Comprehensive resource and integrative model for functional genomics of the adult brain (~4464 w/o figure captions)

Abstract (214)

Understanding how a variety of genomic variants influences brain phenotypes and disorders remains a key challenge. To this end, the PsychENCODE consortium has generated large-scale sequencing datasets on the adult brain at the population level over 2000 people with various phenotypes and diseases, including genotype, transcriptomics, chromatin and single-cell. By uniformly processing and consistently analyzing the datasets, we developed a comprehensive resource on functional genomics of adult brain. We deconvolved the bulk tissue expression using single cell data and found the cell fraction changes can explain >85% tissue covariance and also associate with psychiatric phenotypes. By comparing others, we found that the brain differs from other tissues more notably in gene expression than chromatin actives. We used the chromatin and HiC data for reference brain samples to generate ~80k active enhancers on adult prefrontal cortex, and link them to genes, transcription factors and identified ~2M eQTLs and xx cQTLs to build a full gene regulatory network, which can discover 112 more genes associated with GWAS variants with psychiatric disorders. Finally, we developed a deep-learning model embedding our regulatory network to predict genotype-phenotype associations, with ~3.5X accuracy improvement over previous approaches. This model enables highlighting intermediate genes and functional modules implying potential mechanisms, and also imputation of missing transcriptome and epigenome from genotype data only.

Introduction (423)

Disorders of the brain affect nearly a fifth of the world's population \cite{19507169}. Decades of research has led to little progress in our fundamental understanding of the molecular causes of psychiatric disorders. This contrasts with cardiac disease, for which lifestyle and pharmacological modification of environmental risk factors has had a profound effect on disease morbidity , or cancer, which is now understood to be a direct disorder of the genome \cite{9603539, 24071849, JK}. Though GWAS studies have identified many genomic variants associated with psychiatric disease risk, a detailed understanding of the precise molecular mechanisms behind these associations remains elusive \cite{19339359, JK}.

To this end, a number of genomic studies have recently focused on discovering genomic variants relating to psychiatric diseases. As such, a variety of common and rare variants and linked genes have been found to be associated with brain and psychiatric disorders \cite{26404826}. For instance, the Psychiatric Genomics Consortium (PGC) identified 142 GWAS loci associated with schizophrenia \cite{29483656}. Many variants for psychiatric disorders have been found to lie in non-coding regions \cite{26404826}, suggesting potential roles in gene regulation. Other consortia have annotated many of these non-coding elements -- e.g., expression quantitative trait loci (eQTLs) and eGenes in GTEx, and enhancers from ENCODE and Epigenomics Roadmap that are associated with various cells and tissues. Though some of these elements relate to the brain, none of these groups have specifically tailored their efforts toward comprehensively identifying functional elements in the brain. Pioneering work from the CommonMind consortium \cite{27668389} has attempted to do this but has not reached maximal scale. Moreover, a number of new technologies have recently been developed to better measure aspects of gene regulation. For example, HiC studies have

been used to identify specific chromatin structural and regulatory elements, such as active enhancers in the fetal brain \cite{ 27760116}. Single-cell sequencing techniques also offer great promise for studying the transcriptome \cite{27339989, 26060301}. However, each of the studies that leverage such technologies have generally focused on individual aspects of brain functional genomics. Therefore, these data have not yet been fully integrated at scale.

To this end, we have built a central, publically available resource for adult brain functional genomics, including all the raw and uniformly processed data at both tissue and single cell levels from PsychENCODE and other related major projects \cite{27339989, 26060301} with nearly ~12k data samples in total. By leveraging this resource, our analyses identified various functional genomic elements and QTLs specific to the adult brain, including novel psychiatric GWAS and gene linkages. We also combined these elements and built an integrated deep-learning model for finding potential genotype to phenotype connections in this data resource.

Comprehensive resource construction (205)

We designed this resource to provide a coherent and comprehensive structure to the data (adult.psychencode.org, details in supplement). Broadly, it organizes a large amount of data for brain functional genomics pyramidally, with a large base of raw data files (much of it restricted-access, such as individual genotyping and raw next-generation sequencing data of transcriptomics and epigenomics), a middle layer of uniformly processed and easily shareable results (such as open chromatin "peaks" and gene expression quantifications), and a compact cap at the top, consisting of an integrative model (based on imputed regulatory networks and QTLs). As shown in Figure 1, to build the base layer, we included all the datasets from PsychENCODE related to the adult and merged these with other relevant data from ENCODE, CommonMind, GTEx, Epigenomics Roadmap, and recent brain single cell studies. In total, this resource constitutes up to ~5.5k data samples derived from 1931 individual adult brains from multiple cohorts of PsychENCODE and ~5k brain related samples from other consortia, which covers a large representation of brain phenotypes and psychiatric disorders. Furthermore, the PsychENCODE project developed a specific "reference brain" project utilizing many assays on the same set of brain tissues, which we used to develop an anchoring annotation for the entire resource (Supplement).

