UNIVERSITY OF CALIFORNIA SAN DIEGO

Obesity and diabetes risk variants affect chromatin and gene regulation in the human brain

A thesis submitted in partial satisfaction of the requirements for the degree
Master of Science

in

Biology

by

Whitney Tan Vien

Committee in charge:

Professor Kyle J Gaulton, Chair
Professor Gulcin Pekkurnaz, Co-Chair
Professor James T Kadonaga

2018

The Thesis of Whitney Tan Vien is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

Co-chair

_____

Chair

University of California San Diego

2018

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTAL FIGURES

ACKNOWLEDGEMENTS

I would like to thank Dr. Kyle Gaulton for all his support as my advisor. I am grateful for his guidance from the past year.

I would like to thank Dr. Gulcin Pekkurnaz and Dr. James Kadonaga for being members of my committee.

I would like to thank the members of the Gaulton lab for all of their support.

I would also like to acknowledge NFS XSEDE for the GPU allocation resources that allowed me to perform the deep learning portion of this thesis.

The material in this thesis is currently being prepared for submission for publication of the material. Vien, WT; Chiou, J; Aylward, AJ; Okino, M; Gaulton, KJ. “Diabetes risk variants affect chromatin and gene regulation in the human brain”. I was the primary researcher and author of this material.

ABSTRACT OF THE THESIS

Obesity and diabetes risk variants affect chromatin and gene regulation in the human brain

by

Whitney Tan Vien

Master of Science in Biology

University of California San Diego, 2018

Professor Kyle J Gaulton, Chair
Professor Gulcin Pekkurnaz, Co-Chair

Obesity and type 2 diabetes (T2D) are complex diseases that are major public health concerns. Genetic variation contributes substantially to risk of developing obesity and T2D. Many previous studies on genetic factors influencing obesity (as measured by body mass index, or BMI) and T2D have focused on the effects of insulin resistance and glucose homeostasis in the pancreas and adipose tissue. A recent study demonstrated that genetic loci associated with BMI are enriched near genes expressed in the brain, but the effects of specific BMI variants on brain function is unknown. Furthermore, despite being a key regulator of energy intake as well as glucose

homeostasis, the brain has not been a focus in the molecular genetics of T2D risk. In this work, we aimed to determine the effects of BMI and T2D variants on regulatory processes in the brain through computational analyses integrating genetic association data and brain epigenome data. We determined the enrichment of enhancers in various regions of the brain for BMI and T2D risk signals, identified BMI and T2D risk signals affecting gene expression in the brain, predicted the allelic effects of variants based on brain chromatin features, and prioritized specific BMI and T2D variants likely affecting brain regulation. These data together suggest that variants associated with BMI and T2D risk are broadly enriched for effects on regulatory processes in the brain. Together with pancreatic and adipose tissue, the brain should be considered when studying the molecular mechanisms of genetic variants affecting obesity and diabetes pathogenesis.

INTRODUCTION

1.1 Obesity and type 2 diabetes

In the United States, more than 1 in 3 adults (36.5%) suffer from obesity, a disease that is characterized by excess body fat from the excess of energy intake over expenditure, and is usually measured by body mass index (BMI) greater than 30[1]. Obesity predisposes to many health conditions, including type 2 diabetes (T2D). The two complex diseases have been genetically associated, despite differences in etiology[2]. In addition, diabetes affects about 30.3 million Americans (9.4%), where T2D accounts for the majority of diabetes cases (93%)[3]. Type 2 diabetes is characterized by the gradual dysfunction of pancreatic beta cells caused by insulin resistance which leads to blood hyperglycemia[4]. Diabetics have a 50% higher mortality rate than non-diabetics, as T2D can lead to complications such as cardiovascular disease and stroke[3]. The prevalence of diabetes has tripled in the last 20 years, and understanding the genetic basis and heritability of T2D and obesity in disease-relevant tissues can identify novel therapeutic avenues to help alleviate the widespread public health concern[5].

1.2 Brain regulation of glucose homeostasis

Studies on the molecular mechanisms of obesity and T2D in the past have mainly focused on the effects of insulin resistance on tissues such as the pancreas, liver, skeletal muscle, and adipose tissue[6]. In humans, glucose levels are maintained by insulin-dependent and insulin-independent mechanisms[7]. Individuals with T2D are not able to maintain glucose homeostasis due to dysfunction of pancreatic islet beta-cells that are responsible for producing insulin, and will gradually develop insulin resistance in peripheral tissues[7].

Although the brain is a key regulator of insulin and glucose across many organs, the brain has not been largely studied in the context of genetic effects particularly on T2D[8]. Many studies have shown that mice have two insulin genes, where *Ins1* is only expressed in pancreatic islets and *Ins2* is expressed more broadly and is more similar to the human *INS* gene[9,10]. In addition, expression of *Ins2* and *INS* is found in multiple regions of both rodent and human postnatal brain, most prominently in the hippocampus[11]. Another study on insulin resistant in the human brain demonstrated that insulin modulates cerebrocortical activity in lean individuals, but insulin levels were significantly reduced in obese individuals; this resistance effect was found in people with a polymorphism at insulin receptor substrate-1 (*IRS1*), a T2D risk locus[12]. Another study used a rat model for diabetes to learn the effects of the brain in glucose homeostasis. They found that the hypothalamic insulin-PI3K signaling is impaired and inhibiting the signaling pathway leads to a glucose-lowering effect of insulin[13]. In addition, insulin signaling in the hypothalamus and those altering pancreatic hormone secretion has also been shown to be disrupted by a fatty diet, indicating that too much food intake may contribute to diabetes[14]. The molecular basis of glucose homeostasis in obesity and T2D are currently being studied with cellular and animal brain models, but learning the risk polymorphisms and how they affect brain regulatory processes through genetic analyses can provide more confidence for where in the genome scientists should target.

1.3 Genetic studies of BMI and T2D

While obesity and T2D are mainly influenced by environmental factors, such as sedentary lifestyle and increased food consumption, not all individuals experiencing these conditions result in these diseases, suggesting that genetics and epigenetics may play a role [2,4]. For example, studies on identical twins and immediate relatives have shown that there is genetic contribution to the

development of T2D[4,15]. More recently, genome-wide association studies (GWASs) and meta-analyses, including fine-mapping analyses and trans-ancestry analyses, have identified over a hundred genetic loci influencing BMI and T2D, with several loci overlapping both traits[16–19]. Many genetic studies of BMI have shown enrichment of variants in active enhancers in not only adipose tissue, but also in various regions of the brain and other immune-related tissues; however the specific brain region(s) most involved in regulation of obesity have not been identified[19,20].

GWASs are used to determine trait-associated regions on the genomic scale, but in some cases, the causal single nucleotide polymorphism (SNP), or variant, at a risk locus may be missed, and instead another variant may be identified as the most strongly associated[21]. To compensate for this problem of linkage disequilibrium (LD), fine-mapping analyses genotype more samples to identify the subset of highly associated variants likely causal for a disease signal[21]. Genetic analyses across different ethnicities are often performed to further improve statistical power to fine-map disease signals by leveraging differences in LD across populations. For instance, while most GWASs have been done in European populations, a recent study identified 112 new loci for BMI in the Japanese population[20]. Data from trans-ethnic studies can then be combined through meta-analysis.

Moreover, novel disease-risk loci and variants from these analyses must be compared to reference population catalogs that can serve as models for genetic variation in the global human population. In particular, the 1000 Genomes Project, which has genotyped thousands of individuals from African, American, East Asian, European, and South Asian populations with whole-genome sequencing, is used in this project as a reference for single nucleotide variants with frequencies of at least 1% in the population[22].

Over the past decade, GWASs have given insight to the genetic architectures of BMI and T2D; however, known loci for these traits still only explain a fraction of total genetic heritability [23]. For example, only about 20% of the estimated genetic risk of T2D is driven by the identified >100 loci[24,25]. The remaining heritability of T2D is unlikely to be explained by large effect variants, suggesting that these traits are extremely polygenic and involve many additional hundreds of loci[16]. Despite having small effects, each BMI and T2D risk locus has a biological function which can inform on the underlying mechanisms and genes involved in disease. Association studies of BMI and T2D have shown that most risk variants map to non-coding regions of the genome, suggesting that gene regulation is an important contributor to these traits[23].

1.4 Mapping the human epigenome

Non-genetic, chemical changes to the genome, like histone modifications and chromatin accessibility, are considered epigenetic[26]. The primary role of eukaryotic chromatin is to tightly package DNA with histone proteins, forming nucleosomes[27]. The positioning of nucleosomes affects the availability of binding sites for transcription factors, thus regulating gene expression[28]. Many histone modifications are involved in the control of transcription[24]. As the cost of high-throughput sequencing has declined over the years, sequenced-based molecular assays of cells and tissues have become increasingly popular for studying gene regulation. The Encyclopedia of DNA Elements (ENCODE) project is a database of various sequence data that reveal information about functional elements of the human genome[29]. Because epigenome information varies across the organism, these sequencing-based assays are done in a variety of tissues, including the human pancreas, hippocampus, and dorsolateral prefrontal cortex[29]. The NIH Roadmap Epigenomics

Project has mapped epigenetic marks in 111 human tissues and cell types, and the data can be found in the ENCODE database[29].

The data used in the present study include chromatin immunoprecipitation sequencing (ChIP-seq) and RNA sequencing (RNA-seq) performed on human brain tissue. ChIP-seq identifies DNA at locations bound by specific proteins, such as transcription factors[30]. Histone modifications can also be ChIP-sequenced by targeting histone acetylation and methylation marks to study activating or repressing effects on gene regulation[31]. The information can be processed using a multivariate Hidden Markov Model (HMM) to reveal chromatin states on the genome in various tissues, which is then used to associate variants to functional annotations[32]. The NIH Roadmap Epigenomics project utilizes the ChromHMM model, which was trained on 40 epigenomes with ChIP-seq data corresponding to H3K4me3 (promoters), H3K4me1 (enhancers), H3K36me3 (active transcribed region), H3K27me3 (repressors), H3K9me3 (repressors), and H3K27ac (enhancers), to determine a total of 18 chromatin states[29,32]. Additionally, RNA-seq measures the genome-wide accumulated RNA transcripts in the sample, which is then used to determine levels of gene expression in certain cells and tissues[33]. Unlike the traditional use of gene expression microarrays, RNA-seq provides more accurate quantification of transcript levels, and this is useful for mapping variants to quantitative trait loci[34]. Another common technique for learning about chromatin is assay for transposase-accessible chromatin using sequencing (ATAC-seq), which captures open chromatin (nucleosome-free) sites and reveals regions of the genome that correspond to regulatory elements[35]. Thus, ATAC-seq is a useful tool for studying transcriptional activity and thereby control of gene expression.

1.5 Analytical approaches and validation studies for identifying risk variants

By integrating BMI and T2D genetic data and epigenome data, disease-risk variants potentially influencing regulatory processes in specific cell types and tissues can be identified. To prioritize the potential regulatory effects of these risk variants, their disease-risk loci can be mapped to identify variants overlapping epigenome annotations in each cell type. More recent approaches such as deep learning modeling can also be utilized to predict variant effects on cell type chromatin features directly.

Correlating genotyped samples to molecular phenotypes such as gene expression is used to identify genetic variants that influence the phenotype, called quantitative trait loci (QTL). The most common type is expression quantitative trait locus (eQTL), which maps variants to a locus that contributes to genetic variation in gene expression levels[36,37]. The GTEx project performed eQTL mapping with RNA-seq for expression data in a large variety of human tissues, including 13 brain tissues[38]. While many eQTLs are shared across tissues and near transcription start sites, other eQTLs are specific to certain tissues; this data is useful for studying the role of gene regulation in disease, since most risk loci are in non-coding regions[37,39]. Furthermore, the association of variants for two traits, such as disease-risk and gene expression, are dependent on each other because of linkage disequilibrium, which is the association of disease phenotype and marker genotype due to proximity of their loci[40]. Approaches such as a Bayesian method for colocalization between two traits can determine whether the association signals of both traits are consistent with a shared causal variant; variants at the same genomic location with similar association statistics are said to be colocalized[41]. Associating molecular traits, such as gene expression, with genetic variants is useful for learning the underlying mechanism of the disease.

Another tool for determining the effects of disease-risk variants is by learning features of sequence around these variants. The large amount of complicated BMI and T2D genetic data and epigenome data is particularly well suited for this application. Deep learning is a machine learning method that learns from hierarchical layers of information and is modeled as a neural network, where connections are transmitted through nodes, creating input and output layers of information[42]. In this project, the algorithm framework, DeepSEA, is used to train a neural network to learn histone modification effects from brain ChIP-seq data to determine if a variant is affecting transcriptional activation[43]. Along with the functional epigenome annotations for enhancers from the hidden Markov models, predictions from DeepSEA can help prioritize BMI and T2D risk variants that may be affecting gene regulation in the brain[32].

In general, multiple approaches and lines of evidence need to be considered to prioritize candidate SNPs at a disease locus for molecular experiments. Before utilizing genome editing in cellular models and animal models to characterize the effects of variants on obesity and T2D risk, the SNPs must be tested for their effects in regulatory processes in specific cells. The functionality of the highly causal candidate variants at risk loci can be validated through, for example, (a) gene reporter assays to show if variants have differential activity on gene regulation and (b) gel shift assays, also known as electrophoretic mobility shift assays (EMSAs), to show which proteins may be differentially binding. The reporter assay allows for detection of promoter, enhancer, or repressor activation near the gene of interest, while EMSAs reveal sequence-specific transcription factor binding[44,45].

MATERIALS AND METHODS

2.1 Type 2 diabetes fine mapping

We combined fine mapping credible sets for 107 known T2D signals from European ancestry association studies from the Metabochip[17], GoT2D[16], and DIAGRAM[46] consortia. We resolved redundant data for signals published in multiple studies by prioritizing Metabochip data over GoT2D and DIAGRAM data. For GoT2D and DIAGRAM credible sets, we extracted effect estimates and standard errors from the summary statistics and calculated Bayes factors[47] for each variant. In total, our combined credible set for T2D contained information for 70,975 unique variants.

2.2 Trans-ancestry BMI fine mapping

We obtained summary statistics for BMI from a published study of BMI in a Japanese cohort (N=158,284) imputed into 1000 Genomes[20] and from the UK BioBank (N=336,107). We first calculated the relatedness matrix between the two studies with mean allele frequency differences. We next used MANTRA[48] to conduct trans-ancestry meta-analysis (combined N=494,391) of the two datasets. We then used published index variants at 163 loci[20] and defined the credible set window as all variants within a 1 Mb window centered on the index variant. We defined 100% and 99% credible sets for each locus as previously described[24]. In total, our trans-ancestry 100% and 99% credible sets for BMI contained 673,418 and 15,824 unique variants, respectively. The 100% credible set was used for colocalization analysis, while the 99% credible set was used for enrichment analyses.

2.3 Functional annotation data and enrichment analyses

To determine enrichment of regulatory elements in the brain for BMI and T2D signals, we integrated trans-ancestry BMI and European ancestry T2D fine-mapping data with brain epigenome data. Epigenome annotation data (ChromHMM 18-state model) for angular gyrus, anterior caudate, cingulate gyrus, dorsolateral prefrontal cortex, hippocampus middle, inferior temporal lobe, and substantia nigra, as well as pancreatic islets, were collected from the Roadmap Epigenomics Project[29,49]. To determine if the regulatory annotations differed across the brain, Jaccard similarity indexes were calculated for each brain element with "bedtools jaccard"[50]. Next, we used fGWAS to test for enrichment of active regulatory elements (EnhA1, EnhA2, EnhG1, EnhG2, TssA, TssFlnk, TssFlnkD, TssFlnkU annotations) in the brain for BMI and T2D signals[51]. Each brain annotation was fit to the BMI or T2D model individually, and a second joint model for each trait was built by combining multiple annotations. We also performed conditional analyses on the combined model to determine how many annotations improved the model. To determine if the brain annotations were enriched relative to pancreatic islets, another combined fGWAS model was built with the addition of islet annotation data. Posterior probabilities of association (PPA) for BMI and T2D risk were computed from the combined fGWAS brain and pancreatic islet model. To determine if T2D variants also had an effect on BMI, fasting glucose level, and fasting insulin level, association values were collected from published GWASs for these traits and compared to our T2D fine-mapping data[19,52].

To further identify chromatin state annotations that are enriched for BMI and T2D signals, we applied stratified LD-score regression on genome-wide BMI and T2D association data from European populations[53]. Stratified LD-score (SLDSC) regression is a multiple regression, where the chi-squared statistics for a trait are regressed on LD-scores computed using variants from each

of a set of functional annotations. The estimated parameters quantify the relative contribution of each annotation to the total heritability. For EnhA1, EnhA2, EnhG1, EnhG2, TssA, TssFlnk, TssFlnkD, TssFlnkU annotations in each brain tissue, we computed LD-scores of HapMap3 variants using 1000 Genomes Project SNPs as a reference panel. For each of these 56 annotations, we applied SLDSC to a model including one brain annotation and 24 annotations from the included "full baseline" model, and reported the z-score of the brain annotation coefficient.

2.4 Colocalization analyses of disease risk loci and eQTLs

To determine variants associated with both disease-risk and gene expression, we applied Bayesian colocalization tests to trans-ancestry BMI or European ancestry T2D fine-mapping data and brain expression quantitative trait loci (eQTL) data[41]. The tissue-specific SNP gene associations for 11 brain tissues (amygdala, anterior cingulate cortex (BA24), caudate, cerebellar hemisphere, cerebellum, cortex, frontal cortex (BA9), hippocampus, hypothalamus, nucleus accumbens, putamen, spinal cord cervical (c-1), and substantia nigra) for cis-eQTL analysis were collected from the GTEx project. The Bayesian method for colocalization between two traits determines whether the association signals of the traits are consistent with a common causal variant. The Bayes factors for all the variants for each trait were used to compute the posterior probability that there is an association with only trait one (PP1), association with only trait two (PP2), association between the two traits but at different variants (PP3), and association between the two traits with a shared causal variant (PP4). The threshold of PP4 > 0.50 was used to determine if a variant was associated with both disease-risk and gene expression. As this is the probability for the hypothesis of interest, PP4 will be referred to as posterior probability (PP) henceforth.

2.5 Deep learning of brain chromatin features

We used DeepSEA to train a convolutional neural network to learn chromatin feature effects from brain ChIP-seq data (6 histone marks over 8 brain tissues, 50 features in total)[43]. The published brain ChIP-seq data included H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me3, and H3K9me3 for angular gyrus, caudate nucleus, cingulate gyrus, fetal brain female, hippocampus middle, middle frontal lobe, substantia nigra, and temporal lobe[29]. We called peaks with MACS2, using the "--broad" flag for broad histone marks, H3K27me3 and H3K36me3; otherwise the default parameters were used for narrow peaks[54].

The input data was processed into a label matrix and a sequence matrix, as outlined by the authors of DeepSEA. The label matrix covered 1,018,562,200 bp of sequences, about 32% of the whole genome, and the sequence matrix used UCSC hg19 assembly as the reference genome[55]. The training, testing, and validation sets were randomly split 88%, 10%, and 2%, respectively. The testing set was excluded from training. All hyperparameters for the model were selected on the basis of log likelihood based on the validation set. To evaluate the performance of the trained model, we computed ROC curves for each chromatin feature with the testing set. The empirical cumulative distribution functions of predicted effects for 300,000 random variants from 1000 Genomes Project were used to calculate the E-value, the significance metric, for each variant for each chromatin feature. The E-value is the product of the relative and absolute differences between chromatin feature probability for the sequence carrying the reference allele and the alternative allele.

To predict the effects of BMI or T2D risk variants, we inputted *de novo* sequence information of the variant of interest to the trained DeepSEA model. The chromatin feature probability absolute difference, calculated as,

$$diff = p_{alt} - p_{ref}$$

where p_{alt} is the chromatin feature probability for the sequence carrying the alternative allele and p_{ref}, is the chromatin feature probability for the sequence carrying the reference allele, was used to determine if the predicted effect is at the reference or alternative allele. The threshold of E-value < 0.10 was used to determine if a variant has a significant DeepSEA effect.

2.6 Disrupted motifs analysis

To identify motifs that were potentially disrupted by risk variants, we compiled a comprehensive set of position frequency matrices (PFM) from JASPAR[56] and ENCODE[57] motif databases. We then used fimo[58] to scan for motifs in a 30 bp window centered on the risk variants. For fimo, we used the default parameters for p-value threshold ($1x10^{-4}$) and a background GC content of 40.9% based on hg19[55]. We further identified variants that were predicted to disrupt each motif. We calculated the entropy score for a variant position in the motif by using the PFMs. For each base at a given position bp and the frequency of the base at that position f, we calculated the entropy as:

$$Entropy = \sum_{bp} f(bp) \times \log_2 f(bp).$$

A motif was considered disrupted if a variant fell in a conserved position (Entropy<1.0).

2.7 Selection of T2D variants for functional study

A variant was selected for *in vitro* assays if (a) variant has probability to be causal for T2D (PPA with brain enrichment > 0.01), (b) variant is associated with gene expression in the brain (PP4 > 0.05), (c) variant is at a brain enhancer(s), (d) variant has DeepSEA allelic predictions for

enhancer activation in the brain, and (e) variant has *in silico* transcription factor binding predictions.

RESULTS

3.1. Enrichment analyses of BMI and T2D association in brain regulatory elements

We sought to determine if active regulatory elements in brain tissues were enriched for BMI and T2D signals. First, to determine the extent to which regulatory annotations differ across the 7 brain tissues in ChromHMM 18-state data, Jaccard similarity indexes were calculated for each brain regulatory element. Most elements differ across brain regions (Jaccard index < 0.50) except for active transcription start sites (TSS), which had Jaccard statistics between 0.50 and 0.60 (**Figure S1**).

We then determined the genome-wide enrichment of variants in brain regulatory annotations on T2D and BMI. We used fGWAS to integrate ChromHMM 18-state brain data with BMI and T2D genome-wide association data. For T2D, we also included pancreatic islet annotations in enrichment analyses as previous studies have shown many T2D loci affect islet function. There was enrichment of active TSSs across various areas of the brain for BMI genome-wide (**Figure S2**) and enrichment of both promoters and enhancers in multiple areas of the brain for T2D genome-wide (**Figure S3**). We further applied stratified LD-score regression on BMI and T2D genome-wide association data and found enrichment of functional annotations, specifically active and genic enhancers for BMI signals and genic enhancers for T2D signals (**Figure S4**).

We next determined the enrichment of variants in brain regulatory annotations using detailed fine-mapping data of known BMI and T2D signals. With both the fine-mapping BMI and T2D data, we observed enrichment of enhancer and promoter elements in the brain (**Figure 1A, 2A**). When considering annotations jointly in the BMI model, we observed enrichment of active enhancers in the dorsolateral prefrontal cortex and active TSSs in the cingulate gyrus (**Figure 1B**). For T2D, the joint annotations model showed enrichment of active enhancers in the dorsolateral

prefrontal cortex, angular gyrus, hippocampus middle, and anterior caudate, as well as active enhancers and TSSs in pancreatic islets (**Figure 2B**). We used these brain enrichments to re-weight the causal evidence of variants across 163 BMI and 107 T2D signals, and from this re-weighted evidence calculated posterior causal probabilities (PPA) for each variant.

