAS2, and IFI44—and their interaction with their respective gene neighbors. After performing an enrichment analysis of the genes involved in each key driver subnetwork, we determined that the AIF1 subnetwork appeared to be mainly involved in immune system functions and FC gamma R mediated phagocytosis, the GMP6B subnetwork was involved with keratan sulfate biosynthesis and degradation, the OAS2 and RTP4 subnetworks were involved with interferon signaling, and the IFI44 subnetwork was involved in interferon an cytokine signaling in the immune system. Superset membership is represented by the color of each node (according to the legend), and magnitude of disease association for either T2D or AD according to GWAS p-values is represented by the size of the node. **B.)** This panel depicts the adipose tissue subnetworks in T2D.

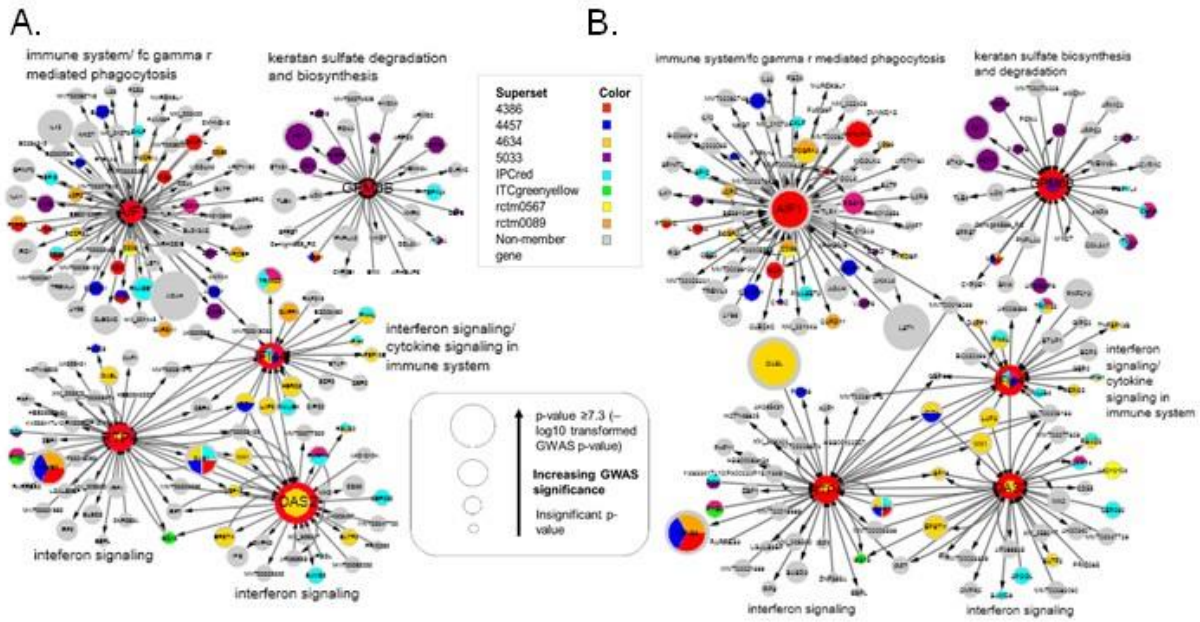


Figure 4. AD and T2D key drivers in brain tissue. A.) This panel depicts the brain subnetworks in AD. The brain subnetworks depict the following key drivers—SERPING1, CYP1B1, OMD, CTGF, and ADORA2A—and their interaction with their respective gene neighbors. The SERPING1, CTGF, and CYP1B1 subnetworks are mainly involved in core matrisome and proteoglycan production, while the ADORA2A subnetwork is involved with G alpha S signaling events and nucleotide-like purinergic receptor function. The OMD subnetwork appears to be involved in keratan sulfate degradation and biosynthesis. **B.)** This panel depicts the brain subnetworks in T2D.

