

## Cross-Species Modeling Identifies Gene Signatures in Type 2 Diabetes Mouse Models Predictive of Inflammatory and Estrogen Signaling Pathways Associated with Alzheimer's Disease Outcomes in Humans

Brendan K. Ball<sup>1</sup>, Elizabeth A. Proctor<sup>2-6</sup>, Douglas K. Brubaker<sup>7-8†</sup>

<sup>1</sup>*Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN, USA.* <sup>2</sup>*Department of Neurosurgery, Penn State College of Medicine, Hershey, PA, USA.* <sup>3</sup>*Department of Pharmacology, Penn State College of Medicine, PA, USA.* <sup>4</sup>*Department of Biomedical Engineering, Penn State University, State College, PA, USA.* <sup>5</sup>*Center for Neuroengineering, Penn State University, State College, PA, USA.*

<sup>6</sup>*Department of Engineering Science & Mechanics, Penn State University, State College, PA, USA.* <sup>7</sup>*Center for Global Health & Diseases, Department of Pathology, School of Medicine, Case Western Reserve University, Cleveland, OH, USA.* <sup>8</sup>*Blood Heart Lung Immunology Research Center, University Hospitals, Cleveland, OH, USA.*

†*Email: dkb50@case.edu*

Alzheimer's disease (AD), the predominant form of dementia, is influenced by several risk factors, including type 2 diabetes (T2D), a metabolic disorder characterized by the dysregulation of blood sugar levels. Despite mouse and human studies reporting this connection between T2D and AD, the mechanism by which T2D contributes to AD pathobiology is not well understood. A challenge in understanding mechanistic links between these conditions is that evidence between mouse and human experimental models must be synthesized, but translating between these systems is difficult due to evolutionary distance, physiological differences, and human heterogeneity. To address this, we employed a computational framework called translatable components regression (TransComp-R) to overcome discrepancies between pre-clinical and clinical studies using omics data. Here, we developed a novel extension of TransComp-R for multi-disease modeling to analyze transcriptomic data from brain samples of mouse models of AD, T2D, and simultaneous occurrence of both disease (ADxT2D) and postmortem human brain data to identify enriched pathways predictive of human AD status. Our TransComp-R model identified inflammatory and estrogen signaling pathways encoded by mouse principal components derived from models of T2D and ADxT2D, but not AD alone, predicted with human AD outcomes. The same mouse PCs predictive of human AD outcomes were able to capture sex-dependent differences in human AD biology, including significant effects unique to female patients, despite the TransComp-R being derived from data from only male mice. We demonstrated that our approach identifies biological pathways of interest at the intersection of the complex etiologies of AD and T2D which may guide future studies into pathogenesis and therapeutic development for patients with T2D-associated AD.

*Keywords:* Alzheimer's disease, type 2 diabetes, preclinical translation, cross-species modeling, systems biology

### 1. Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by memory loss, confusion, and behavioral changes. With more than 6.9 million people living with AD in the

---

© 2024 The Authors. Open Access chapter published by World Scientific Publishing Company and distributed under the terms of the Creative Commons Attribution Non-Commercial (CC BY-NC) 4.0 License.

United States<sup>1</sup>, \$360 billion dollars in health and long-term care costs is expected to be spent in 2024, and projected to rise to \$1 trillion by 2050<sup>1</sup>. As the prevalence of AD is expected to increase with the country's aging population, developing effective therapeutics proven to treat or cure AD becomes urgent. Despite the rapid increase of AD cases, studies to develop therapeutics for AD is difficult<sup>2,3</sup>. This difficulty is in part due to the development of AD occurring decades before diagnosis<sup>4</sup> and the multi-factorial nature of the disease<sup>5-8</sup>.

In efforts to identify risk factors for AD it was observed that individuals with type 2 diabetes (T2D), a metabolic condition distinguished by chronic hyperglycemia, have an elevated risk in developing AD<sup>9,10</sup>. The development of T2D occurs decades before the diagnosis of AD and is reported to increase the risk of dementia<sup>11</sup>. In the United States, more than 39 million people have T2D, and 116 million have pre-diabetes<sup>12</sup>. This population of people diagnosed with or at risk for developing T2D may face a heightened risk for developing AD in light of the comorbidity of the diseases<sup>13,14</sup>. In clinical studies, common features of both AD and T2D include chronic inflammation<sup>15,16</sup>, increased insulin resistance<sup>17</sup>, and alterations to mitochondria and energy metabolism<sup>18,19</sup>. Despite multiple studies supporting a link between T2D and AD risk, the biological mechanisms by which this occurs are not well understood.

A critical challenge in understanding the mechanistic links between these conditions is that evidence must be synthesized and translated between experiments in mouse models and human-based clinical studies. Translating information from pre-clinical models to human clinical contexts is difficult due to discrepancies in interspecies physiology<sup>20</sup>, timeline of disease development<sup>21</sup>, and heterogeneity of the human population<sup>22</sup>. In cases of precision medicine, where complex dependencies between clinical phenotypes are difficult to deconvolute, such as is the case with AD and T2D, there is an important role for computational approaches to resolve this heterogeneity into testable mechanistic hypotheses to guide therapeutic development<sup>23-25</sup>.

To overcome this challenge, we developed a computational framework termed translatable components regression (TransComp-R) to identify omics-based signatures in mouse models predictive of AD conditions in human<sup>26-28</sup>. The TransComp-R model works by projecting human omics data into a mouse principal component analysis (PCA) space, followed by linear regression of mouse principal components (PCs) against human disease outcomes to identify translatable mouse PCs. The gene signatures encoded within mouse PCs that best separate conditions between human AD and control outcomes can be interpreted using biological pathway analyses such as gene set enrichment analysis (GSEA). These informed pathways can then be validated through literature and experimental studies.

Here, we aimed to perform a cross-species analysis using publicly available mouse and human transcriptomic data to determine biological pathways by which T2D contributes to AD. We developed a novel extension of TransComp-R that integrated PCs from multiple murine disease models: AD, T2D, and co-occurrence of both diseases (ADxT2D) in a single computational model to compare the predictive power of different murine models of disease and identify mouse-specific features predictive of human AD status. We also modified the existing TransComp-R method by incorporating human demographic variables such as sex and age variables into our model to inform the selection of translatable mouse PCs and better position the insights from the cross-species model to specific human patient subsets, an important goal of precision medicine. Our method synthesizes

mouse models with multiple disease etiologies with human information to prioritize biological pathways affected in disease and prospectively evaluate therapeutic avenues from pre-clinical to clinical contexts with high-throughput omics data.

## 2. Results

### 2.1. Selected mouse and human transcriptomic data were pre-processed for TransComp-R

Publicly available mouse (GSE152539)<sup>29</sup> and human (GSE48350)<sup>30,31</sup> datasets of microarrayed brain tissue samples were selected from Gene Expression Omnibus (GEO). The mouse dataset uniquely included conditions of solely AD, only T2D, and simultaneous occurrence of both diseases from the hippocampus. The mouse models consisted of six-month-old male *App<sup>NL-F/NL-F</sup>* knock-ins responsible for heightened amyloid-beta in the brain (Swedish KM670/671NL, Iberian I716F) and wild type (C57BL/6J) mice were fed with either a high-fat diet (custom diet, 40% kcal from fat, and 0.15% from cholesterol) or regular diet (CA-1, 18.8% kcal from fat) for 12 months ( $n = 3$  per condition) for the respective disease groups. The human dataset contains demographic variables of sex and age along with the transcriptomic data of AD ( $n = 80$ ) and control ( $n = 173$ ) subjects from four brain regions: hippocampus, entorhinal cortex, superior frontal cortex, and post-central gyrus.