We then identified specific T2D/BMI signals with likely causal variants in brain regulatory elements. We calculated the cumulative PPA of variants at each signal in elements for each brain regions. Across 163 BMI signals, 39 have at least 10% cumulative PPA (**Figure 3**). The PPA value for BMI of the leading variants of 16 loci improved >1% when integrated with functional brain annotation data (**Figure 4**). The PPA of rs4790841 near nonsense mediated MRNA decay factor (*SMG6*) increased 43.2% and the PPA of rs4867732 near cytoplasmic polyadenylation element binding protein 4 (*CPEB4*) increased 21.4% (**Figure 4**). Across 107 T2D signals, 69 loci have at least 10% cumulative PPA in the brain and/or pancreatic islets, including breast cancer anti-estrogen resistance protein 1 (*BCAR1*) and pleckstrin homology domain containing A1 (*PLEKHA1*) where variants in active enhancers have >50% cumulative PPA for 6 and 7 brain regions, respectively (**Figure 5, Table 1**). The PPA values for T2D of the top variant at 37 loci improved >1% with the brain and pancreatic islet enrichments (**Figure 7**). After excluding the loci with strong pancreatic enrichments (and thus likely to affect islet function), 41 loci have at least 10% cumulative PPA for T2D in the brain (**Figure 6**), where the PPA values for T2D of the leading variants for 32 loci increased >1% when integrated with only functional brain annotation data (**Figure 7**).

Among the set of T2D loci with likely causal variants in brain elements, we determined which were associated with BMI/obesity or other T2D-relevant endophenotypes. We collected p-values of the top variant at each of the 41 brain loci for T2D, BMI, fasting glucose level, and

fasting insulin level from the respective published GWAS data for these traits. Among these 41 T2D loci, only 6 were associated with BMI (ADCY5, CENTD2, NRXN3, APOE, JAZF1, KCNJ11) while 4 additional loci were associated with fasting glycemia (ADCY5, CENTD2, GCK, PROX1, KCNQ1, ZBED3) (**Figure 8)**.

3.2 Colocalization between BMI and T2D risk loci and various eQTLs in the brain

We next sought to determine if BMI and T2D signals with causal variants in brain regulatory elements were further associated with gene expression level in the brain. We applied Bayesian colocalization tests to determine colocalization between BMI and T2D fine-mapping data and gene expression data of 13 different brain tissues from the GTEx project[38]. The colocalization tests yielded 42 BMI and gene expression quantitative trait locus (eQTL) associations across 20 genes and 28 T2D and eQTL associations across 11 different genes, where the cutoff for posterior probability (PP) for shared association between the two traits was >50% (**Table 2, 3**).

The gene with the strongest colocalization between BMI and expression was RNA binding motif protein 6 (*RBM6*) at rs62262093 for 11 brain regions, where the posterior probabilities of colocalization for BMI and *RBM6* expression in the brain ranged from 64.4% to 98.9% (**Table 2**). Other associations between BMI and gene expression in the brain include adenylate cyclase 3 (*ADCY3*) in the putamen (PP = 83.5%) and frontal cortex (PP = 64.0%) colocalized at rs6749422, and regulator of MON1-CCZ1 (*RMC1/C18orf8*) in frontal cortex (PP = 94.7%), cerebellum (90.3%), amygdala (PP = 78.2%), and caudate (PP = 61.8%) colocalized at rs1367083.

The gene with the strongest colocalization between T2D risk and expression was hydroxysteroid 17-beta dehydrogenase 12 (*HSD17B12*), which was associated with gene

expression across all 13 brain regions. The posterior probabilities for T2D and *HSD17B12* expression in these regions ranged from 64.8% to 97.7% (**Table 3**). Other associations between T2D risk and gene expression include gastric inhibitory polypeptide receptor (*GIPR*) in cortex (PP = 97.0%) and cerebellum (PP = 76.1%) colocalized at rs2238689, protein regulatory of cytokinesis 1 (*PRC1*) in nucleus accumbens (PP = 83.0%) and in cortex (PP = 58.1%) colocalized at rs3803563, *BCAR1* in caudate (PP = 66.2%), anterior cingulate cortex (BA24) (PP = 61.4%), and putamen (PP = 56.5%) colocalized at rs8056814 (**Figure 9**), and *PLEKHA1* in hippocampus (PP = 54.2%) and cerebellar hemisphere (PP = 52.0%) colocalized at rs4311997 (**Figure 10**), as well as other variants near *PLEKHA1* (**Table 3**).

3.3 Deep learning of histone markers in the brain

We next sought to determine the allelic effects of BMI and T2D variants on brain regulatory element activity. To predict the effects of BMI and T2D risk variants on chromatin features, DeepSEA[43] was used to train a convoluted neural network to learn ChIP-seq data for histone modifications in the brain. The model was trained on data for 50 chromatin features (6 histone marks over 8 brain tissues). The area under the ROC curves on the testing set were between 0.605 and 0.995 (**Figure 11**). Of the 15,468 BMI fine-mapped variants, DeepSEA predicted allelic effects for an average of 480 variants for H3K27ac, 316 variants for H3K4me1, 441 variants for H3K4me3, and 411 variants for H3K9ac per brain region (**Figure 12**). In addition, of the 68,254 T2D fine-mapped variants, DeepSEA made predictions for an average of 3873 variants for H3K27ac, 3203 variants for H3K4me1, 3737 variants for H3K4me3, and 3520 variants for H3K9ac per brain region (**Figure 13**). rs8056814 near *BCAR1* and rs4311997 near *PLEKHA1* were predicted to have active transcriptional effect on the alternative allele (**Figure 14)**.

3.4 Prioritizing of candidate T2D risk variants for functional study

We prioritized specific T2D variants affecting brain regulation for further molecular validation. In addition to the functional annotations for enhancers from the ChromHMM model and eQTL mapping from the colocalization tests, predictions from DeepSEA helped prioritize T2D risk variants that may be directly affecting gene regulation in the brain (**Table 4**). From this prioritized list of T2D candidate variants, we further predicted whether they disrupted transcription factor binding motifs.

The variant rs8056814 near *BCAR1* overlaps active enhancers in angular gyrus, anterior caudate, cingulate gyrus, hippocampus middle, inferior temporal lobe, substantia nigra, and pancreatic islets, is associated with both T2D risk and gene expression of *BCAR1* in three brain regions, has DeepSEA prediction for alternative allele affecting transcriptional activation, and has disrupted motif predictions for multiple transcription factors including ARNT, CLOCK, FOXN1, HES5, DBD, HES7, HEY1, HEY2, MAX, SEF1, and ZBTB14.

rs2421016 near *PLEKHA1* overlaps active enhancers in all 7 brain regions and TSS in pancreatic islets and has disrupted motif predictions for EGR1, EGR2, EGR3, and EGR4. rs4311997 at the same locus is in an enhancer specific to only anterior caudate, and has a disrupted motif prediction for SP100. DeepSEA predicted the alternative allele as the effect allele for transcriptional activation for both variants. This locus is also associated with T2D risk and gene expression of *PLEKHA1* in hippocampus and cerebellar hemisphere.

rs113144510 near *HNF1A* also overlaps active enhancers in all 7 brain regions, is not at any pancreatic islet regulatory elements, and has disrupted motif prediction of NR5A1. *HNF1A* locus is associated with both T2D risk and gene expression of *HNF1A* in the cerebellum.

DISCUSSION

In this study, we tested whether genetic variants influencing BMI and T2D risk affect chromatin and gene regulation in the brain. Using a Bayesian model of T2D genetic data and brain and pancreatic islet epigenome data, we observed enrichment of active regulatory elements in many regions of the brain for BMI signals, as well as for T2D signals, independent of regulatory elements active in pancreatic islets. The combined models showed substantial enrichment of variants in enhancers that regulate expression in dorsolateral prefrontal cortex, anterior caudate, cingulate gyrus, and hippocampus, suggesting that these variants may play a role in BMI/T2D risk in these regions of the brain. Previous studies have found that BMI and T2D associated variants from the >100 known loci preferentially overlap regulatory elements in specific cell types such as pancreatic islets and adipocytes [24,59,60]. While it has been previously shown that BMI variants overlap brain regulatory elements, our results demonstrate that T2D variants are also enriched in brain regulatory elements and that many of these loci are also associated with fasting glycemia.

Next, we determined whether BMI and T2D associated variants affected gene expression levels in the brain. The colocalization tests yielded 42 BMI and gene expression associations across 20 different genes and 28 T2D risk and gene expression associations across 11 different genes. As expected given the established role of the brain in BMI, we observed more associations with BMI than T2D. The deep neural network model trained on histone modification sequence data in the brain predicted significant allelic effects for 3,567 T2D risk variants on average for each chromatin feature. These results suggest that our candidate variants for BMI/T2D risk are also associated with transcriptional activation and gene expression in the brain, and that altered brain activity of the genes affected by these variants might contribute to the mechanistic basis of T2D.

ADCY3 is a known locus associated with both obesity and T2D risk[19,46]. Our results show ADCY3 is also associated with gene expression in putamen and frontal cortex. rs6749422 has a PPA value for BMI of 0.84 and overlaps active enhancers at anterior caudate, cingulate gyrus, and dorsolateral prefrontal cortex. A recent study showed that loss-of-function mutations in *ADCY3* negatively affect cyclic AMP production, impairing G-protein-coupled receptor mechanisms in the leptin–melanocortin pathway in the hypothalamus, and thus promoting obesity[61].

One T2D locus of interest was *BCAR1*, where rs8056814 was colocalized for T2D risk and *BCAR1* expression in anterior cingulate cortex, caudate, and putamen. This variant also has a high posterior probability of being causal for T2D risk and overlaps several brain regulatory regions. From the deep neural network model trained on sequence data for histone modification markers, rs8056814 had prediction for enhancer activity in the alternative allele. *BCAR1* is a substrate for Src family kinase involved in many cellular events including proliferation and apoptosis, and overexpression of *BCAR1* contributes to anti-estrogen resistance on breast cancer cells[62]. A previous study has found that individuals carrying risk alleles near *BCAR1* had decreased insulin release, and suggested that the T2D risk may be through impairment of islet beta cell function[63]. Our results suggest that this may be a new, previously unknown role of *BCAR1* activity in the brain in the context of T2D.

Another locus with multiple candidate T2D risk variants was *PLEKHA1*, where the variants were also associated with gene expression in hippocampus and cerebellar hemisphere. The cumulative PPA values for variants at active enhancer elements in the *PLEKHA1* region were >50% in all 7 brain regions. The deep learning model predicted transcriptional activation allelic effects for 4 variants near *PLEKHA1*. The gene encodes for PLEKHA1 protein, which is localized to the plasma membrane, and is thus also an eQTL in whole blood[46]. Besides T2D, the *PLEKHA1*

gene is prominently associated with age-related macular degeneration, where the retina becomes damaged, leading to vision loss[64]. The loss or impairment of movement is usually caused by the dysfunction of oligodendrocytes and loss of myelin sheaths in the central nervous system[65]. A study demonstrated that *PLEKHA1* was expressed in differentiating oligodendrocyte progenitor cells, and knocking down *PLEKHA1* improved oligodendrocyte differentiation and production of myelin, implying that *PLEKHA1* is a down-regulator of oligodendrocytes[66].

To validate that our candidate BMI/T2D risk variants affect regulatory processes in the brain, molecular functional studies must be performed. Our *in silico* predictions can be tested with luciferase reporter assays to determine the allelic effects and electrophoretic mobility shift assays (EMSAs) to determine which proteins may be differentially binding to the effect allele. Since most of our candidate variants are predicted to have an effect across many brain regions, neural progenitor cells (NPCs) that express pan-neuronal genes[67] can be used for the initial studies to show the activity of BMI/T2D variants in neuronal specific cells. However, most of our computational analyses used adult T2D-relevant tissues, so the proper model should involve differentiated neuronal cells and other cell types.

As a next step, we plan to determine the specific cell-types that are affecting T2D risk in disease-relevant regions of the brain. For instance, recently single-nucleus ATAC-seq assays have identified 20 different cell types in the mouse forebrain[68]. Performing more single cell assays on human brain tissues of interest can provide insight on which cell populations to target for studying obesity and T2D. Moreover, other methods such as genome editing on cell and animal models are needed to further characterize the effects of variants for T2D risk. Studying the neural mechanisms involving the specific-cell types and transcription factors affecting BMI and T2D pathogenesis

and regulation, in additional to mechanisms involving adipocytes and pancreatic islets, will guide researchers towards identifying novel therapeutic and disease prevention strategies.

The material in this thesis is currently being prepared for submission for publication of the material. Vien, WT; Chiou, J; Aylward, AJ; Okino, M; Gaulton, KJ. “Diabetes risk variants affect chromatin and gene regulation in the human brain”. I was the primary researcher and author of this material.

FIGURES

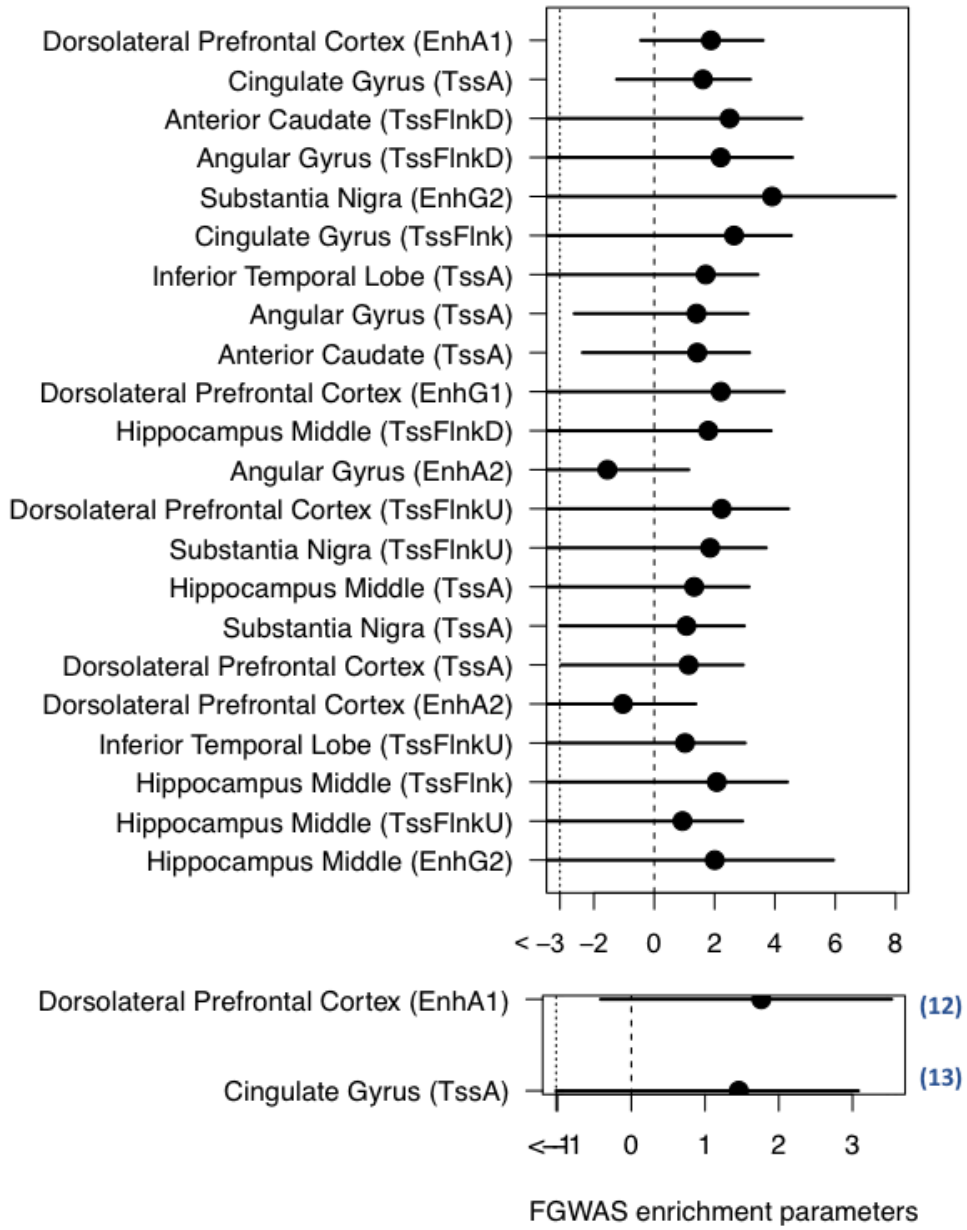


Figure 1. Enrichment of enhancers and promoters in different brain regions for BMI signals.

Trans-ancestry BMI fine-mapping data (99% credible set) and functional brain epigenome data were integrated with fGWAS. The BMI model is fit to each brain annotation individually (top) and the second model is built by combining multiple annotations (bottom). The plots show the maximum likelihood estimates and the 95% confidence intervals of the enrichment parameters for each annotation, in order of how much each annotation improved the likelihood of the model(s). In parentheses for annotations in the joint model is the number of annotations that improve the model from conditional analysis. (EnhA1, active enhancers 1; EnhA2, active enhancers 2; EnhG1, genic enhancers 1; EnhG2, genic enhancers 2; TssA, active transcription start site; TssFlnk, TssFlnkD, TssFlnkU, flanking transcription start site, downstream, upstream)

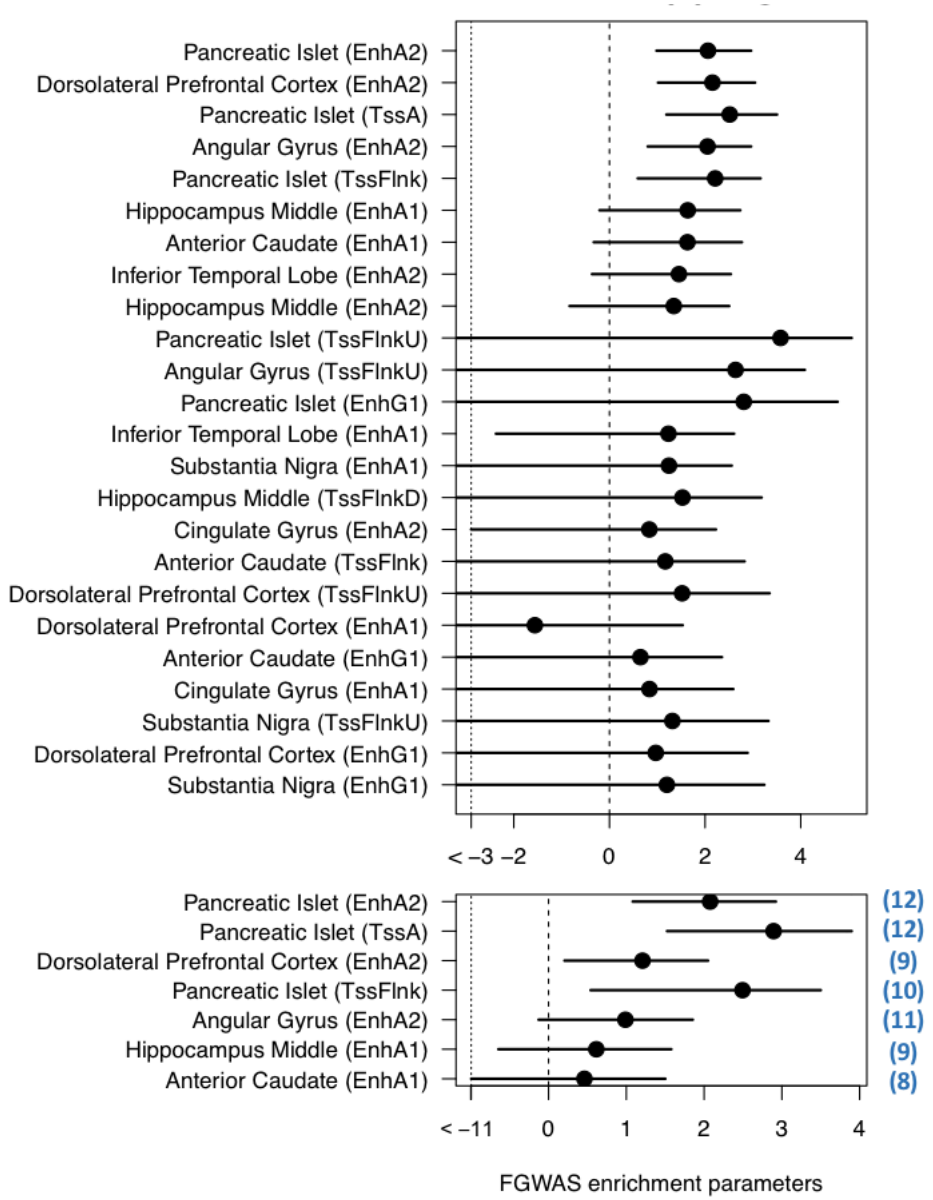


Figure 2. Enrichment of enhancers and promoters in different brain regions for T2D signals, relative to pancreatic islets.

T2D fine-mapping data and functional brain and pancreatic islet epigenome data were integrated with fGWAS. The T2D model is fit to each brain and islet annotation individually (top) and the second model is built by combining multiple annotations (bottom). The plots show the maximum likelihood estimates and the 95% confidence intervals of the enrichment parameters for each annotation, in order of how much each annotation improved the likelihood of the model(s). In parentheses for annotations in the joint model is the number of annotations that improve the model from conditional analysis. (EnhA1, active enhancers 1; EnhA2, active enhancers 2; EnhG1, genic enhancers 1; EnhG2, genic enhancers 2; TssA, active transcription start site; TssFlnk, TssFlnkD, TssFlnkU, flanking transcription start site, downstream, upstream)

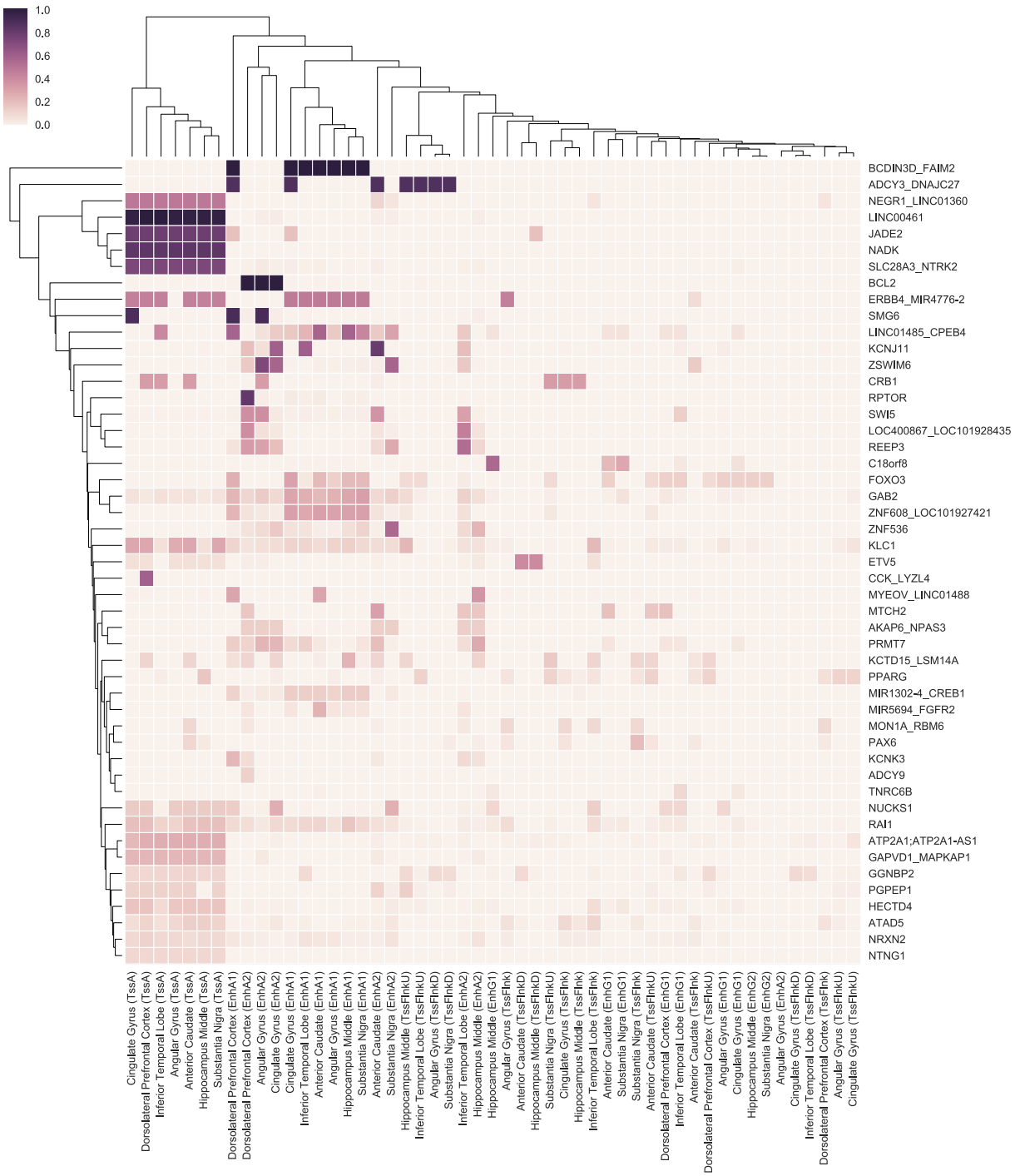


Figure 3. Posterior probabilities of association (PPA) for BMI were computed from the combined fGWAS brain model.