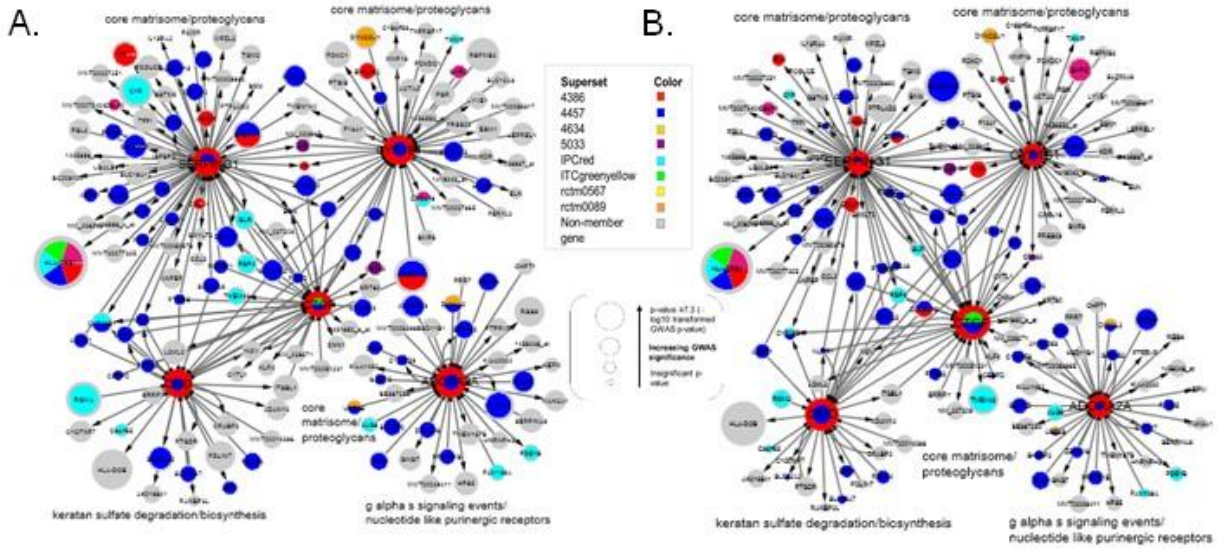


Figure 5. AD and T2D key drivers in liver tissue. A.) This panel shows the liver subnetworks in AD. The liver subnetworks depict the following key drivers—C1QC, CSF1R, HLA-F, IFIT1, and LILRB3—and their interaction with their respective gene neighbors. The C1QC subnetwork is mainly involved in the complement pathway of immune regulation, the HLA-F subnetwork is involved with endosomal vacuolar pathways as well as antigen processing and presentation, the CSF1R subnetwork is involved with FC gamma R mediated phagocytosis and natural killer cell-mediated cytotoxicity, the IFIT1 subnetwork is involved with interferon signaling, and the LILRB3 subnetwork is involved in adaptive immune system function and interactions between lymphoid and non-lymphoid cells. **B.)** This panel shows the liver subnetworks in T2D.