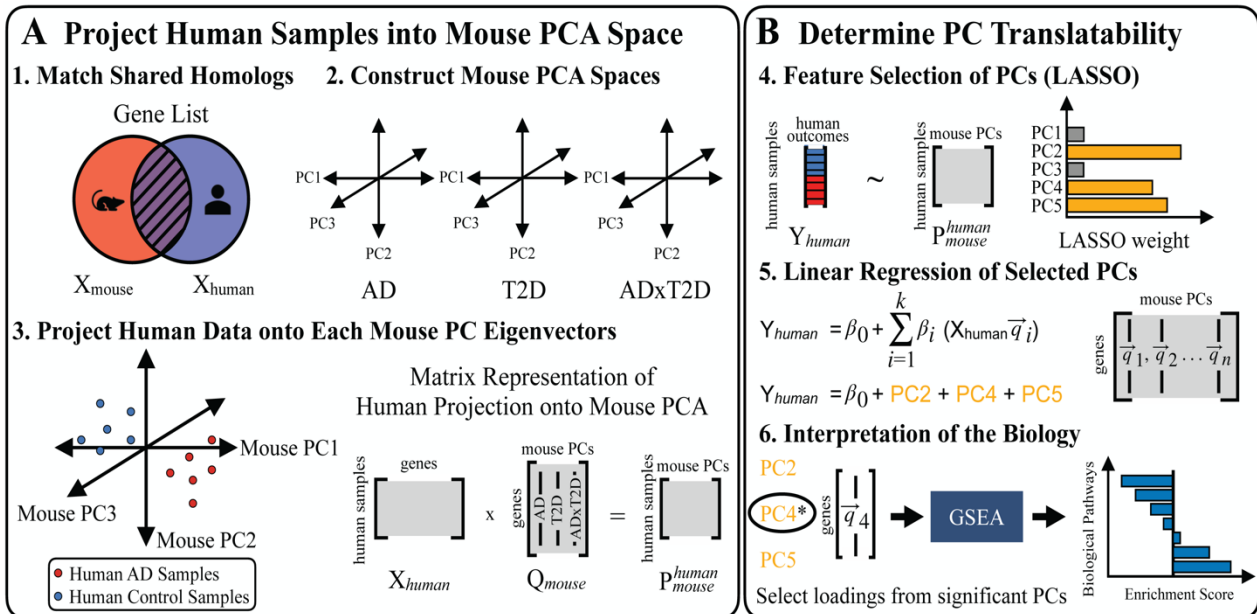
To prepare the data for the TransComp-R framework, both mouse and human transcriptomics datasets were matched for one-to-one homologs. From homolog matching, 13,428 genes were identified, and all other genes that did not have a matching homolog pair were excluded from the analysis. The human data was next filtered for the hippocampal region to account for brain-region variability. Any subjects below the age of 65 were removed from the study to reduce age bias (**Table 1**). Both datasets were individually  $\log_2$  transformed and normalized by z-score per gene.

**Table 1.** Summary of the processed human data across disease condition, age, and sex.

Condition	Age (years) (Mean $\pm$ SD)	Sex (%)		Total Sample Size ( $n$ )
		Female	Male	
Control	82.7 $\pm$ 9.5	11 (46%)	13 (54%)	24
Alzheimer's Disease	84.3 $\pm$ 6.6	9 (50%)	9 (50%)	18

### 2.2. TransComp-R modeling separates human samples in mouse principal component space

Here, we applied the TransComp-R methodology, with the incorporation of LASSO to select PCs most predictive of AD outcomes<sup>26</sup>. The TransComp-R model begins with the projection of human data into the mouse PCA space (**Figure 1A**), followed by the evaluation of mouse PC translatability through LASSO and generalized linear model (GLM) regression. The significant mouse PCs that can distinguish between human AD and control are interpreted by GSEA of the gene loading coefficients on each PC (**Figure 1B**). The biological pathways identified from GSEA can provide insight on human biology translated by mouse PCs, which can then be validated through follow-up experiments and literature review.



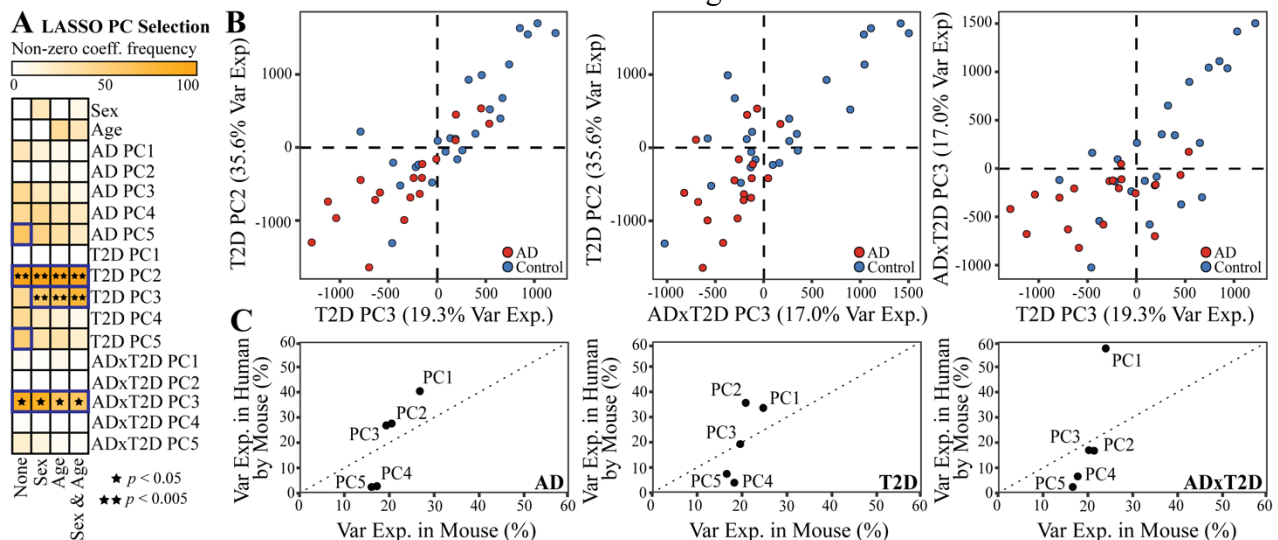
**Figure 1. The TransComp-R computational approach. (A)** Homolog gene pairs between human and mouse datasets are selected for analysis. Human samples are projected into mouse PCA spaces to combine mouse and human information. **(B)** Principal component translatability from mouse to human is determined by performing a GLM regression against human AD outcomes with PCs selected from LASSO. The loadings from the significant PCs are analyzed via GSEA to identify enriched biological pathways.

Implementing this approach, mouse data were separated into AD, T2D, and ADxT2D with controls prior to constructing separate PCA models, such that three groups of PCs encoded transcriptomic variation between healthy controls and AD, T2D, or ADxT2D mice. To avoid overfitting the mouse data, a threshold of 80% cumulative variance explained was set for each PCA, and as a result, a total of five PCs per disease group were selected (**Supplementary Figure S1**). Next, the human data was projected on the mouse PCA space.

We then trained four separate LASSO models to identify PCs most predictive of binarized human disease outcomes. Using the combined dataset containing rows of human and columns of T2D PCs, we progressively included human demographic variables associated with the respective human subjects in LASSO such that we examined: models of only mouse PCs, PCs with human sex, PCs with human age, and PCs with both human sex and age main effects. This approach allows us to include human demographic variables in a cross-species translation model, prioritizing not just mouse PCs, but also how mouse PCs capture the heterogeneity of human sex and age when predicting AD outcomes. The PCs were next selected based on 100 rounds of 5-fold cross-validation, where PCs with a significant LASSO coefficient in greater than half of the models were carried forward to the GLM.

