

# Comprehensive transcriptomic analysis indicates brain regional specific alterations in type 2 diabetes

Zhe Zhou<sup>1,2</sup>, Yizhang Zhu<sup>1,2</sup>, Yang Liu<sup>2</sup>, Yuxin Yin<sup>1,2</sup>

<sup>1</sup>Institute of Systems Biomedicine, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100191, China

<sup>2</sup>Department of Pathology, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100191, China

**Correspondence to:** Yuxin Yin; **email:** [yinyuxin@hsc.pku.edu.cn](mailto:yinyuxin@hsc.pku.edu.cn)

**Keywords:** brain, transcriptome, type 2 diabetes, co-expression network, caudate

**Received:** July 03, 2019

**Accepted:** August 10, 2019

**Published:** August 26, 2019

**Copyright:** Zhou et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

**Type 2 diabetes (T2D) can result in a number of comorbidities involving various organs including heart, eye, kidney and even the brain. However, little is known about the molecular basis of T2D associated brain disorders. In this study, we performed a comprehensive transcriptomic analysis in a total of 304 T2D samples and 608 matched control samples from thirteen distinct brain regions. We observed prominent difference among transcriptomic profiles of diverse brain regions in T2D. The most striking change was found in caudate with thousands of T2D-associated genes identified, followed by hippocampus, while nearly no transcriptomic change was observed in other brain regions. Functional analysis of T2D-associated genes revealed impaired synaptic functions and an association with neurodegenerative diseases. Co-expression analysis of caudate transcriptomic profiles unveiled a core regional specific module that was disorganized in T2D. Sub-modules consisting of regional markers were enriched in T2D risk single nucleotide polymorphisms (SNPs) and implied a causal link with T2D. Hub genes of this module include *NSF* and *ADD2*, the former of which has been associated with T2D and neurodegenerative diseases. Thus, our work provides useful information for further studies in T2D associated brain disorders.**

## INTRODUCTION

Type 2 diabetes (T2D), which is defined as a group of metabolic disorders characterized by both insufficient insulin secretion and insulin resistance, is a major subtype of diabetes and accounts for more than 90% of it [1]. Globally, an estimated 422 million adults are living with diabetes and over 1.5 million deaths are caused by it every year [2]. T2D is a complex disease with numerous systemic complications including heart attack, kidney failure, vision loss and peripheral nerve damage [3]. Specific cell types and molecular pathways involved in majority of these complications have been well discussed [4]. Nonetheless, scant attention has been paid to the influence of T2D on the structure and function of the central nervous system

(CNS), which is the most important system in the body.

Recent evidence suggests that T2D doubles the risk of vascular dementia and neurodegenerative diseases in older age [5–8]. Longitudinal cohort studies have also linked T2D with significant decline in processing speed [9, 10], executive function [9, 11, 12], memory [9–11] and verbal fluency [10]. Imaging studies in diabetic brains reported global brain atrophy [13] and microstructural lesions in the cerebral gray and white matter [14]. Given the above observations, it is of great need to understand the mechanism of these disorders and to identify molecular targets and pathways involved. Several studies reported the potential role of inflammation, defective insulin signaling and mito-

chondrial dysfunction in T2D-associated CNS disorders [15–17]. Nonetheless, few of them were performed on human brains with sufficient samples. Moreover, human brain is composed of diverse regions which execute different functions, whilst majority of these studies focused on specific brain regions. To the best of our knowledge, the difference between diverse brain regions in T2D remains elusive.

Disruption in the normal gene expression profile of various tissues is an important link between T2D and its complications [18]. With the advent of high-throughput sequencing, it is feasible to study the brain transcriptomic changes associated with T2D in various brain regions. The Genotype-Tissue Expression (GTEx) project [19] provides whole transcriptomic profiles of 13 brain regions derived from T2D subjects and healthy controls, thus making it one of the most comprehensive datasets for studying region-specific T2D-associated transcriptomic changes.

In the present study, we first performed differential expression analysis in 13 brain regions samples and identified regional specific T2D-associated genes (DAGs). We found that majority of the brain was immune to T2D, while great transcriptomic changes were observed in caudate and hippocampus. Next, we explored the distributions and functions of these regional specific DAGs. Rather than analyzed at

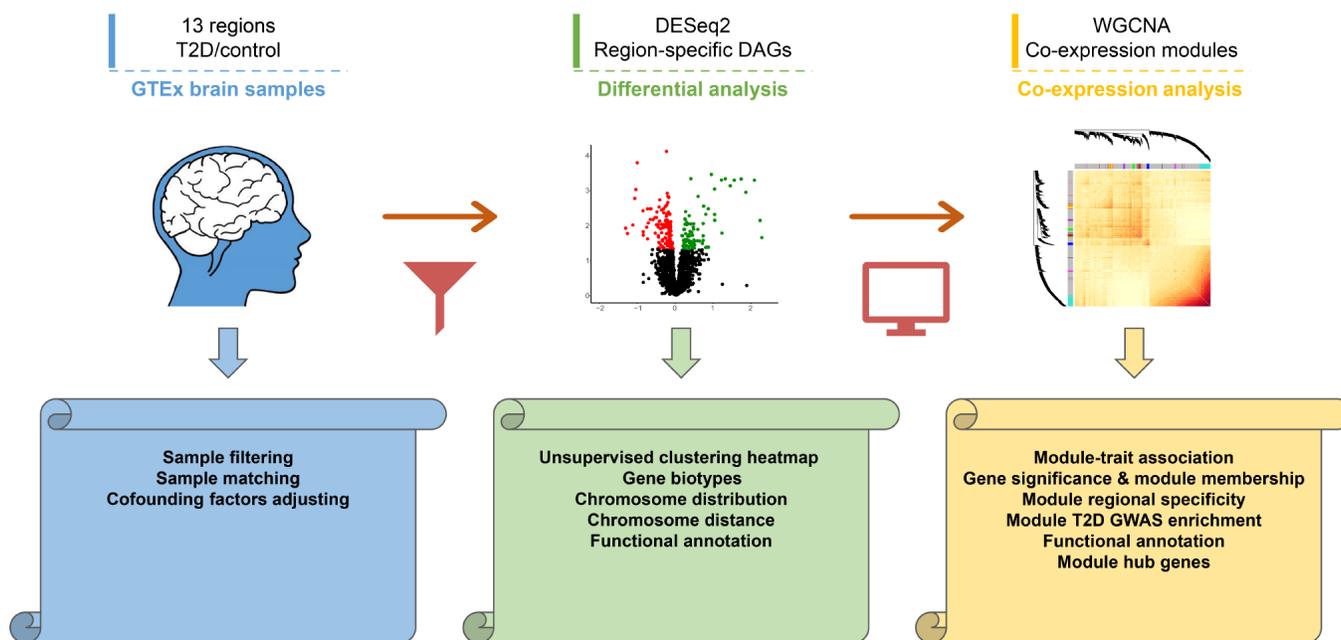
individual gene level, we also performed co-expression analysis on T2D-involved brain regions. Core modules and hub genes which might help unravel the underlying mechanisms were identified by systemic analysis. Hereby, our analysis provides a basis for further researches of T2D-associated brain alterations.

## RESULTS

### Identification of DAGs in 13 brain regions

The major workflow of present study was shown in Figure 1. Control samples were matched with T2D cases to avoid bias resulting from confounding factors. A total of 304 T2D samples and 608 matched control samples from thirteen distinct brain regions were then obtained. No significant differences were detected between matched samples for most of the covariates (Supplementary Figure 1 and Supplementary Table 1). Sample distribution after matching was plotted for each brain region, as shown in Supplementary Figure 2.

To identify DAGs, we performed generalized linear regression for genes against T2D status in each brain region using DESeq2 [20], with known and surrogate variables adjusted. At a 5% false discovery rate (FDR), less than 10 or even no DAGs were identified in most regions, indicating that T2D had no notable effect on transcriptome of the majority of human brain.



**Figure 1. Workflow diagram of the study design.** First, transcriptomic profiles of 13 human brain regions were derived from GTEx dataset. Second, differential analysis was performed to investigate regional specific changes. Distributions and functional annotations of DAGs were analyzed subsequently. Finally, Co-expression analysis was performed in caudate and hippocampus to study core modules and hub genes. The volcano plot and heatmap were generated using random sampling data of caudate transcriptome.

**Table 1. Numbers of DAGs in 13 brain regions.**

Regions	FDR < 10%			FDR < 5%			FDR < 1%		
	total	up	down	total	up	down	total	up	down
Amygdala	59	7	52	6	0	6	1	0	1
Anterior Cingulate Cortex	3	0	3	1	0	1	1	0	1
Caudate	4548	1910	2638	2939	1186	1753	1049	448	601
Cerebellar Hemisphere	12	4	8	2	0	2	0	0	0
Cerebellum	1	0	1	0	0	0	0	0	0
Cortex	0	0	0	0	0	0	0	0	0
Frontal Cortex	0	0	0	0	0	0	0	0	0
Hippocampus	1143	527	616	486	223	263	82	31	51
Hypothalamus	1	0	1	0	0	0	0	0	0
Nucleus Accumbens	66	26	40	0	0	0	0	0	0
Putamen	0	0	0	0	0	0	0	0	0
Spinal Cord	3	1	2	1	0	1	0	0	0
Substantia Nigra	0	0	0	0	0	0	0	0	0

Note: Columns “total”, “up” and “down” list the number of total, up- and down-regulated DAGs respectively. Results derived from using three different FDR cutoffs (10%, 5% and 1%) are shown.

Nevertheless, regulation of a considerable number of genes were disturbed in two regions, namely caudate (2939 DAGs) and hippocampus (486 DAGs) (Table 1). Among all the DAGs, only 178 genes were dysregulated in more than one tissue, and expression of all the ‘multi-hit’ DAGs responded to T2D status in the same direction except for 3 of them. This result indicated that different brain regions might associate with T2D in a region-specific manner. The details of DAGs in each brain region are provided in Supplementary Table 2. Control cases and T2D samples can be distinguished roughly in unsupervised clustering heatmap (P-value: 2.19eE-03 for caudate and 7.78E-07 for hippocampus, Fisher’s exact test) using identified DAGs (Figure 2A and 2B).