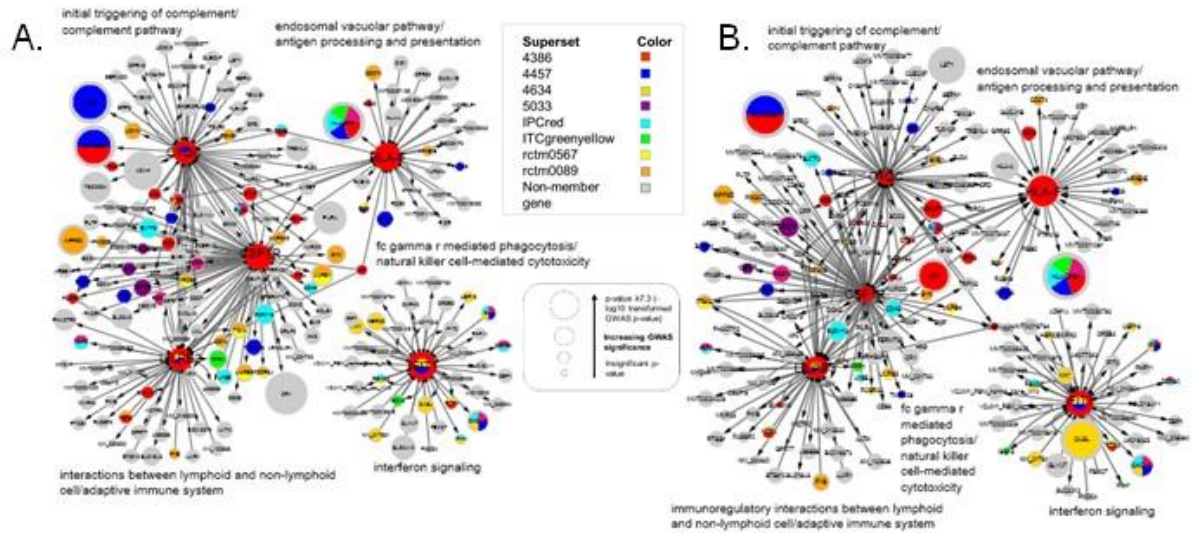


Figure 6. AD and T2D key drivers in skeletal muscle tissue. A.) This panel shows the muscle subnetworks in AD. The muscle subnetworks depict the following key drivers—HLA-DRB1, MMT00074772, PCOLCE2, PLAC8, and POSTN—and their interaction with their respective gene neighbors. The HLA-DRB1 subnetwork is involved with antigen processing and presentation as well as allograft rejection, the PCOLCE2 and MMT00074772 subnetworks are involved with matrisome and collagen formation, the POSTN subnetwork is involved with core matrisome and glycoprotein production, and the PLAC8 subnetwork is involved with chemokine signaling pathways as well as matrisome formation. **B.)** This panel shows the muscle subnetworks in T2D.

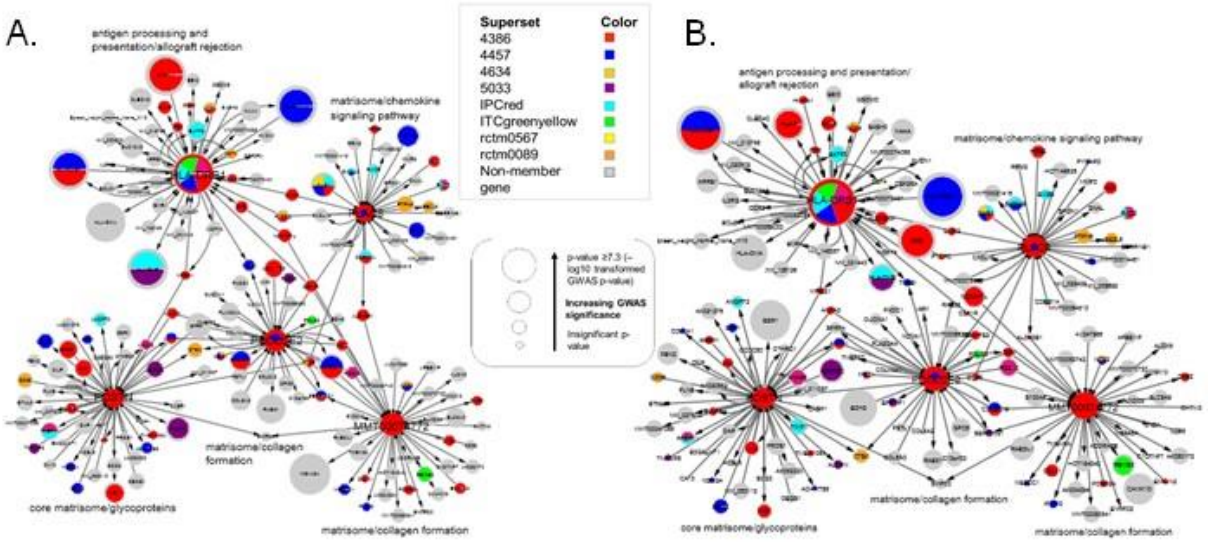
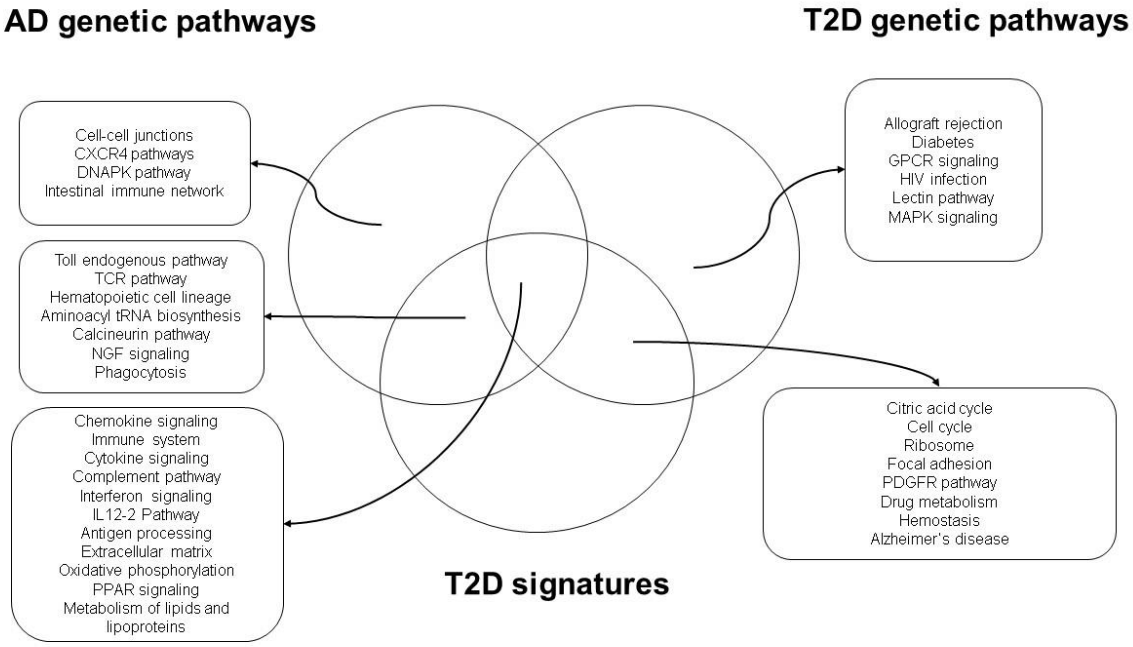