From the LASSO models, we found T2D PC2 and ADxT2D PC3 to be consistently selected across all four LASSO models, while T2D PC3 was selected from all models except for the model with only mouse PC main effects (**Figure 2A**). Additionally, AD PC5 and T2D PC5 were selected from the LASSO model with only mouse PCs as main effect variables, but not in other LASSO models that included human demographic variables. The PCs identified by the LASSO model,

which were encoded with transcriptomic variance, were next evaluated for their respective ability to discern between human AD and control status through GLM.



**Figure 2. TransComp-R identifies translatable PCs predictive for AD outcomes in human.** (A) Selection of PCs using LASSO across 100 rounds of 5-fold cross-validation. The four LASSO models included terms with just mouse PCs, PCs and human sex, PCs and human age, and PCs and human age and sex. PCs with a coefficient frequency greater than 50 rounds of 100 were selected for the GLM and regressed against binarized human disease outcomes (significance defined by simple regression model  $p$  value). (B) A principal component plot of human scores on the selected mouse T2D PC2, T2D PC3, and ADxT2D PC3 separating human control and AD outcomes (C) Mouse PCs were separated by disease cohort, comparing the variance explained in mice to the variance in humans explained by mouse respective mouse PCs.

To evaluate the predictability of the selected mouse PCs for human AD, we constructed GLMs with all selected PCs predicting AD status in humans, but these multi-PC models were not significantly predictive due to multi-collinearity (**Supplementary Figure S2**). As a result, we constructed GLMs for each individual PC regressed against human disease outcomes. We found the three mouse PCs consistently selected from LASSO to be predictive of human AD outcomes individually (T2D PC2  $p = 0.0047$ , T2D PC3  $p = 0.0042$ , and ADxT2D PC3  $p = 0.0130$ ) (**Figure 2A**). We also note that although AD PC5 and T2D PC5 satisfied the non-zero frequency greater than 50 in the LASSO model with only mouse PCs, the regression against human outcomes was not significant, and was excluded from further analysis (AD PC5  $p = 0.275$ , T2D PC5  $p = 0.443$ ). Consistent LASSO selection of T2D PC2, T2D PC3, and ADxT2D PC3 as significant PCs indicates the importance of including human clinical and demographic variables in the TransComp-R model to detect translatable cross-species biology while controlling for clinical covariates.

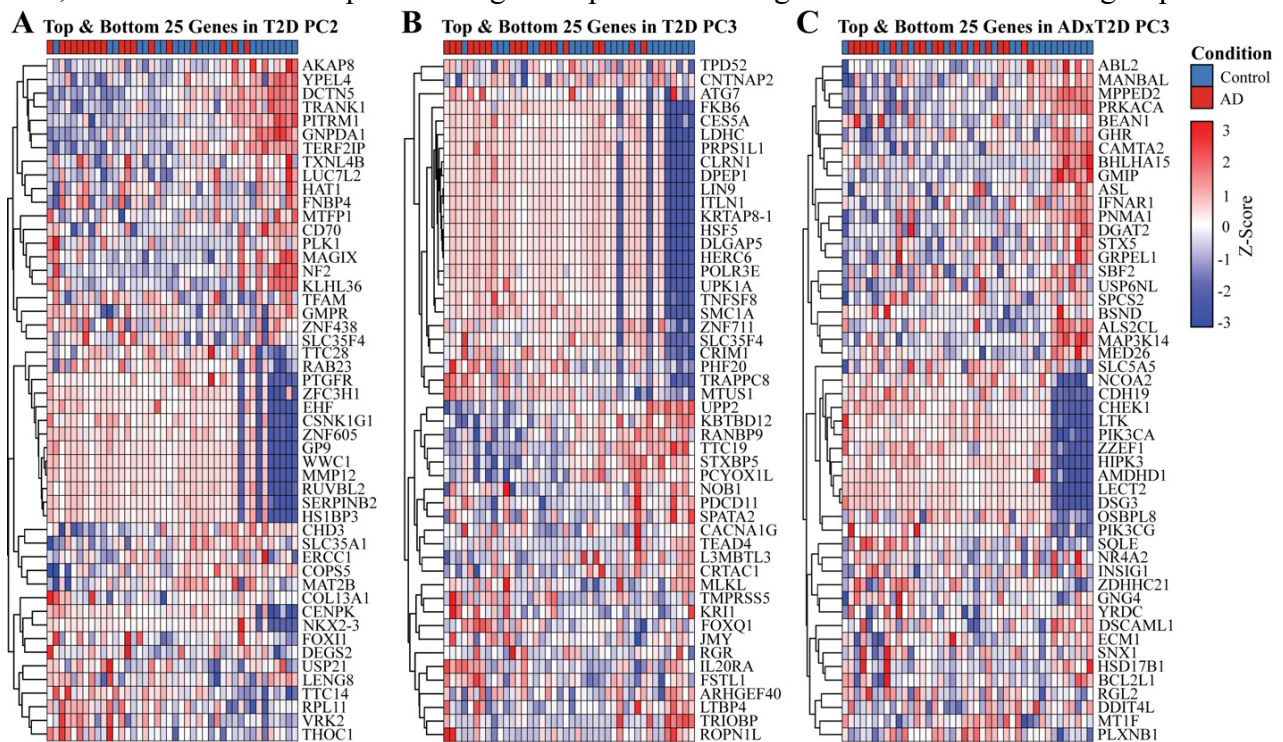
We visualized the two T2D mouse PCs and one ADxT2D mouse PC that were identified by TransComp-R as predictive of human AD status (**Figure 2B**). In all three PCs, there was visible separation between the control and AD groups. We next compared the translatability of the selected mouse PCs to their ability to explain the variance in human data (**Figure 2C**). Comparing the proportion of PC variance explained in mouse to the variance explained in human by the same mouse PC, we found that T2D PC3 and ADxT2D PC3 explained a similar ratio, whereas mouse T2D PC2 explained almost double the variance in human by mouse than the mouse PCs alone. This



could imply that certain pathways represented by mouse T2D PC3 and ADxT2D PC3 were conserved consistently across mice and humans, whereas mouse T2D PC2 may have had a more pronounced effect in capturing information cross-species.

### 2.3. Mouse principal components selected genes contribute to human disease separation

Having identified three mouse PCs predictive of human AD versus control status from TransCompR, we were interested in isolating genes that were contributing to the separation between human AD and control subjects. Filtering for human genes ranked with the top and bottom 25 loadings within their respective PCs, we identified genes in the model predictive of AD and control in humans (**Figure 3A-C**). While no genes were shared across the top and bottom 25 ranked on the three mouse PCs, we observed distinct patterns of gene expression among human AD and control groups.