In view of the fact that caudate had the maximum sample size among all the 13 brain regions and the marked influence of sample size on differential analysis, we cannot evaluate the impact of T2D on diverse brain regions according to the number of DAGs directly. Therefore, we randomly selected samples with size ranging from 10 T2D samples versus 20 matched controls (the minimal size in 13 brain regions) to maximal number of each brain regions, with five T2D samples and ten matched controls added each time, and then bootstrapped 100 times. As expected, caudate got the largest number of DAGs at a 5% FDR level, followed by hippocampus, regardless of the sample size (Supplementary Figure 3). This result indicated that diverse brain regions were altered differentially in T2D. Furthermore, caudate and

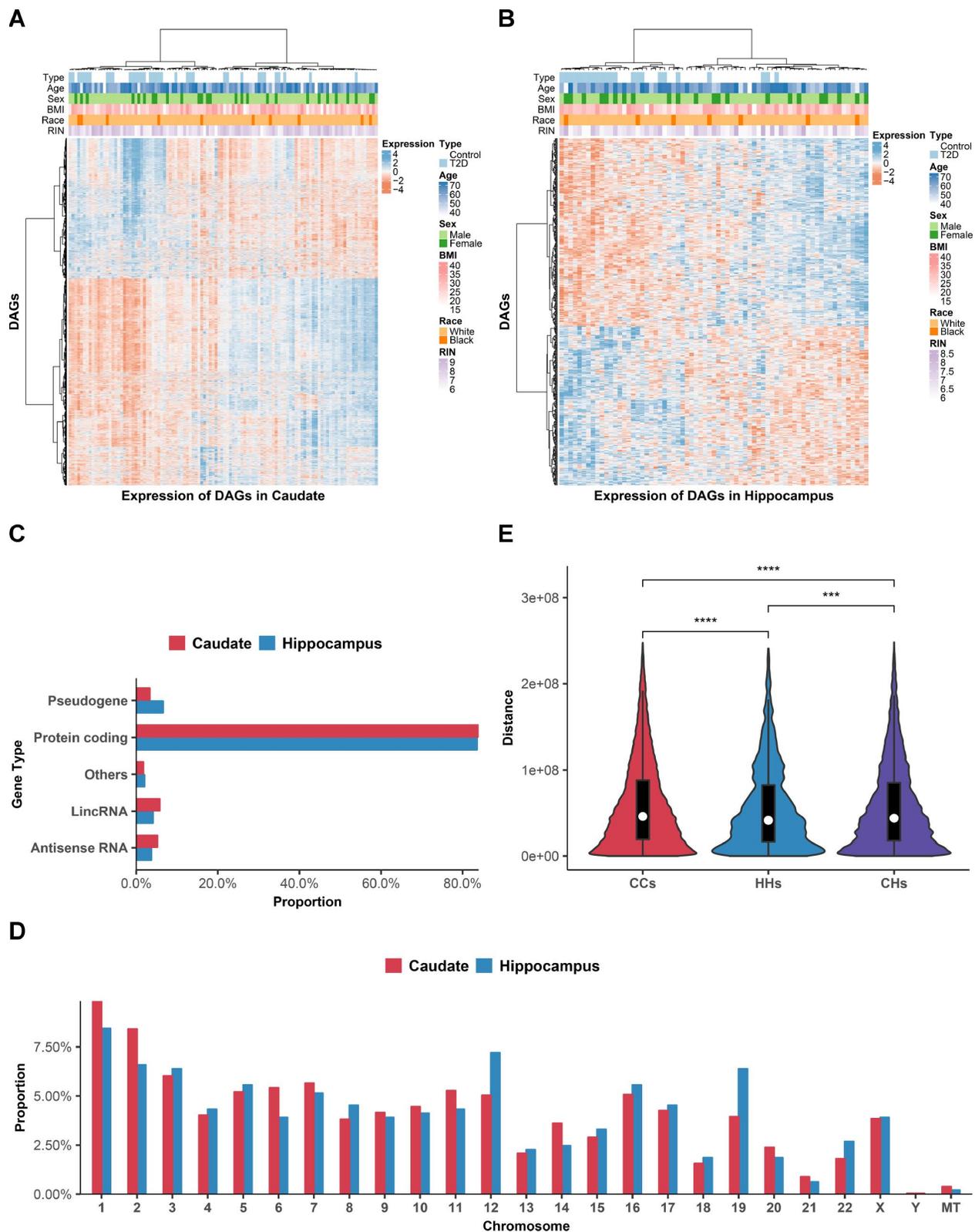
hippocampus might be the brain regions with the most significant transcriptomic changes in T2D.

#### Distributions and functional annotations of DAGs in various brain regions

To study the distributions of DAGs in various brain regions, we first categorized the DAGs according to gene biotypes in GENCODE [21]. Protein coding genes account for majority of DAGs (83.9% of caudate DAGs and 83.5% of hippocampus DAGs, Figure 2C). Whilst there was also a small subset of DAGs comprised of non-coding genes, including antisense RNA, pseudogene, lincRNA and others, implying the function of these non-coding genes in pathological process of diabetic brain.

Next, we investigated the chromosomal distribution patterns of DAGs, which showed that they were widespread across chromosomes (Figure 2D). More than 95% of DAGs located on autosomes and the overall distribution of caudate and hippocampus DAGs were similar. However, the proportion of DAGs located on small, gene-rich chromosomes (chr16-22) was slightly higher in hippocampus than in caudate.

As genes involved with the same disease process tend to locate adjacently [22], we then calculated the genomic distance of each pair of DAGs on the same chromosome (Figure 2E). Results showed that distances of DAGs



**Figure 2. Distributions of DAGs.** (A and B) Heat maps of DAGs (row) on samples (column) were shown for caudate (A) and hippocampus (B). (C) Gene biotype categories of DAGs according to GENCODE. X-axis shows the proportion to the total DAGs in each brain region. (D) Chromosomal distribution of DAGs. Y-axis shows the proportion to the total DAGs in each brain region. (E) The violin plot comparing the intra-chromosomal distances between DAGs in different brain regions.  $***P < 0.001$ ;  $****P < 0.0001$ , Wilcoxon's test. CCs, distance of DAGs within caudate group; HHs, distance of DAGs within hippocampus group; CHs, distance of DAGs between caudate and hippocampus group.

within hippocampus group were the smallest compared with distances of DAGs within caudate group (P-value=8.95E-14) and between groups (P-value=1.11E-4). Surprisingly, distances of DAGs between groups were also smaller than distances of DAGs within caudate group (P-value = 1.82E-27). In general, above results suggested regional specific distribution of DAGs in various brain regions.

To obtain a functional overview of DAGs in different brain regions, we dissected functional annotations of the up- and down-regulated DAGs respectively. As a result, top Gene Ontology (GO) biological process terms enriched with down-regulated DAGs in caudate and hippocampus both related to synaptic functions (Supplementary Table 3), which have been reported to impaired in diabetic brain [23]. Nonetheless, functions of up-regulated DAGs tend to be more regional specific. In caudate, up-regulated DAGs enriched in terms like cilium movement and cilium organization, implying the hidden association of T2D and ciliopathy [24]. However, for up-regulated DAGs in hippocampus, terms related to muscle function and muscle tissue morphogenesis were enriched. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis revealed that DAGs in caudate also enriched in neurodegenerative diseases pathways such as Alzheimer disease, Huntington's disease and Parkinson's disease. In consideration of aforementioned association between T2D and neurodegenerative diseases [6–8], the caudate nucleus and related DAGs might play a vital role in these processes.

### Co-expression analysis of caudate and hippocampus

Although numerous DAGs were identified in caudate and hippocampus, limited information could be provided by individual genes. Moreover, DAGs determined by differential expression analysis methods were biased to genes with large change in expression, while overlooking the coordination of gene expression. To integrate the expression difference into a systems-level context, we performed weighted gene co-expression network analysis (WGCNA) on caudate and hippocampus. We detected 14 modules in caudate (Supplementary Figure 4 and Supplementary Table 4) and 40 modules in hippocampus (Supplementary Figure 5 and Supplementary Table 5). The average size of caudate modules was larger than hippocampus (P-value: 9.43E-03, Wilcoxon test), which implied a stronger co-expression of transcriptomic profiles in caudate.

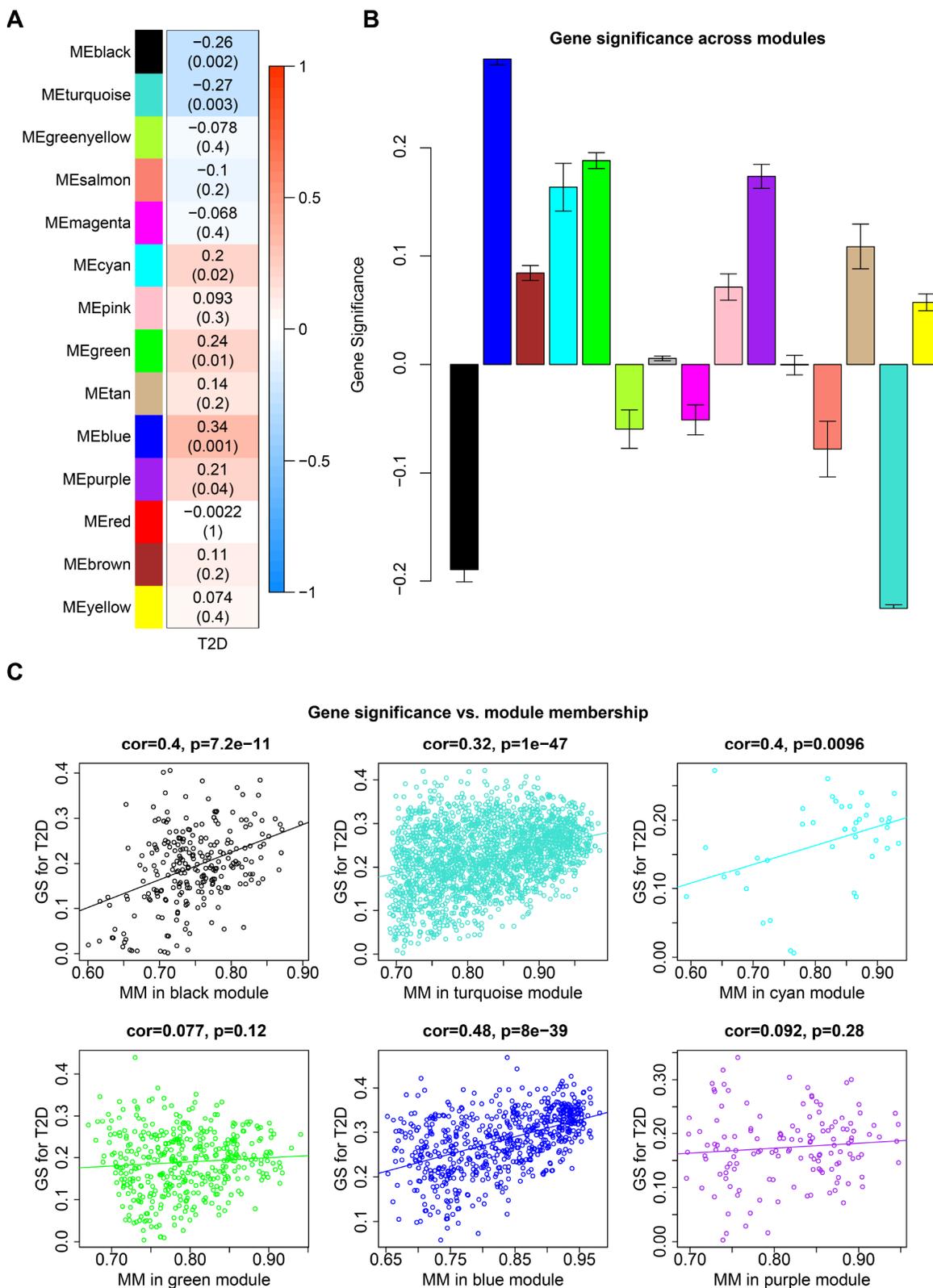
Among the 14 co-expression modules in caudate, eigengenes of 4 modules (cyan, green, blue and purple) were positively correlated (up-regulated) with

T2D, while black and turquoise module were negatively correlated (down-regulated) with T2D (Figure 3A, P-value < 0.05, Student t test). No modules correlated with any of the potential confounding variables which have been regressed out. Gene significance (GS) varied dramatically across modules, which also indicated the different levels of correlation between co-expression modules and T2D (Figure 3B). To examine gene constitution of particular modules, we plotted GS against module membership (MM) for T2D-associated modules (Figure 3C). Significant correlations were detected in black, turquoise and blue modules, illustrating that genes significantly associated with T2D status are often also the important elements of these modules. This result confirmed the crucial role of these modules in type 2 diabetes process.

In hippocampus, five of the 40 co-expression modules positively associated with T2D and another five modules negatively associated (Supplementary Figure 6A, P-value < 0.05, Student t test). The mean GS values across modules were shown in Supplementary Figure 6C. Among the 10 T2D-associated modules, significant correlations of GS and MM were detected in 7 modules (Supplementary Figure 6B).

### Cell-type and regional specificity of co-expression modules

Disorder of regional specific gene networks usually results in impairment of regional specific functions. Hence, it is of great value to determine whether brain regional specific modules were disturbed in T2D, which might help unravel the underlying mechanisms of T2D-associated brain disorders. By testing the enrichment of distinct regional and cell-type specific expression markers [25], several modules in caudate were found to be significant at pSI 0.0001 (Figure 4). For instance, the green module, positively correlated with T2D, was significantly enriched in cortical astrocytes markers. The red module was significantly enriched in expression markers of oligodendrocyte of cortex and cerebellum, whilst eigengene of this module did not show prominent correlation with the T2D status. The most surprising result is the turquoise module, which was found to be striatum specific at every pSI (Supplementary Table 6). Cell-type markers of two major neuronal populations in striatum: Drd1-expressing and Drd2-expressing medium spiny neurons (D1- and D2-MSNs) were also enriched in turquoise module at every pSI. In consideration of the transcriptional association between turquoise module and T2D, this module could be determined as a striatum-specific module which was disturbed in T2D status.