Figure 7. Overlap between AD/T2D-associated biological pathways and coexpression networks and T2D signatures. This figure depicts the overlap of AD and T2D associated biological pathways and gene coexpression networks identified by our computational analysis with the gene signatures identified from the T2D transcriptome profiles of previously performed studies. Of the 48 total AD modules and 314 T2D modules found to be significantly enriched for disease GWAS genes, 27 AD modules and 70 T2D modules were found to be significantly enriched for T2D signatures, as identified from tissue-specific transcriptome profiles. 16 AD modules in total are enriched for T2D signatures, and 59 T2D modules that were identified through analysis of GWAS studies were enriched for T2D DEGs. Representative unique AD-associated genetically-driven pathways and coexpression networks include cell-cell junctions, CSCR4 pathways, DNAPK pathways, and intestinal immune network. Representative T2D-specific genetically-driven pathways and coexpression networks include allograft rejection, diabetes, GPCR signaling, PPAR signaling, and MAPK signaling. Modules that are downstream of T2D based on transcriptome signature but are putatively causal for AD based on genetic evidence include annotations such as toll endogenous pathway, T cell receptor (TCR) pathway, aminoacyl tRNA biosynthesis, hematopoietic cell lineage, phagocytosis, NGF signaling, and calcineurin signaling. Representative modules enriched for both T2D signatures and T2D GWAS signals include cell cycle processes, citric acid cycle, focal adhesion, ribosome function, hemostasis, and Alzheimer's disease. Modules detected in all analyses (T2D vs AD; genetic vs transcriptome) are involved in immune pathways such as interferon signaling, antigen processing, chemokine binding, and cytokine signaling, as well as extracellular matrix functions and oxidative phosphorylation.



References:

1. Yarchoan, M., and Arnold, S.E. (2014). Repurposing diabetes drugs for brain insulin resistance in Alzheimer disease. Diabetes 63, 2253-2261.

2. Kroner, Z. (2009). The relationship between Alzheimer's disease and diabetes: Type 3 diabetes? Alternative medicine review : a journal of clinical therapeutic 14, 373-379.

3. Vicario, A., Del Sueldo, M., Fernandez, R.A., Enders, J., Zilberman, J., and Cerezo, G.H. (2012). Cognition and vascular risk factors: an epidemiological study. International journal of hypertension 2012, 783696.

4. Cheng, C., Lin, C.H., Tsai, Y.W., Tsai, C.J., Chou, P.H., and Lan, T.H. (2014). Type 2 diabetes and antidiabetic medications in relation to dementia diagnosis. The journals of gerontology Series A, Biological sciences and medical sciences 69, 1299-1305.

5. Strachan, M.W., Reynolds, R.M., Marioni, R.E., and Price, J.F. (2011). Cognitive function, dementia and type 2 diabetes mellitus in the elderly. Nature reviews Endocrinology 7, 108-114.

6. Launer, L.J., Miller, M.E., Williamson, J.D., Lazar, R.M., Gerstein, H.C., Murray, A.M., Sullivan, M., Horowitz, K.R., Ding, J., Marcovina, S., et al. (2011). Effects of intensive glucose lowering on brain structure and function in people with type 2 diabetes (ACCORD MIND): a randomised open-label substudy. The Lancet Neurology 10, 969-977.