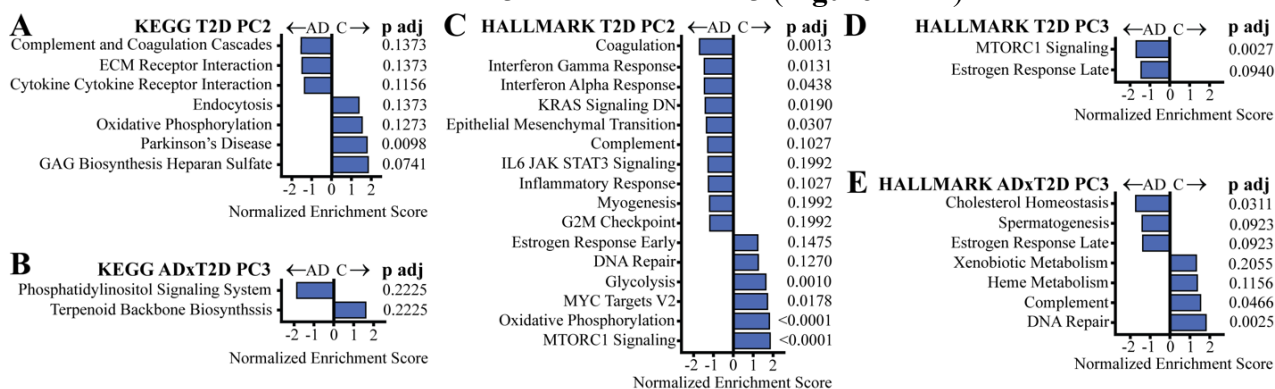


**Figure 3.** The top and bottom 25 genes of translatable PCs. Z-scored AD human transcriptomic data were filtered by genes with the 25 largest and smallest scores on (A) T2D PC2, (B) T2D PC3, and (C) ADxT2D PC3. Human samples were sorted by their respective PC scores with the most negative (left) to the most positive (right).

### 2.4. Gene set enrichment analysis identifies inflammatory and estrogen signaling pathways enriched in human Alzheimer's disease outcome

We performed GSEA on the selected T2D and ADxT2D PCs and identified pathways associated with inflammatory and estrogen signaling. From the KEGG database, we identified “Complement and Coagulation Cascades” and “Cytokine-Cytokine Receptor Interaction” on T2D PC2 (**Figure 4A**). On T2D PC3, the “Phosphatidylinositol Signaling System” was the only pathway found to be enriched for AD (**Figure 4B**). There were no significant KEGG pathways on T2D PC3. On the Hallmark database, we identified “Interferon Gamma Response,” “Interferon Alpha Response,”

“IL6 JAK STAT Signaling,” and “Inflammatory Response” to be enriched for AD conditions by T2D PC2 (**Figure 4C**). Interestingly in the Hallmark database, we identified “Estrogen Response Early” was enriched for the control group in T2D PC2 (**Figure 4C**), while “Estrogen Response Late” was enriched for AD in T2D PC3 and ADxT2D PC3 (**Figure 4D-E**).

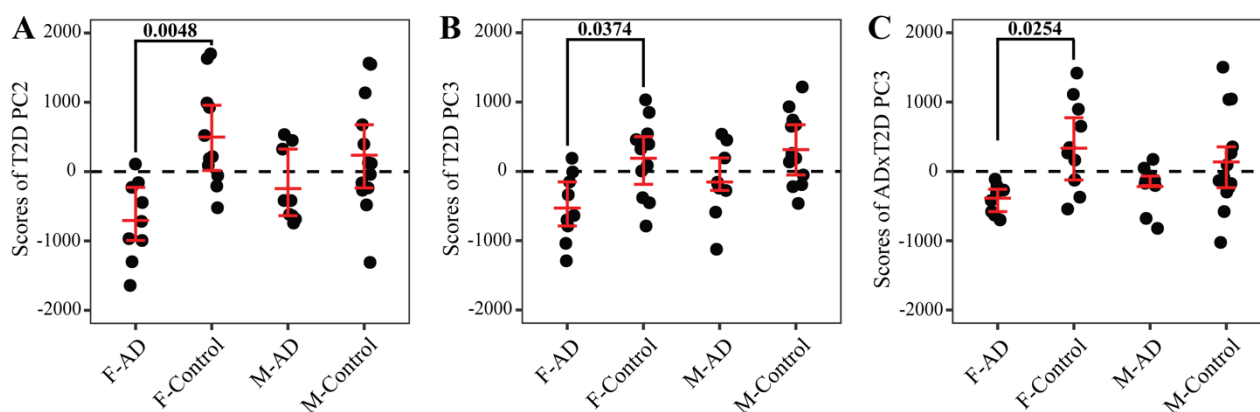


**Figure 4. Enriched biological pathways identified from GSEA.** Significant KEGG pathways from (A) T2D PC2 and (B) ADxT2D PC3. No significant pathways were enriched in T2D PC3. Significant Hallmark pathways were identified for (C) T2D PC2, (D) T2D PC3, and (E) ADxT2D PC3. Enriched pathways were defined by a Benjamini-Hochberg adjusted  $p$  value  $< 0.25$ . Pathways enriched for AD are displayed with a negative normalized enrichment score.

Based on our findings with the estrogen pathways, we were interested in distinguishing the genes that contributed to “Estrogen Response Early” and “Estrogen Response Late.” From GSEA, we identified 76, 77, and 49 core enrichment genes contributing to the estrogen-associated pathways in mouse T2D PC2, T2D PC3, and ADxT2D PC3, respectively. Comparing the genes that were contributing to the estrogen response, we found 23 shared genes between T2D PC3 and ADxT2D PC3 (*PDZK1*, *LLGL2*, *KLK11*, *TOP2A*, *PTGES*, *FARP1*, *NAB2*, *CISH*, *MEST*, *KIF20A*, *LTF*, *ISG20*, *IMPA2*, *DUSP2*, *PLAC1*, *PRKAR2B*, *TNNC1*, *OPN3*, *AREG*, *ATP2B4*, *AGR2*, *CALCR*, and *RABEP1*), 2 genes between T2D PC2 and ADxT2D PC3 (*DHCR7* and *MAPT*), 10 genes between T2D PC2 and T2D PC3 (*TPBG*, *FKBP4*, *GLA*, *NXT1*, *CD44*, *PGR*, *RAB31*, *AFF1*, *TFAP2C*, and *TJP3*), and 5 genes shared across all three mouse PCs (*SULT2B1*, *OVOL2*, *SIAH2*, *FDFT1*, and *RBBP8*) (**Supplementary Figure S3**). Additionally, 19 genes enriched in ADxT2D PC3, 39 genes enriched in T2D PC3, and 59 genes enriched in T2D PC2 did not overlap with any other mouse PCs.

## 2.5. Male mouse-derived principal components significantly stratify female Alzheimer’s disease and control groups in human subjects

Expanding upon the potential sex-based predictability, we were curious to see if the model was able to distinguish sex and disease status by the PC scores. Here, we separated the scores of each mouse PC by human sex and AD status and found that mouse T2D PC2, T2D PC3, and ADxT2D PC3 significantly stratified human female AD and control groups, and not male AD and control groups, despite the mouse data originating from all male mice (**Figure 5A-C**). The ability of these PCs to distinguish between female AD and control groups shows the model’s ability in identifying human sex-based differences in the context of disease development. This is supported by the significance of the separation between the two groups ( $p$  value  $< 0.05$ ).



**Figure 5. Comparison of sex and disease status among the translatable PCs.** Scores of each PC were separated by female (F) AD, female control, male (M) AD, and male control for (A) T2D PC2, (B) T2D PC3, and (C) ADxT2D PC3. A Mann-Whitney pair-wise test corrected by the Benjamini-Hochberg method (FDR  $q$  value < 0.05) was used to determine the significance among the groups. The mean of the distribution is labeled with the interquartile range.