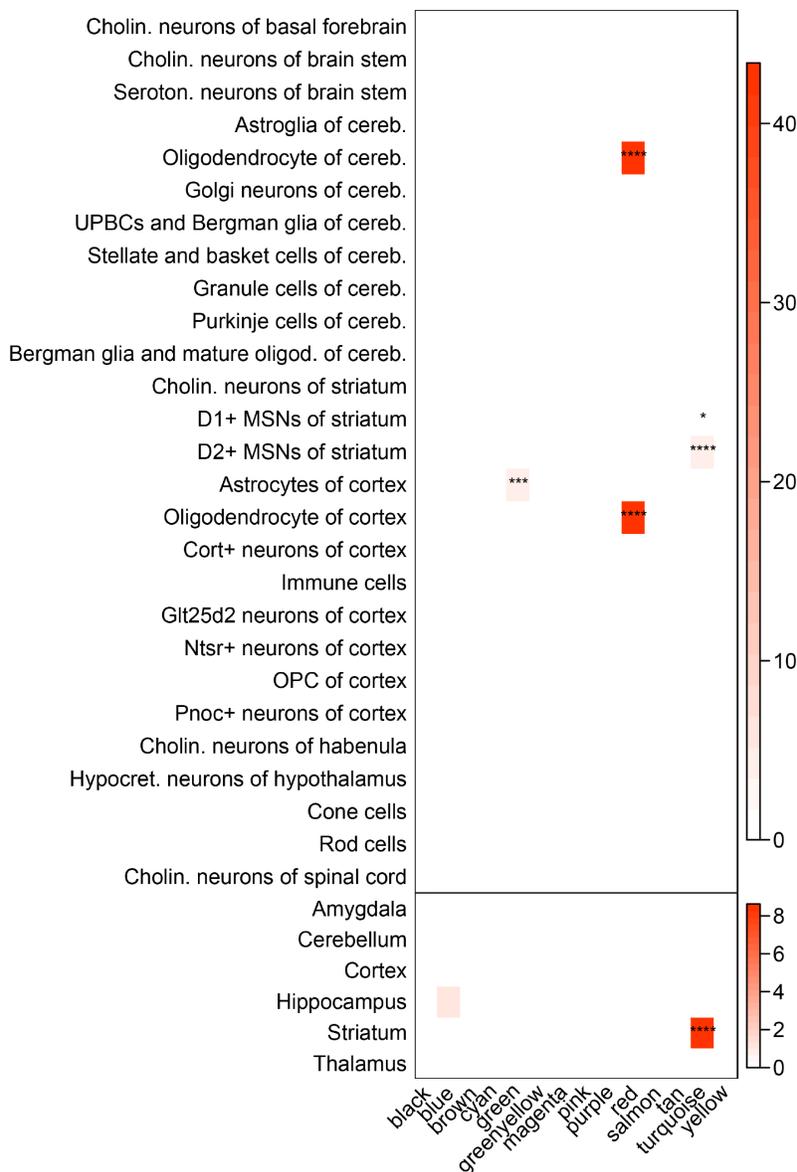


**Figure 3. Co-expression modules in caudate.** (A) Each row corresponds to a module eigengene and column indicate T2D status. The table were colored by correlation according to the legend. Each cell contains the corresponding correlation and P-value. (B) Each bar indicates the average of gene significance measure for all genes in a given module. (C) GS vs. MM plot for modules significantly correlated with T2D status. Each point corresponds to an individual gene within a given module, which was plotted by GS on the y-axis and MM on the x-axis. The regression line, correlation value and P-value were shown for each module.

However, in hippocampus, there were no T2D-associated modules enriched in any regional or cell-type markers at pSI 0.0001 (Supplementary Figure 7 and Supplementary Table 7). The darkolivegreen module was the only module enriched in hippocampus markers. Nevertheless, eigengene of this module had a weak correlation with T2D. Therefore, the darkolivegreen module may not account for the alteration of hippocampus in T2D. Next, we focused only on caudate to further study the correlation between T2D and these putative core modules.

### Identification of modules genetically associated with T2D in caudate

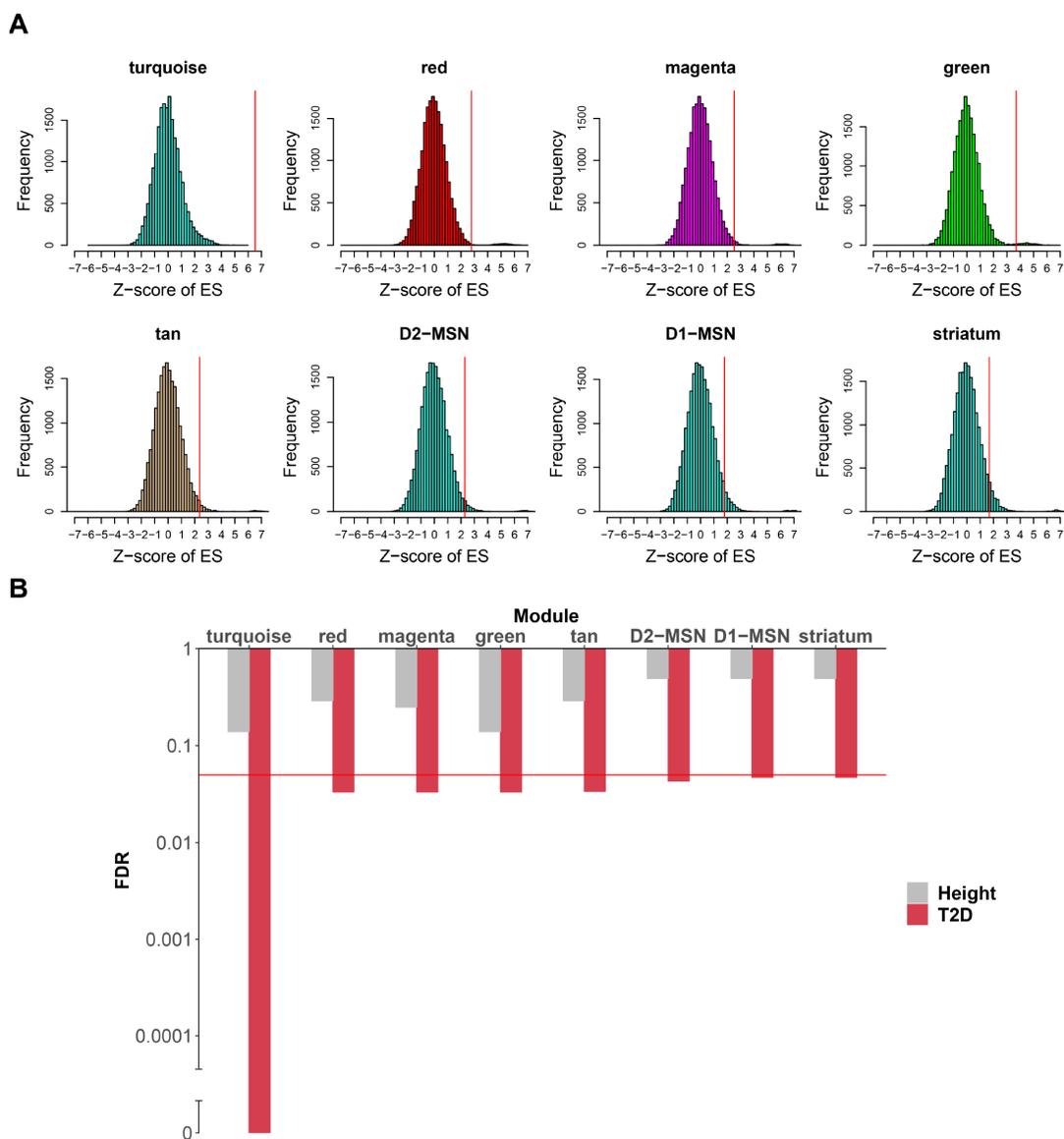
Though genes in modules specific to caudate were dysregulated in T2D, it was unclear whether these impairments were only consequences of T2D or genetically associated with T2D. To identify the causal link of modules to T2D, we tested for their enrichment in T2D and height (as a negative control) associated single nucleotide polymorphisms (SNPs) from large genome-wide association study (GWAS) data sets.



**Figure 4. Enrichment of brain regional and cell-type markers in caudate modules.** (Top) Enrichment of caudate modules in markers of various neuronal and glial cell types. (Bottom) Same as above, but using markers for different brain regions. Asterisks indicate significant enrichment after Bonferroni adjustment. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; \*\*\*\* $P < 0.0001$ , Fisher's Exact test. Cholin., cholinergic; seroton., serotonergic; cereb., cerebellum; UPBCs, unipolar brush cells; oligod., oligodendrocyte; MSNs, medium spiny neurons; hypocret., hypocretinergic.

In caudate, the turquoise module exhibited highly significant enrichment in T2D risk SNPs, while no such significant result was found in height-associated SNPs (Figure 5A, 5B and Supplementary Table 8). As the turquoise module was also highly enriched in striatum and MSNs markers, it was of great interest whether the T2D risk SNPs signal of this module was in connection with these specific markers. To explore this, we extract sub-modules of the turquoise module which only contain striatum, D1-MSNs or D2-MSNs specific markers at pSI 0.01 (sub-module derived at pSI 0.001 and 0.0001 was too small), respectively,

and then performed the same enrichment method. All the three sub-modules were significantly enriched in T2D risk SNPs compared with height, especially for D2-MSNs. It can be inferred that the genetic risk for T2D associated with the turquoise module might distribute mainly across striatum and MSNs specific genes. Another 4 modules (red, magenta, green and tan module) were also relatively significant compared to negative control. However, only eigengene of green module was correlated with T2D and this module was enriched in cell-type markers of astrocytes of cortex.



**Figure 5. Caudate modules enriched in T2D genetic signals.** (A) For each module (or sub-module), the null distribution of T2D SNPs enrichment scores generated by 20,000 random permutations is shown. Real enrichment scores were depicted by red vertical lines. Modules were considered significant if  $FDR < 0.05$ . (B) For each module in (A), enrichment FDR for T2D SNPs are shown by histogram compared to height SNPs. Y-axis was log10 transformed and broken axis was used to show zero value. The red horizontal line marks the FDR threshold for significance, which is 0.05. FDR, false discovery rate.

## Functional annotations and hub genes of caudate modules

Previous studies have shown that co-expressed genes tend to be functional related [26], we therefore studied the functional annotations of interested modules. Actually, although more than half of input genes were unassigned to specific co-expression modules, there were still large overlaps between DAGs and several modules. For instance, the blue module accounted for nearly 40% of up-regulated DAGs in caudate and a half of down-regulated DAGs were assigned to turquoise module (Supplementary Table 9). Nonetheless, the co-expression modules also harbored non-T2D-associated genes and had a greater power to delineate T2D-relevant transcriptional changes compared with DAGs. As an illustration, similar but more significant pathways were enriched for the blue module than up-regulated DAGs (Supplementary Table 10). The turquoise was also enriched in terms related to synaptic functions more significantly (Figure 6A), compared with down-regulated DAGs. In terms of other aforementioned T2D-associated modules, the cyan module associated with ribosome and the purple module is an immune module, revealing corresponding dysfunction in T2D brain. It was worth mentioning that the green module enriched in many metabolic processes and axonogenesis, in accordance with its enrichment in cortical astrocytes marker.

Other than taking the turquoise module as a whole, we further investigated enriched pathways of striatum and MSNs sub-modules. All of these were enriched in pathways related to learning, memory and cognition (Supplementary Table 10). Taken together, the turquoise module was considered as a core functional module of caudate and its related synaptic impairments may contribute to the cognition decline in T2D.