Bulk & single-cell transcriptome analysis (701)

To identify the genomic elements that exhibit transcriptional activities specific to the adult brain, we used the ENCODE pipeline to uniformly process the RNA-seq data of all available samples from PsychENCODE and GTEx. Using these data, we identified interpretable functional elements, such as non-coding regions of transcription and sets of differentially expressed and co-expressed genes, which are provided as part of our resource (Supplement). In particular, the co-expressed genes are summarized as a list of modules for psychiatric disorders \cite{cap1} and for brain regions after clustering together with other tissues (Supplement).

Brain tissues have been found to comprise a variety of cell types, including neuronal and non-neuronal cells. Previous studies have suggested that gene expression changes at the tissue level can be associated with changing proportions of cell types \cite{21614001, 29439242, 18849986, 27409810}. However, they have not systematically revealed how different cell types can quantitatively contribute to the tissue-expression variation. Here, we address the question whether expression changes over our population of 1931 individuals are driven by gene expression changes in a specific cell type or whether they result from changing proportions of various cell-types.

We used two complementary strategies. First, we used the standard pipeline to uniformly process single cell RNA-seq data in PsychENCODE, in conjunction with a number of other single-cell studies on the brain, in order to assemble a list of basic cell types for PsychENCODE in the brain. [[{Q for Dan & Jim}]] This list includes 16 previously identified neuronal types, 5 major non-neuronal types and a number of additional types in brain development (Supplement), and constitutes a matrix (C) of expression signatures, which are mostly concordant with what has been published (Figure S5 and Discussion). Across these, we found that a number of genes whose expression levels vary more substantially than they do across individuals in a population (e.g., the dopamine receptor gene DRD3, Figure 2A). This implies that small variations in cell types can readily give rise to substantial changes in bulk gene expression at the tissue level.

To explore this further, we performed an unsupervised analysis for bulk tissue expression data to identify the primary components as they relate to different single cell types. In particular, we decomposed the bulk gene expression matrix (B) from our resource using non-negative matrix factorization (NMF, Figure 2B, see Methods); i.e., $B \sim = VH$, and then determined whether the top components (TCs) of the NMF (ie, NMF-TCs) that capture the majority of covariance are consistent with the single cell signatures. As shown in Figure 2B, we found a number of NMF-TCs highly correlated with neuronal, non-neuronal and developmental-related cell types (Figure 2C). This demonstrates that an unsupervised analysis derived from bulk data roughly matches the single cell data, partially corroborating our basic cell types.

We then want to understand how the differing cell types contribute to variation in bulk gene expression and relate to different brain phenotypes. In particular, as shown in Figure 2B, we de-convolved the expression matrix of tissue, B using the single cell signatures (C) to estimate the cell fractions W, solving the equation "$B \sim = W*C$" (See methods). As a validation, we found that our estimated fractions of NEU+/- cells match the experimentally determined fractions for the reference brain samples (Median error = 0.04 , Figure S6). Moreover, we found that using the single-cell expression signatures can explain much of the population-level expression variation (i.e., across tissue samples of the same brain region taken from different individuals); i.e., $1-||B-WC||_{2}/||B||_{2} > 85\%$ (Methods).

Furthermore, we found cell fraction changes were associated with different phenotypes and psychiatric disorders (Figure 2E, SXXX). For example, particular excitatory and inhibitory neurons exhibit significantly different fractions between male and female samples (Ex3 and In6). The fraction of Ex3 is also significantly reduced in ASD samples ($p = 2.73e-11$, ANOVA test), while non-neuronal cells (e.g., oligodendrocytes) are represented in greater abundance. Another interesting association was that cell fractions change with age. In particular, the fractions of neuronal types Ex3 and Ex4 significantly increase with age, but some non-neuronal types such as oligodendrocyte are found to decrease. Furthermore, these age-related cell changes are potentially associated with differentially expressed genes across age groups – e.g., the gene involved in early growth response is down-regulated in older age groups, whereas ceruloplasmin is down-regulated among middle-aged groups (Figure 2E).