### 3. Discussion

In this work, we aimed to uncover potential biological mechanisms that connected T2D as a risk factor for AD development using mouse and human transcriptomic data. An obstacle in understanding the links between these diseases, in which multifactorial mechanisms interact in humans and biological mechanisms are isolated in animal studies, is that information from mouse models and human-based studies must be synthesized to inform clinical and therapeutic decisions. Currently, translating information from pre-clinical models to patient-specific contexts is often difficult due to discrepancies in interspecies physiology<sup>20</sup>, timeline of disease development<sup>21</sup>, and heterogeneity of the human population<sup>22</sup>. To overcome these challenges, we innovated on TransComp-R to identify potential biological pathways from mouse PCs that are predictable for AD outcomes. In the TransComp-R workflow, we fused multiple mouse disease models in a single computational model together with human data containing demographic sex and age variables to predict outcomes in AD. With our computational model, we pinpointed potential biological pathways associated with AD, and identified sex-specific differences, despite the mouse disease models being representative of only males.

We identified inflammatory pathways that may link T2D as a risk factor for AD development. These links have the potential translational utility in bridging mouse and human biology to understand and develop therapeutic strategies for AD with T2D exacerbating factors. The mouse T2D PC2 identified several pathways on both KEGG and hallmark databases. From KEGG, “Complement and Coagulation Cascades” and “Cytokine-Cytokine Receptor Interaction” were enriched for AD. From the literature, studies report complement activation to be associated with insulin resistance and T2D<sup>32–34</sup>. Likewise, high complement levels are contributed by neurons and glial cells in AD<sup>35,36</sup>. In both T2D and AD, cytokines are found to actively participate in the progression of disease<sup>37,38</sup>.

Using complementary pathway databases, we identified “Interferon Gamma Response,” “Interferon Alpha Response,” “IL6 JAK STAT Signaling,” and “Inflammatory Response” pathways on mouse T2D PC2 enriched in human AD. Interferon gamma<sup>39</sup> and alpha<sup>40</sup>, key cytokines in the



innate immune response and response to viral infections, are altered in AD. However, we notice that interferon gamma<sup>41</sup> is more associated with T2D, whereas interferon alpha<sup>42,43</sup> is found to be elevated in subjects with type 1 diabetes instead<sup>44</sup>. IL6 JAK-STAT signaling has been reported to impair the insulin-degrading enzyme, a protein found to be associated with obesity and T2D<sup>45</sup>. In AD, IL6 signaling has been linked with cognitive impairment and metabolic alterations<sup>46</sup>. Collectively, these results may indicate that chronic inflammation could lead to downstream insulin resistance and cognitive deficits<sup>47</sup>.

Our results also indicate that estrogen signaling may serve as a potential connection between T2D and AD. From GSEA, ranked genes in ADxT2D PC3 and T2D PC3 both identified “Estrogen Response Late” as pathways enriched for AD, whereas “Estrogen Response Early” was enriched for human control by T2D PC2. Among the three PCs, 49 genes were enriched for ADxT2D PC3, 77 genes were enriched for T2D PC2, and 76 genes were enriched for T2D PC2. Of these, 59 were enriched in T2D PC2, and 23 were shared between ADxT2D PC3 and T2D PC3, in which we compared with previously published literature to potential associations with AD and T2D. Associated with AD in the mouse T2D PC2, we identified *MEDI3L*<sup>48</sup> and *XBPI*<sup>49</sup> connected to cognitive deficits, changes in mitochondrial metabolism (*PMAIP1*)<sup>50</sup>, inflammation (*RASGRP1*)<sup>51</sup>, and the expression of *NRIP1*<sup>52</sup> reduced in AD. Similarly, we identified genes associated with insulin resistance (*FASN* and *FKBP5*)<sup>53,54</sup>, genetic variances of *RAPGEF1*<sup>55</sup> and increased expression of *AQP3*<sup>56</sup> related to T2D development. Interpreting genes shared across ADxT2D PC3 and T2D PC3, both PCs, we found *MEST*<sup>57</sup> reported to alter Wnt signaling in AD, and *KIF20A*<sup>58</sup>, a gene found to be differentially expressed in AD. Likewise in T2D, we found *CISH*<sup>59</sup> to be involved with gluconeogenesis, whereas beta-cells were preserved with upregulated *AGR2*<sup>60</sup>.

There were five genes shared across the three mouse PCs identifying estrogen signaling as a potential biological pathway, which included *SULT2B1*, *OVOL2*, *SIAH2*, *FDFT1*, and *RBBP8*. Of the five genes, all but *RBBP8* were reported to have connections to AD or T2D in literature. *SULT2B1*, part of the sulfotransferase family that catalyzes the sulfate conjunction of hormones and neurotransmitters, was found to be upregulated in AD rat models<sup>61</sup>. In a T2D study, *SULT2B1* overexpressed in the liver inhibited hepatic gluconeogenesis in two separate diabetic mouse models: one induced by high-fat diet, and another via leptin-deficiency (*ob/ob*)<sup>62</sup>. Other genes related to T2D include *OVOL2* and *SIAH2*. The presence of *OVOL2* was found to be linked with beta cell dedifferentiation, a mechanism linked with pancreatic dysfunction<sup>63</sup>, and *SIAH2* deficiency improved glucose and insulin tolerance<sup>64</sup>. Related to AD, inhibition of squalene synthase (*FDFT1*) inhibited by squalostatatin reduced cellular prion protein in ScN2a, SMB, and ScGT1 (prion-infected cell lines)<sup>65</sup>, and protection against amyloid beta-induced synapse damage<sup>66</sup>. Further examination of these genes may be of potential interest to connect biological pathways between T2D and AD.

Interestingly in both diseases, previous studies report that estrogen may play a protective role in AD<sup>67</sup> and T2D<sup>68</sup>. In AD, estrogen provides protection from amyloid-beta toxicity, a hallmark of AD pathology<sup>69,70</sup>. In females that experienced menopause, hormone therapy with estrogen has been found to reduce the risk of T2D onset<sup>68</sup>. Although studies indicate estrogen to be protective, others report that estrogen may be deleterious depending on the timing and onset of T2D<sup>71,72</sup>. These differences could be a result of the varying roles that different genes may have: some genes may contribute to disease when upregulated, while others may serve a protective role that can lead to

disease if downregulated. This variability in genes could further explain the possible observation of estrogen appearing to have both harmful and protective effects. Therefore, further investigations are encouraged to further understand the role of estrogen as a shared pathway between AD and T2D.

Finally, we found that the mouse PCs defined by T2D (PC2 and PC3) and ADxT2D (PC3) were able to distinguish between female AD and control subjects. Despite the mouse groups being entirely male, our model detected sex-based differences in females. This is interesting because females are at a higher risk of developing AD than males<sup>73</sup>. Observing this result, as well as PCs showing enrichment for estrogen, may suggest that despite the widespread lack of female animals in preclinical research, our model is able to detect biological signals in male mice predictive of female human disease biology, thereby enhancing the retrospective utility of prior animal studies that fell short of equitable design. In the specific case of our models, the pathways we identified on the male mouse PC's predictive of human female AD pathology implicate our model's ability to translate transcriptomic signatures across human sex demographics.

There are limitations and opportunities to expand this study. Few research groups have explored the T2D-AD axis, and as a result, there are limited sample sizes available for mouse and human omics data. The incorporation of additional studies that satisfy the criteria of our selection process into the model may improve the confidence of these results. Second, our TransComp-R model only considers homologous gene pairs shared across mice and humans. As a result, we potentially omit genes in pre-processing that may be involved in the development of AD. Additionally, the GLM in our model only regresses against control or AD status without the incorporation of transient phases such as mild cognitive impairment. Finally, the TransComp-R framework has the opportunity to consider other clinical variables that may predict disease outcomes. Some additional factors include information on race, clinical neuropathological scores for AD severity, and current T2D biomarkers. Considering these potential factors may further enhance future cross-species modeling.