39 loci have >10% cumulative PPA for BMI with at least one brain annotation.

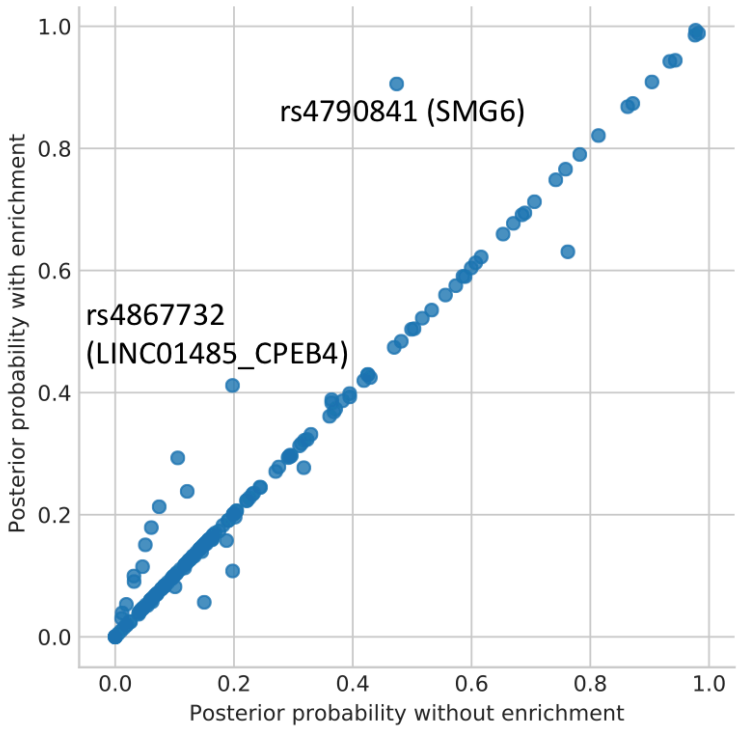


Figure 4. Comparison of PPA for BMI of the top variant of each locus with and without enrichment information from fGWAS analysis.
without enrichment (x-axis) and with brain enrichment (y-axis). The PPA value for BMI of the leading variants of 16 loci improved >1% when integrated with functional brain annotation data. The PPA of rs4790841 near nonsense mediated MRNA decay factor (*SMG6*) increased 43.2% and the PPA of rs4867732 near cytoplasmic polyadenylation element binding protein 4 (*CPEB4*) increased 21.4%.

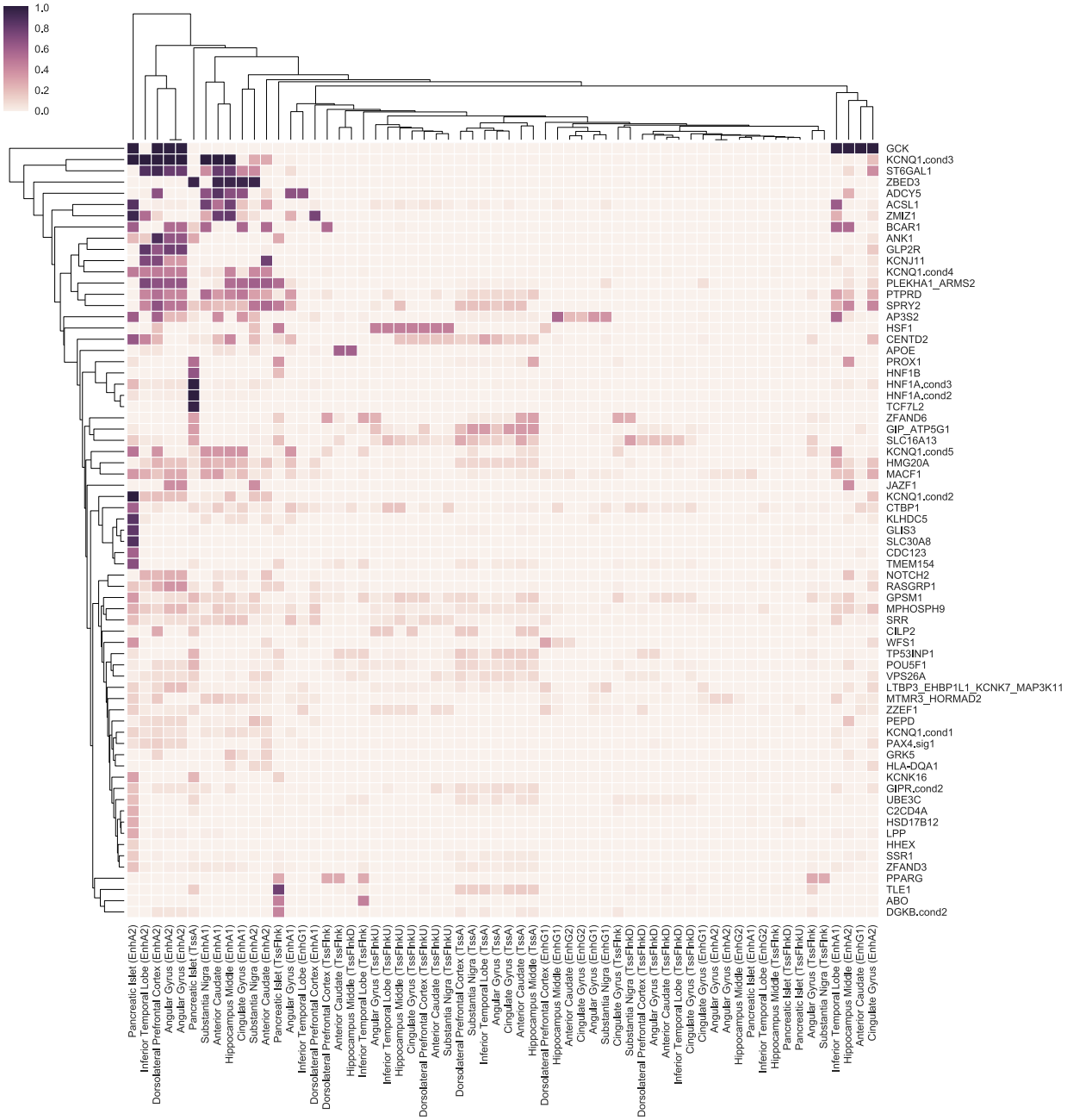


Figure 5. Posterior probabilities of association (PPA) for T2D were computed from the combined fGWAS brain and pancreatic islet model.

69 loci have >10% accumulated PPA for T2D with at least one brain annotation.

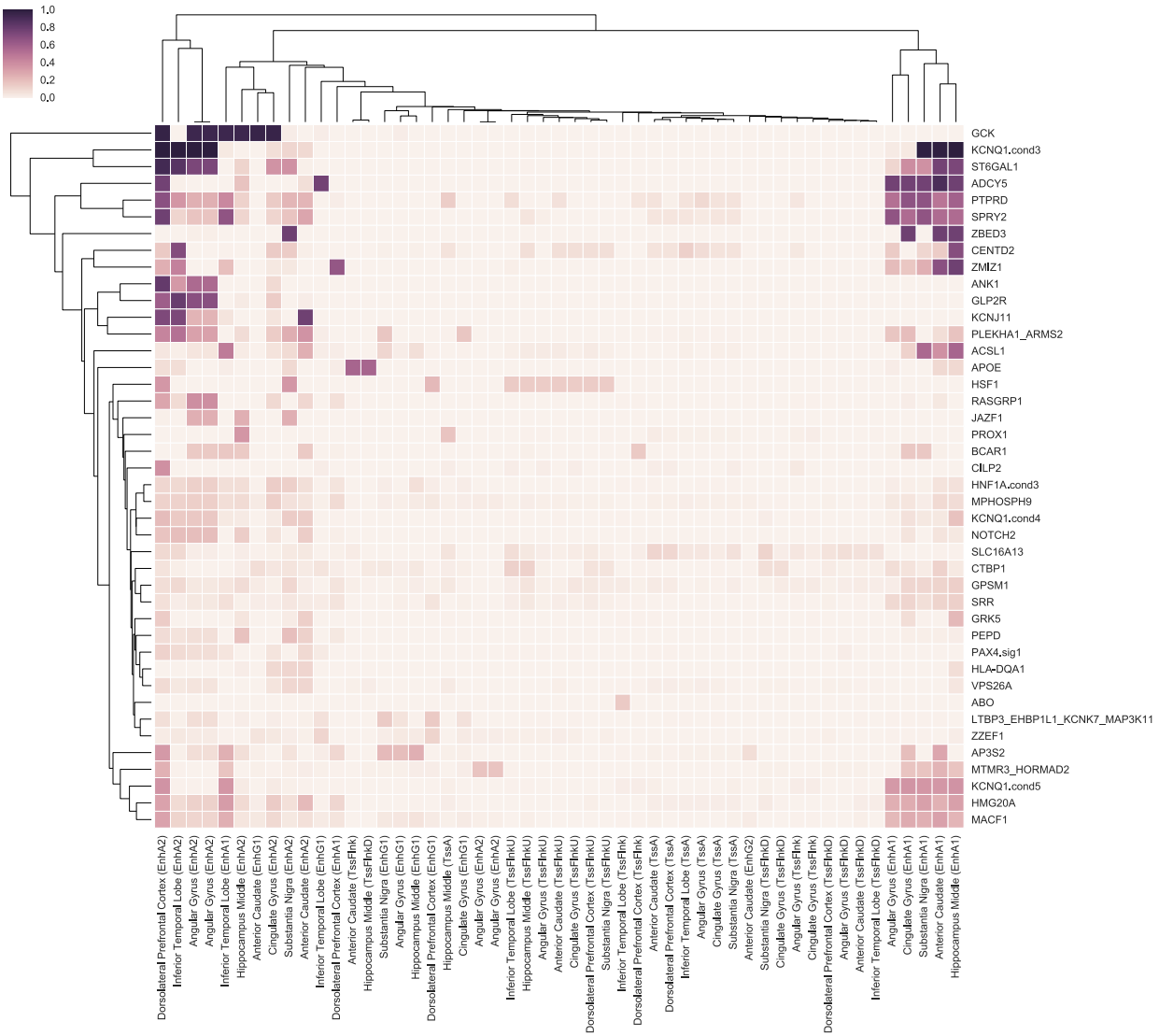


Figure 6. Posterior probabilities of association (PPA) for T2D were computed from the combined fGWAS brain model.

41 loci have >10% accumulated PPA for T2D with at least one brain annotation.

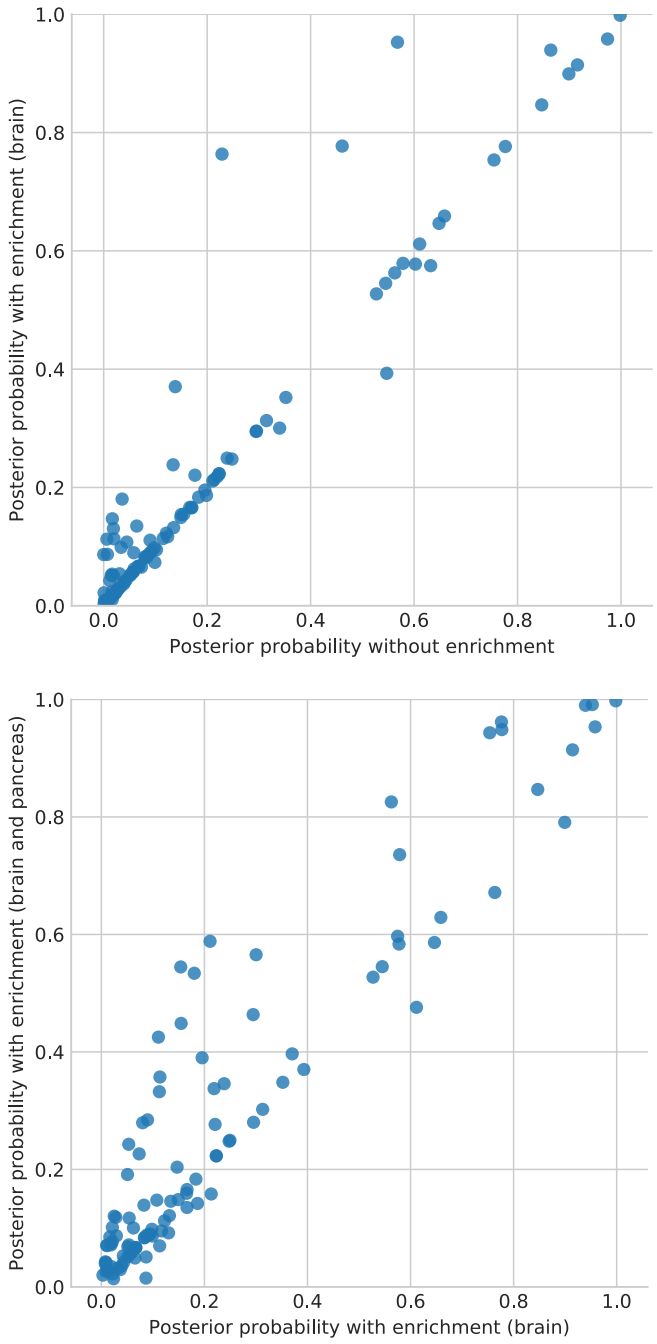


Figure 7. Comparison of PPA for T2D of the top variant of each locus with and without enrichment information from fGWAS analysis.
Top: without enrichment (x-axis) and with brain enrichment (y-axis); bottom: with brain (x-axis) and with addition of pancreatic islet (y-axis) enrichment. The PPA value for T2D of the leading variants for 32 loci increased >1% when integrated with functional brain annotation data, while 20 leading variants increased >1% when integrated with both functional brain and pancreatic islet annotation data.

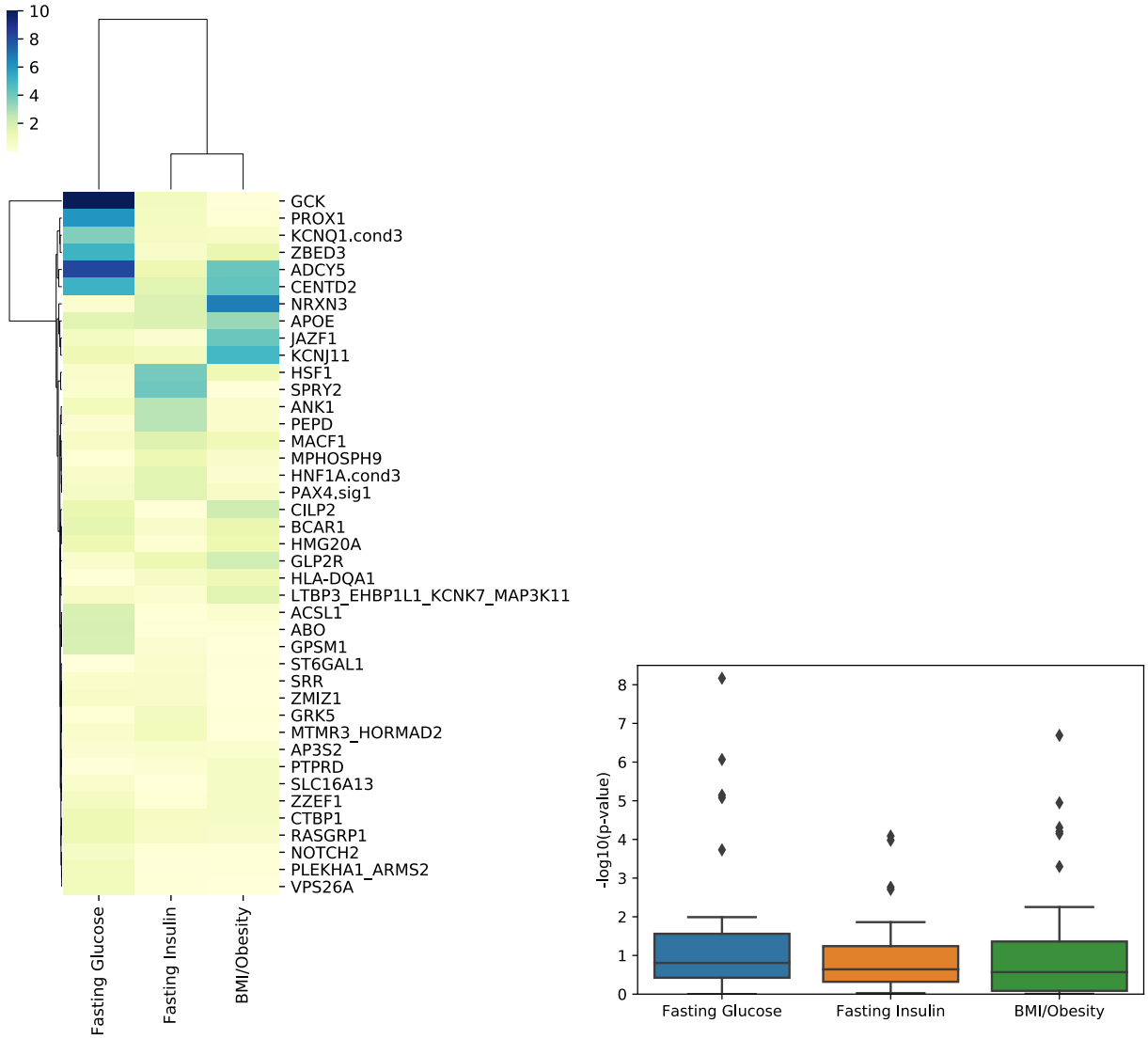


Figure 8. Association of T2D risk loci with variants in brain regulatory elements with fasting glucose, fasting insulin, and BMI.
30 of the 41 loci with PPA >10% accumulated for T2D with at least one brain annotation. The top variant at each of the 30 loci has -log10(p-value) < 2 for fasting glucose, fasting insulin, and BMI. Subsets of T2D loci are associated with BMI (e.g. *NRXN3*) while others are associated with glucose levels (e.g. *GCK*) and not BMI.

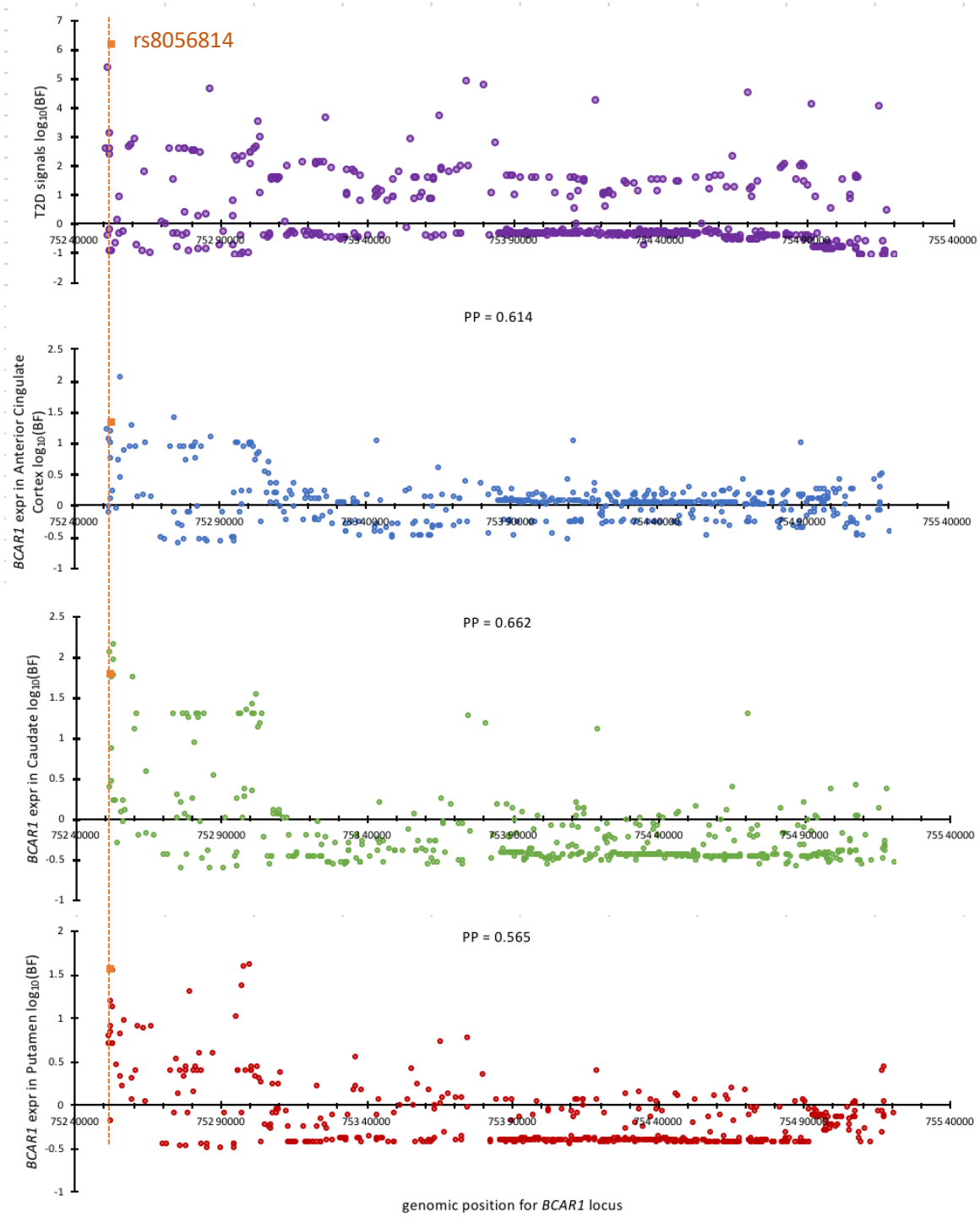


Figure 9. Colocalization between T2D and *BCAR1* gene expression in the human brain at shared variant rs8056814.

The plot shows the association signals (y-axis; BF, Bayes factor) for T2D risk (purple) and expression in anterior cingulate cortex (blue; BA2, Brodmann area 24), caudate (green), and putamen (red) of the brain for each variant at the *BCAR1* locus (x-axis). The shared variant between T2D and gene expression in those three areas of the brain is rs8056814 (orange) at genomic position 75252327 on chromosome 16 near *BCAR1*.