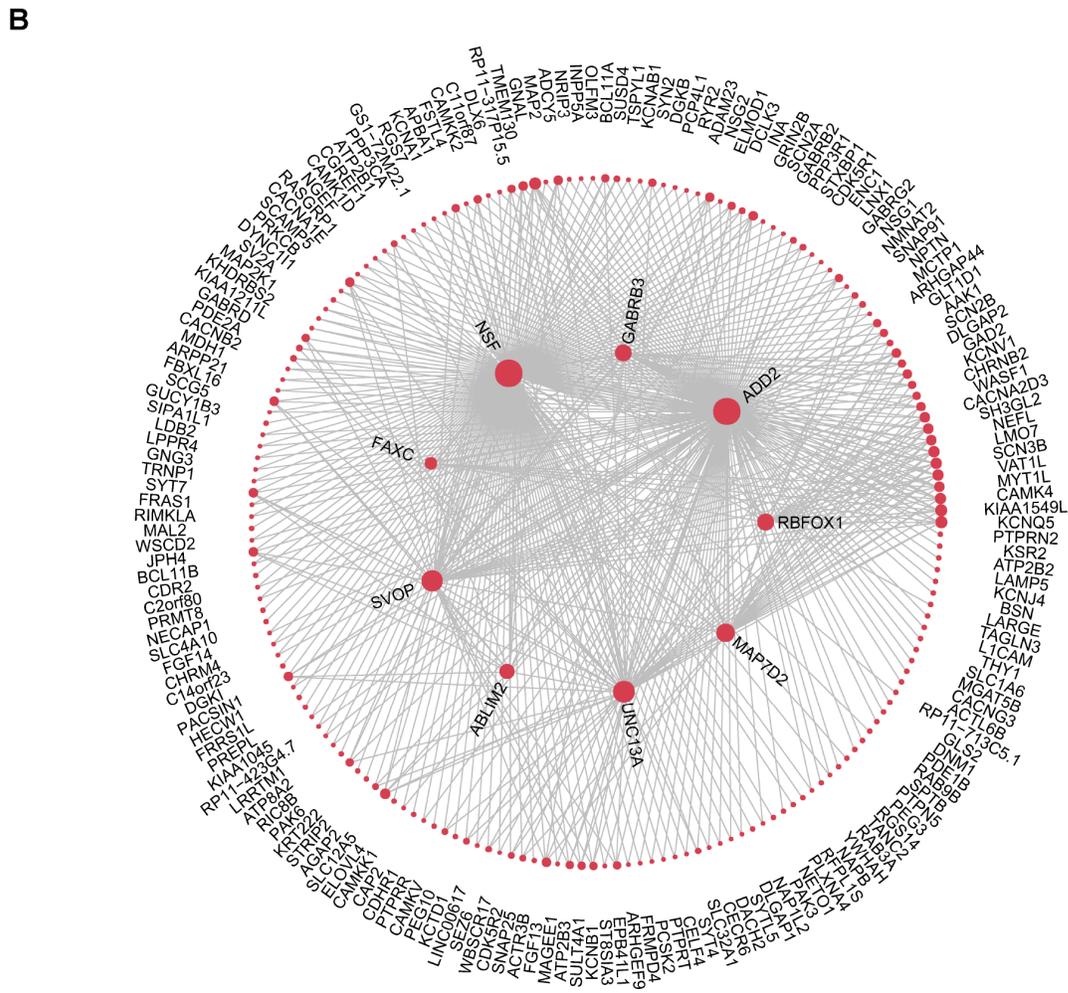
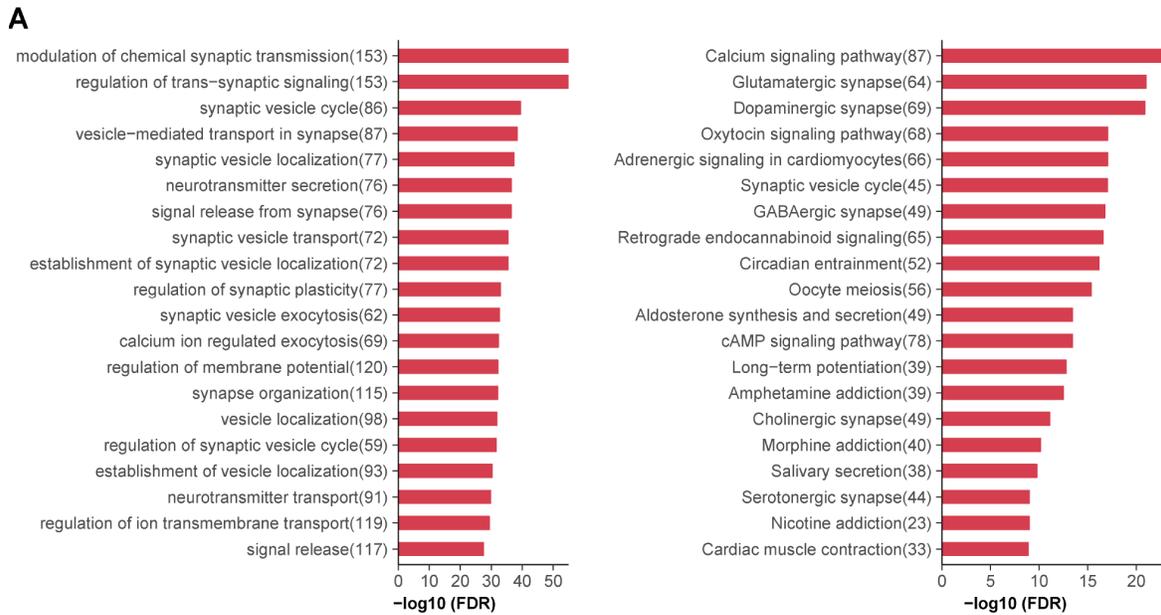
Finally, we tried to identify highly connected hub genes for the turquoise module (Figure 6B). The top two hub genes were *NSF* (N-ethylmaleimide-sensitive factor, Entrez ID 4905) and *ADD2* (adducin 2, Entrez ID 119), both of which were highly expressed in brain. NSF is essential for neurotransmitter release, together with other key factors of synaptic fusion machinery, such as SNAREs (soluble NSF attachment protein receptors) and SNAP (soluble NSF adaptor protein). They have been reported to play an important role in both diabetes and neurodegenerative disorders [27]. *ADD2* encodes the beta subunit of adducins, which are a family of cytoskeletal proteins. Polymorphism of *ADD1* gene (encoding the alpha subunit, Entrez ID 118) has been reported to associated with T2D [28], while little is known about the relationship between *ADD2* and T2D. Thus, further studies are needed to explore the roles of these hub genes in T2D brain.

## DISCUSSION

Identification of genes and pathways altered in diabetic brains may provide insights into the mechanism and prognosis of T2D-associated CNS disorders. In the present study, we delineated transcriptomic changes of 13 brain regions in T2D and observed prominent difference among diverse brain regions. Although transcriptome of majority of the human brain was stable in T2D, a considerable number of DAGs were identified in caudate and hippocampus. Even anatomically adjacent regions might also show dramatic difference in T2D. A notable example was caudate and putamen, the former of which has the most abundant number of DAGs, while none are identified in the latter. Functional annotations indicate that the down-regulated DAGs in caudate and hippocampus are both enriched in synaptic pathways, whilst the up-regulated DAGs have regional specific functions.

We also performed co-expression analysis on these two regions and observed different co-expression patterns. We identified a turquoise module that harbors a half of down-regulated DAGs in caudate while performed better in delineating transcriptomic changes of caudate. This module is enriched in regional markers of striatum and cell-type markers of its two major neuronal populations, consistent with previous evidence that GABAergic neurons in striatum is negatively affected in T2D rats [29]. Of particular interest, the turquoise module is enriched in T2D risk SNPs, implying their potential role in etiology of diabetes. Dissection of functions of turquoise modules and its sub-modules has revealed their core role in synaptic transmission and cognition. Moreover, the identified hub genes of the turquoise modules might play a vital role in coordination of involved genes. For instance, the top hub gene *NSF* can link T2D with neurodegenerative diseases together with SNAREs. It is of note that although no T2D-associated CNS complications were observed in GTEEx donors, remarkable alterations have been existed in diabetic brain. Hence, early interventions to prevent diabetes-related CNS complication were recommended. There is no such global gene expression profiling of multiple brain regions in T2D has been reported. As type 2 diabetes and CNS disorders getting prevailing, our study provides a broader horizon for further research.

Given the association between T2D and its related brain alterations, much attention has been directed to the hippocampus and cognitive decline [30]. However, pooled analysis showed that the hippocampus was not more severely affected than the rest of the brain [13]. In our study, large transcriptomic changes of hippocampus were observed, whilst the changes seemed weaker than caudate and we were not able to identify convergent core



**Figure 6. Functional annotations and hub genes of caudate modules.** (A) (Left) Top 20 GO biological processes significantly enriched in turquoise module. (Right) Top 20 KEGG pathways significantly enriched in turquoise module. Numbers in the parenthesis indicate the numbers of genes associated with the respective terms. (B) Network plots showing top 500 connections in turquoise module; genes with most connections (hub genes) are shown in center. The size of each dot is proportional to  $\log_2$  (number of connections for each gene).

modules for it. Further researches were still required for better understanding the role of hippocampus in T2D related brain alterations.

On the other hand, the caudate has also been associated with cognitive impairment [31, 32]. The caudate and putamen together constitute the striatum, which is a part of basal ganglia. Various nuclei of basal ganglia are functionally delineated along corticostriatal lines. The caudate is associated with selection of appropriate sub-goals based on an evaluation of action-outcomes, whereas its nearest neighbor, the putamen, appears to be involved in simpler motor control [33]. This might help explain the striking difference in transcriptomic changes of caudate and putamen in T2D. The cognitive function of caudate, along with caudate abnormality observed in T2D [34, 35], also provides support for the potential role of caudate in T2D associated brain alterations.

Nevertheless, there are also limitations of the present study. First, due to the rarity of T2D postmortem brain samples, the sample size of GTEx database is still relatively small. Further investigation with larger sample size is recommended to reduce noise and draw more reliable conclusions. Second, gene expression is a complex trait influenced by various factors. Although we have tried our best to match samples as well as control known and hidden factors in our pipeline, bias might still exist. Sample pairing in the experiment design stage can further reduce possible bias. Third, there are no quantitative indicators of hyperglycemia, such as glycated hemoglobin (HbA1c), in GTEx dataset. Hence, transcriptomic changes associated with blood glucose level could not be determined, which may provide useful information for explanation of etiology of T2D complications. Ultimately, more comprehensive studies are expected in the future to deepen our understanding on this topic.

## **MATERIALS AND METHODS**

### **GTEx tissues and subjects**

The GTEx project (v7, released in June 2017) provides expression data of 13 human brain regions from 752 post-mortem donors. Two of the brain regions were initially sampled from cerebellum and cortex preserved using the PAXgene tissue preservation system, and another 11 regions were subsequently sampled from frozen brains, including amygdala, anterior cingulate cortex, caudate, cerebellar hemisphere, frontal cortex, hippocampus, hypothalamus, nucleus accumbens, putamen, spinal cord and substantia nigra. Details regarding the sample collection, RNA sequencing and data processing are available at GTEx consortium paper [36].

To draw reliable conclusion, we only keep high-quality sequencing samples with RINs > 6.0. Cases with type 1 diabetes or unknown T2D status, and races other than black or white were excluded from this study.

Confounding factors including age [37], gender [38], race, BMI [39] and RIN [40] have been reported to be correlated with gene expression. To avoid bias, we used an optimal matching algorithm in R package MatchIt [41] to balance them between control and T2D groups, with optimal ratio of 2:1. The number of remaining matched samples of 13 brain regions ranged from 30 to 108.