Our work expanded upon the existing TransComp-R framework to identify potential biological pathways in which T2D may exacerbate AD development. We show that mouse PCs from T2D and ADxT2D were most predictive of AD outcomes in human. Interestingly, mouse PC's derived from mice with AD alone were not predictive of human AD, which may indicate that metabolic dysfunction encoded on the mouse T2D and T2DxAD PCs plays a more significant role in human AD biology than is typically accounted for. Indeed, these results encourage future applications of TransComp-R to overcome barriers of pre-clinical to human studies and identify affected biological pathways in AD or different diseases. The implications of this work for precision medicine can be expanded to other disease models that may be difficult to synthesize between pre-clinical experiments and clinical studies. This platform could synthesize various omics data from pre-clinical and patient-specific data to rationally select potential pathways to target, which may further enhance clinical studies or possible therapeutic avenues.

## 4. Materials and Methods

### 4.1. Data selection

Mouse and human datasets were selected with the criteria of matching hippocampal brain region, information containing AD and T2D conditions in the mouse dataset, human sample size greater

than 12 per condition, and at least sex and age information in the human dataset. Additionally, datasets derived from similar sequencing platforms were prioritized. Search terms on GEO included phrases such as “hippocampus Alzheimer’s disease in human,” “mouse Alzheimer’s disease hippocampus,” and “mouse diabetes hippocampus.” Additional searches included the term “gene expression” on the GEO repository.

#### 4.2. *Pre-processing and normalization*

Publicly available transcriptomic human and mouse data were obtained from the GEO repository using Bioconductor tools in R (*GEOquery* 2.70.0, *limma* 3.58.1, and *Biobase* 2.62.0)<sup>74–76</sup>. Before processing, all human subjects with a reported age below 65 years old were removed from the analysis to prevent bias from younger age groups. The imported datasets were  $\log_2$  transformed, then human and mouse gene lists were matched for homologous pairs (*orthogene* 1.8.0)<sup>77</sup>. The two datasets were filtered for the hippocampal brain region. The genes were then internally normalized by z-score prior to TransComp-R modeling.

#### 4.3. *Cross-species modeling and variable selection*

We applied TransComp-R by conducting PCA on the mouse data separated in AD, T2D, and ADxT2D groups with controls, such that three groups of PCs encoded transcriptomic variation between healthy controls and AD, T2D, or ADxT2D mice. To avoid overfitting, the number of PCs in its respective group was limited to an 80% cumulative variance explained. Human AD and control subjects were projected into mouse PCA space. Mouse PCs associated with AD outcomes in human were selected by performing LASSO across 100 rounds of 5-fold cross-validation regressing the human positions in mouse PC space against human disease status. Four sets of LASSO models were trained, including main effects of mouse PCs, PCs and human sex, PCs and human age, and PCs and human age and sex. PCs with a coefficient frequency greater than 50 of the 100 rounds were selected for GLMs with individual PCCs and human clinical covariates regressed against human AD outcomes. The significance of the PC was determined if the model  $p$  value was less than 0.05.

#### 4.4. *Variance explained in human by mouse principal components*

Human data containing subject information and gene lists, as well as mouse PCs with a matching gene list, was used to calculate the variance explained by mouse in human. Using mouse PCs in the columns of  $Q$ , we projected the human data matrix  $X$  onto the PCs via matrix multiplication and calculated the percent variance of mouse in  $X$  explained by a given column  $q_i$  of  $Q$  (with  $T$  representing the matrix transpose) as:

$$\text{VarExpHuman}(q_i) = \frac{q_i^T [X^T X] q_i}{\sum \text{diag}(Q^T X^T X Q)} \quad (1)$$

#### 4.5. *Identifying genes contributing to human separation by mouse principal components*

Genes contributing to the most positive and negative scores were identified by selecting loaded genes with the top 25 and bottom 25 scores in each of the selected PCs. The selected genes were then used to filter the gene list of the human dataset containing z-scored gene expression data. A

heatmap, with the human subjects, clustered by their scores from the TransComp-R model, and the 50 total genes were visualized to compare gene expression between AD and control.

#### **4.6. Gene set enrichment analysis**

GSEA was performed on the loadings of selected PCs from the GLM in R (*msigdb* 7.5.1, *fgsea* 1.28.0, and *clusterProfiler* 4.10.1)<sup>78–80</sup>. From the Molecular Signatures Database, two human collections to perform GSEA included the KEGG and Hallmark databases. The parameters for the minimum gene set size and the maximum gene set size were set to 5 and 500, respectively. The tuning constant, epsilon, was established at 0. For both KEGG and Hallmarks databases, enriched biological pathways were determined significant if the Benjamini-Hochberg adjusted  $p$  value was less than 0.25.

#### **4.7. Sex-based comparison across principal component scores**

As an approach to compare predictability across sex, scores of selected PCs were separated by sex and disease categories. A Mann-Whitney pair-wise test was used to determine significance among four groups (AD females, control females, AD males, and control males). To correct for multiple comparisons,  $p$  values were adjusted with the Benjamini-Hochberg factor. An adjusted  $p$  value less than 0.05 was considered significant for the analysis.

### **Acknowledgments**

This work is supported by an award from the Good Ventures Foundation and Open Philanthropy (DKB and BKB). BKB is supported by the National Science Foundation through the Graduate Research Fellowship program under grant DGE-1842166. BKB also acknowledges the NIH T32 predoctoral fellowship T32DK101001 from the National Institute of Diabetes and Digestive and Kidney Diseases. This work is also supported by the National Institute on Aging under R01AG072513 (EAP).

### **Author Contributions**

BKB: Conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, visualization, writing-original draft, writing-review & editing. EAP: Conceptualization, methodology, writing-review & editing. DKB: Conceptualization, funding acquisition, methodology, project administration, resources, writing review & editing.

### **Competing Interests**

The authors declare no competing interests.

### **Supplementary Material & Code Availability**

All supplementary figures, documents, and code for this study are made publicly available at <https://github.com/Brubaker-Lab/MouseHuman-TransCompR-T2D-AD>.