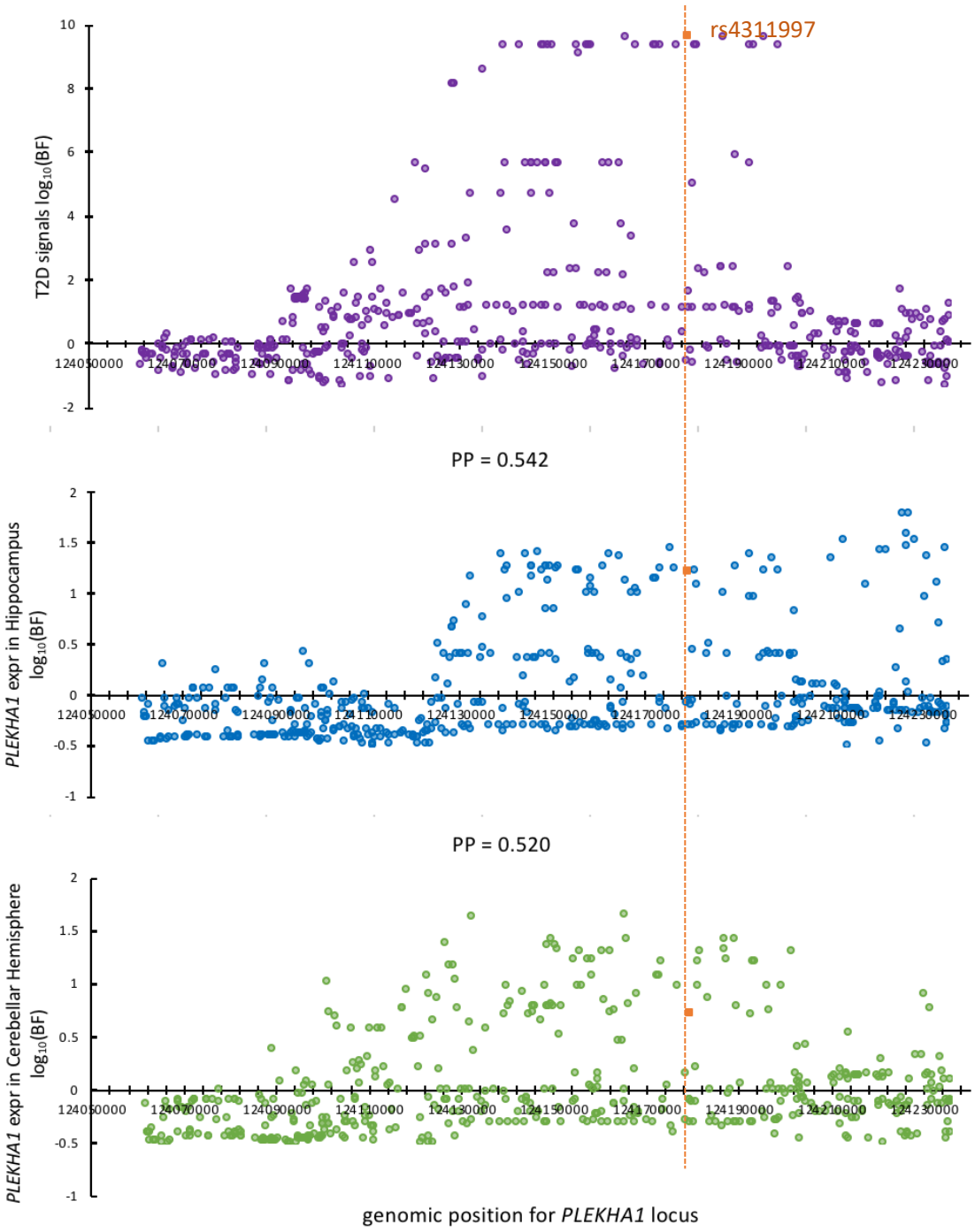


Figure 10. Colocalization between T2D and *PLEKHA1* gene expression in the human brain at shared variant rs4311997.

The plot shows the association signals (y-axis; BF, Bayes factor) for T2D risk (purple) and expression in hippocampus (blue), and cerebellar hemisphere (green)of the brain for each variant at the *PLEKHA1* locus (x-axis). The shared variant between T2D and gene expression in those three areas of the brain is rs4311997 (orange) at genomic position 124179299 on chromosome 10 near *PLEKHA1*.

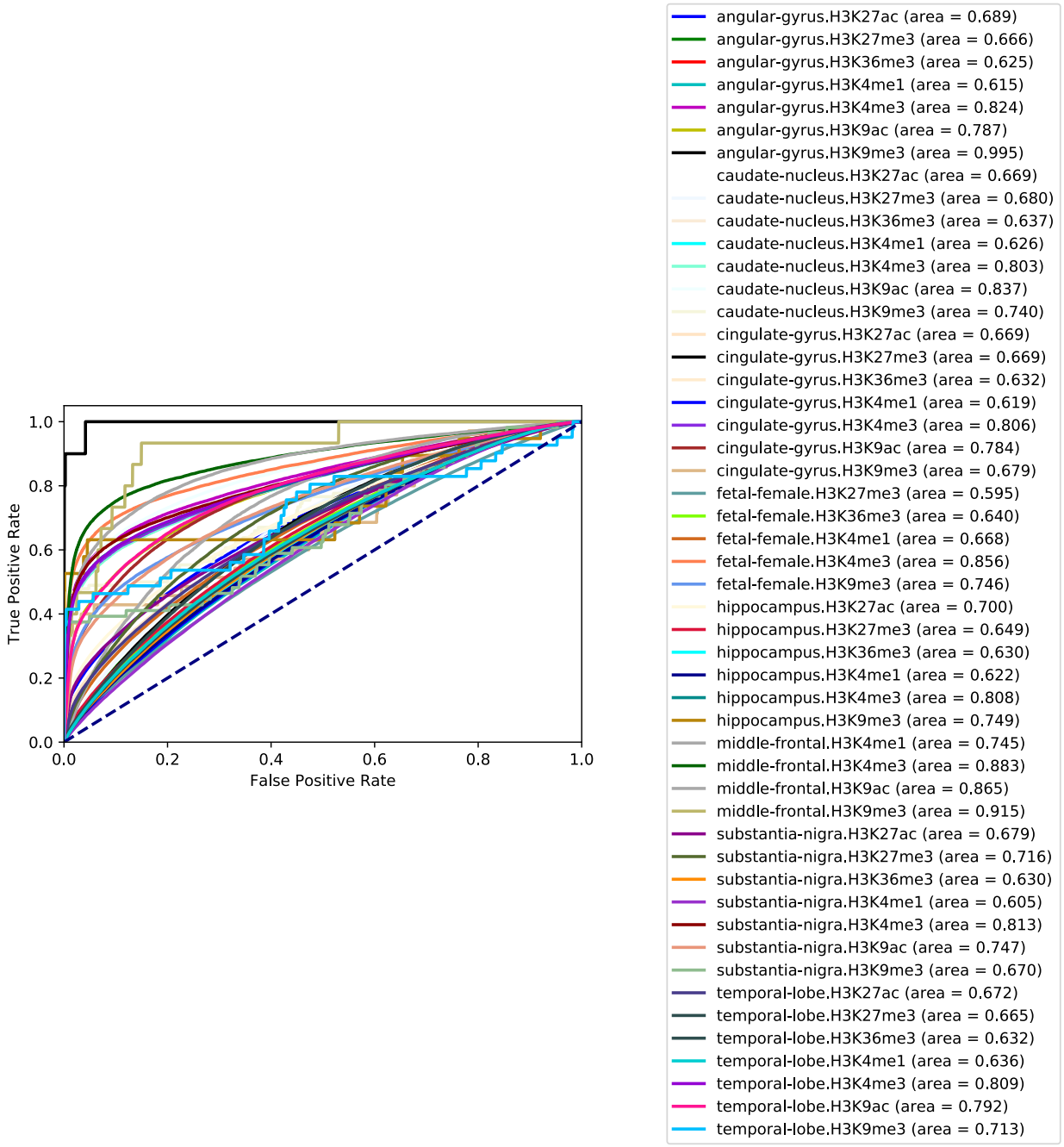


Figure 11. DeepSEA performance on the test dataset left out of training was between 60.5% and 99.5% accuracy for each chromatin feature.
The convoluted neural network was trained on data from 50 ChIP-seq assays correlating to both active and repressive histone markers across 8 different brain tissues.

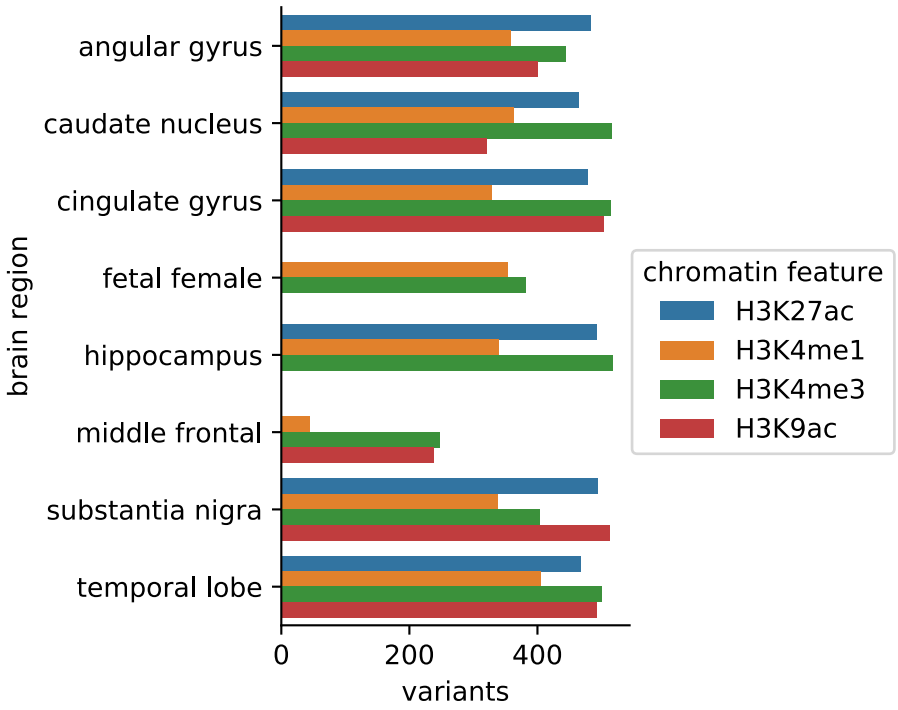


Figure 12. BMI variants with predicted effects (chromatin probability difference >|0.001| and E-value >0.10.
Of 15,468 BMI variants, DeepSEA made predictions for an average of 480 variants for H3K27ac, 316 variants for H3K4me1, 441 variants for H3K4me3, and 411 variants for H3K9ac per brain region.

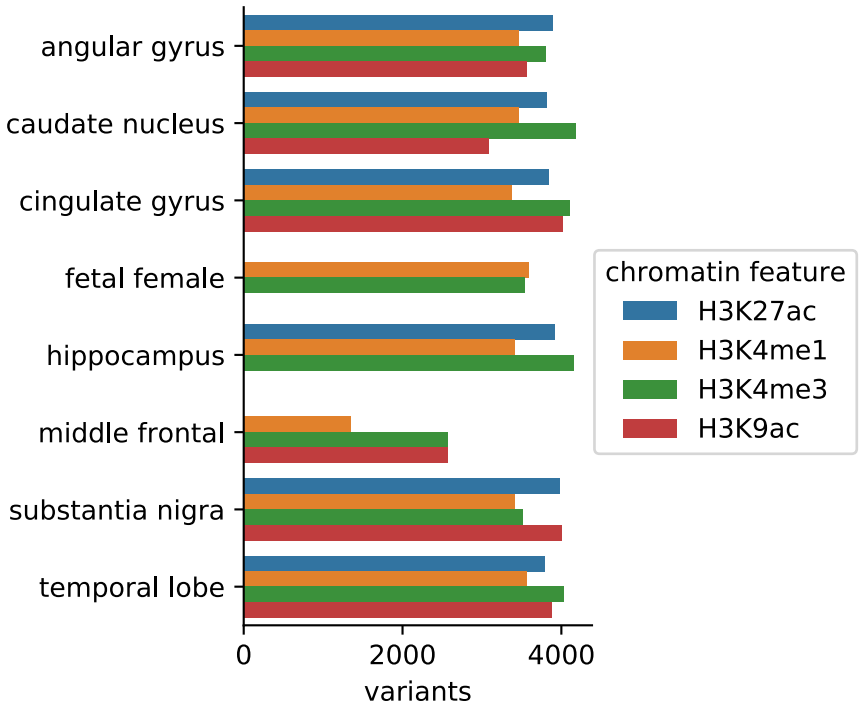


Figure 13. T2D variants with predicted effects (chromatin probability difference >|0.001| and E-value >0.10.
Of 68,254 T2D variants, DeepSEA made predictions for an average of 3873 variants for H3K27ac, 3203 variants for H3K4me1, 3737 variants for H3K4me3, and 3520 variants for H3K9ac per brain region

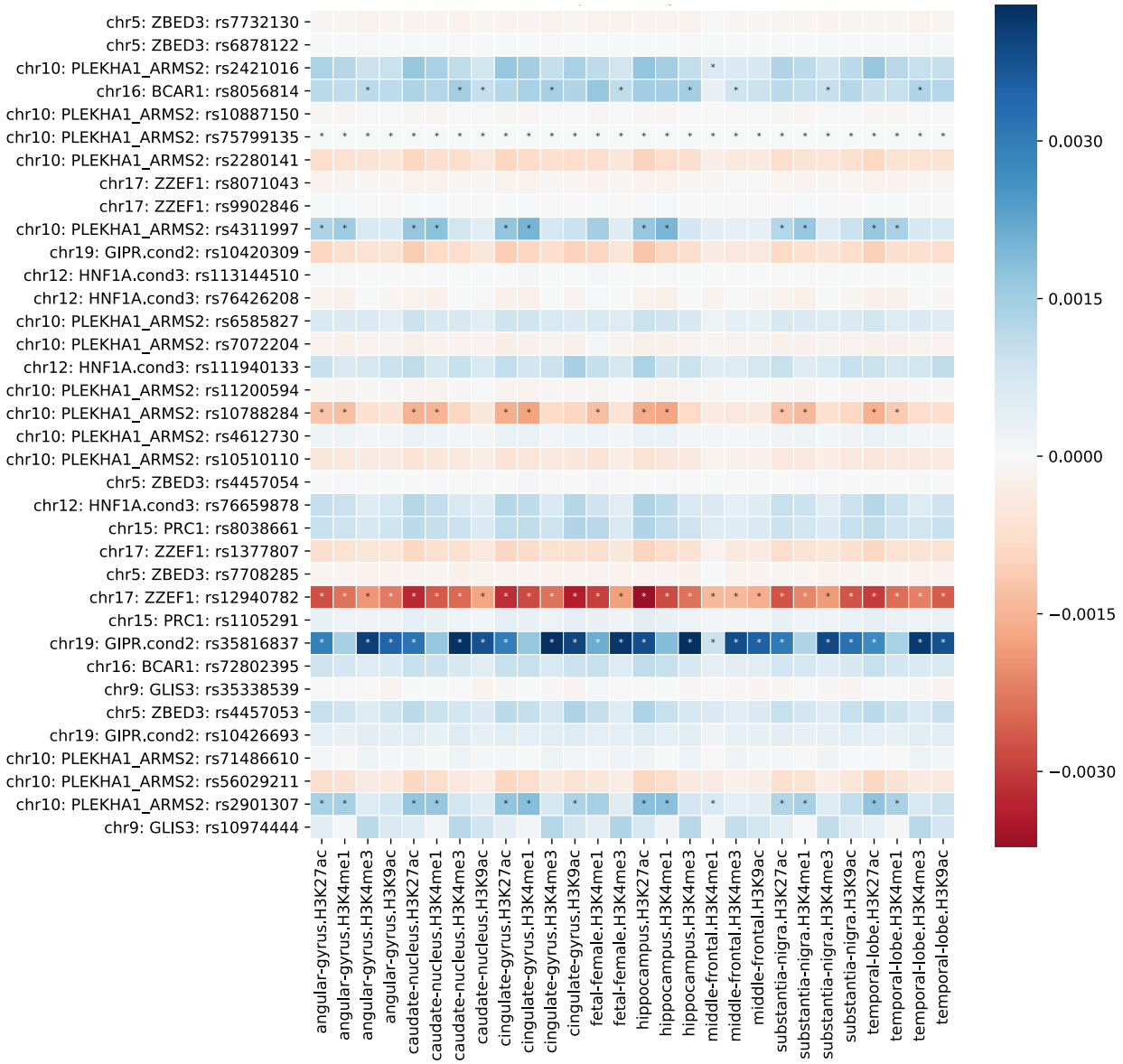


Figure 14. DeepSEA predicts the allelic effects of T2D risk variants on histone modifications in the brain.

The plot includes variants with posterior probability > 0.01 to be causal for T2D (x-axis; chromosome: locus: reference SNP identification number) and shows the predicted histone markers (y-axis) effects for each variant. The red color corresponds to higher probability for the variant carrying the alternative allele for that histone marker, and the blue color corresponds to higher probability for the variant carrying the reference allele. The starred values are considered significant values based on the empirical distribution of running the trained model on 100,000 random variants.

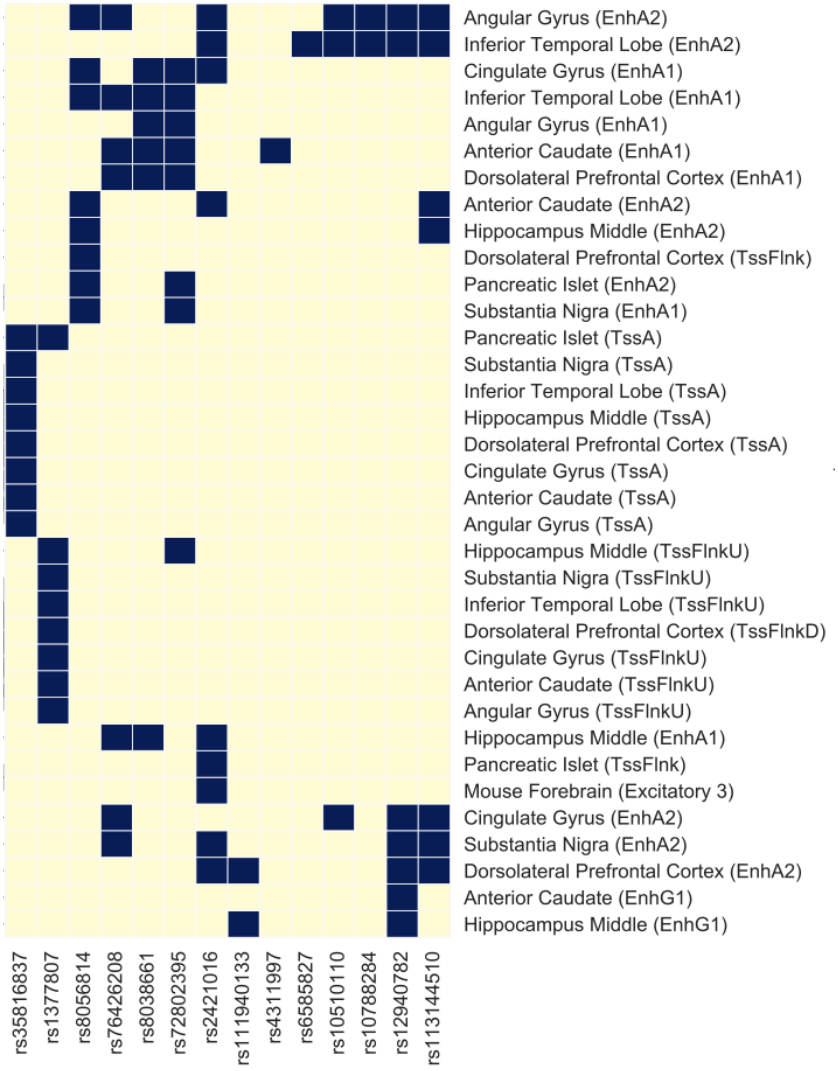


Figure 15. Candidate T2D risk variants and their brain annotations.
The variants shown here were prioritized with (a) enrichment analyses on functional annotations, (b) T2D risk and brain gene expression colocalization tests, and (c) predictions of brain histone markers allelic effects. Several variants, like rs8056814 and rs72802395 near *BCAR1*, have annotations for enhancers and transcription start sites across various areas of the brain, while other variants, like rs80019595 near *HNF1A* which only overlaps enhancers in the hippocampus, are specific to one brain annotation. The functional annotations are based off of chromatin immunoprecipitation of histone markers in human brain tissues and single-nucleus analysis of accessible chromatin in mouse forebrain. (EnhA1, active enhancers 1; EnhA2, active enhancers 2; EnhG1, genic enhancers 1; EnhG2, genic enhancers 2; TssA, active transcription start site; TssFlnk, TssFlnkD, TssFlnkU, flanking transcription start site, downstream, upstream)

TABLES

Table 1. Cumulative PPA for T2D of BCAR1 and PLEKHA1 loci that are enriched for active enhancers in different brain regions.

Locus	Brain region (regulatory element)	Cumulative PPA for T2D
BCAR1	Angular gyrus (EnhA2)	0.5446654794
	Anterior caudate (EnhA2)	0.5447744999
	Caudate nucleus (EnhA1)	0.5599113158
	Dorsolateral prefrontal cortex (TssFlnk)	0.5446470637
	Hippocampus (EnhA2)	0.5446475444
	Inferior temporal lobe (EnhA1)	0.5598202537
	Pancreatic islet (EnhA2)	0.602738449
	Substantia nigra (EnhA1)	0.5592605506
PLEKHA1	Angular gyrus (EnhA2)	0.6456088851
	Anterior caudate (EnhA2)	0.6146043423
	Caudate nucleus (EnhA1)	0.5480257327
	Dorsolateral prefrontal cortex (EnhA2)	0.6680523736
	Hippocampus middle (EnhA1)	0.5363408342
	Inferior temporal lobe (EnhA2)	0.7331566202
	Substantia nigra (EnhA2)	0.5990532849

Table 2. Association between BMI and expression quantitative trait loci (eQTLs) across 13 brain tissues and 20 different genes.
Bayesian colocalization tests were performed on BMI genetic data and RNA-seq data for 13 different brain tissues. There are 41 associations between BMI and gene expression in the brain that are colocalized a shared variant. The cutoff was 50% posterior probability for association of the two traits.