### **Identification of regional specific DAGs**

DESeq2 [20] (v1.22.2) was employed on all of the 13 brain regions to identify regional specific DAGs using raw read counts. Independent filtering of genes with low read counts was performed automatically by DESeq2 with  $\alpha=0.05$ , and genes remained were referred to as ‘detectable genes’. To correct the known covariates as well as remove inferred hidden confounders in GTEx expression data, we employed “svaseq” function in sva R package [42] to identify 3 surrogate variables. The known confounding factors and surrogate variables were then added to the DESeq2 design, and negative binomial (NB) generalized linear regression model (GLM) was performed. Benjamini-Hochberg (BH) algorithm was used to adjust the Wald test P-values for multiple testing. Raw counts data were transformed to continuous, homoscedastic regularized logarithm transformed (r-log) values for further analysis.

### **Genomic distance and functional annotations**

Genomic distance of each pair of DAGs on the same chromosome was calculated according to their genome coordinate positions. Wilcoxon rank-sum test was used to compare the distributions of intra- and inter-groups.

Functional enrichment analysis was performed using R package clusterProfiler [43] to identify significant GO biological process terms and KEGG terms. The FDR adjustment for P-value was made using Benjamin-Hochberg procedure and an FDR cutoff of 0.05 was used.

### **Weighted gene co-expression network analysis**

To regress out uninterested sources of large variation, linear models containing age, gender, race, BMI, RIN and surrogate variables were fitted on the r-log data for all of the ‘detectable genes’ in each brain region, respectively. Then co-expression analysis was performed on residuals using WGCNA R package [44]. The soft-

thresholding power were picked according to the scale-free topology criterion, and a signed gene network was constructed using blockwiseModules function in a single block with parameters mergeCutHeight = 0.15, minModSize = 40 and minKMEtostay = 0.7.

The module eigengene (ME) is defined as the first principal component of a given module, which can be considered as a representative of the gene expression profiles in a module. Module membership (MM) is a measure of gene-to-module membership by correlating its gene expression profile with the module eigengene of a given module. Gene significance (GS) was defined as the correlation between the gene and the T2D status.

Module graphs showing the top 500 connections were plotted using the iGraph [45] R package. Hub genes were defined as top 5% in number of connections.

### Cell-type and regional specificity analysis

R package pSI [46], which provides lists of expression markers for diverse brain regions and cell types, was used to perform Fisher's Exact test for regional or cell-type marker enrichment in co-expression modules at different specificity index thresholds (pSI 0.01, 0.001 and 0.0001; pSI 0.05 was deprecated for increased false positives). Regional markers were derived from Atlas of the developing Human Brain ([www.brainspan.org](http://www.brainspan.org)), while cell-type markers were originally identified using translational profiling of genetically tagged cell lines purified from mouse brain [25].

### GWAS set enrichment analysis

Enrichment for GWAS signal was conducted as previously described [47]. SNPs and their associated P-values were from large GWAS data sets of T2D [48] and height [49]. SNPs were assigned to genes if located within gene boundaries with additional 20kb on 5' end and 10kb on 3' end. The most significant SNP of each gene was selected, and then all of the genes were ranked with associated scores ( $-\log_{10}$  P-value) to calculate gene set enrichment score (ES) based on the Kolmogorov-Smirnov running-sum statistic using GSEA-P [50]. The null distribution of ES was generated by 20,000 random permutations of gene labels and associated scores. Enrichment scores were scaled by subtracting the mean and dividing by the standard deviation of permutation scores to correct for the gene set size, and empirical P-values were determined by the resulting z-scores.

### Data and resource availability

The GTEx gene expression data and phenotype data were downloaded from dbGaP (<http://www.ncbi.nlm.nih.gov/gap.nih.gov/gap>)

under accession number phs000424.v7.p2. Statistical analysis was performed on R (v3.5.0). Scripts are available at [https://github.com/ZedekiahZhou/T2D\\_Brain](https://github.com/ZedekiahZhou/T2D_Brain). All data generated or analyzed during this study are included in the published article (and its online supplementary files).

### AUTHOR CONTRIBUTIONS

Z.Z. and Y.Y. conceived the project. Z.Z. collected datasets, performed computational analysis and drafted the manuscript. Y.Z. participated in the design of the study and helped revise the manuscript. Y.L. provided general assistance. Y.Y. thoroughly revised the manuscript. All authors read and approved the final manuscript.

### ACKNOWLEDGMENTS

We are grateful to the GTEx program and the GTEx Consortium for providing enormous database and resources.

### CONFLICTS OF INTEREST

The authors have declared no competing interests.

### FUNDING

This work was supported by grants to Y. Yin, including the National Key Research and Development Program of China (Grant 2016YFA0500302), the National Natural Science Foundation of China (Key grants 81430056, 31420103905, 81874235 and 81621063), the Beijing Natural Science Foundation (Key grant 7161007), and the Lam Chung Nin Foundation for Systems Biomedicine.

### REFERENCES

1. Chatterjee S, Khunti K, Davies MJ. Type 2 diabetes. *Lancet*. 2017; 389:2239–51. [https://doi.org/10.1016/S0140-6736\(17\)30058-2](https://doi.org/10.1016/S0140-6736(17)30058-2) PMID:28190580
2. World Health Organization. Global report on diabetes. Geneva: World Health Organization; 2016.
3. Malone JI. Diabetic Central Neuropathy: CNS Damage Related to Hyperglycemia. *Diabetes*. 2016; 65:355–57. <https://doi.org/10.2337/dbi15-0034> PMID:26798123
4. Brownlee M. The pathobiology of diabetic complications: a unifying mechanism. *Diabetes*. 2005; 54:1615–25. <https://doi.org/10.2337/diabetes.54.6.1615>

PMID:[15919781](https://pubmed.ncbi.nlm.nih.gov/15919781/)

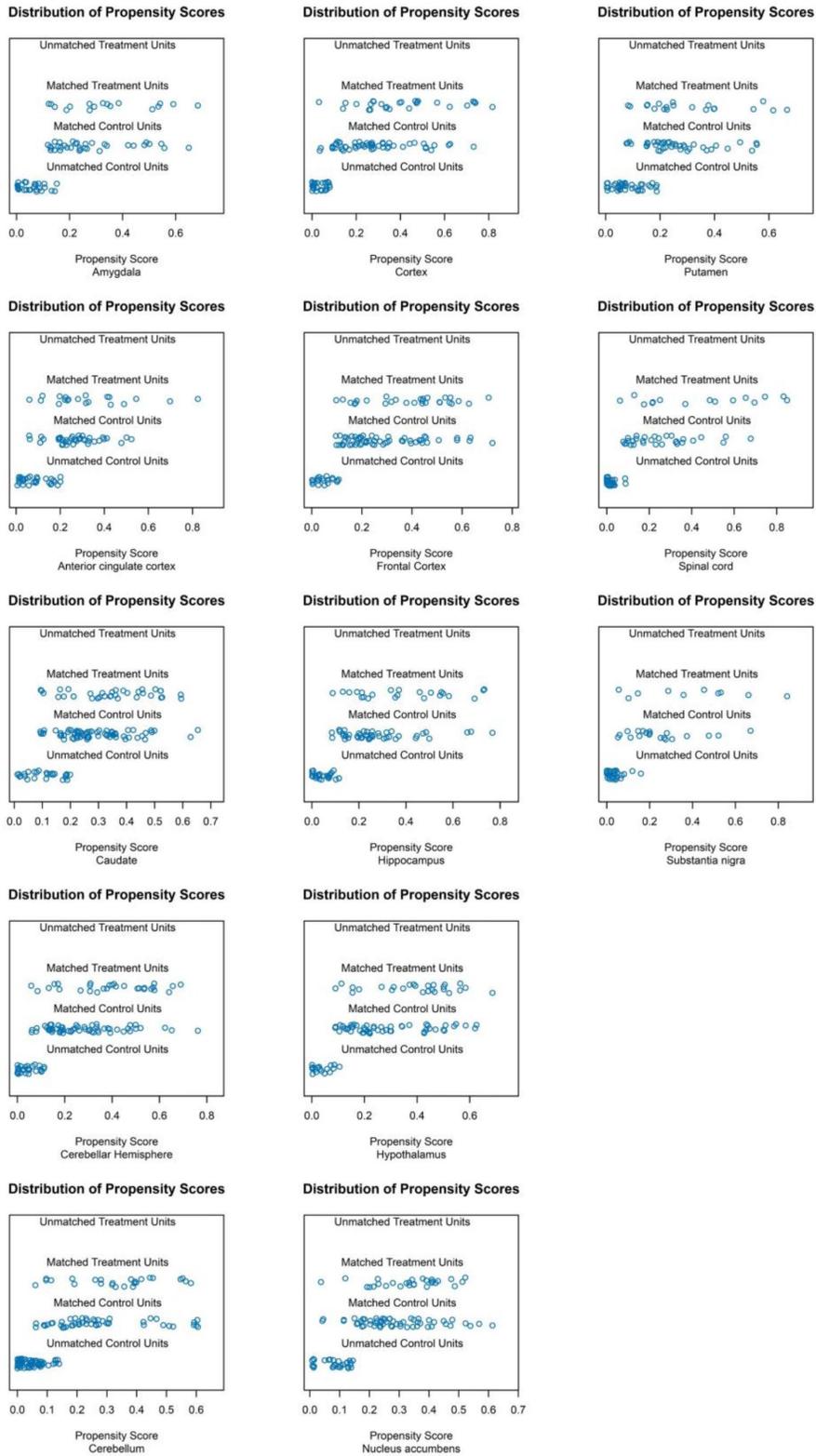
5. Cheng G, Huang C, Deng H, Wang H. Diabetes as a risk factor for dementia and mild cognitive impairment: a meta-analysis of longitudinal studies. *Intern Med J*. 2012; 42:484–91. <https://doi.org/10.1111/j.1445-5994.2012.02758.x> PMID:[22372522](https://pubmed.ncbi.nlm.nih.gov/22372522/)
6. Cereda E, Barichella M, Pedrolli C, Klersy C, Cassani E, Caccialanza R, Pezzoli G. Diabetes and risk of Parkinson's disease. *Mov Disord*. 2013; 28:257–61. <https://doi.org/10.1002/mds.25211> PMID:[23032425](https://pubmed.ncbi.nlm.nih.gov/23032425/)
7. Baglietto-Vargas D, Shi J, Yaeger DM, Ager R, LaFerla FM. Diabetes and Alzheimer's disease crosstalk. *Neurosci Biobehav Rev*. 2016; 64:272–87. <https://doi.org/10.1016/j.neubiorev.2016.03.005> PMID:[26969101](https://pubmed.ncbi.nlm.nih.gov/26969101/)
8. Montojo MT, Aganzo M, González N. Huntington's Disease and Diabetes: Chronological Sequence of its Association. *J Huntingtons Dis*. 2017; 6:179–88. <https://doi.org/10.3233/JHD-170253> PMID:[28968242](https://pubmed.ncbi.nlm.nih.gov/28968242/)
9. Spauwen PJ, Köhler S, Verhey FR, Stehouwer CD, van Boxtel MP. Effects of type 2 diabetes on 12-year cognitive change: results from the Maastricht Aging Study. *Diabetes Care*. 2013; 36:1554–61. <https://doi.org/10.2337/dc12-0746> PMID:[23275366](https://pubmed.ncbi.nlm.nih.gov/23275366/)
10. Rawlings AM, Sharrett AR, Schneider AL, Coresh J, Albert M, Couper D, Griswold M, Gottesman RF, Wagenknecht LE, Windham BG, Selvin E. Diabetes in midlife and cognitive change over 20 years: a cohort study. *Ann Intern Med*. 2014; 161:785–93. <https://doi.org/10.7326/M14-0737> PMID:[25437406](https://pubmed.ncbi.nlm.nih.gov/25437406/)
11. Tuligenga RH, Dugravot A, Tabák AG, Elbaz A, Brunner EJ, Kivimäki M, Singh-Manoux A. Midlife type 2 diabetes and poor glycaemic control as risk factors for cognitive decline in early old age: a post-hoc analysis of the Whitehall II cohort study. *Lancet Diabetes Endocrinol*. 2014; 2:228–35. [https://doi.org/10.1016/S2213-8587\(13\)70192-X](https://doi.org/10.1016/S2213-8587(13)70192-X) PMID:[24622753](https://pubmed.ncbi.nlm.nih.gov/24622753/)
12. Rajan KB, Arvanitakis Z, Lynch EB, McAninch EA, Wilson RS, Weuve J, Barnes LL, Bianco AC, Evans DA. Cognitive decline following incident and preexisting diabetes mellitus in a population sample. *Neurology*. 2016; 87:1681–87. <https://doi.org/10.1212/WNL.0000000000003226> PMID:[27655734](https://pubmed.ncbi.nlm.nih.gov/27655734/)
13. Wisse LE, de Bresser J, Geerlings MI, Reijmer YD, Portegies ML, Brundel M, Kappelle LJ, van der Graaf Y, Biessels GJ, and Utrecht Diabetic Encephalopathy Study Group, and SMART-MR Study Group. Global brain atrophy but not hippocampal atrophy is related to type 2 diabetes. *J Neurol Sci*. 2014; 344:32–36. <https://doi.org/10.1016/j.jns.2014.06.008> PMID:[24958596](https://pubmed.ncbi.nlm.nih.gov/24958596/)
14. Biessels GJ, Reijmer YD. Brain changes underlying cognitive dysfunction in diabetes: what can we learn from MRI? *Diabetes*. 2014; 63:2244–52. <https://doi.org/10.2337/db14-0348> PMID:[24931032](https://pubmed.ncbi.nlm.nih.gov/24931032/)
15. De Felice FG, Ferreira ST. Inflammation, defective insulin signaling, and mitochondrial dysfunction as common molecular denominators connecting type 2 diabetes to Alzheimer disease. *Diabetes*. 2014; 63:2262–72. <https://doi.org/10.2337/db13-1954> PMID:[24931033](https://pubmed.ncbi.nlm.nih.gov/24931033/)
16. Simó R, Ciudin A, Simó-Servat O, Hernández C. Cognitive impairment and dementia: a new emerging complication of type 2 diabetes-The diabetologist's perspective. *Acta Diabetol*. 2017; 54:417–24. <https://doi.org/10.1007/s00592-017-0970-5> PMID:[28210868](https://pubmed.ncbi.nlm.nih.gov/28210868/)
17. Zilliox LA, Chadrasekaran K, Kwan JY, Russell JW. Diabetes and Cognitive Impairment. *Curr Diab Rep*. 2016; 16:87. <https://doi.org/10.1007/s11892-016-0775-x> PMID:[27491830](https://pubmed.ncbi.nlm.nih.gov/27491830/)
18. Olokoba AB, Obateru OA, Olokoba LB. Type 2 diabetes mellitus: a review of current trends. *Oman Med J*. 2012; 27:269–73. <https://doi.org/10.5001/omj.2012.68> PMID:[23071876](https://pubmed.ncbi.nlm.nih.gov/23071876/)
19. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, Foster B, Moser M, Karasik E, et al, and GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013; 45:580–85. <https://doi.org/10.1038/ng.2653> PMID:[23715323](https://pubmed.ncbi.nlm.nih.gov/23715323/)
20. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014; 15:550. <https://doi.org/10.1186/s13059-014-0550-8> PMID:[25516281](https://pubmed.ncbi.nlm.nih.gov/25516281/)
21. Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, Barnes I, Berry A, Bignell A, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*. 2019; 47:D766–73. <https://doi.org/10.1093/nar/gky955> PMID:[30357393](https://pubmed.ncbi.nlm.nih.gov/30357393/)
22. Cohen BA, Mitra RD, Hughes JD, Church GM. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat Genet*. 2000; 26:183–86. <https://doi.org/10.1038/79896> PMID:[11017073](https://pubmed.ncbi.nlm.nih.gov/11017073/)