## References

1. Alzheimer's Association. 2024 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* **n/a**, (2024).
2. Asher, S. & Priefer, R. Alzheimer's disease failed clinical trials. *Life Sciences* **306**, 120861 (2022).
3. Cummings, J., Feldman, H. H. & Scheltens, P. The “rights” of precision drug development for Alzheimer's disease. *Alzheimer's Research & Therapy* **11**, 76 (2019).
4. Rajan, K. B., Wilson, R. S., Weuve, J., Barnes, L. L. & Evans, D. A. Cognitive impairment 18 years before clinical diagnosis of Alzheimer disease dementia. *Neurology* **85**, 898–904 (2015).
5. Niu, H., Álvarez-Álvarez, I., Guillén-Grima, F. & Aguinaga-Ontoso, I. Prevalence and incidence of Alzheimer's disease in Europe: A meta-analysis. *Neurología (English Edition)* **32**, 523–532 (2017).
6. Ferretti, M. T. *et al.* Sex differences in Alzheimer disease — the gateway to precision medicine. *Nat Rev Neurol* **14**, 457–469 (2018).
7. Huang, Y.-W. A., Zhou, B., Wernig, M. & Südhof, T. C. ApoE2, ApoE3, and ApoE4 Differentially Stimulate APP Transcription and A $\beta$  Secretion. *Cell* **168**, 427–441.e21 (2017).
8. Nedelec, T. *et al.* Identifying health conditions associated with Alzheimer's disease up to 15 years before diagnosis: an agnostic study of French and British health records. *The Lancet Digital Health* **4**, e169–e178 (2022).
9. Chatterjee, S. *et al.* Type 2 Diabetes as a Risk Factor for Dementia in Women Compared With Men: A Pooled Analysis of 2.3 Million People Comprising More Than 100,000 Cases of Dementia. *Diabetes Care* **39**, 300–307 (2015).
10. Janson, J. *et al.* Increased Risk of Type 2 Diabetes in Alzheimer Disease. *Diabetes* **53**, 474–481 (2004).
11. Zilkens, R. R., Davis, W. A., Spilbury, K., Semmens, J. B. & Bruce, D. G. Earlier Age of Dementia Onset and Shorter Survival Times in Dementia Patients With Diabetes. *American Journal of Epidemiology* **177**, 1246–1254 (2013).
12. Martin, S. S. *et al.* 2024 Heart Disease and Stroke Statistics: A Report of US and Global Data From the American Heart Association. *Circulation* **149**, e347–e913 (2024).
13. Jash, K. *et al.* Cognitive dysfunction: A growing link between diabetes and Alzheimer's disease. *Drug Development Research* **81**, 144–164 (2020).
14. Kubis-Kubiak, A., Dyba, A. & Piwowar, A. The Interplay between Diabetes and Alzheimer's Disease—In the Hunt for Biomarkers. *International Journal of Molecular Sciences* **21**, 2744 (2020).
15. Donath, M. Y. & Shoelson, S. E. Type 2 diabetes as an inflammatory disease. *Nat Rev Immunol* **11**, 98–107 (2011).
16. Akiyama, H. *et al.* Inflammation and Alzheimer's disease. *Neurobiology of Aging* **21**, 383–421 (2000).
17. Yoon, J. H. *et al.* How Can Insulin Resistance Cause Alzheimer's Disease? *International Journal of Molecular Sciences* **24**, 3506 (2023).
18. Bingham, E. M. *et al.* The Role of Insulin in Human Brain Glucose Metabolism. *Diabetes* **51**, 3384–3390 (2002).
19. Hirvonen, J. *et al.* Effects of Insulin on Brain Glucose Metabolism in Impaired Glucose Tolerance. *Diabetes* **60**, 443–447 (2011).
20. Brubaker, D. K. & Lauffenburger, D. A. Translating preclinical models to humans. *Science* **367**, 742–743 (2020).
21. Saraceno, C., Musardo, S., Marcello, E., Pelucchi, S. & Diluca, M. Modeling Alzheimer's disease: from past to future. *Front. Pharmacol.* **4**, (2013).
22. Gauthier, S. *et al.* Why has therapy development for dementia failed in the last two decades? *Alzheimer's & Dementia* **12**, 60–64 (2016).
23. Brubaker, D. K., Proctor, E. A., Haigis, K. M. & Lauffenburger, D. A. Computational translation of genomic responses from experimental model systems to humans. *PLoS Comput Biol* **15**, e1006286 (2019).
24. Brubaker, D. K. *et al.* Proteogenomic Network Analysis of Context-Specific KRAS Signaling in Mouse-to-Human Cross-Species Translation. *Cell Systems* **9**, 258–270.e6 (2019).
25. Holland, C. H., Szalai, B. & Saez-Rodriguez, J. Transfer of regulatory knowledge from human to mouse for functional genomics analysis. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1863**, 194431 (2020).
26. Brubaker, Douglas. K. *et al.* An interspecies translation model implicates integrin signaling in infliximab-resistant inflammatory bowel disease. *Sci Signal* **13**, eaay3258 (2020).
27. Suarez-Lopez, L. *et al.* Cross-species transcriptomic signatures predict response to MK2 inhibition in mouse models of chronic inflammation. *iScience* **24**, 103406 (2021).
28. Lee, M. J. *et al.* Computational Interspecies Translation Between Alzheimer's Disease Mouse Models and Human Subjects Identifies Innate Immune Complement, TYROBP, and TAM Receptor Agonist Signatures, Distinct From Influences of Aging. *Frontiers in Neuroscience* **15**, (2021).



29. Mazzei, G. *et al.* A high-fat diet exacerbates the Alzheimer's disease pathology in the hippocampus of the AppNL-F/NL-F knock-in mouse model. *Aging Cell* **20**, e13429 (2021).
30. Berchtold, N. C. *et al.* Synaptic genes are extensively downregulated across multiple brain regions in normal human aging and Alzheimer's disease. *Neurobiol Aging* **34**, 1653–1661 (2013).
31. Cribbs, D. H. *et al.* Extensive innate immune gene activation accompanies brain aging, increasing vulnerability to cognitive decline and neurodegeneration: a microarray study. *J Neuroinflammation* **9**, 179 (2012).
32. Shim, K., Begum, R., Yang, C. & Wang, H. Complement activation in obesity, insulin resistance, and type 2 diabetes mellitus. *World J Diabetes* **11**, 1–12 (2020).
33. Tan, S. M., Snelson, M., Østergaard, J. A. & Coughlan, M. T. The Complement Pathway: New Insights into Immunometabolic Signaling in Diabetic Kidney Disease. *Antioxid Redox Signal* **37**, 781–801 (2022).
34. Yudkin, J. S., Stehouwer, C. D. A., Emeis, J. J. & Coppack, S. W. C-Reactive Protein in Healthy Subjects: Associations With Obesity, Insulin Resistance, and Endothelial Dysfunction. *Arteriosclerosis, Thrombosis, and Vascular Biology* **19**, 972–978 (1999).
35. Luchena, C., Zuazo-Ibarra, J., Alberdi, E., Matute, C. & Capetillo-Zarate, E. Contribution of Neurons and Glial Cells to Complement-Mediated Synapse Removal during Development, Aging and in Alzheimer's Disease. *Mediators Inflamm* **2018**, 2530414 (2018).
36. Goetzl, E. J., Schwartz, J. B., Abner, E. L., Jicha, G. A. & Kapogiannis, D. High complement levels in astrocyte-derived exosomes of Alzheimer's disease. *Ann Neurol* **83**, 544–552 (2018).
37. Sankar, S. B. *et al.* Amyloid beta and diabetic pathology cooperatively stimulate cytokine expression in an Alzheimer's mouse model. *J Neuroinflammation* **17**, 38 (2020).
38. Ball, B. K., Kuhn, M. K., Fleeman Bechtel, R. M., Proctor, E. A. & Brubaker, D. K. Differential responses of primary neuron-secreted MCP-1 and IL-9 to type 2 diabetes and Alzheimer's disease-associated metabolites. *Sci Rep* **14**, 12743 (2024).
39. Browne, T. C. *et al.* IFN- $\gamma$  Production by Amyloid  $\beta$ -Specific Th1 Cells Promotes Microglial Activation and Increases Plaque Burden in a Mouse Model of Alzheimer's Disease. *The Journal of Immunology* **190**, 2241–2251 (2013).
40. Taylor, J. M. *et al.* Type-1 interferon signaling mediates neuro-inflammatory events in models of Alzheimer's disease. *Neurobiology of Aging* **35**, 1012–1023 (2014).
41. Taylor, S. *et al.* Suppressing Interferon- $\gamma$  Stimulates Microglial Responses and Repair of Microbleeds in the Diabetic Brain. *J. Neurosci.* **38**, 8707–8722 (2018).
42. Newby, B. N. & Mathews, C. E. Type I Interferon Is a Catastrophic Feature of the Diabetic Islet Microenvironment. *Front Endocrinol (Lausanne)* **8**, 232 (2017).
43. Marroqui, L. *et al.* Interferon- $\alpha$  mediates human beta cell HLA class I overexpression, endoplasmic reticulum stress and apoptosis, three hallmarks of early human type 1 diabetes. *Diabetologia* **60**, 656–667 (2017).
44. Wang, L. *et al.* Characteristics of Interferon-Associated Diabetes Mellitus in Past 30 Years: A Review. *Horm Metab Res* **54**, 145–152 (2022).
45. Kurauti, M. A. *et al.* Interleukin-6 increases the expression and activity of insulin-degrading enzyme. *Sci Rep* **7**, 46750 (2017).
46. Lyra e Silva, N. M. *et al.* Pro-inflammatory interleukin-6 signaling links cognitive impairments and peripheral metabolic alterations in Alzheimer's disease. *Transl Psychiatry* **11**, 1–15 (2021).
47. Vinuesa, A. *et al.* Inflammation and Insulin Resistance as Risk Factors and Potential Therapeutic Targets for Alzheimer's Disease. *Front Neurosci* **15**, 653651 (2021).
48. Schiano, C., Luongo, L., Maione, S. & Napoli, C. Mediator complex in neurological disease. *Life Sciences* **329**, 121986 (2023).
49. Cissé, M., Duplan, E. & Checler, F. The Transcription Factor XBP1 in Memory and Cognition: implications in Alzheimer's Disease. *Mol Med* **22**, 905–917 (2016).
50. Ling, Y., Hu, L., Chen, J., Zhao, M. & Dai, X. The mechanism of mitochondrial metabolic gene *PMAIP1* involved in Alzheimer's disease process based on bioinformatics analysis and experimental validation. *Clinics* **79**, 100373 (2024).
51. Oh, M. *et al.* Nurr1 performs its anti-inflammatory function by regulating RasGRP1 expression in neuro-inflammation. *Sci Rep* **10**, 10755 (2020).
52. Blonrath, K. *et al.* The nuclear cofactor receptor interacting protein-140 (RIP140) regulates the expression of genes involved in A $\beta$  generation. *Neurobiology of Aging* **47**, 180–191 (2016).
53. Menendez, J. A., Vazquez-Martin, A., Ortega, F. J. & Fernandez-Real, J. M. Fatty Acid Synthase: Association with Insulin Resistance, Type 2 Diabetes, and Cancer. *Clinical Chemistry* **55**, 425–438 (2009).