Gene	Brain tissue(s)	Highest posterior probability
RBM6	Frontal cortex, putamen, spinal cord (c-1), caudate, substantia nigra, hypothalamus, cerebellar hemisphere, hippocampus, nucleus accumbens (9)	0.98868
MTIF3	Nucleus accumbens, putamen (2)	0.98391
KCNK3	Frontal cortex, cortex (2)	0.96769
GGNBP2	Caudate, nucleus accumbens, hippocampus, cortex, frontal cortex, amygdala (6)	0.96012
C18orf8	Frontal cortex (1)	0.94674
GAPVD1	Cerebellum, cerebellar hemisphere (2)	0.88401
TRIM66	Caudate, putamen, amygdala, substantia nigra (4)	0.85332
POC5	Substantia nigra (1)	0.85282
ADCY3	Putamen, frontal cortex (2)	0.83509
ZBTB10	Frontal cortex, cortex (2)	0.77357
STK39	Putamen, nucleus accumbens (2)	0.72521
ETV5	Substantia nigra (1)	0.71369
MTCH2	Cerebellar hemisphere (1)	0.63668
RIT1	Caudate (1)	0.63296
SKAP1	Hypothalamus (1)	0.60337
NUCKS1	Frontal cortex (1)	0.58962
MAP2K5	Caudate (1)	0.56547
LMOD1	Substantia nigra (1)	0.55274
STK39	Nucleus accumbens (1)	0.548
LSM14A	Anterior cingulate cortex (1)	0.50078

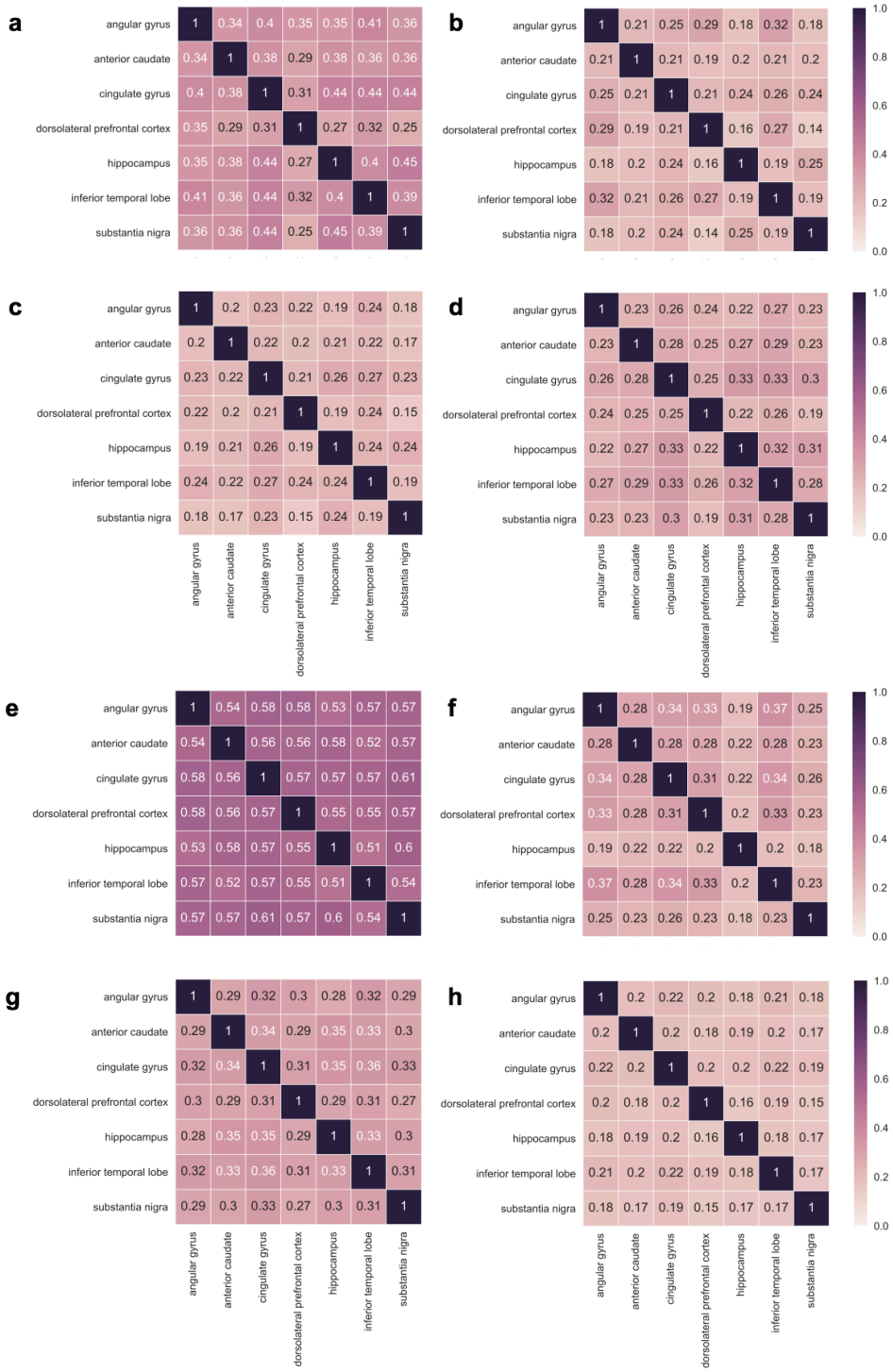
Table 3. Association between T2D and expression quantitative trait loci (eQTLs) across 13 brain tissues and 11 different genes.
Bayesian colocalization tests were performed on T2D genetic data and RNA-seq data for 13 different brain tissues. There are 28 associations between T2D and gene expression in the brain that are colocalized a shared variant. *HSD17B12* gene expression in all 13 brain tissues have a shared variant with T2D risk. The cutoff was 50% posterior probability for association of the two traits.

Gene	Brain tissue(s)	Highest posterior probability
HSD17B12	Anterior cingulate cortex, hypothalamus, cortex, amygdala, cerebellum, caudate, substania nigra, spinal cord (c-1), putamen, frontal cortex, nucleus accumbens, cerebellar hemisphere, hippocampus (13)	0.97708
GIPR	Cortex, cerebellum (2)	0.96965
PRC1	Nucleus accumbens, cortex(2)	0.82978
ZBED3	Spinal cord cervical (1)	0.70643
HNF1A	Cerebellum (1)	0.70005
BCAR1	Caudate, anterior cingulate cortex, putamen (3)	0.66189
CDKAL1	Frontal cortex (1)	0.63598
ZZEF1	Cortex (1)	0.60827
GLIS3	Frontal cortex (1)	0.56004
PLEKHA1	Hippocampus, cerebellar hemisphere (2)	0.54185
CENPW	Putamen (1)	0.5049

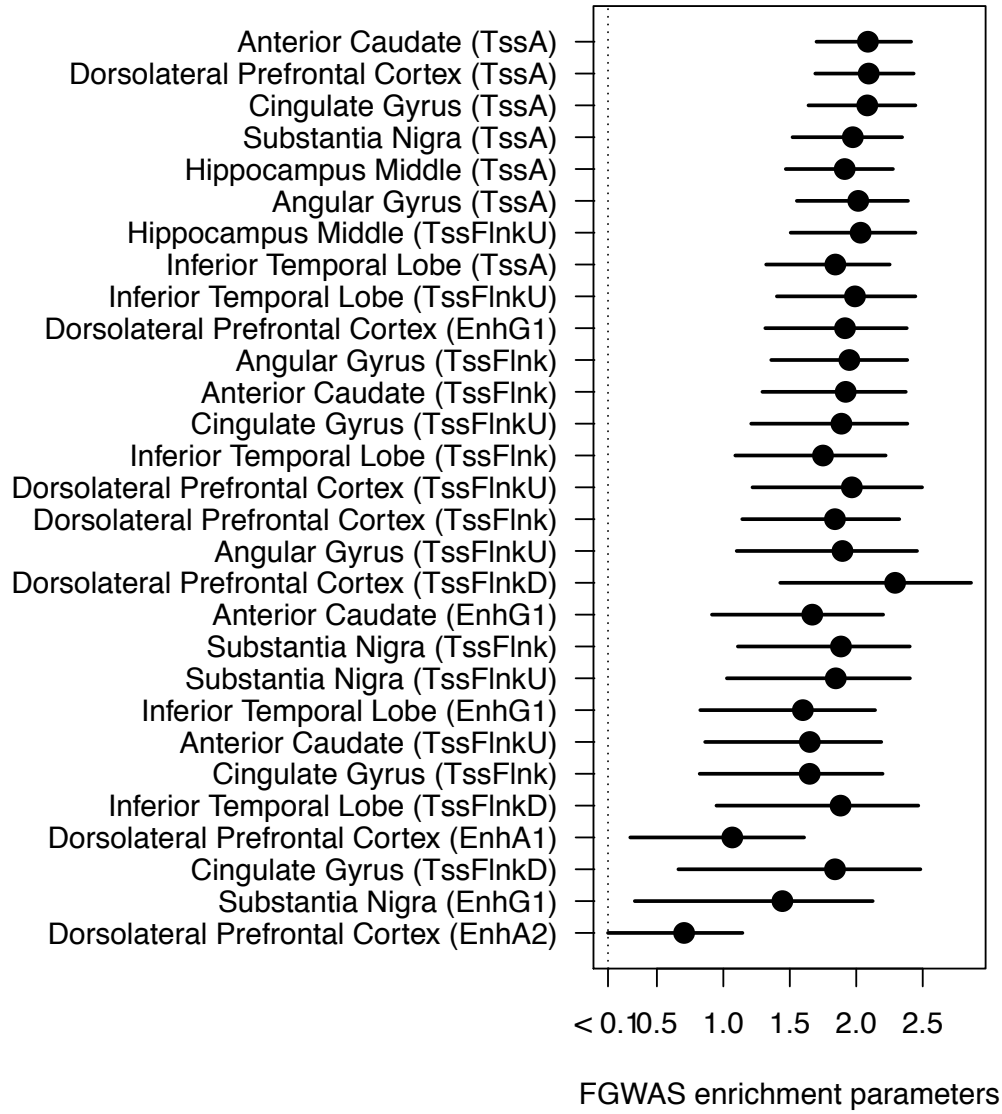
Table 4. Top candidate T2D risk variants in brain regulatory elements.
Table shows summary of top candidate variants sorted by posterior probability of association (PPA) for T2D risk. The variants shown here were prioritized with (a) enrichment analyses on functional annotations, (b) T2D risk and brain gene expression colocalization tests, and (c) predictions of brain histone markers allelic effects. The PPA for each T2D with no enrichment are from GoT2D and DIAGRAM consortia and the PPAs with enrichment were computed from the Bayesian statistical enrichment analysis (fGWAS) of combining brain and pancreatic islet annotations for different chromatin states. (chr, chromosome; rsid, reference SNP identification number, ln(BF), natural log Bayes Factor).

CHR	POSITION	RSID	LOCUS	LN(BF)	PPA (NO ENRICHMENT)	PPA (BRAIN AND ISLETS)	PPA (BRAIN)	DISRUPTED MOTIF PREDICTIONS
10	**124167512**	**rs2421016**	**PLEKHA1**	**21.597**	**0.036**	**0.534**	**0.180**	**EGR1, EGR2, EGR3, EGR4**
16	**75252327**	**rs8056814**	**BCAR1**	**14.306**	**0.150**	**0.545**	**0.154**	**ARNT, CLOCK, FOXN1, HES5, DBD, HES7, HEY1, HEY2, MAX, SEF1, ZBTB14**
10	**124179299**	**rs4311997**	**PLEKHA1**	**22.182**	**0.064**	**0.036**	**0.075**	**SP100**
12	**121261196**	**rs113144510**	**HNF1A**	**-0.203**	**6.39E-07**	**0.020**	**0.052**	**NR5A1**
12	121273991	rs76426208	HNF1A	-0.081	7.21E-07	0.018	0.048	
10	124165615	rs6585827	PLEKHA1	22.182	0.064	0.021	0.039	
12	121197628	rs111940133	HNF1A	-0.142	6.79E-07	0.007	0.033	
10	124149917	rs10788284	PLEKHA1	21.597	0.036	0.022	0.022	
10	124192430	rs10510110	PLEKHA1	21.597	0.036	0.022	0.022	
15	91502849	rs8038661	PRC1	14.209	0.004	0.013	0.016	
17	4045440	rs1377807	ZZEF1	8.396	0.014	0.081	0.014	AP3
17	3992712	rs12940782	ZZEF1	6.711	0.003	0.012	0.010	FOXA, HMX3
19	46148903	rs35816837	GIPR	4.427	0.009	0.046	0.008	DNAJC21, EGR1, GLI2, GLIS1, GLIS2, GLIS3, KLF11, KLF5, MZF1, OVOL2, PRDM4, SP1, SP2, SP4, SPZ1, VDR, ZBTB7A, ZIC3, ZNF219, ZNF281, ZNF740, Zfp740
16	75286484	rs72802395	BCAR1	10.710	0.004	0.014	0.008	

SUPPLEMENTAL FIGURES

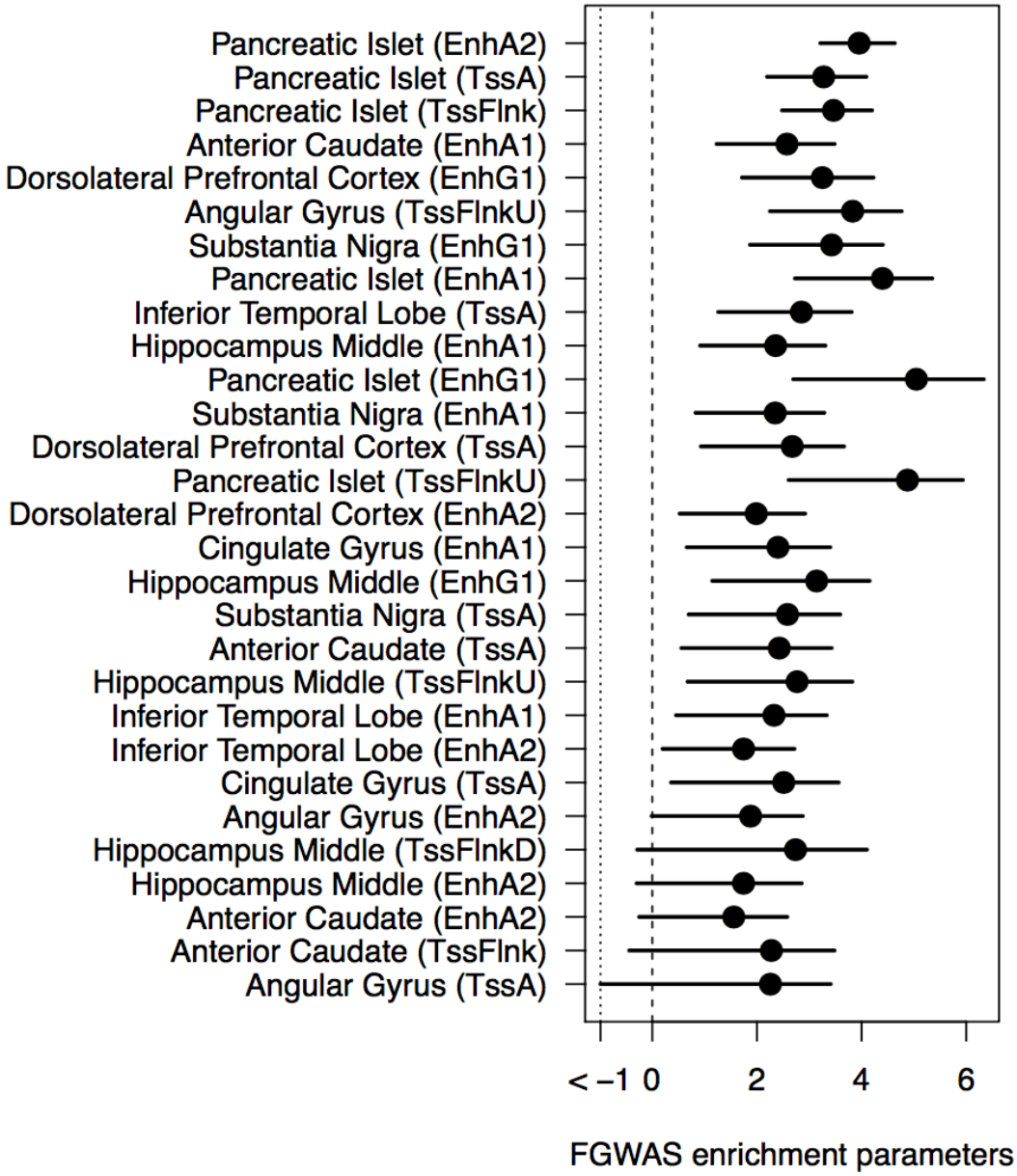


Supplemental Figure 1. Jaccard similarity analyses show that most enhancers and promoters across brain tissues are different (Jaccard index < 0.50) except active TSS
(**a**) Active enhancer 1 (**b**) Active enhancer 2 (**c**) Genic enhancer 1 (**d**) Genic enhancer 2 (**e**) Active transcription start site (TSS) (**f**) Flanking TSS (**g**) Flanking TSS upstream (**h**) Flanking TSS downstream

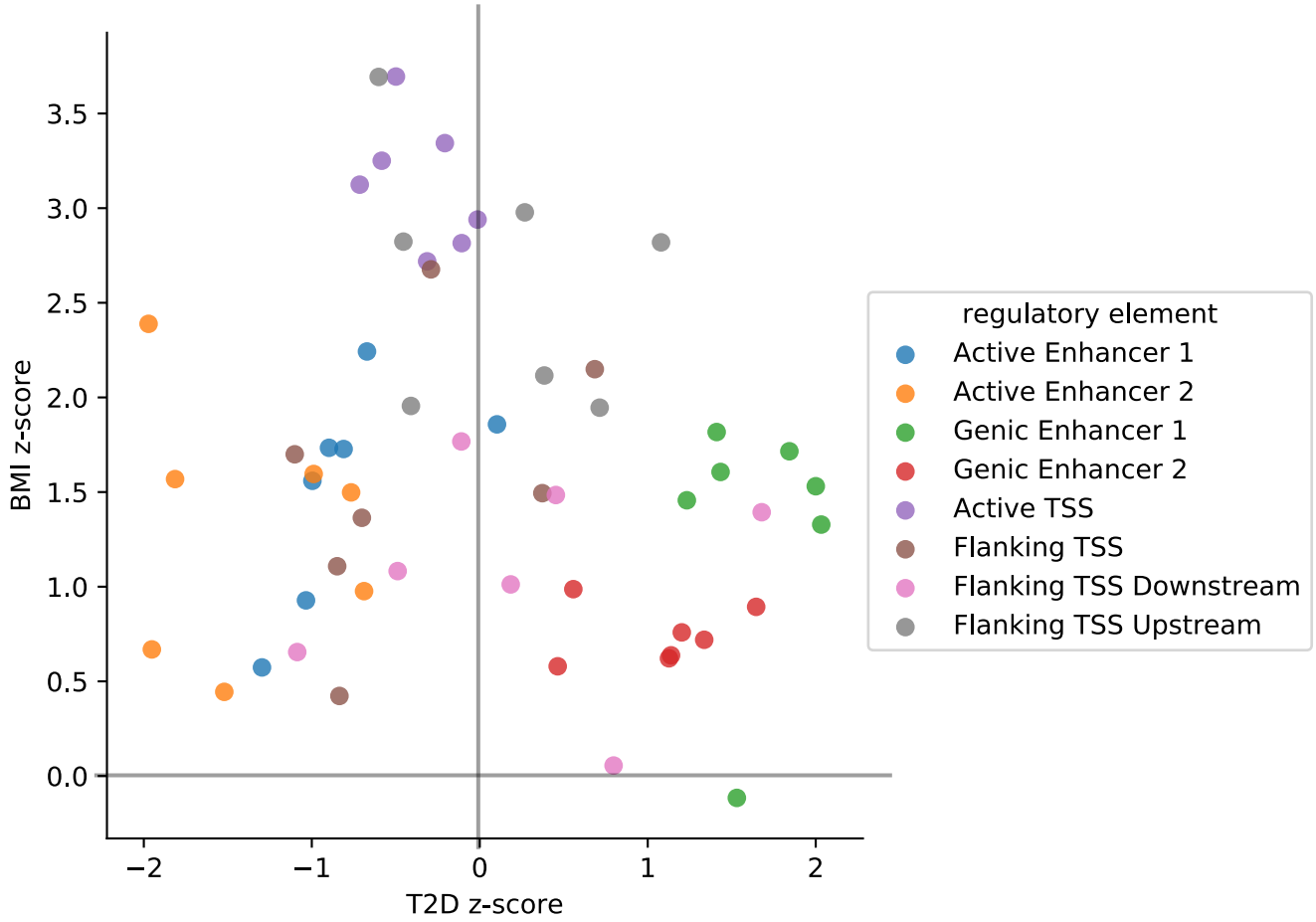


Supplemental Figure 2. Enrichment of promoters in various areas of the brain for BMI signals genome-wide.

Genome-wide BMI data and functional brain and pancreatic islet epigenome data were integrated with fGWAS. The T2D risk variants model is fit to each brain annotation individually. (EnhA1, active enhancers 1; EnhA2, active enhancers 2; EnhG1, genic enhancers 1; EnhG2, genic enhancers 2; TssA, active transcription start site; TssFlnk, TssFlnkD, TssFlnkU, flanking transcription start site, downstream, upstream)



Supplemental Figure 3. Enrichment of enhancers and promoters in various areas of the brain for T2D signals genome-wide, relative to pancreatic islets.
Genome-wide T2D data and functional brain and pancreatic islet epigenome data were integrated with fGWAS. The T2D risk variants model is fit to each brain and islet annotation individually. (EnhA1, active enhancers 1; EnhA2, active enhancers 2; EnhG1, genic enhancers 1; EnhG2, genic enhancers 2; TssA, active transcription start site; TssFlnk, TssFlnkD, TssFlnkU, flanking transcription start site, downstream, upstream)



Supplemental Figure 4. Stratified LD-score regression identifies functional annotations enriched for T2D signals genome-wide.
Z-scores (y-axis) for the LD-score regression coefficients for each annotation for each brain region (x-axis) were computed. (EnhA1, active enhancers 1; EnhA2, active enhancers 2; EnhG1, genic enhancers 1; EnhG2, genic enhancers 2; TssA, active transcription start site; TssFlnk, TssFlnkD, TssFlnkU, flanking transcription start site, downstream, upstream, respectively)

REFERENCES

1. Cynthia L. Ogden, Margaret D. Carroll, Cheryl D. Fryar & Katherine M. Flegal. Prevalence of obesity among adults and youth: United States, 2015–2016. *NCHS Data Brief* 1–8 (2015).

2. Karaderi, T., Drong, A. W. & Lindgren, C. M. Insights into the Genetic Susceptibility to Type 2 Diabetes from Genome-Wide Association Studies of Obesity-Related Traits. *Current Diabetes Reports* **15,** (2015).

3. National Diabetes Statistics Report, 2017. 20

4. McCarthy, M. & Menzel, S. The genetics of type 2 diabetes. 5 (2001).

5. Menke, A., Casagrande, S., Geiss, L. & Cowie, C. C. Prevalence of and Trends in Diabetes Among Adults in the United States, 1988-2012. *JAMA* **314,** 1021 (2015).

6. Ruud, J., Steculorum, S. M. & Brüning, J. C. Neuronal control of peripheral insulin sensitivity and glucose metabolism. *Nature Communications* **8,** 15259 (2017).

7. Schwartz, M. W., Seeley, R. J., Tschöp, M. H., Woods, S. C., Morton, G. J., Myers, M. G. & D'Alessio, D. Cooperation between brain and islet in glucose homeostasis and diabetes. *Nature* **503,** 59–66 (2013).

8. Roh, E., Song, D. K. & Kim, M.-S. Emerging role of the brain in the homeostatic regulation of energy and glucose metabolism. *Experimental & Molecular Medicine* **48,** e216 (2016).

9. Deltour, L., Leduque, P., Blume, N., Madsen, O., Dubois, P., Jami, J. & Bucchini, D. Differential expression of the two nonallelic proinsulin genes in the developing mouse embryo. *Proceedings of the National Academy of Sciences* **90,** 527–531 (1993).

10. Hay, C. W. & Docherty, K. Comparative Analysis of Insulin Gene Promoters: Implications for Diabetes Research. *Diabetes* **55,** 3201–3213 (2006).