23. Wang J, Gong B, Zhao W, Tang C, Varghese M, Nguyen T, Bi W, Bilski A, Begum S, Vempati P, Knable L, Ho L, Pasinetti GM. Epigenetic mechanisms linking diabetes and synaptic impairments. *Diabetes*. 2014; 63:645–54. <https://doi.org/10.2337/db13-1063> PMID:[24154559](https://pubmed.ncbi.nlm.nih.gov/24154559/)
24. Lee H, Song J, Jung JH, Ko HW. Primary cilia in energy balance signaling and metabolic disorder. *BMB Rep*. 2015; 48:647–54. <https://doi.org/10.5483/BMBRep.2015.48.12.229> PMID:[26538252](https://pubmed.ncbi.nlm.nih.gov/26538252/)
25. Xu X, Wells AB, O'Brien DR, Nehorai A, Dougherty JD. Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. *J Neurosci*. 2014; 34:1420–31. <https://doi.org/10.1523/JNEUROSCI.4488-13.2014> PMID:[24453331](https://pubmed.ncbi.nlm.nih.gov/24453331/)
26. van Dam S, Cordeiro R, Craig T, van Dam J, Wood SH, de Magalhães JP. GeneFriends: an online co-expression analysis tool to identify novel gene targets for aging and complex diseases. *BMC Genomics*. 2012; 13:535. <https://doi.org/10.1186/1471-2164-13-535> PMID:[23039964](https://pubmed.ncbi.nlm.nih.gov/23039964/)
27. Aslamy A, Thurmond DC. Exocytosis proteins as novel targets for diabetes prevention and/or remediation? *Am J Physiol Regul Integr Comp Physiol*. 2017; 312:R739–52. <https://doi.org/10.1152/ajpregu.00002.2017> PMID:[28356294](https://pubmed.ncbi.nlm.nih.gov/28356294/)
28. Yazdanpanah M, Sayed-Tabatabaei FA, Hofman A, Aulchenko YS, Oostra BA, Stricker BH, Pols HA, Lamberts SW, Witteman JC, Janssen JA, van Duijn CM. The alpha-adducin gene is associated with macrovascular complications and mortality in patients with type 2 diabetes. *Diabetes*. 2006; 55:2922–27. <https://doi.org/10.2337/db06-0302> PMID:[17003363](https://pubmed.ncbi.nlm.nih.gov/17003363/)
29. Larsson M, Lietzau G, Nathanson D, Östenson CG, Mallard C, Johansson ME, Nyström T, Patrone C, Darsalia V. Diabetes negatively affects cortical and striatal GABAergic neurons: an effect that is partially counteracted by exendin-4. *Biosci Rep*. 2016; 36:36. <https://doi.org/10.1042/BSR20160437> PMID:[27780892](https://pubmed.ncbi.nlm.nih.gov/27780892/)
30. Gold SM, Dziobek I, Sweat V, Tirsi A, Rogers K, Bruehl H, Tsui W, Richardson S, Javier E, Convit A. Hippocampal damage and memory impairments as possible early brain complications of type 2 diabetes. *Diabetologia*. 2007; 50:711–19. <https://doi.org/10.1007/s00125-007-0602-7> PMID:[17334649](https://pubmed.ncbi.nlm.nih.gov/17334649/)
31. Rubin DC. Frontal-striatal circuits in cognitive aging: evidence for caudate involvement. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn*. 1999; 6:241–59. [https://doi.org/10.1076/1382-5585\(199912\)06:04;1-B;FT241](https://doi.org/10.1076/1382-5585(199912)06:04;1-B;FT241)
32. Grahn JA, Parkinson JA, Owen AM. The cognitive functions of the caudate nucleus. *Prog Neurobiol*. 2008; 86:141–55. <https://doi.org/10.1016/j.pneurobio.2008.09.004> PMID:[18824075](https://pubmed.ncbi.nlm.nih.gov/18824075/)
33. Haber SN. Corticostriatal circuitry. *Dialogues Clin Neurosci*. 2016; 18:7–21. [https://doi.org/10.1007/978-1-4614-6434-1\\_135-1](https://doi.org/10.1007/978-1-4614-6434-1_135-1) PMID:[27069376](https://pubmed.ncbi.nlm.nih.gov/27069376/)
34. Lucassen EB, Delfyett WT, Stahl MC. Persistent Hemichorea and Caudate Atrophy in Untreated Diabetic Striatopathy: A Case Report. *Case Rep Neurol*. 2017; 9:299–303. <https://doi.org/10.1159/000484201> PMID:[29422853](https://pubmed.ncbi.nlm.nih.gov/29422853/)
35. Moran C, Phan TG, Chen J, Blizzard L, Beare R, Venn A, Münch G, Wood AG, Forbes J, Greenaway TM, Pearson S, Srikanth V. Brain atrophy in type 2 diabetes: regional distribution and influence on cognition. *Diabetes Care*. 2013; 36:4036–42. <https://doi.org/10.2337/dc13-0143> PMID:[23939539](https://pubmed.ncbi.nlm.nih.gov/23939539/)
36. Ardlie KG, Deluca DS, Segre AV, Sullivan TJ, Young TR, Gelfand ET, Trowbridge CA, Maller JB, Tukiainen T, Lek M, Ward LD, Kheradpour P, Iriarte B, et al, and GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015; 348:648–60. <https://doi.org/10.1126/science.1262110> PMID:[25954001](https://pubmed.ncbi.nlm.nih.gov/25954001/)
37. Yang J, Huang T, Petralia F, Long Q, Zhang B, Argmann C, Zhao Y, Mobbs CV, Schadt EE, Zhu J, Tu Z, and GTEx Consortium. Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases. *Sci Rep*. 2015; 5:15145. <https://doi.org/10.1038/srep15145> PMID:[26477495](https://pubmed.ncbi.nlm.nih.gov/26477495/)
38. Gershoni M, Pietrokovski S. The landscape of sex-differential transcriptome and its consequent selection in human adults. *BMC Biol*. 2017; 15:7. <https://doi.org/10.1186/s12915-017-0352-z> PMID:[28173793](https://pubmed.ncbi.nlm.nih.gov/28173793/)
39. Hao RH, Yang TL, Rong Y, Yao S, Dong SS, Chen H, Guo Y. Gene expression profiles indicate tissue-specific obesity regulation changes and strong obesity relevant tissues. *Int J Obes*. 2018; 42:363–69. <https://doi.org/10.1038/ijo.2017.283> PMID:[29151593](https://pubmed.ncbi.nlm.nih.gov/29151593/)
40. Zhu Y, Wang L, Yin Y, Yang E. Systematic analysis of gene expression patterns associated with