54. Sidibeh, C. O. *et al.* FKBP5 expression in human adipose tissue: potential role in glucose and lipid metabolism, adipogenesis and type 2 diabetes. *Endocrine* **62**, 116–128 (2018).
55. Hong, K.-W. *et al.* RAPGEF1 gene variants associated with type 2 diabetes in the Korean population. *Diabetes Research and Clinical Practice* **84**, 117–122 (2009).
56. Nejsum, L. N. *et al.* Compensatory increase in AQP2, p-AQP2, and AQP3 expression in rats with diabetes mellitus. *American Journal of Physiology-Renal Physiology* **280**, F715–F726 (2001).
57. Prasad, R. *et al.* Hypermethylation of Mest promoter causes aberrant Wnt signaling in patients with Alzheimer's disease. *Sci Rep* **11**, 20075 (2021).
58. Petralia, M. C. *et al.* Computational Analysis of Pathogenetic Pathways in Alzheimer's Disease and Prediction of Potential Therapeutic Drugs. *Brain Sci* **12**, 827 (2022).
59. Xiao, F. *et al.* Hepatic cytokine-inducible SH2-containing protein (CISH) regulates gluconeogenesis via cAMP-responsive element binding protein (CREB). *The FASEB Journal* **36**, e22541 (2022).
60. Kanno, A. *et al.* Early administration of dapagliflozin preserves pancreatic  $\beta$ -cell mass through a legacy effect in a mouse model of type 2 diabetes. *Journal of Diabetes Investigation* **10**, 577–590 (2019).
61. Chik, M. W., Hazalin, N. A. M. N. & Singh, G. K. S. Regulation of phase I and phase II neurosteroid enzymes in the hippocampus of an Alzheimer's disease rat model: A focus on sulphotransferases and UDP-glucuronosyltransferases. *Steroids* **184**, 109035 (2022).
62. Shi, X. *et al.* Cholesterol Sulfate and Cholesterol Sulfotransferase Inhibit Gluconeogenesis by Targeting Hepatocyte Nuclear Factor 4 $\alpha$ . *Molecular and Cellular Biology* **34**, 485–497 (2014).
63. de Jesus, D. S. *et al.* Dysregulation of the *Pdx1/Ovol2/Zeb2* axis in dedifferentiated  $\beta$ -cells triggers the induction of genes associated with epithelial–mesenchymal transition in diabetes. *Molecular Metabolism* **53**, 101248 (2021).
64. Ghosh, S. *et al.* Siah2 modulates sex-dependent metabolic and inflammatory responses in adipose tissue to a high-fat diet challenge. *Biol Sex Differ* **10**, 19 (2019).
65. Bate, C., Salmona, M., Diomede, L. & Williams, A. Squalastatin Cures Prion-infected Neurons and Protects Against Prion Neurotoxicity \*. *Journal of Biological Chemistry* **279**, 14983–14990 (2004).
66. Bate, C. & Williams, A. Squalastatin protects neurons and reduces the activation of cytoplasmic phospholipase A2 by A $\beta$ 1–42. *Neuropharmacology* **53**, 222–231 (2007).
67. Li, R., Cui, J. & Shen, Y. Brain Sex Matters: estrogen in cognition and Alzheimer's disease. *Mol Cell Endocrinol* **389**, 13–21 (2014).
68. Mauvais-Jarvis, F., Manson, J. E., Stevenson, J. C. & Fonseca, V. A. Menopausal Hormone Therapy and Type 2 Diabetes Prevention: Evidence, Mechanisms, and Clinical Implications. *Endocr Rev* **38**, 173–188 (2017).
69. Nilsen, J., Chen, S., Irwin, R. W., Iwamoto, S. & Brinton, R. D. Estrogen protects neuronal cells from amyloid beta-induced apoptosis via regulation of mitochondrial proteins and function. *BMC Neurosci* **7**, 74 (2006).
70. Marin, R. *et al.* Estradiol prevents amyloid- $\beta$  peptide-induced cell death in a cholinergic cell line via modulation of a classical estrogen receptor. *Neuroscience* **121**, 917–926 (2003).
71. Carcaillon, L. *et al.* High plasma estradiol interacts with diabetes on risk of dementia in older postmenopausal women. *Neurology* **82**, 504–511 (2014).
72. Pereira, R. I. *et al.* Timing of Estradiol Treatment After Menopause May Determine Benefit or Harm to Insulin Action. *The Journal of Clinical Endocrinology & Metabolism* **100**, 4456–4462 (2015).
73. O'Neal, M. A. Women and the risk of Alzheimer's disease. *Front. Glob. Womens Health* **4**, (2024).
74. Davis, S. & Meltzer, P. S. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846–1847 (2007).
75. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47 (2015).
76. Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* **12**, 115–121 (2015).
77. Schilder, B. & Skene, N. orthogene: an R package for easy mapping of orthologous genes across hundreds of species. *Bioconductor* (2022).
78. Dolgalev, I. msigdb: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format. (2022).
79. Korotkevich, G. *et al.* Fast gene set enrichment analysis. 060012 Preprint at <https://doi.org/10.1101/060012> (2021).
80. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology* **16**, 284–287 (2012).