7. Miklossy, J., Qing, H., Radenovic, A., Kis, A., Vileno, B., Laszlo, F., Miller, L., Martins, R.N., Waeber, G., Mooser, V., et al. (2010). Beta amyloid and hyperphosphorylated tau deposits in the pancreas in type 2 diabetes. Neurobiology of aging 31, 1503-1515.

8. Macauley, S.L., Stanley, M., Caesar, E.E., Yamada, S.A., Raichle, M.E., Perez, R., Mahan, T.E., Sutphen, C.L., and Holtzman, D.M. (2015). Hyperglycemia modulates extracellular amyloid-beta concentrations and neuronal activity in vivo. The Journal of clinical investigation 125, 2463-2467.

9. Alexandraki, K., Piperi, C., Kalofoutis, C., Singh, J., Alaveras, A., and Kalofoutis, A. (2006). Inflammatory process in type 2 diabetes: The role of cytokines. Annals of the New York Academy of Sciences 1084, 89-117.

10. Rubio-Perez, J.M., and Morillas-Ruiz, J.M. (2012). A review: inflammatory process in Alzheimer's disease, role of cytokines. TheScientificWorldJournal 2012, 756357.

11. Holscher, C. (2014). Drugs developed for treatment of diabetes show protective effects in Alzheimer's and Parkinson's diseases. Sheng li xue bao : [Acta physiologica Sinica] 66, 497-510.

12. Craft, S., Baker, L.D., Montine, T.J., Minoshima, S., Watson, G.S., Claxton, A., Arbuckle, M., Callaghan, M., Tsai, E., Plymate, S.R., et al. (2012). Intranasal insulin therapy for Alzheimer disease and amnestic mild cognitive impairment: a pilot clinical trial. Archives of neurology 69, 29-38.

13. Stoothoff, W.H., and Johnson, G.V. (2005). Tau phosphorylation: physiological and pathological consequences. Biochimica et biophysica acta 1739, 280-297.

14. Rivera, E.J., Goldin, A., Fulmer, N., Tavares, R., Wands, J.R., and de la Monte, S.M. (2005). Insulin and insulin-like growth factor expression and function deteriorate with progression of Alzheimer's disease: link to brain reductions in acetylcholine. Journal of Alzheimer's disease : JAD 8, 247-268.

15. Shu L, Z.Y., Kurt Z, Byars SG, Tukiainen T, Kettunen J, Ripatti S, Zhang B, Inouye M, Makinen VP*, Yang X. . (2016). Mergeomics: integration of diverse genomics resources to identify pathogenic perturbations to biological systems. BioRxiv

16. Desikan, R.S., Schork, A.J., Wang, Y., Thompson, W.K., Dehghan, A., Ridker, P.M., Chasman, D.I., McEvoy, L.K., Holland, D., Chen, C.H., et al. (2015). Polygenic Overlap Between C-Reactive Protein, Plasma Lipids, and Alzheimer Disease. Circulation 131, 2061-2069.

17. Morris, A.P., Voight, B.F., Teslovich, T.M., Ferreira, T., Segre, A.V., Steinthorsdottir, V., Strawbridge, R.J., Khan, H., Grallert, H., Mahajan, A., et al. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nature genetics 44, 981-990.

18. Haw, R., and Stein, L. (2012). Using the reactome database. Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al] Chapter 8, Unit8 7.

19. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research 28, 27-30.

20. Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M., Pletikos, M., Meyer, K.A., Sedmak, G., et al. (2011). Spatio-temporal transcriptome of the human brain. Nature 478, 483-489.

21. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics 9, 559.

22. Ihmels, J., Bergmann, S., and Barkai, N. (2004). Defining transcription modules using large-scale gene expression data. Bioinformatics 20, 1993-2003.

23. Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Statistical applications in genetics and molecular biology 4, Article17.

24. (2013). The Genotype-Tissue Expression (GTEx) project. Nature genetics 45, 580-585.

25. Kondro, W. (2004). Molecular biology. Consortium tackles mouse regulome. Science 304, 942.

26. (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 306, 636-640.

27. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research 13, 2498-2504.

28. Dubois, R.N. (2015). The Jeremiah Metzger Lecture: Inflammation, Immune Modulators, and Chronic Disease. Transactions of the American Clinical and Climatological Association 126, 230-236.