11. Mehran, A. E., Templeman, N. M., Brigidi, G. S., Lim, G. E., Chu, K.-Y., Hu, X., Botezelli, J. D., Asadi, A., Hoffman, B. G., Kieffer, T. J., Bamji, S. X., Clee, S. M. & Johnson, J. D. Hyperinsulinemia Drives Diet-Induced Obesity Independently of Brain Insulin Production. *Cell Metabolism* **16,** 723–737 (2012).

12. Tschritter, O., Preissl, H., Hennige, A. M., Stumvoll, M., Porubska, K., Frost, R., Marx, H., Klosel, B., Lutzenberger, W., Birbaumer, N., Haring, H.-U. & Fritsche, A. The cerebrocortical response to hyperinsulinemia is reduced in overweight humans: A magnetoencephalographic study. *Proceedings of the National Academy of Sciences* **103,** 12103–12108 (2006).

13. Gelling, R. W., Morton, G. J., Morrison, C. D., Niswender, K. D., Myers, M. G., Rhodes, C. J. & Schwartz, M. W. Insulin action in the brain contributes to glucose lowering during insulin treatment of diabetes. *Cell Metabolism* **3,** 67–73 (2006).

14. Ono, H., Pocai, A., Wang, Y., Sakoda, H., Asano, T., Backer, J. M., Schwartz, G. J. & Rossetti, L. Activation of hypothalamic S6 kinase mediates diet-induced hepatic insulin resistance in rats. *Journal of Clinical Investigation* (2008). doi:10.1172/JCI34277

15. Barnett, A. H., Eff, C., Leslie, R. D. G. & Pyke, D. A. Diabetes in identical twins: A study of 200 pairs. *Diabetologia* **20,** 87–93 (1981).

16. Fuchsberger, C., Flannick, J., Teslovich, T. M., Mahajan, A., Agarwala, V., Gaulton, K. J., Ma, C., Fontanillas, P., Moutsianas, L., McCarthy, D. J., Rivas, M. A., Perry, J. R. B., Sim, X., Blackwell, T. W., Robertson, N. R., Rayner, N. W., Cingolani, P., Locke, A. E., Tajes, J. F., Highland, H. M., Dupuis, J., Chines, P. S., Lindgren, C. M., Hartl, C., Jackson, A. U., Chen, H., Huyghe, J. R., van de Bunt, M., Pearson, R. D., Kumar, A., Müller-Nurasyid, M., Grarup, N., Stringham, H. M., Gamazon, E. R., Lee, J., Chen, Y., Scott, R. A., Below, J. E., Chen, P., Huang, J., Go, M. J., Stitzel, M. L., Pasko, D., Parker, S. C. J., Varga, T. V., Green, T., Beer, N. L., Day-Williams, A. G., Ferreira, T., Fingerlin, T., Horikoshi, M., Hu, C., Huh, I., Ikram, M. K., Kim, B.-J., Kim, Y., Kim, Y. J., Kwon, M.-S., Lee, J., Lee, S., Lin, K.-H., Maxwell, T. J., Nagai, Y., Wang, X., Welch, R. P., Yoon, J., Zhang, W., Barzilai, N., Voight, B. F., Han, B.-G., Jenkinson, C. P., Kuulasmaa, T., Kuusisto, J., Manning, A., Ng, M. C. Y., Palmer, N. D., Balkau, B., Stančáková, A., Abboud, H. E., Boeing, H., Giedraitis, V., Prabhakaran, D., Gottesman, O., Scott, J., Carey, J., Kwan, P., Grant, G., Smith, J. D., Neale, B. M., Purcell, S., Butterworth, A. S., Howson, J. M. M., Lee, H. M., Lu, Y., Kwak, S.-H., Zhao, W., Danesh, J., Lam, V. K. L., Park, K. S., Saleheen, D., So, W. Y., Tam, C. H. T., Afzal, U., Aguilar, D., Arya, R., Aung, T., Chan, E., Navarro, C., Cheng, C.-Y., Palli, D., Correa, A., Curran, J. E., Rybin, D., Farook, V. S., Fowler, S. P., Freedman, B. I., Griswold, M., Hale, D. E., Hicks, P. J., Khor, C.-C., Kumar, S., Lehne, B., Thuillier, D., Lim, W. Y., Liu, J., van der Schouw, Y. T., Loh, M., Musani, S. K., Puppala, S., Scott, W. R., Yengo, L., Tan, S.-T., Taylor, H. A., Thameem, F., Wilson, G., Wong, T. Y., Njølstad, P. R., Levy, J. C., Mangino, M., Bonnycastle, L. L., Schwarzmayr, T., Fadista, J., Surdulescu, G. L., Herder, C., Groves, C. J., Wieland, T., Bork-Jensen, J., Brandslund, I., Christensen, C., Koistinen, H. A., Doney, A. S. F., Kinnunen, L., Esko, T., Farmer, A. J., Hakaste, L., Hodgkiss, D., Kravic, J., Lyssenko, V., Hollensted, M., Jørgensen, M. E., Jørgensen, T., Ladenvall, C., Justesen, J. M., Käräjämäki, A., Kriebel, J., Rathmann, W., Lannfelt, L., Lauritzen, T., Narisu, N., Linneberg, A., Melander, O., Milani, L., Neville, M., Orho-Melander, M., Qi, L., Qi, Q., Roden, M., Rolandsson, O., Swift, A., Rosengren, A. H., Stirrups, K., Wood, A. R., Mihailov, E., Blancher, C., Carneiro, M. O., Maguire, J., Poplin, R., Shakir, K., Fennell, T., DePristo, M., Hrabé de Angelis, M., Deloukas, P., Gjesing, A. P., Jun, G., Nilsson, P., Murphy, J., Onofrio, R., Thorand, B., Hansen, T., Meisinger, C., Hu, F. B., Isomaa, B., Karpe, F., Liang, L., Peters, A., Huth, C., O'Rahilly, S. P., Palmer, C. N. A., Pedersen, O., Rauramaa, R., Tuomilehto, J., Salomaa, V., Watanabe, R. M., Syvänen, A.-C., Bergman, R. N., Bharadwaj, D., Bottinger, E. P., Cho, Y. S., Chandak, G. R., Chan, J. C. N., Chia, K. S., Daly, M. J., Ebrahim, S. B., Langenberg, C., Elliott, P., Jablonski, K. A., Lehman, D. M., Jia, W., Ma, R. C. W., Pollin, T. I., Sandhu, M., Tandon, N., Froguel, P., Barroso, I., Teo, Y. Y., Zeggini, E., Loos, R. J. F., Small, K. S., Ried, J. S., DeFronzo, R. A., Grallert, H., Glaser, B., Metspalu, A., Wareham, N. J., Walker, M., Banks, E., Gieger, C., Ingelsson, E., Im, H. K., Illig, T., Franks, P. W., Buck, G., Trakalo, J., Buck, D., Prokopenko, I., Mägi, R., Lind, L., Farjoun, Y., Owen, K. R., Gloyn, A. L., Strauch, K., Tuomi, T., Kooner, J. S., Lee, J.-Y., Park, T., Donnelly, P., Morris, A. D., Hattersley, A. T.,

Bowden, D. W., Collins, F. S., Atzmon, G., Chambers, J. C., Spector, T. D., Laakso, M., Strom, T. M., Bell, G. I., Blangero, J., Duggirala, R., Tai, E. S., McVean, G., Hanis, C. L., Wilson, J. G., Seielstad, M., Frayling, T. M., Meigs, J. B., Cox, N. J., Sladek, R., Lander, E. S., Gabriel, S., Burtt, N. P., Mohlke, K. L., Meitinger, T., Groop, L., Abecasis, G., Florez, J. C., Scott, L. J., Morris, A. P., Kang, H. M., Boehnke, M., Altshuler, D. & McCarthy, M. I. The genetic architecture of type 2 diabetes. *Nature* **536,** 41–47 (2016).

17. Gaulton, K. J., Ferreira, T., Lee, Y., Raimondo, A., the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, Mägi, R., Reschen, M. E., Mahajan, A., Locke, A., William Rayner, N., Robertson, N., Scott, R. A., Prokopenko, I., Scott, L. J., Green, T., Sparso, T., Thuillier, D., Yengo, L., Grallert, H., Wahl, S., Frånberg, M., Strawbridge, R. J., Kestler, H., Chheda, H., Eisele, L., Gustafsson, S., Steinthorsdottir, V., Thorleifsson, G., Qi, L., Karssen, L. C., van Leeuwen, E. M., Willems, S. M., Li, M., Chen, H., Fuchsberger, C., Kwan, P., Ma, C., Linderman, M., Lu, Y., Thomsen, S. K., Rundle, J. K., Beer, N. L., van de Bunt, M., Chalisey, A., Kang, H. M., Voight, B. F., Abecasis, G. R., Almgren, P., Baldassarre, D., Balkau, B., Benediktsson, R., Blüher, M., Boeing, H., Bonnycastle, L. L., Bottinger, E. P., Burtt, N. P., Carey, J., Charpentier, G., Chines, P. S., Cornelis, M. C., Couper, D. J., Crenshaw, A. T., van Dam, R. M., Doney, A. S. F., Dorkhan, M., Edkins, S., Eriksson, J. G., Esko, T., Eury, E., Fadista, J., Flannick, J., Fontanillas, P., Fox, C., Franks, P. W., Gertow, K., Gieger, C., Gigante, B., Gottesman, O., Grant, G. B., Grarup, N., Groves, C. J., Hassinen, M., Have, C. T., Herder, C., Holmen, O. L., Hreidarsson, A. B., Humphries, S. E., Hunter, D. J., Jackson, A. U., Jonsson, A., Jørgensen, M. E., Jørgensen, T., Kao, W.-H. L., Kerrison, N. D., Kinnunen, L., Klopp, N., Kong, A., Kovacs, P., Kraft, P., Kravic, J., Langford, C., Leander, K., Liang, L., Lichtner, P., Lindgren, C. M., Lindholm, E., Linneberg, A., Liu, C.-T., Lobbens, S., Luan, J., Lyssenko, V., Männistö, S., McLeod, O., Meyer, J., Mihailov, E., Mirza, G., Mühleisen, T. W., Müller-Nurasyid, M., Navarro, C., Nöthen, M. M., Oskolkov, N. N., Owen, K. R., Palli, D., Pechlivanis, S., Peltonen, L., Perry, J. R. B., Platou, C. G. P., Roden, M., Ruderfer, D., Rybin, D., van der Schouw, Y. T., Sennblad, B., Sigurðsson, G., Stančáková, A., Steinbach, G., Storm, P., Strauch, K., Stringham, H. M., Sun, Q., Thorand, B., Tikkanen, E., Tonjes, A., Trakalo, J., Tremoli, E., Tuomi, T., Wennauer, R., Wiltshire, S., Wood, A. R., Zeggini, E., Dunham, I., Birney, E., Pasquali, L., Ferrer, J., Loos, R. J. F., Dupuis, J., Florez, J. C., Boerwinkle, E., Pankow, J. S., van Duijn, C., Sijbrands, E., Meigs, J. B., Hu, F. B., Thorsteinsdottir, U., Stefansson, K., Lakka, T. A., Rauramaa, R., Stumvoll, M., Pedersen, N. L., Lind, L., Keinanen-Kiukaanniemi, S. M., Korpi-Hyövälti, E., Saaristo, T. E., Saltevo, J., Kuusisto, J., Laakso, M., Metspalu, A., Erbel, R., Jöcke, K.-H., Moebus, S., Ripatti, S., Salomaa, V., Ingelsson, E., Boehm, B. O., Bergman, R. N., Collins, F. S., Mohlke, K. L., Koistinen, H., Tuomilehto, J., Hveem, K., Njølstad, I., Deloukas, P., Donnelly, P. J., Frayling, T. M., Hattersley, A. T., de Faire, U., Hamsten, A., Illig, T., Peters, A., Cauchi, S., Sladek, R., Froguel, P., Hansen, T., Pedersen, O., Morris, A. D., Palmer, C. N. A., Kathiresan, S., Melander, O., Nilsson, P. M., Groop, L. C., Barroso, I., Langenberg, C., Wareham, N. J., O'Callaghan, C. A., Gloyn, A. L., Altshuler, D., Boehnke, M., Teslovich, T. M., McCarthy, M. I. & Morris, A. P. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nature Genetics* **47,** 1415–1425 (2015).

18. Scott, R. A., Scott, L. J., Mägi, R., Marullo, L., Gaulton, K. J., Kaakinen, M., Pervjakova, N., Pers, T. H., Johnson, A. D., Eicher, J. D., Jackson, A. U., Ferreira, T., Lee, Y., Ma, C.,

Steinthorsdottir, V., Thorleifsson, G., Qi, L., Van Zuydam, N. R., Mahajan, A., Chen, H., Almgren, P., Voight, B. F., Grallert, H., Müller-Nurasyid, M., Ried, J. S., Rayner, N. W., Robertson, N., Karssen, L. C., van Leeuwen, E. M., Willems, S. M., Fuchsberger, C., Kwan, P., Teslovich, T. M., Chanda, P., Li, M., Lu, Y., Dina, C., Thuillier, D., Yengo, L., Jiang, L., Sparso, T., Kestler, H. A., Chheda, H., Eisele, L., Gustafsson, S., Frånberg, M., Strawbridge, R. J., Benediktsson, R., Hreidarsson, A. B., Kong, A., Sigurðsson, G., Kerrison, N. D., Luan, J., Liang, L., Meitinger, T., Roden, M., Thorand, B., Esko, T., Mihailov, E., Fox, C., Liu, C.-T., Rybin, D., Isomaa, B., Lyssenko, V., Tuomi, T., Couper, D. J., Pankow, J. S., Grarup, N., Have, C. T., Jørgensen, M. E., Jørgensen, T., Linneberg, A., Cornelis, M. C., van Dam, R. M., Hunter, D. J., Kraft, P., Sun, Q., Edkins, S., Owen, K. R., Perry, J. R. B., Wood, A. R., Zeggini, E., Tajes-Fernandes, J., Abecasis, G. R., Bonnycastle, L. L., Chines, P. S., Stringham, H. M., Koistinen, H. A., Kinnunen, L., Sennblad, B., Mühleisen, T. W., Nöthen, M. M., Pechlivanis, S., Baldassarre, D., Gertow, K., Humphries, S. E., Tremoli, E., Klopp, N., Meyer, J., Steinbach, G., Wennauer, R., Eriksson, J. G., Männistö, S., Peltonen, L., Tikkanen, E., Charpentier, G., Eury, E., Lobbens, S., Gigante, B., Leander, K., McLeod, O., Bottinger, E. P., Gottesman, O., Ruderfer, D., Blüher, M., Kovacs, P., Tonjes, A., Maruthur, N. M., Scapoli, C., Erbel, R., Jöckel, K.-H., Moebus, S., de Faire, U., Hamsten, A., Stumvoll, M., Deloukas, P., Donnelly, P. J., Frayling, T. M., Hattersley, A. T., Ripatti, S., Salomaa, V., Pedersen, N. L., Boehm, B. O., Bergman, R. N., Collins, F. S., Mohlke, K. L., Tuomilehto, J., Hansen, T., Pedersen, O., Barroso, I., Lannfelt, L., Ingelsson, E., Lind, L., Lindgren, C. M., Cauchi, S., Froguel, P., Loos, R. J. F., Balkau, B., Boeing, H., Franks, P. W., Barricarte Gurrea, A., Palli, D., van der Schouw, Y. T., Altshuler, D., Groop, L. C., Langenberg, C., Wareham, N. J., Sijbrands, E., van Duijn, C. M., Florez, J. C., Meigs, J. B., Boerwinkle, E., Gieger, C., Strauch, K., Metspalu, A., Morris, A. D., Palmer, C. N. A., Hu, F. B., Thorsteinsdottir, U., Stefansson, K., Dupuis, J., Morris, A. P., Boehnke, M., McCarthy, M. I. & Prokopenko, I. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes* **66,** 2888–2902 (2017).

19. The LifeLines Cohort Study, The ADIPOGen Consortium, The AGEN-BMI Working Group, The CARDIOGRAMplusC4D Consortium, The CKDGen Consortium, The GLGC, The ICBP, The MAGIC Investigators, The MuTHER Consortium, The MIGen Consortium, The PAGE Consortium, The ReproGen Consortium, The GENIE Consortium, The International Endogene Consortium, Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., Powell, C., Vedantam, S., Buchkovich, M. L., Yang, J., Croteau-Chonka, D. C., Esko, T., Fall, T., Ferreira, T., Gustafsson, S., Kutalik, Z., Luan, J., Mägi, R., Randall, J. C., Winkler, T. W., Wood, A. R., Workalemahu, T., Faul, J. D., Smith, J. A., Hua Zhao, J., Zhao, W., Chen, J., Fehrmann, R., Hedman, Å. K., Karjalainen, J., Schmidt, E. M., Absher, D., Amin, N., Anderson, D., Beekman, M., Bolton, J. L., Bragg-Gresham, J. L., Buyske, S., Demirkan, A., Deng, G., Ehret, G. B., Feenstra, B., Feitosa, M. F., Fischer, K., Goel, A., Gong, J., Jackson, A. U., Kanoni, S., Kleber, M. E., Kristiansson, K., Lim, U., Lotay, V., Mangino, M., Mateo Leach, I., Medina-Gomez, C., Medland, S. E., Nalls, M. A., Palmer, C. D., Pasko, D., Pechlivanis, S., Peters, M. J., Prokopenko, I., Shungin, D., Stančáková, A., Strawbridge, R. J., Ju Sung, Y., Tanaka, T., Teumer, A., Trompet, S., van der Laan, S. W., van Setten, J., Van Vliet-Ostaptchouk, J. V., Wang, Z., Yengo, L., Zhang, W., Isaacs, A., Albrecht, E., Ärnlöv, J., Arscott, G. M., Attwood, A. P., Bandinelli, S., Barrett, A., Bas, I. N., Bellis, C., Bennett, A. J., Berne, C., Blagieva, R., Blüher, M., Böhringer, S., Bonnycastle, L. L., Böttcher, Y., Boyd, H. A., Bruinenberg, M., Caspersen, I.

H., Ida Chen, Y.-D., Clarke, R., Warwick Daw, E., de Craen, A. J. M., Delgado, G., Dimitriou, M., Doney, A. S. F., Eklund, N., Estrada, K., Eury, E., Folkersen, L., Fraser, R. M., Garcia, M. E., Geller, F., Giedraitis, V., Gigante, B., Go, A. S., Golay, A., Goodall, A. H., Gordon, S. D., Gorski, M., Grabe, H.-J., Grallert, H., Grammer, T. B., Gräßler, J., Grönberg, H., Groves, C. J., Gusto, G., Haessler, J., Hall, P., Haller, T., Hallmans, G., Hartman, C. A., Hassinen, M., Hayward, C., Heard-Costa, N. L., Helmer, Q., Hengstenberg, C., Holmen, O., Hottenga, J.-J., James, A. L., Jeff, J. M., Johansson, Å., Jolley, J., Juliusdottir, T., Kinnunen, L., Koenig, W., Koskenvuo, M., Kratzer, W., Laitinen, J., Lamina, C., Leander, K., Lee, N. R., Lichtner, P., Lind, L., Lindström, J., Sin Lo, K., Lobbens, S., Lorbeer, R., Lu, Y., Mach, F., Magnusson, P. K. E., Mahajan, A., McArdle, W. L., McLachlan, S., Menni, C., Merger, S., Mihailov, E., Milani, L., Moayyeri, A., Monda, K. L., Morken, M. A., Mulas, A., Müller, G., Müller-Nurasyid, M., Musk, A. W., Nagaraja, R., Nöthen, M. M., Nolte, I. M., Pilz, S., Rayner, N. W., Renstrom, F., Rettig, R., Ried, J. S., Ripke, S., Robertson, N. R., Rose, L. M., Sanna, S., Scharnagl, H., Scholtens, S., Schumacher, F. R., Scott, W. R., Seufferlein, T., Shi, J., Vernon Smith, A., Smolonska, J., Stanton, A. V., Steinthorsdottir, V., Stirrups, K., Stringham, H. M., Sundström, J., Swertz, M. A., Swift, A. J., Syvänen, A.-C., Tan, S.-T., Tayo, B. O., Thorand, B., Thorleifsson, G., Tyrer, J. P., Uh, H.-W., Vandenput, L., Verhulst, F. C., Vermeulen, S. H., Verweij, N., Vonk, J. M., Waite, L. L., Warren, H. R., Waterworth, D., Weedon, M. N., Wilkens, L. R., Willenborg, C., Wilsgaard, T., Wojczynski, M. K., Wong, A., Wright, A. F., Zhang, Q., Brennan, E. P., Choi, M., Dastani, Z., Drong, A. W., Eriksson, P., Franco-Cereceda, A., Gådin, J. R., Gharavi, A. G., Goddard, M. E., Handsaker, R. E., Huang, J., Karpe, F., Kathiresan, S., Keildson, S., Kiryluk, K., Kubo, M., Lee, J.-Y., Liang, L., Lifton, R. P., Ma, B., McCarroll, S. A., McKnight, A. J., Min, J. L., Moffatt, M. F., Montgomery, G. W., Murabito, J. M., Nicholson, G., Nyholt, D. R., Okada, Y., Perry, J. R. B., Dorajoo, R., Reinmaa, E., Salem, R. M., Sandholm, N., Scott, R. A., Stolk, L., Takahashi, A., Tanaka, T., van't Hooft, F. M., Vinkhuyzen, A. A. E., Westra, H.-J., Zheng, W., Zondervan, K. T., Heath, A. C., Arveiler, D., Bakker, S. J. L., Beilby, J., Bergman, R. N., Blangero, J., Bovet, P., Campbell, H., Caulfield, M. J., Cesana, G., Chakravarti, A., Chasman, D. I., Chines, P. S., Collins, F. S., Crawford, D. C., Adrienne Cupples, L., Cusi, D., Danesh, J., de Faire, U., den Ruijter, H. M., Dominiczak, A. F., Erbel, R., Erdmann, J., Eriksson, J. G., Farrall, M., Felix, S. B., Ferrannini, E., Ferrières, J., Ford, I., Forouhi, N. G., Forrester, T., Franco, O. H., Gansevoort, R. T., Gejman, P. V., Gieger, C., Gottesman, O., Gudnason, V., Gyllensten, U., Hall, A. S., Harris, T. B., Hattersley, A. T., Hicks, A. A., Hindorff, L. A., Hingorani, A. D., Hofman, A., Homuth, G., Kees Hovingh, G., Humphries, S. E., Hunt, S. C., Hyppönen, E., Illig, T., Jacobs, K. B., Jarvelin, M.-R., Jöckel, K.-H., Johansen, B., Jousilahti, P., Wouter Jukema, J., Jula, A. M., Kaprio, J., Kastelein, J. J. P., Keinanen-Kiukaanniemi, S. M., Kiemeney, L. A., Knekt, P., Kooner, J. S., Kooperberg, C., Kovacs, P., Kraja, A. T., Kumari, M., Kuusisto, J., Lakka, T. A., Langenberg, C., Le Marchand, L., Lehtimäki, T., Lyssenko, V., Männistö, S., Marette, A., Matise, T. C., McKenzie, C. A., McKnight, B., Moll, F. L., Morris, A. D., Morris, A. P., Murray, J. C., Nelis, M., Ohlsson, C., Oldehinkel, A. J., Ong, K. K., Madden, P. A. F., Pasterkamp, G., Peden, J. F., Peters, A., Postma, D. S., Pramstaller, P. P., Price, J. F., Qi, L., Raitakari, O. T., Rankinen, T., Rao, D. C., Rice, T. K., Ridker, P. M., Rioux, J. D., Ritchie, M. D., Rudan, I., Salomaa, V., Samani, N. J., Saramies, J., Sarzynski, M. A., Schunkert, H., Schwarz, P. E. H., Sever, P., Shuldiner, A. R., Sinisalo, J., Stolk, R. P., Strauch, K., Tönjes, A., Trégouët, D.-A., Tremblay, A., Tremoli, E., Virtamo, J.,

Vohl, M.-C., Völker, U., Waeber, G., Willemsen, G., Witteman, J. C., Carola Zillikens, M., Adair, L. S., Amouyel, P., Asselbergs, F. W., Assimes, T. L., Bochud, M., Boehm, B. O., Boerwinkle, E., Bornstein, S. R., Bottinger, E. P., Bouchard, C., Cauchi, S., Chambers, J. C., Chanock, S. J., Cooper, R. S., de Bakker, P. I. W., Dedoussis, G., Ferrucci, L., Franks, P. W., Froguel, P., Groop, L. C., Haiman, C. A., Hamsten, A., Hui, J., Hunter, D. J., Hveem, K., Kaplan, R. C., Kivimaki, M., Kuh, D., Laakso, M., Liu, Y., Martin, N. G., März, W., Melbye, M., Metspalu, A., Moebus, S., Munroe, P. B., Njølstad, I., Oostra, B. A., Palmer, C. N. A., Pedersen, N. L., Perola, M., Pérusse, L., Peters, U., Power, C., Quertermous, T., Rauramaa, R., Rivadeneira, F., Saaristo, T. E., Saleheen, D., Sattar, N., Schadt, E. E., Schlessinger, D., Eline Slagboom, P., Snieder, H., Spector, T. D., Thorsteinsdottir, U., Stumvoll, M., Tuomilehto, J., Uitterlinden, A. G., Uusitupa, M., van der Harst, P., Walker, M., Wallaschofski, H., Wareham, N. J., Watkins, H., Weir, D. R., Wichmann, H.-E., Wilson, J. F., Zanen, P., Borecki, I. B., Deloukas, P., Fox, C. S., Heid, I. M., O'Connell, J. R., Strachan, D. P., Stefansson, K., van Duijn, C. M., Abecasis, G. R., Franke, L., Frayling, T. M., McCarthy, M. I., Visscher, P. M., Scherag, A., Willer, C. J., Boehnke, M., Mohlke, K. L., Lindgren, C. M., Beckmann, J. S., Barroso, I., North, K. E., Ingelsson, E., Hirschhorn, J. N., Loos, R. J. F. & Speliotes, E. K. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518,** 197–206 (2015).