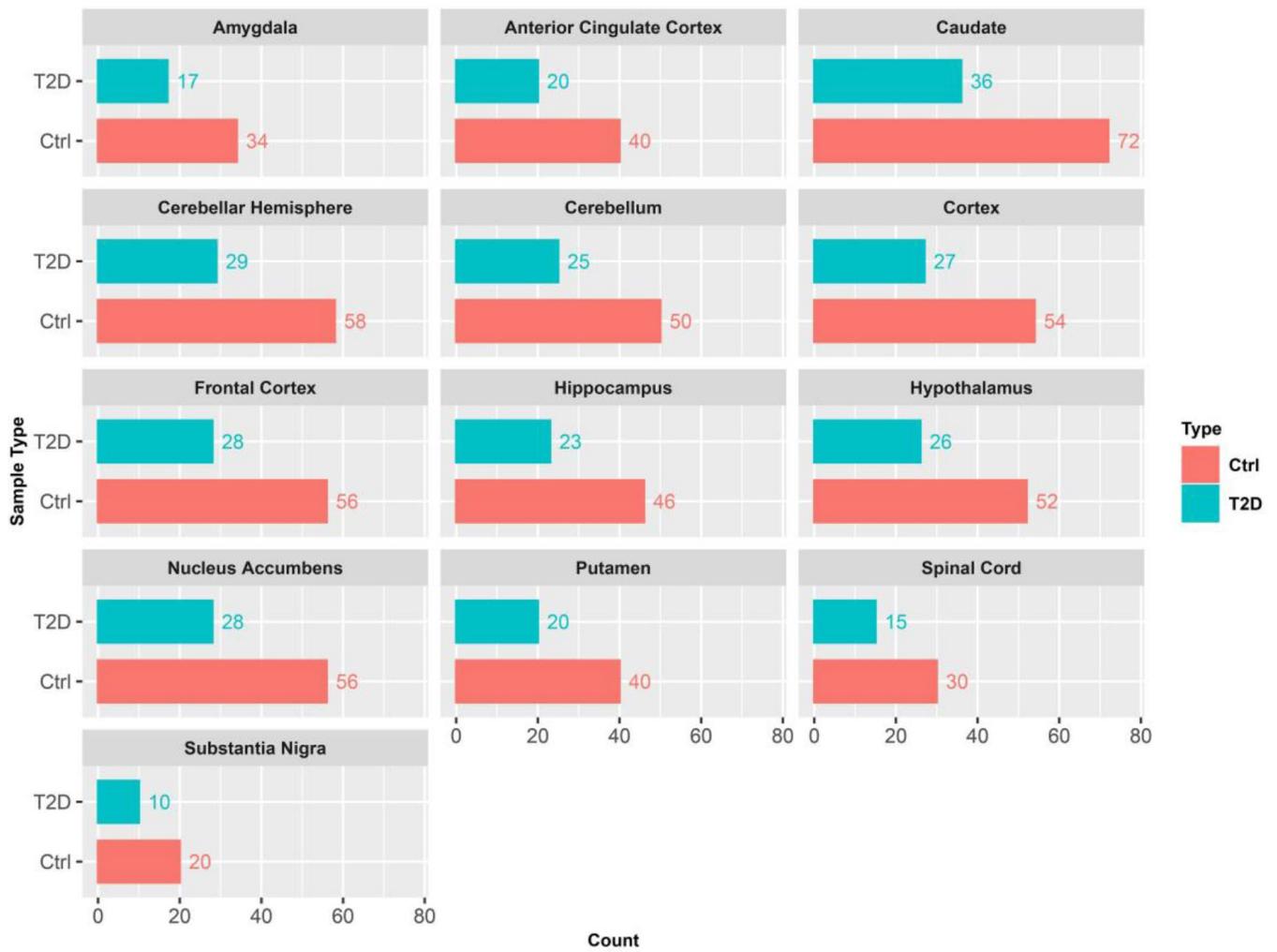
- postmortem interval in human tissues. *Sci Rep.* 2017; 7:5435. <https://doi.org/10.1038/s41598-017-05882-0> PMID:[28710439](https://pubmed.ncbi.nlm.nih.gov/28710439/)
41. Ho D, Imai K, King G, Stuart EA. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *J Stat Softw.* 2011; 42:1–28. <https://doi.org/10.18637/jss.v042.i08>
  42. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics.* 2012; 28:882–83. <https://doi.org/10.1093/bioinformatics/bts034> PMID:[22257669](https://pubmed.ncbi.nlm.nih.gov/22257669/)
  43. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012; 16:284–87. <https://doi.org/10.1089/omi.2011.0118> PMID:[22455463](https://pubmed.ncbi.nlm.nih.gov/22455463/)
  44. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008; 9:559. <https://doi.org/10.1186/1471-2105-9-559> PMID:[19114008](https://pubmed.ncbi.nlm.nih.gov/19114008/)
  45. Nepusz GCaT. The igraph software package for complex network research. *InterJournal.* 2006. *Complex Syst.* 1695.
  46. Dougherty JD, Schmidt EF, Nakajima M, Heintz N. Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. *Nucleic Acids Res.* 2010; 38:4218–30. <https://doi.org/10.1093/nar/gkq130> PMID:[20308160](https://pubmed.ncbi.nlm.nih.gov/20308160/)
  47. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, Mill J, Cantor RM, Blencowe BJ, Geschwind DH. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature.* 2011; 474:380–84. <https://doi.org/10.1038/nature10110> PMID:[21614001](https://pubmed.ncbi.nlm.nih.gov/21614001/)
  48. Xue A, Wu Y, Zhu Z, Zhang F, Kemper KE, Zheng Z, Yengo L, Lloyd-Jones LR, Sidorenko J, Wu Y, McRae AF, Visscher PM, Zeng J, Yang J, and eQTLGen Consortium. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat Commun.* 2018; 9:2941. <https://doi.org/10.1038/s41467-018-04951-w> PMID:[30054458](https://pubmed.ncbi.nlm.nih.gov/30054458/)
  49. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan J, Kutalik Z, Amin N, Buchkovich ML, Croteau-Chonka DC, et al, and Electronic Medical Records and Genomics (eMEMERGE) Consortium, and MIGen Consortium, and PAGEGE Consortium, and LifeLines Cohort Study. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet.* 2014; 46:1173–86. <https://doi.org/10.1038/ng.3097> PMID:[25282103](https://pubmed.ncbi.nlm.nih.gov/25282103/)
  50. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics.* 2007; 23:3251–53. <https://doi.org/10.1093/bioinformatics/btm369> PMID:[17644558](https://pubmed.ncbi.nlm.nih.gov/17644558/)

# SUPPLEMENTARY MATERIALS

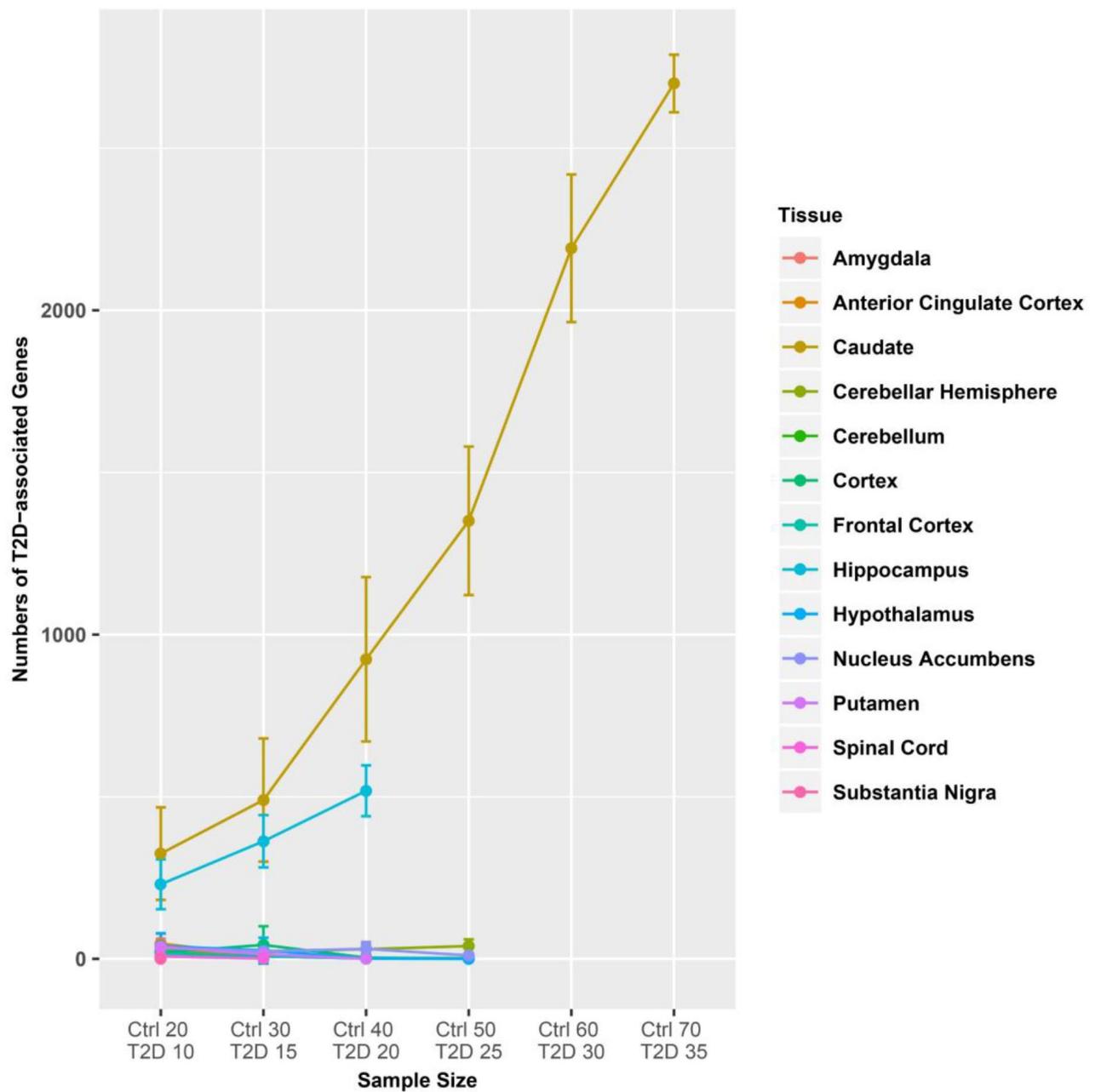
## Supplementary Figures



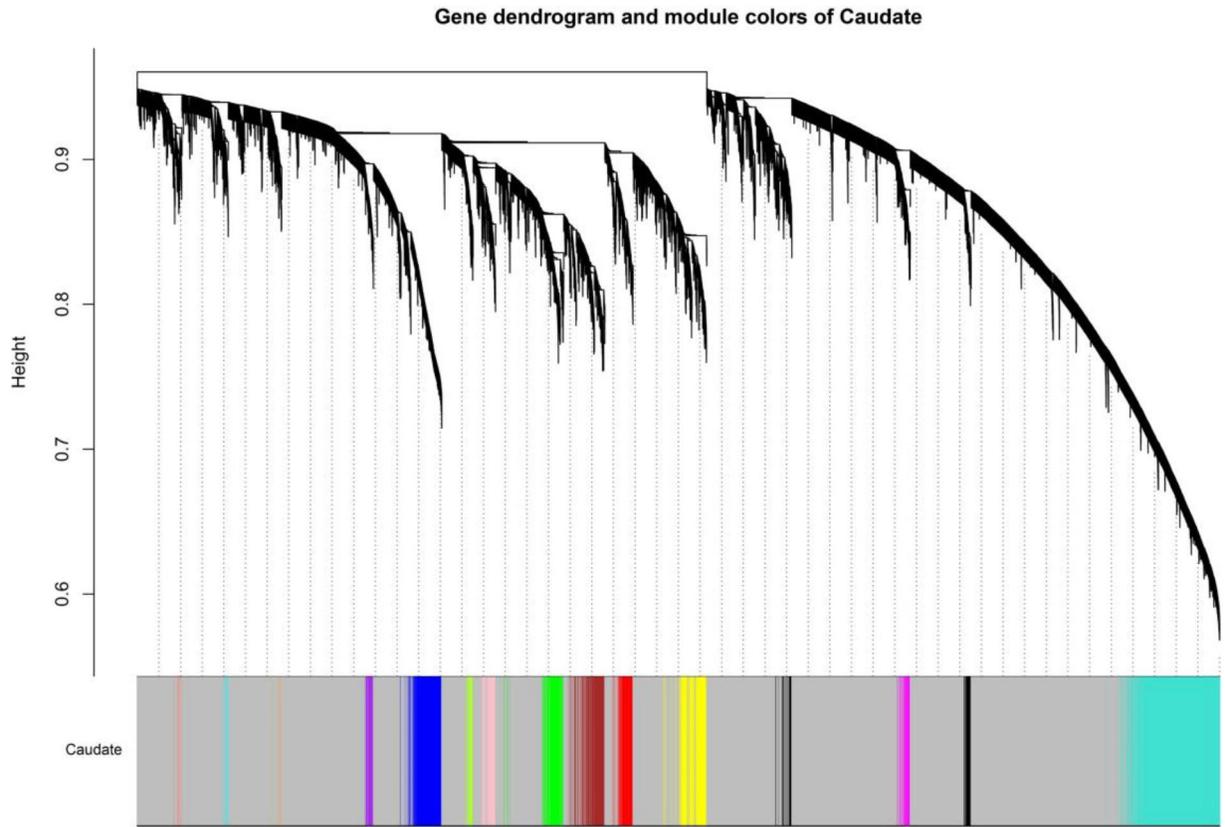
**Supplementary Figure 1. Distribution of propensity scores.** Jitter plot showing the distribution of propensity scores in T2D samples and matched controls for each brain regions respectively.