29. Watson, G.S., and Craft, S. (2003). The role of insulin resistance in the pathogenesis of Alzheimer's disease: implications for treatment. CNS drugs 17, 27-45.

30. Feinstein, D.L. (2003). Therapeutic potential of peroxisome proliferator-activated receptor agonists for neurological disease. Diabetes technology & therapeutics 5, 67-73.

31. Combs, C.K., Johnson, D.E., Karlo, J.C., Cannady, S.B., and Landreth, G.E. (2000). Inflammatory mechanisms in Alzheimer's disease: inhibition of beta-amyloid-stimulated proinflammatory responses and neurotoxicity by PPARgamma agonists. The Journal of neuroscience : the official journal of the Society for Neuroscience 20, 558-567.

32. Li, X., Song, D., and Leng, S.X. (2015). Link between type 2 diabetes and Alzheimer's disease: from epidemiology to mechanism and treatment. Clinical interventions in aging 10, 549-560.

33. Horan, M.P., Pichaud, N., and Ballard, J.W. (2012). Review: quantifying mitochondrial dysfunction in complex diseases of aging. The journals of gerontology Series A, Biological sciences and medical sciences 67, 1022-1035.

34. Schaefer, A.M., McFarland, R., Blakely, E.L., He, L., Whittaker, R.G., Taylor, R.W., Chinnery, P.F., and Turnbull, D.M. (2008). Prevalence of mitochondrial DNA disease in adults. Annals of neurology 63, 35-39.

35. Gong, H., Yang, X., Zhao, Y., Petersen, R.B., Liu, X., Liu, Y., and Huang, K. (2014). Amyloidogenicity of p53: A Hidden Link Between Protein Misfolding and Cancer. Current protein & peptide science.

36. de la Monte, S.M., and Wands, J.R. (2008). Alzheimer's disease is type 3 diabetes-evidence reviewed. Journal of diabetes science and technology 2, 1101-1113.

37. Tornovsky-Babeay, S., Dadon, D., Ziv, O., Tzipilevich, E., Kadosh, T., Schyr-Ben Haroush, R., Hija, A., Stolovich-Rain, M., Furth-Lavi, J., Granot, Z., et al. (2014). Type 2 diabetes and congenital hyperinsulinism cause DNA double-strand breaks and p53 activity in beta cells. Cell metabolism 19, 109-121.

38. MacPherson, R.E., Baumeister, P., Peppler, W.T., Wright, D.C., and Little, J.P. (2015). Reduced cortical BACE1 content with one bout of exercise is accompanied by declines in AMPK, AKT, and MAPK signaling in obese, glucose intolerant mice. J Appl Physiol (1985), jap 00299 02015.

39. Sridhar, G.R., Lakshmi, G., and Nagamani, G. (2015). Emerging links between type 2 diabetes and Alzheimer's disease. World journal of diabetes 6, 744-751.

40. Wilson, C.M., Magnaudeix, A., Yardin, C., and Terro, F. (2014). Autophagy dysfunction and its link to Alzheimer's disease and type II diabetes mellitus. CNS & neurological disorders drug targets 13, 226-246.

41. Carvalho, C., Santos, M.S., Oliveira, C.R., and Moreira, P.I. (2015). Alzheimer's disease and type 2 diabetes-related alterations in brain mitochondria, autophagy and synaptic markers. Biochimica et biophysica acta 1852, 1665-1675.

42. Meng, Q., Makinen, V.P., Luk, H., and Yang, X. (2013). Systems Biology Approaches and Applications in Obesity, Diabetes, and Cardiovascular Diseases. Current cardiovascular risk reports 7, 73-83.

43. Chan, K.H., Huang, Y.T., Meng, Q., Wu, C., Reiner, A., Sobel, E.M., Tinker, L., Lusis, A.J., Yang, X., and Liu, S. (2014). Shared molecular pathways and gene networks for cardiovascular disease and type 2 diabetes mellitus in women across diverse ethnicities. Circulation Cardiovascular genetics 7, 911-919.

44. Zhao, Y., Chen, J., Freudenberg, J.M., Meng, Q., Rajpal, D.K., and Yang, X. (2016). Network-Based Identification and Prioritization of Key Regulators of Coronary Artery Disease Loci. Arteriosclerosis, thrombosis, and vascular biology 36, 928-941.