20. Akiyama, M., Okada, Y., Kanai, M., Takahashi, A., Momozawa, Y., Ikeda, M., Iwata, N., Ikegawa, S., Hirata, M., Matsuda, K., Iwasaki, M., Yamaji, T., Sawada, N., Hachiya, T., Tanno, K., Shimizu, A., Hozawa, A., Minegishi, N., Tsugane, S., Yamamoto, M., Kubo, M. & Kamatani, Y. Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat. Genet.* **49,** 1458–1467 (2017).

21. Spain, S. L. & Barrett, J. C. Strategies for fine-mapping complex traits. *Human Molecular Genetics* **24,** R111–R119 (2015).

22. Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., Lehrach, H., Mardis, E. R., Marth, G. T., McVean, G. A., Nickerson, D. A., Schmidt, J. P., Sherry, S. T., Wang, J., Wilson, R. K., Gibbs, R. A., Boerwinkle, E., Doddapaneni, H., Han, Y., Korchina, V., Kovar, C., Lee, S., Muzny, D., Reid, J. G., Zhu, Y., Wang, J., Chang, Y., Feng, Q., Fang, X., Guo, X., Jian, M., Jiang, H., Jin, X., Lan, T., Li, G., Li, J., Li, Y., Liu, S., Liu, X., Lu, Y., Ma, X., Tang, M., Wang, B., Wang, G., Wu, H., Wu, R., Xu, X., Yin, Y., Zhang, D., Zhang, W., Zhao, J., Zhao, M., Zheng, X., Lander, E. S., Altshuler, D. M., Gabriel, S. B., Gupta, N., Gharani, N., Toji, L. H., Gerry, N. P., Resch, A. M., Flicek, P., Barker, J., Clarke, L., Gil, L., Hunt, S. E., Kelman, G., Kulesha, E., Leinonen, R., McLaren, W. M., Radhakrishnan, R., Roa, A., Smirnov, D., Smith, R. E., Streeter, I., Thormann, A., Toneva, I., Vaughan, B., Zheng-Bradley, X., Bentley, D. R., Grocock, R., Humphray, S., James, T., Kingsbury, Z., Lehrach, H., Sudbrak, R., Albrecht, M. W., Amstislavskiy, V. S., Borodina, T. A., Lienhard, M., Mertes, F., Sultan, M., Timmermann, B., Yaspo, M.-L., Mardis, E. R., Wilson, R. K., Fulton, L., Fulton, R., Sherry, S. T., Ananiev, V., Belaia, Z., Beloslyudtsev, D., Bouk, N., Chen, C., Church, D., Cohen, R., Cook, C., Garner, J., Hefferon, T., Kimelman, M., Liu, C., Lopez, J., Meric, P., O'Sullivan, C., Ostapchuk, Y., Phan, L., Ponomarov, S., Schneider, V., Shekhtman, E., Sirotkin, K., Slotta, D., Zhang, H.,

McVean, G. A., Durbin, R. M., Balasubramaniam, S., Burton, J., Danecek, P., Keane, T. M., Kolb-Kokocinski, A., McCarthy, S., Stalker, J., Quail, M., Schmidt, J. P., Davies, C. J., Gollub, J., Webster, T., Wong, B., Zhan, Y., Auton, A., Campbell, C. L., Kong, Y., Marcketta, A., Gibbs, R. A., Yu, F., Antunes, L., Bainbridge, M., Muzny, D., Sabo, A., Huang, Z., Wang, J., Coin, L. J. M., Fang, L., Guo, X., Jin, X., Li, G., Li, Q., Li, Y., Li, Z., Lin, H., Liu, B., Luo, R., Shao, H., Xie, Y., Ye, C., Yu, C., Zhang, F., Zheng, H., Zhu, H., Alkan, C., Dal, E., Kahveci, F., Marth, G. T., Garrison, E. P., Kural, D., Lee, W.-P., Fung Leong, W., Stromberg, M., Ward, A. N., Wu, J., Zhang, M., Daly, M. J., DePristo, M. A., Handsaker, R. E., Altshuler, D. M., Banks, E., Bhatia, G., del Angel, G., Gabriel, S. B., Genovese, G., Gupta, N., Li, H., Kashin, S., Lander, E. S., McCarroll, S. A., Nemesh, J. C., Poplin, R. E., Yoon, S. C., Lihm, J., Makarov, V., Clark, A. G., Gottipati, S., Keinan, A., Rodriguez-Flores, J. L., Korbel, J. O., Rausch, T., Fritz, M. H., Stütz, A. M., Flicek, P., Beal, K., Clarke, L., Datta, A., Herrero, J., McLaren, W. M., Ritchie, G. R. S., Smith, R. E., Zerbino, D., Zheng-Bradley, X., Sabeti, P. C., Shlyakhter, I., Schaffner, S. F., Vitti, J., Cooper, D. N., Ball, E. V., Stenson, P. D., Bentley, D. R., Barnes, B., Bauer, M., Keira Cheetham, R., Cox, A., Eberle, M., Humphray, S., Kahn, S., Murray, L., Peden, J., Shaw, R., Kenny, E. E., Batzer, M. A., Konkel, M. K., Walker, J. A., MacArthur, D. G., Lek, M., Sudbrak, R., Amstislavskiy, V. S., Herwig, R., Mardis, E. R., Ding, L., Koboldt, D. C., Larson, D., Ye, K., Gravel, S., Swaroop, A., Chew, E., Lappalainen, T., Erlich, Y., Gymrek, M., Frederick Willems, T., Simpson, J. T., Shriver, M. D., Rosenfeld, J. A., Bustamante, C. D., Montgomery, S. B., De La Vega, F. M., Byrnes, J. K., Carroll, A. W., DeGorter, M. K., Lacroute, P., Maples, B. K., Martin, A. R., Moreno-Estrada, A., Shringarpure, S. S., Zakharia, F., Halperin, E., Baran, Y., Lee, C., Cerveira, E., Hwang, J., Malhotra, A., Plewczynski, D., Radew, K., Romanovitch, M., Zhang, C., Hyland, F. C. L., Craig, D. W., Christoforides, A., Homer, N., Izatt, T., Kurdoglu, A. A., Sinari, S. A., Squire, K., Sherry, S. T., Xiao, C., Sebat, J., Antaki, D., Gujral, M., Noor, A., Ye, K., Burchard, E. G., Hernandez, R. D., Gignoux, C. R., Haussler, D., Katzman, S. J., James Kent, W., Howie, B., Ruiz-Linares, A., Dermitzakis, E. T., Devine, S. E., Abecasis, G. R., Min Kang, H., Kidd, J. M., Blackwell, T., Caron, S., Chen, W., Emery, S., Fritsche, L., Fuchsberger, C., Jun, G., Li, B., Lyons, R., Scheller, C., Sidore, C., Song, S., Sliwerska, E., Taliun, D., Tan, A., Welch, R., Kate Wing, M., Zhan, X., Awadalla, P., Hodgkinson, A., Li, Y., Shi, X., Quitadamo, A., Lunter, G., McVean, G. A., Marchini, J. L., Myers, S., Churchhouse, C., Delaneau, O., Gupta-Hinch, A., Kretzschmar, W., Iqbal, Z., Mathieson, I., Menelaou, A., Rimmer, A., Xifara, D. K., Oleksyk, T. K., Fu, Y., Liu, X., Xiong, M., Jorde, L., Witherspoon, D., Xing, J., Eichler, E. E., Browning, B. L., Browning, S. R., Hormozdiari, F., Sudmant, P. H., Khurana, E., Durbin, R. M., Hurles, M. E., Tyler-Smith, C., Albers, C. A., Ayub, Q., Balasubramaniam, S., Chen, Y., Colonna, V., Danecek, P., Jostins, L., Keane, T. M., McCarthy, S., Walter, K., Xue, Y., Gerstein, M. B., Abyzov, A., Balasubramanian, S., Chen, J., Clarke, D., Fu, Y., Harmanci, A. O., Jin, M., Lee, D., Liu, J., Jasmine Mu, X., Zhang, J., Zhang, Y., Li, Y., Luo, R., Zhu, H., Alkan, C., Dal, E., Kahveci, F., Marth, G. T., Garrison, E. P., Kural, D., Lee, W.-P., Ward, A. N., Wu, J., Zhang, M., McCarroll, S. A., Handsaker, R. E., Altshuler, D. M., Banks, E., del Angel, G., Genovese, G., Hartl, C., Li, H., Kashin, S., Nemesh, J. C., Shakir, K., Yoon, S. C., Lihm, J., Makarov, V., Degenhardt, J., Korbel, J. O., Fritz, M. H., Meiers, S., Raeder, B., Rausch, T., Stütz, A. M., Flicek, P., Paolo Casale, F., Clarke, L., Smith, R. E., Stegle, O., Zheng-Bradley, X., Bentley, D. R., Barnes, B., Keira Cheetham, R., Eberle, M., Humphray, S., Kahn, S., Murray, L., Shaw, R., Lameijer, E.-W.,

Batzer, M. A., Konkel, M. K., Walker, J. A., Ding, L., Hall, I., Ye, K., Lacroute, P., Lee, C., Cerveira, E., Malhotra, A., Hwang, J., Plewczynski, D., Radew, K., Romanovitch, M., Zhang, C., Craig, D. W., Homer, N., Church, D., Xiao, C., Sebat, J., Antaki, D., Bafna, V., Michaelson, J., Ye, K., Devine, S. E., Gardner, E. J., Abecasis, G. R., Kidd, J. M., Mills, R. E., Dayama, G., Emery, S., Jun, G., Shi, X., Quitadamo, A., Lunter, G., McVean, G. A., Chen, K., Fan, X., Chong, Z., Chen, T., Witherspoon, D., Xing, J., Eichler, E. E., Chaisson, M. J., Hormozdiari, F., Huddleston, J., Malig, M., Nelson, B. J., Sudmant, P. H., Parrish, N. F., Khurana, E., Hurles, M. E., Blackburne, B., Lindsay, S. J., Ning, Z., Walter, K., Zhang, Y., Gerstein, M. B., Abyzov, A., Chen, J., Clarke, D., Lam, H., Jasmine Mu, X., Sisu, C., Zhang, J., Zhang, Y., Gibbs, R. A., Yu, F., Bainbridge, M., Challis, D., Evani, U. S., Kovar, C., Lu, J., Muzny, D., Nagaswamy, U., Reid, J. G., Sabo, A., Yu, J., Guo, X., Li, W., Li, Y., Wu, R., Marth, G. T., Garrison, E. P., Fung Leong, W., Ward, A. N., del Angel, G., DePristo, M. A., Gabriel, S. B., Gupta, N., Hartl, C., Poplin, R. E., Clark, A. G., Rodriguez-Flores, J. L., Flicek, P., Clarke, L., Smith, R. E., Zheng-Bradley, X., MacArthur, D. G., Mardis, E. R., Fulton, R., Koboldt, D. C., Gravel, S., Bustamante, C. D., Craig, D. W., Christoforides, A., Homer, N., Izatt, T., Sherry, S. T., Xiao, C., Dermitzakis, E. T., Abecasis, G. R., Min Kang, H., McVean, G. A., Gerstein, M. B., Balasubramanian, S., Habegger, L., Yu, H., Flicek, P., Clarke, L., Cunningham, F., Dunham, I., Zerbino, D., Zheng-Bradley, X., Lage, K., Berg Jespersen, J., Horn, H., Montgomery, S. B., DeGorter, M. K., Khurana, E., Tyler-Smith, C., Chen, Y., Colonna, V., Xue, Y., Gerstein, M. B., Balasubramanian, S., Fu, Y., Kim, D., Auton, A., Marcketta, A., Desalle, R., Narechania, A., Wilson Sayres, M. A., Garrison, E. P., Handsaker, R. E., Kashin, S., McCarroll, S. A., Rodriguez-Flores, J. L., Flicek, P., Clarke, L., Zheng-Bradley, X., Erlich, Y., Gymrek, M., Frederick Willems, T., Bustamante, C. D., Mendez, F. L., David Poznik, G., Underhill, P. A., Lee, C., Cerveira, E., Malhotra, A., Romanovitch, M., Zhang, C., Abecasis, G. R., Coin, L., Shao, H., Mittelman, D., Tyler-Smith, C., Ayub, Q., Banerjee, R., Cerezo, M., Chen, Y., Fitzgerald, T. W., Louzada, S., Massaia, A., McCarthy, S., Ritchie, G. R., Xue, Y., Yang, F., Gibbs, R. A., Kovar, C., Kalra, D., Hale, W., Muzny, D., Reid, J. G., Wang, J., Dan, X., Guo, X., Li, G., Li, Y., Ye, C., Zheng, X., Altshuler, D. M., Flicek, P., Clarke, L., Zheng-Bradley, X., Bentley, D. R., Cox, A., Humphray, S., Kahn, S., Sudbrak, R., Albrecht, M. W., Lienhard, M., Larson, D., Craig, D. W., Izatt, T., Kurdoglu, A. A., Sherry, S. T., Xiao, C., Haussler, D., Abecasis, G. R., McVean, G. A., Durbin, R. M., Balasubramaniam, S., Keane, T. M., McCarthy, S., Stalker, J., Chakravarti, A., Knoppers, B. M., Abecasis, G. R., Barnes, K. C., Beiswanger, C., Burchard, E. G., Bustamante, C. D., Cai, H., Cao, H., Durbin, R. M., Gerry, N. P., Gharani, N., Gibbs, R. A., Gignoux, C. R., Gravel, S., Henn, B., Jones, D., Jorde, L., Kaye, J. S., Keinan, A., Kent, A., Kerasidou, A., Li, Y., Mathias, R., McVean, G. A., Moreno-Estrada, A., Ossorio, P. N., Parker, M., Resch, A. M., Rotimi, C. N., Royal, C. D., Sandoval, K., Su, Y., Sudbrak, R., Tian, Z., Tishkoff, S., Toji, L. H., Tyler-Smith, C., Via, M., Wang, Y., Yang, H., Yang, L., Zhu, J., Bodmer, W., Bedoya, G., Ruiz-Linares, A., Cai, Z., Gao, Y., Chu, J., Peltonen, L., Garcia-Montero, A., Orfao, A., Dutil, J., Martinez-Cruzado, J. C., Oleksyk, T. K., Barnes, K. C., Mathias, R. A., Hennis, A., Watson, H., McKenzie, C., Qadri, F., LaRocque, R., Sabeti, P. C., Zhu, J., Deng, X., Sabeti, P. C., Asogun, D., Folarin, O., Happi, C., Omoniwa, O., Stremlau, M., Tariyal, R., Jallow, M., Sisay Joof, F., Corrah, T., Rockett, K., Kwiatkowski, D., Kooner, J., Tịnh Hiê`n, T., Dunstan, S. J., Thuy Hang, N., Fonnie, R., Garry, R., Kanneh, L., Moses, L., Sabeti, P. C., Schieffelin, J., Grant, D. S., Gallo, C., Poletti, G., Saleheen, D., Rasheed, A., Brooks, L. D.,

Felsenfeld, A. L., McEwen, J. E., Vaydylevich, Y., Green, E. D., Duncanson, A., Dunn, M., Schloss, J. A., Wang, J., Yang, H., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Min Kang, H., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A. & Abecasis, G. R. A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).

23. Mohlke, K. L. & Boehnke, M. Recent advances in understanding the genetic architecture of type 2 diabetes. *Human Molecular Genetics* **24,** R85–R92 (2015).

24. Gaulton, K. J., Ferreira, T., Lee, Y., Raimondo, A., Mägi, R., Reschen, M. E., Mahajan, A., Locke, A., Rayner, N. W., Robertson, N., Scott, R. A., Prokopenko, I., Scott, L. J., Green, T., Sparso, T., Thuillier, D., Yengo, L., Grallert, H., Wahl, S., Frånberg, M., Strawbridge, R. J., Kestler, H., Chheda, H., Eisele, L., Gustafsson, S., Steinthorsdottir, V., Thorleifsson, G., Qi, L., Karssen, L. C., van Leeuwen, E. M., Willems, S. M., Li, M., Chen, H., Fuchsberger, C., Kwan, P., Ma, C., Linderman, M., Lu, Y., Thomsen, S. K., Rundle, J. K., Beer, N. L., van de Bunt, M., Chalisey, A., Kang, H. M., Voight, B. F., Abecasis, G. R., Almgren, P., Baldassarre, D., Balkau, B., Benediktsson, R., Blüher, M., Boeing, H., Bonnycastle, L. L., Bottinger, E. P., Burtt, N. P., Carey, J., Charpentier, G., Chines, P. S., Cornelis, M. C., Couper, D. J., Crenshaw, A. T., van Dam, R. M., Doney, A. S. F., Dorkhan, M., Edkins, S., Eriksson, J. G., Esko, T., Eury, E., Fadista, J., Flannick, J., Fontanillas, P., Fox, C., Franks, P. W., Gertow, K., Gieger, C., Gigante, B., Gottesman, O., Grant, G. B., Grarup, N., Groves, C. J., Hassinen, M., Have, C. T., Herder, C., Holmen, O. L., Hreidarsson, A. B., Humphries, S. E., Hunter, D. J., Jackson, A. U., Jonsson, A., Jørgensen, M. E., Jørgensen, T., Kao, W.-H. L., Kerrison, N. D., Kinnunen, L., Klopp, N., Kong, A., Kovacs, P., Kraft, P., Kravic, J., Langford, C., Leander, K., Liang, L., Lichtner, P., Lindgren, C. M., Lindholm, E., Linneberg, A., Liu, C.-T., Lobbens, S., Luan, J., Lyssenko, V., Männistö, S., McLeod, O., Meyer, J., Mihailov, E., Mirza, G., Mühleisen, T. W., Müller-Nurasyid, M., Navarro, C., Nöthen, M. M., Oskolkov, N. N., Owen, K. R., Palli, D., Pechlivanis, S., Peltonen, L., Perry, J. R. B., Platou, C. G. P., Roden, M., Ruderfer, D., Rybin, D., van der Schouw, Y. T., Sennblad, B., Sigurðsson, G., Stančáková, A., Steinbach, G., Storm, P., Strauch, K., Stringham, H. M., Sun, Q., Thorand, B., Tikkanen, E., Tonjes, A., Trakalo, J., Tremoli, E., Tuomi, T., Wennauer, R., Wiltshire, S., Wood, A. R., Zeggini, E., Dunham, I., Birney, E., Pasquali, L., Ferrer, J., Loos, R. J. F., Dupuis, J., Florez, J. C., Boerwinkle, E., Pankow, J. S., van Duijn, C., Sijbrands, E., Meigs, J. B., Hu, F. B., Thorsteinsdottir, U., Stefansson, K., Lakka, T. A., Rauramaa, R., Stumvoll, M., Pedersen, N. L., Lind, L., Keinanen-Kiukaanniemi, S. M., Korpi-Hyövälti, E., Saaristo, T. E., Saltevo, J., Kuusisto, J., Laakso, M., Metspalu, A., Erbel, R., Jöcke, K.-H., Moebus, S., Ripatti, S., Salomaa, V., Ingelsson, E., Boehm, B. O., Bergman, R. N., Collins, F. S., Mohlke, K. L., Koistinen, H., Tuomilehto, J., Hveem, K., Njølstad, I., Deloukas, P., Donnelly, P. J., Frayling, T. M., Hattersley, A. T., de Faire, U., Hamsten, A., Illig, T., Peters, A., Cauchi, S., Sladek, R., Froguel, P., Hansen, T., Pedersen, O., Morris, A. D., Palmer, C. N. A., Kathiresan, S., Melander, O., Nilsson, P. M., Groop, L. C., Barroso, I., Langenberg, C., Wareham, N. J., O'Callaghan, C. A., Gloyn, A. L., Altshuler, D., Boehnke, M., Teslovich, T. M., McCarthy, M. I., Morris, A. P. & DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat. Genet.* **47,** 1415–1425 (2015).

25. Noble, J. A. & Valdes, A. M. Genetics of the HLA Region in the Prediction of Type 1 Diabetes. *Current Diabetes Reports* **11,** 533–542 (2011).

26. Dupont, C., Armant, D. & Brenner, C. Epigenetics: Definition, Mechanisms and Clinical Perspective. *Seminars in Reproductive Medicine* **27,** 351–357 (2009).

27. Richmond, T. J. & Davey, C. A. The structure of DNA in the nucleosome core. *Nature* **423,** 145–150 (2003).

28. Radman-Livaja, M. & Rando, O. J. Nucleosome positioning: How is it established, and why does it matter? *Developmental Biology* **339,** 258–266 (2010).

29. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., Amin, V., Whitaker, J. W., Schultz, M. D., Ward, L. D., Sarkar, A., Quon, G., Sandstrom, R. S., Eaton, M. L., Wu, Y.-C., Pfenning, A., Wang, X., ClaussnitzerYaping Liu, M., Coarfa, C., Alan Harris, R., Shoresh, N., Epstein, C. B., Gjoneska, E., Leung, D., Xie, W., David Hawkins, R., Lister, R., Hong, C., Gascard, P., Mungall, A. J., Moore, R., Chuah, E., Tam, A., Canfield, T. K., Scott Hansen, R., Kaul, R., Sabo, P. J., Bansal, M. S., Carles, A., Dixon, J. R., Farh, K.-H., Feizi, S., Karlic, R., Kim, A.-R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T. R., Neph, S. J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R. C., Siebenthall, K. T., Sinnott-Armstrong, N. A., Stevens, M., Thurman, R. E., Wu, J., Zhang, B., Zhou, X., Abdennur, N., Adli, M., Akerman, M., Barrera, L., Antosiewicz-Bourget, J., Ballinger, T., Barnes, M. J., Bates, D., Bell, R. J. A., Bennett, D. A., Bianco, K., Bock, C., Boyle, P., Brinchmann, J., Caballero-Campo, P., Camahort, R., Carrasco-Alfonso, M. J., Charnecki, T., Chen, H., Chen, Z., Cheng, J. B., Cho, S., Chu, A., Chung, W.-Y., Cowan, C., Athena Deng, Q., Deshpande, V., Diegel, M., Ding, B., Durham, T., Echipare, L., Edsall, L., Flowers, D., Genbacev-Krtolica, O., Gifford, C., Gillespie, S., Giste, E., Glass, I. A., Gnirke, A., Gormley, M., Gu, H., Gu, J., Hafler, D. A., Hangauer, M. J., Hariharan, M., Hatan, M., Haugen, E., He, Y., Heimfeld, S., Herlofsen, S., Hou, Z., Humbert, R., Issner, R., Jackson, A. R., Jia, H., Jiang, P., Johnson, A. K., Kadlecek, T., Kamoh, B., Kapidzic, M., Kent, J., Kim, A., Kleinewietfeld, M., Klugman, S., Krishnan, J., Kuan, S., Kutyavin, T., Lee, A.-Y., Lee, K., Li, J., Li, N., Li, Y., Ligon, K. L., Lin, S., Lin, Y., Liu, J., Liu, Y., Luckey, C. J., Ma, Y. P., Maire, C., Marson, A., Mattick, J. S., Mayo, M., McMaster, M., Metsky, H., Mikkelsen, T., Miller, D., Miri, M., Mukame, E., Nagarajan, R. P., Neri, F., Nery, J., Nguyen, T., O'Geen, H., Paithankar, S., Papayannopoulou, T., Pelizzola, M., Plettner, P., Propson, N. E., Raghuraman, S., Raney, B. J., Raubitschek, A., Reynolds, A. P., Richards, H., Riehle, K., Rinaudo, P., Robinson, J. F., Rockweiler, N. B., Rosen, E., Rynes, E., Schein, J., Sears, R., Sejnowski, T., Shafer, A., Shen, L., Shoemaker, R., Sigaroudinia, M., Slukvin, I., Stehling-Sun, S., Stewart, R., Subramanian, S. L., Suknuntha, K., Swanson, S., Tian, S., Tilden, H., Tsai, L., Urich, M., Vaughn, I., Vierstra, J., Vong, S., Wagner, U., Wang, H., Wang, T., Wang, Y., Weiss, A., Whitton, H., Wildberg, A., Witt, H., Won, K.-J., Xie, M., Xing, X., Xu, I., Xuan, Z., Ye, Z., Yen, C., Yu, P., Zhang, X., Zhang, X., Zhao, J., Zhou, Y., Zhu, J., Zhu, Y., Ziegler, S., Beaudet, A. E., Boyer, L. A., De Jager, P. L., Farnham, P. J., Fisher, S. J., Haussler, D., Jones, S. J. M., Li, W., Marra, M. A., McManus, M. T., Sunyaev, S., Thomson, J. A., Tlsty, T. D., Tsai, L.-H., Wang, W., Waterland, R. A., Zhang, M. Q., Chadwick, L. H., Bernstein, B. E., Costello, J. F., Ecker, J. R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B.,

Stamatoyannopoulos, J. A., Wang, T., Kellis, M., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., Amin, V., Whitaker, J. W., Schultz, M. D., Ward, L. D., Sarkar, A., Quon, G., Sandstrom, R. S., Eaton, M. L., Wu, Y.-C., Pfenning, A. R., Wang, X., Claussnitzer, M., Liu, Y., Coarfa, C., Harris, R. A., Shoresh, N., Epstein, C. B., Gjoneska, E., Leung, D., Xie, W., Hawkins, R. D., Lister, R., Hong, C., Gascard, P., Mungall, A. J., Moore, R., Chuah, E., Tam, A., Canfield, T. K., Hansen, R. S., Kaul, R., Sabo, P. J., Bansal, M. S., Carles, A., Dixon, J. R., Farh, K.-H., Feizi, S., Karlic, R., Kim, A.-R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T. R., Neph, S. J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R. C., Siebenthall, K. T., Sinnott-Armstrong, N. A., Stevens, M., Thurman, R. E., Wu, J., Zhang, B., Zhou, X., Beaudet, A. E., Boyer, L. A., De Jager, P. L., Farnham, P. J., Fisher, S. J., Haussler, D., Jones, S. J. M., Li, W., Marra, M. A., McManus, M. T., Sunyaev, S., Thomson, J. A., Tlsty, T. D., Tsai, L.-H., Wang, W., Waterland, R. A., Zhang, M. Q., Chadwick, L. H., Bernstein, B. E., Costello, J. F., Ecker, J. R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J. A., Wang, T. & Kellis, M. Integrative analysis of 111 reference human epigenomes. *Nature* **518,** 317–330 (2015).

30. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-Wide Mapping of in Vivo Protein-DNA Interactions. *Science* **316,** 1497–1502 (2007).

31. O'Geen, H., Echipare, L. & Farnham, P. J. in *Epigenetics Protocols* (ed. Tollefsbol, T. O.) **791,** 265–286 (Humana Press, 2011).

32. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nature Protocols* **12,** 2478–2492 (2017).

33. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10,** 57–63 (2009).

34. Sun, W. & Hu, Y. eQTL Mapping Using RNA-seq Data. *Statistics in Biosciences* **5,** 198–219 (2013).

35. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods* **10,** 1213–1218 (2013).

36. Brem, R. B. Genetic Dissection of Transcriptional Regulation in Budding Yeast. *Science* **296,** 752–755 (2002).

37. Nica, A. C. & Dermitzakis, E. T. Expression quantitative trait loci: present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368,** 20120362–20120362 (2013).

38. The GTEx Consortium, Ardlie, K. G., Deluca, D. S., Segre, A. V., Sullivan, T. J., Young, T. R., Gelfand, E. T., Trowbridge, C. A., Maller, J. B., Tukiainen, T., Lek, M., Ward, L. D., Kheradpour, P., Iriarte, B., Meng, Y., Palmer, C. D., Esko, T., Winckler, W., Hirschhorn, J. N., Kellis, M., MacArthur, D. G., Getz, G., Shabalin, A. A., Li, G., Zhou, Y.-H., Nobel, A. B., Rusyn, I., Wright, F. A., Lappalainen, T., Ferreira, P. G., Ongen, H., Rivas, M. A., Battle,

A., Mostafavi, S., Monlong, J., Sammeth, M., Mele, M., Reverter, F., Goldmann, J. M., Koller, D., Guigo, R., McCarthy, M. I., Dermitzakis, E. T., Gamazon, E. R., Im, H. K., Konkashbaev, A., Nicolae, D. L., Cox, N. J., Flutre, T., Wen, X., Stephens, M., Pritchard, J. K., Tu, Z., Zhang, B., Huang, T., Long, Q., Lin, L., Yang, J., Zhu, J., Liu, J., Brown, A., Mestichelli, B., Tidwell, D., Lo, E., Salvatore, M., Shad, S., Thomas, J. A., Lonsdale, J. T., Moser, M. T., Gillard, B. M., Karasik, E., Ramsey, K., Choi, C., Foster, B. A., Syron, J., Fleming, J., Magazine, H., Hasz, R., Walters, G. D., Bridge, J. P., Miklos, M., Sullivan, S., Barker, L. K., Traino, H. M., Mosavel, M., Siminoff, L. A., Valley, D. R., Rohrer, D. C., Jewell, S. D., Branton, P. A., Sobin, L. H., Barcus, M., Qi, L., McLean, J., Hariharan, P., Um, K. S., Wu, S., Tabor, D., Shive, C., Smith, A. M., Buia, S. A., Undale, A. H., Robinson, K. L., Roche, N., Valentino, K. M., Britton, A., Burges, R., Bradbury, D., Hambright, K. W., Seleski, J., Korzeniewski, G. E., Erickson, K., Marcus, Y., Tejada, J., Taherian, M., Lu, C., Basile, M., Mash, D. C., Volpi, S., Struewing, J. P., Temple, G. F., Boyer, J., Colantuoni, D., Little, R., Koester, S., Carithers, L. J., Moore, H. M., Guan, P., Compton, C., Sawyer, S. J., Demchok, J. P., Vaught, J. B., Rabiner, C. A., Lockhart, N. C., Ardlie, K. G., Getz, G., Wright, F. A., Kellis, M., Volpi, S. & Dermitzakis, E. T. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348,** 648–660 (2015).

39. Battle, A., Mostafavi, S., Zhu, X., Potash, J. B., Weissman, M. M., McCormick, C., Haudenschild, C. D., Beckman, K. B., Shi, J., Mei, R., Urban, A. E., Montgomery, S. B., Levinson, D. F. & Koller, D. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Research* **24,** 14–24 (2014).

40. Dermitzakis, E. T. From gene expression to disease risk. *Nature Genetics* **40,** 492–493 (2008).

41. Giambartolomei, C., Vukcevic, D., Schadt, E. E., Franke, L., Hingorani, A. D., Wallace, C. & Plagnol, V. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genetics* **10,** e1004383 (2014).

42. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521,** 436–444 (2015).

43. Zhou, J. & Troyanskaya, O. G. Predicting effects of noncoding variants with deep learning–based sequence model. *Nature Methods* **12,** 931–934 (2015).

44. Edwards, S. L., Beesley, J., French, J. D. & Dunning, A. M. Beyond GWASs: Illuminating the Dark Road from Association to Function. *The American Journal of Human Genetics* **93,** 779–797 (2013).

45. Garner, M. M. & Revzin, A. The use of gel electrophoresis to detect and study nucleic acid— protein interactions. *Trends in Biochemical Sciences* **11,** 395–396 (1986).

46. Scott, R. A., Scott, L. J., Mägi, R., Marullo, L., Gaulton, K. J., Kaakinen, M., Pervjakova, N., Pers, T. H., Johnson, A. D., Eicher, J. D., Jackson, A. U., Ferreira, T., Lee, Y., Ma, C., Steinthorsdottir, V., Thorleifsson, G., Qi, L., Van Zuydam, N. R., Mahajan, A., Chen, H., Almgren, P., Voight, B. F., Grallert, H., Müller-Nurasyid, M., Ried, J. S., Rayner, W. N., Robertson, N., Karssen, L. C., van Leeuwen, E. M., Willems, S. M., Fuchsberger, C., Kwan, P., Teslovich, T. M., Chanda, P., Li, M., Lu, Y., Dina, C., Thuillier, D., Yengo, L., Jiang, L.,

Sparso, T., Kestler, H. A., Chheda, H., Eisele, L., Gustafsson, S., Frånberg, M., Strawbridge, R. J., Benediktsson, R., Hreidarsson, A. B., Kong, A., Sigurðsson, G., Kerrison, N. D., Luan, J., Liang, L., Meitinger, T., Roden, M., Thorand, B., Esko, T., Mihailov, E., Fox, C., Liu, C.-T., Rybin, D., Isomaa, B., Lyssenko, V., Tuomi, T., Couper, D. J., Pankow, J. S., Grarup, N., Have, C. T., Jørgensen, M. E., Jørgensen, T., Linneberg, A., Cornelis, M. C., van Dam, R. M., Hunter, D. J., Kraft, P., Sun, Q., Edkins, S., Owen, K. R., Perry, J. R., Wood, A. R., Zeggini, E., Tajes-Fernandes, J., Abecasis, G. R., Bonnycastle, L. L., Chines, P. S., Stringham, H. M., Koistinen, H. A., Kinnunen, L., Sennblad, B., Mühleisen, T. W., Nöthen, M. M., Pechlivanis, S., Baldassarre, D., Gertow, K., Humphries, S. E., Tremoli, E., Klopp, N., Meyer, J., Steinbach, G., Wennauer, R., Eriksson, J. G., Männistö, S., Peltonen, L., Tikkanen, E., Charpentier, G., Eury, E., Lobbens, S., Gigante, B., Leander, K., McLeod, O., Bottinger, E. P., Gottesman, O., Ruderfer, D., Blüher, M., Kovacs, P., Tonjes, A., Maruthur, N. M., Scapoli, C., Erbel, R., Jöckel, K.-H., Moebus, S., de Faire, U., Hamsten, A., Stumvoll, M., Deloukas, P., Donnelly, P. J., Frayling, T. M., Hattersley, A. T., Ripatti, S., Salomaa, V., Pedersen, N. L., Boehm, B. O., Bergman, R. N., Collins, F. S., Mohlke, K. L., Tuomilehto, J., Hansen, T., Pedersen, O., Barroso, I., Lannfelt, L., Ingelsson, E., Lind, L., Lindgren, C. M., Cauchi, S., Froguel, P., Loos, R. J., Balkau, B., Boeing, H., Franks, P. W., Gurrea, A. B., Palli, D., van der Schouw, Y. T., Altshuler, D., Groop, L. C., Langenberg, C., Wareham, N. J., Sijbrands, E., van Duijn, C. M., Florez, J. C., Meigs, J. B., Boerwinkle, E., Gieger, C., Strauch, K., Metspalu, A., Morris, A. D., Palmer, C. N., Hu, F. B., Thorsteinsdottir, U., Stefansson, K., Dupuis, J., Morris, A. P., Boehnke, M., McCarthy, M. I., Prokopenko, I. & DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes* (2017). doi:10.2337/db16-1253

47. Wakefield, J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet. Epidemiol.* **33,** 79–86 (2009).

48. Morris, A. P. Transethnic Meta-Analysis of Genomewide Association Studies. *Genet Epidemiol* **35,** 809–822 (2011).

49. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods* **9,** 215–216 (2012).

50. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).

51. Pickrell, J. K. Joint Analysis of Functional Genomic Data and Genome-wide Association Studies of 18 Human Traits. *The American Journal of Human Genetics* **94,** 559–573 (2014).

52. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, The Multiple Tissue Human Expression Resource (MUTHER) Consortium, Manning, A. K., Hivert, M.-F., Scott, R. A., Grimsby, J. L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C.-T., Bielak, L. F., Prokopenko, I., Amin, N., Barnes, D., Cadby, G., Hottenga, J.-J., Ingelsson, E., Jackson, A. U., Johnson, T., Kanoni, S., Ladenvall, C., Lagou, V., Lahti, J., Lecoeur, C., Liu, Y., Martinez-Larrad, M. T., Montasser, M. E., Navarro, P., Perry, J. R. B., Rasmussen-Torvik, L. J., Salo, P., Sattar, N., Shungin, D., Strawbridge, R. J., Tanaka, T., van Duijn, C. M., An,

P., de Andrade, M., Andrews, J. S., Aspelund, T., Atalay, M., Aulchenko, Y., Balkau, B., Bandinelli, S., Beckmann, J. S., Beilby, J. P., Bellis, C., Bergman, R. N., Blangero, J., Boban, M., Boehnke, M., Boerwinkle, E., Bonnycastle, L. L., Boomsma, D. I., Borecki, I. B., Böttcher, Y., Bouchard, C., Brunner, E., Budimir, D., Campbell, H., Carlson, O., Chines, P. S., Clarke, R., Collins, F. S., Corbatón-Anchuelo, A., Couper, D., de Faire, U., Dedoussis, G. V., Deloukas, P., Dimitriou, M., Egan, J. M., Eiriksdottir, G., Erdos, M. R., Eriksson, J. G., Eury, E., Ferrucci, L., Ford, I., Forouhi, N. G., Fox, C. S., Franzosi, M. G., Franks, P. W., Frayling, T. M., Froguel, P., Galan, P., de Geus, E., Gigante, B., Glazer, N. L., Goel, A., Groop, L., Gudnason, V., Hallmans, G., Hamsten, A., Hansson, O., Harris, T. B., Hayward, C., Heath, S., Hercberg, S., Hicks, A. A., Hingorani, A., Hofman, A., Hui, J., Hung, J., Jarvelin, M.-R., Jhun, M. A., Johnson, P. C. D., Jukema, J. W., Jula, A., Kao, W. H., Kaprio, J., Kardia, S. L. R., Keinanen-Kiukaanniemi, S., Kivimaki, M., Kolcic, I., Kovacs, P., Kumari, M., Kuusisto, J., Kyvik, K. O., Laakso, M., Lakka, T., Lannfelt, L., Lathrop, G. M., Launer, L. J., Leander, K., Li, G., Lind, L., Lindstrom, J., Lobbens, S., Loos, R. J. F., Luan, J., Lyssenko, V., Mägi, R., Magnusson, P. K. E., Marmot, M., Meneton, P., Mohlke, K. L., Mooser, V., Morken, M. A., Miljkovic, I., Narisu, N., O'Connell, J., Ong, K. K., Oostra, B. A., Palmer, L. J., Palotie, A., Pankow, J. S., Peden, J. F., Pedersen, N. L., Pehlic, M., Peltonen, L., Penninx, B., Pericic, M., Perola, M., Perusse, L., Peyser, P. A., Polasek, O., Pramstaller, P. P., Province, M. A., Räikkönen, K., Rauramaa, R., Rehnberg, E., Rice, K., Rotter, J. I., Rudan, I., Ruokonen, A., Saaristo, T., Sabater-Lleal, M., Salomaa, V., Savage, D. B., Saxena, R., Schwarz, P., Seedorf, U., Sennblad, B., Serrano-Rios, M., Shuldiner, A. R., Sijbrands, E. J. G., Siscovick, D. S., Smit, J. H., Small, K. S., Smith, N. L., Smith, A. V., Stančáková, A., Stirrups, K., Stumvoll, M., Sun, Y. V., Swift, A. J., Tönjes, A., Tuomilehto, J., Trompet, S., Uitterlinden, A. G., Uusitupa, M., Vikström, M., Vitart, V., Vohl, M.-C., Voight, B. F., Vollenweider, P., Waeber, G., Waterworth, D. M., Watkins, H., Wheeler, E., Widen, E., Wild, S. H., Willems, S. M., Willemsen, G., Wilson, J. F., Witteman, J. C. M., Wright, A. F., Yaghootkar, H., Zelenika, D., Zemunik, T., Zgaga, L., Wareham, N. J., McCarthy, M. I., Barroso, I., Watanabe, R. M., Florez, J. C., Dupuis, J., Meigs, J. B. & Langenberg, C. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nature Genetics* **44,** 659–669 (2012).

53. Finucane, H. K., Bulik-Sullivan, B., ReproGen Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium, The RACI Consortium, Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., Ripke, S., Day, F. R., Purcell, S., Stahl, E., Lindstrom, S., Perry, J. R. B., Okada, Y., Raychaudhuri, S., Daly, M. J., Patterson, N., Neale, B. M. & Price, A. L. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics* **47,** 1228–1235 (2015).

54. Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W. & Liu, X. S. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9,** R137 (2008).

55. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A.,

Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.-F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H.-C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G. R., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F. A., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S.-P., Yeh, R.-F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Patrinos, A. & Morgan, M. J. Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).

56. Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J. A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S. R., Tan, G., Baranasic, D., Arenillas, D. J., Sandelin, A., Vandepoele, K., Lenhard, B., Ballester, B., Wasserman, W. W., Parcy, F. & Mathelier, A. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Research* **46,** D260–D266 (2018).

57. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).

58. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27,** 1017–1018 (2011).

60. Parker, S. C. J., Stitzel, M. L., Taylor, D. L., Orozco, J. M., Erdos, M. R., Akiyama, J. A., van Bueren, K. L., Chines, P. S., Narisu, N., NISC Comparative Sequencing Program, Black, B. L., Visel, A., Pennacchio, L. A., Collins, F. S., National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program Authors, NISC Comparative Sequencing Program Authors., Becker, J., Benjamin, B., Blakesley, R., Bouffard, G., Brooks, S., Coleman, H., Dekhtyar, M., Gregory, M., Guan, X., Gupta, J., Han, J., Hargrove, A., Ho, S. -l., Johnson, T., Legaspi, R., Lovett, S., Maduro, Q., Masiello, C., Maskeri, B., McDowell, J., Montemayor, C., Mullikin, J., Park, M., Riebow, N., Schandler, K., Schmidt, B., Sison, C., Stantripop, M., Thomas, J., Thomas, P., Vemulapalli, M. & Young, A. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proceedings of the National Academy of Sciences* **110,** 17921–17926 (2013).

61. Saeed, S., Bonnefond, A., Tamanini, F., Mirza, M. U., Manzoor, J., Janjua, Q. M., Din, S. M., Gaitan, J., Milochau, A., Durand, E., Vaillant, E., Haseeb, A., De Graeve, F., Rabearivelo, I., Sand, O., Queniat, G., Boutry, R., Schott, D. A., Ayesha, H., Ali, M., Khan, W. I., Butt, T. A., Rinne, T., Stumpel, C., Abderrahmani, A., Lang, J., Arslan, M. & Froguel, P. Loss-of-function mutations in ADCY3 cause monogenic severe obesity. *Nature Genetics* **50,** 175–179 (2018).

62. Dorssers, L. C. J. Induction of Antiestrogen Resistance in Human Breast Cancer Cells by Random Insertional Mutagenesis Using Defective Retroviruses: Identification of bear-I, a Common Integration Site. 9 (2018).

63. Harder, M. N., Ribel-Madsen, R., Justesen, J. M., Sparsø, T., Andersson, E. A., Grarup, N., Jørgensen, T., Linneberg, A., Hansen, T. & Pedersen, O. Type 2 Diabetes Risk Alleles Near *BCAR1* and in *ANK1* Associate With Decreased β-Cell Function Whereas Risk Alleles Near *ANKRD55* and *GRB14* Associate With Decreased Insulin Sensitivity in the Danish Inter99 Cohort. *The Journal of Clinical Endocrinology & Metabolism* **98,** E801–E806 (2013).

64. Bressler, N. M., Bressler, S. B. & Fine, S. L. Age-related macular degeneration. *Survey of Ophthalmology* **32,** 375–413 (1988).

65. Emery, B. Regulation of Oligodendrocyte Differentiation and Myelination. *Science* **330,** 779–782 (2010).

66. Chen, Y., Mei, R., Teng, P., Yang, A., Hu, X., Zhang, Z., Qiu, M. & Zhao, X. TAPP1 inhibits the differentiation of oligodendrocyte precursor cells via suppressing the Mek/Erk pathway. *Neuroscience Bulletin* **31,** 517–526 (2015).

67. Keung, A. J., Dong, M., Schaffer, D. V. & Kumar, S. Pan-neuronal maturation but not neuronal subtype differentiation of adult neural stem cells is mechanosensitive. *Scientific Reports* **3,** (2013).

68. Preissl, S., Fang, R., Huang, H., Zhao, Y., Raviram, R., Gorkin, D. U., Zhang, Y., Sos, B. C., Afzal, V., Dickel, D. E., Kuan, S., Visel, A., Pennacchio, L. A., Zhang, K. & Ren, B. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-

specific transcriptional regulation. *Nature Neuroscience* (2018). doi:10.1038/s41593-018-0079-3