**Supplementary Figure 2. Sample distribution of 13 brain regions after matching.** Histograms denote the distribution of matched samples in each brain regions respectively.

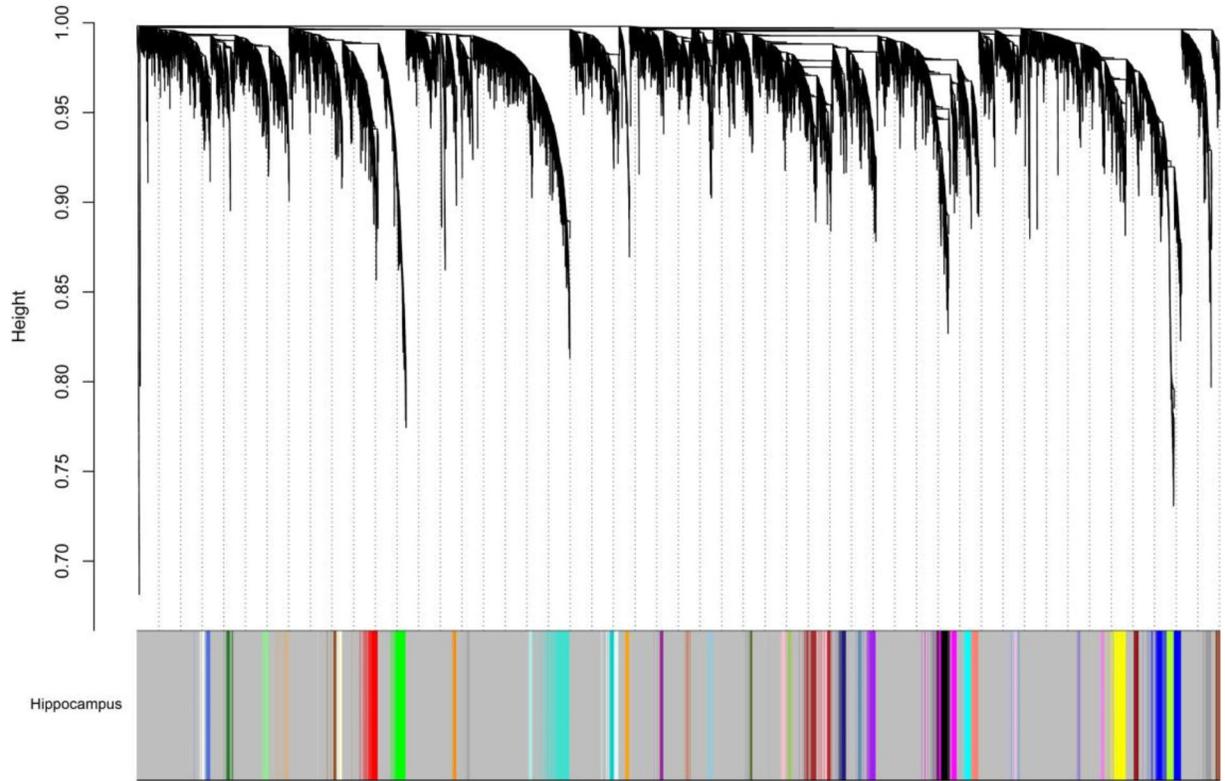


**Supplementary Figure 3. Effect of sample size on detected DAGs in 13 brain regions.** The x-axis indicates the sample size of T2D and matched control groups. The y-axis indicates the average number of DAGs in 100 repetitions of bootstrapping at a 5% FDR level. The error bar indicates 95% confidence interval of group means.

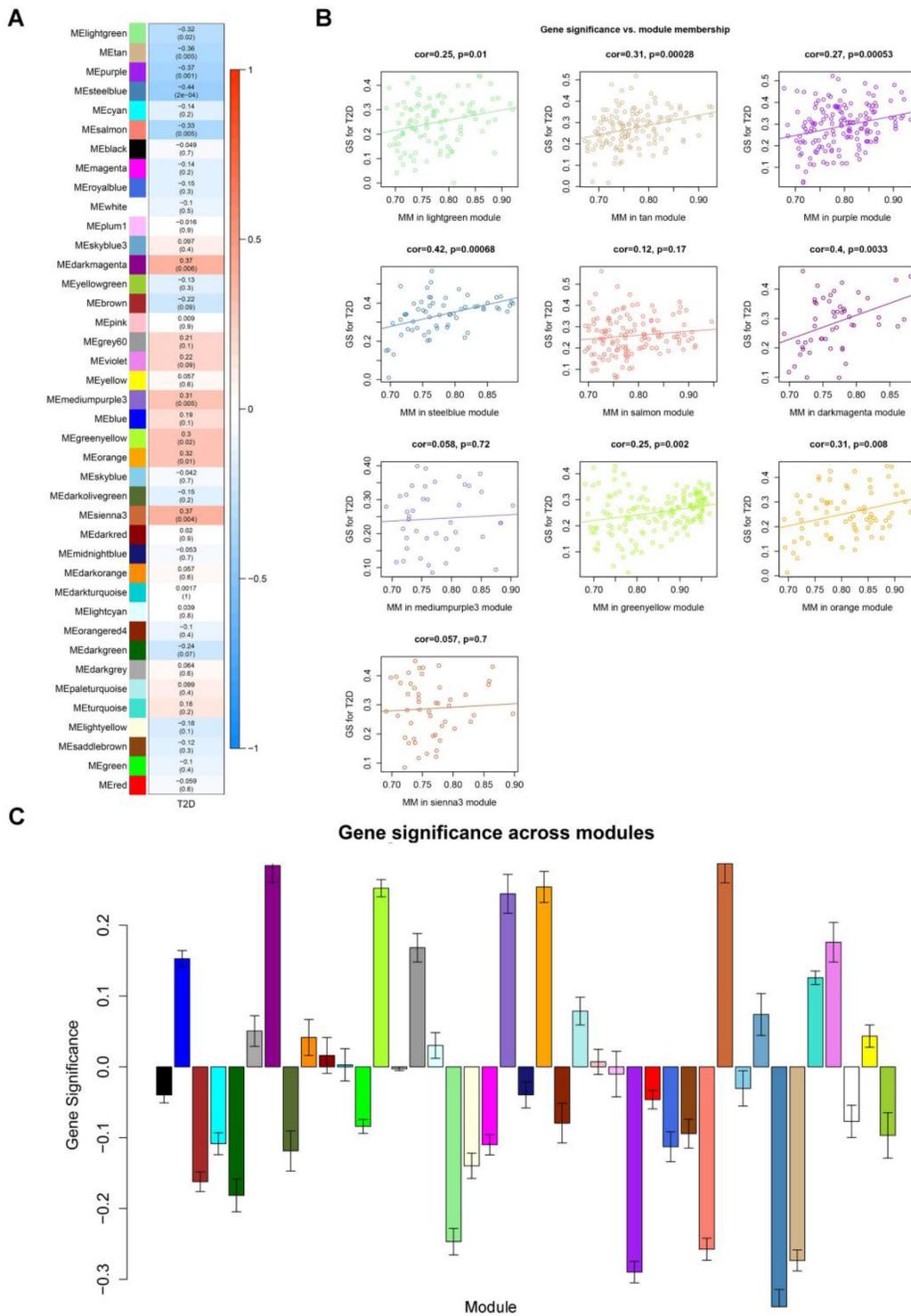


**Supplementary Figure 4. Dendrogram for module assignment in caudate.** The branches correspond to modules of highly interconnected groups of genes. Colors in the horizontal bar represent the modules and grey indicates unassigned genes.

Gene dendrogram and module colors of Hippocampus



**Supplementary Figure 5. Dendrogram for module assignment in hippocampus.** The branches correspond to modules of highly interconnected groups of genes. Colors in the horizontal bar represent the modules and grey indicates unassigned genes.



**Supplementary Figure 6. Co-expression modules in hippocampus.** (A) Each row corresponds to a module eigengene and column indicate T2D status. The table were colored by correlation according to the legend. Each cell contains the corresponding correlation and P-value. (B) GS vs MM plot for modules significantly correlated with T2D status. Each point corresponds to an individual gene within a given module, which was plotted by GS on the y-axis and MM on the x-axis. The regression line, correlation value and P-value were shown for each module. (C) Each bar indicates the average of gene significance measure for all genes in a given module.



## Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1–7.

**Supplementary Table 1. Statistics of known covariates after sample matching.**

**Supplementary Table 2. Details of DAGs in each brain region.**

**Supplementary Table 3. Functional annotations of DAGs.**

**Supplementary Table 4. Membership information for all assigned genes in caudate.**

**Supplementary Table 5. Membership information for all assigned genes in hippocampus.**

**Supplementary Table 6. Enrichment of regional and cell-type markers in caudate modules at various specificity indices.**

**Supplementary Table 7. Enrichment of regional and cell-type markers in hippocampus modules at various specificity indices.**

**Supplementary Table 8. SNP-based enrichment analysis for caudate modules and sub-modules.**

Module	T2D SNPs				Height SNPs			
	Size	ES	P-value	FDR	Size	ES	P-value	FDR
turquoise	1576	0.415712	0	0	1576	0.4111951	0.0497	0.13916
red	260	0.425652	0.00815	0.03325	260	0.42124715	0.2414	0.289756
magenta	156	0.444919	0.0091	0.03325	156	0.45935112	0.12445	0.2489
green	373	0.435294	0.0095	0.03325	374	0.4505773	0.04135	0.13916
tan	61	0.504785	0.01205	0.03374	61	0.46593708	0.248362	0.289756
yellow	481	0.368788	0.03835	0.089483	481	0.49281818	3.00E-04	0.0042
purple	118	0.400124	0.09525	0.1905	118	0.38965136	0.56555	0.56555
salmon	39	0.439662	0.163975	0.286956	39	0.56709814	0.069181	0.161423
greenyellow	95	0.384014	0.20735	0.306483	95	0.57951885	0.00415	0.019367
black	228	0.351542	0.24705	0.306483	228	0.42270425	0.24785	0.289756
blue	468	0.337558	0.25135	0.306483	468	0.40408793	0.2989	0.321892
brown	503	0.335818	0.2627	0.306483	503	0.47241223	0.00295	0.019367
pink	215	0.345973	0.30135	0.324531	215	0.42818555	0.22605	0.289756
cyan	37	0.315055	0.746361	0.746361	37	0.49608	0.228706	0.289756

Sub-module	T2D SNPs				Height SNPs			
	Size	ES	P-value	FDR	Size	ES	P-value	FDR
D2-MSN	71	0.485125	0.0144	0.0432	71	0.4155968	0.45245	0.49075
D1-MSN	75	0.453621	0.0355	0.04715	75	0.47593623	0.1792	0.49075
striatum	83	0.4398	0.04715	0.04715	83	0.40625075	0.49075	0.49075

**Supplementary Table 9. Modules enriched in DAGs.**

Up-regulated DAGs					Down-regulated DAGs				
Module	GeneRatio	P-value	FDR	Count	Module	GeneRatio	P-value	FDR	Count
blue	460/1186	0	0	460	turquoise	887/1753	0	0	887
green	57/1186	5.81E-12	1.60E-11	57	black	54/1753	6.01E-13	1.35E-12	54
purple	26/1186	5.44E-09	1.20E-08	26					

Note: GeneRatio refers to Number of intersection genes / Size of Up- or Down-regulated DAG set.

Please browse Full Text version to see the data of Supplementary Table 10.

**Supplementary Table 10. Functional annotations of caudate modules (sub-modules).**