45. Meng Q, Y.Z., Noble E, Zhao Y, Agrawal R, Mikhail A, Zhuang Y, Tyagi E, Zhang Q, Lee JH, Morselli M, Orozco L, Guo Q, Kilts TM, Zhu J, Zhang B, Pellegrini M, Xiao X, Young MF, Gomez-Pinilla F, Yang X. (2016). Systems Nutrigenomics Reveals Brain Gene Networks Linking Metabolic and Brain Disorders. EBioMedcine.

46. Frischknecht, R., and Gundelfinger, E.D. (2012). The brain's extracellular matrix and its role in synaptic plasticity. Advances in experimental medicine and biology 970, 153-171.

47. Funderburgh, J.L. (2000). Keratan sulfate: structure, biosynthesis, and function. Glycobiology 10, 951-958.

48. Wyatt, A.R., Yerbury, J.J., Dabbs, R.A., and Wilson, M.R. (2012). Roles of extracellular chaperones in amyloidosis. Journal of molecular biology 421, 499-516.

49. Ahn, J.M., Kim, B.G., Yu, M.H., Lee, I.K., and Cho, J.Y. (2010). Identification of diabetic nephropathy-selective proteins in human plasma by multi-lectin affinity chromatography and LC-MS/MS. Proteomics Clinical applications 4, 644-653.

50. Todd, J.A. (1990). Genetic control of autoimmunity in type 1 diabetes. Immunol Today 11, 122-129.

51. Todd, J.A. (1997). Genetics of type 1 diabetes. Pathologie-biologie 45, 219-227.

52. Redondo, M.J., Fain, P.R., and Eisenbarth, G.S. (2001). Genetics of type 1A diabetes. Recent Prog Horm Res 56, 69-89.

53. Syed, M.A., Barinas-Mitchell, E., Pietropaolo, S.L., Zhang, Y.J., Henderson, T.S., Kelley, D.E., Korytkowski, M.T., Donahue, R.P., Tracy, R.P., Trucco, M., et al. (2002). Is type 2 diabetes a chronic inflammatory/autoimmune disease? Diabetes Nutr Metab 15, 68-83.

54. D'Andrea, M.R. (2005). Add Alzheimer's disease to the list of autoimmune diseases. Medical hypotheses 64, 458-463.

55. Pickup, J.C. (2004). Inflammation and activated innate immunity in the pathogenesis of type 2 diabetes. Diabetes care 27, 813-823.

56. Fulop, T., Jr., Larbi, A., Dupuis, G., and Pawelec, G. (2003). Ageing, autoimmunity and arthritis: Perturbations of TCR signal transduction pathways with ageing - a biochemical paradigm for the ageing immune system. Arthritis research & therapy 5, 290-302.

57. Son, S.M., Song, H., Byun, J., Park, K.S., Jang, H.C., Park, Y.J., and Mook-Jung, I. (2012). Accumulation of autophagosomes contributes to enhanced amyloidogenic APP processing under insulin-resistant conditions. Autophagy 8, 1842-1844.

58. Deng, H., and Mi, M.T. (2016). Resveratrol Attenuates Abeta Caused Neurotoxicity by Inducing Autophagy Through the TyrRS-PARP1-SIRT1 Signaling Pathway. Neurochemical research.

59. Ozbay, L.A., Moller, N., Juhl, C., Bjerre, M., Carstens, J., Rungby, J., and Jorgensen, K.A. (2012). Calcineurin inhibitors acutely improve insulin sensitivity without affecting insulin secretion in healthy human volunteers. British journal of clinical pharmacology 73, 536-545.

60. Wu, H.Y., Tomizawa, K., and Matsui, H. (2007). Calpain-calcineurin signaling in the pathogenesis of calcium-dependent disorder. Acta medica Okayama 61, 123-137.

61. Sima, A.A., and Li, Z.G. (2006). Diabetes and Alzheimer's disease - is there a connection? The review of diabetic studies : RDS 3, 161-168.

62. Holscher, C. (2011). Diabetes as a risk factor for Alzheimer's disease: insulin signalling impairment in the brain as an alternative model of Alzheimer's disease. Biochemical Society transactions 39, 891-897.

63. Jabir, N.R., Firoz, C.K., Baeesa, S.S., Ashraf, G.M., Akhtar, S., Kamal, W., Kamal, M.A., and Tabrez, S. (2015). Synopsis on the linkage of Alzheimer's and Parkinson's disease with chronic diseases. CNS neuroscience & therapeutics 21, 1-7.