Enhancers in adult brain (417)

We use chromatin modification signals to identify active enhancers in the brain. The reference brain is valuable here as it allows integration of many assays, including ATAC-seq, ChIP-seq and Hi-C. Combining the data from the reference brain with that DNase and ChIP-seq of the same brain region from ENCODE and Roadmap Epigenomics, we identified a set of ~80k brain

enhancers, active in PFC, consistent with that approach in ENCODE - i.e. enriched in H3K27ac and depleted in H3K4me3 [[{check}]] (Figure 3A).[[{see meth?}]]

Assessing the variability enhancers across individuals and tissues is more difficult than looking at gene expression variability. This is obvious from looking at figure XXX. As one can see across a population, the chromatin signal level associated with enhancers goes up and down. The boundaries of enhancers grow and shrink and sometimes disappear altogether -- or re-appear in different places in a few individuals. These 1D signal changes, of course, fundamentally reflect change in 3D chromatin structure.

To start measuring variability, we developed simple aggregated sets of K27 peaks across different tissues. We uniformly processed H3K27ac "peaks" (enriched regions) for PFC, temporal cortex (TC) and cerebellum (CBC) for a cohort of 50 individuals. We aggregated the peaks to generate a set of potential regulatory regions (see supplement), giving in total 37,761 H3K27ac in PFC, 42,683 in TC, and 26,631 in CBC. [[{ we need to work on this - why less than 80K}]][[{ move to meth}]] Each of these sets is in more than half of the samples of the corresponding brain region. The overlap of the pooled peaks of PFC with TC is ~90% but it is only 34% with CBC. [[{correct?}]]

Next, to explore the variability of enhancers across the population, we examined the overlap of the H3K27ac peaks in each of the individuals in the cohort. As expected, the active enhancers in the reference brain are not active in every individual in the cohort. In fact, on average ~55K out of 80K active enhancers in the reference brain are active in another individual (Figure 3B). As expected, there is a core set of enhancers ubiquitously active across the cohort. Indeed, a majority (~68%) of the enhancers identified in the reference brain are active in more than half of the population. On the other hand, some enhancers are more variable. The cumulative number of active enhancers increases dramatically for the first 20 individuals sampled, but becomes saturated at the 30th sample. We hypothesize that pooling together the identified prefrontal cortex enhancers from 20~30 individuals is enough to cover all potential regulatory enhancers in PFC, which we estimate to be around 120K.

Consistent comparison of transcriptomic and epigenomic activity (309)

A key aspect of our analysis is that we uniformly processed the transcriptomic and epigenomic data across PsychENCODE, ENCODE, GTEx and Roadmap. This allows us to compare the brain to other organs in a consistent fashion and also to compare the consistency of this over transcriptome and epigenome. We attempted several methods including PCA and tSNE for an appropriate comparison, and finally used Reference Component Analysis (RCA). PCA, though popular, tends to capture global structures, ignoring most of the local structure, but it can easily be influenced by outliers. On the other hand, t-SNE analysis preserves local structure but "shatters" global structure (see discussions in supplement). RCA, however is capable of capturing local structure while maintaining meaningful distances in global structure space. It projects the gene expression in an individual sample against a reference panel, and then essentially reduces dimensionality of the individual projections.

Our comparative analysis for gene expression shows that the brain tends to separate from the other tissues in the first component. Inter-tissue differences are much larger than intra-tissue ones. A different picture emerges when one looks at our comparison using chromatin data. It

shows that the chromatin levels at enhancers are much less distinguishable between brain and other tissues (Figure 3E).

Our RCA analysis focuses on inter-tissue differences in well-annotated regions (i.e. genes, promoters and enhancers). In addition to the expression differences in protein-coding genes, we also found transcriptional diversity across tissues in intergenic and noncoding regions. For protein-coding regions, it has previously been demonstrated that testes and lung tend to have the largest transcriptional diversity in terms of the percentage of transcribed regions of protein-coding genes (Figure 3F). However, when we shift to non-coding regions, we find that brain tissues (such as cortex and cerebellum) do, to some degree, stand out by exhibiting a greater extent of non-coding transcription than other tissues (Figure 3F).

QTL analysis (453)

We used the PsychENCODE resource data to identify quantitative trait loci (QTLs) affecting gene expression and chromatin activity. In particular, we calculated: expression QTLs (eQTLs), chromatin QTLs (cQTLs), splicing QTLs (sQTLs) and cell fraction QTLs. For eQTLs, we adopted a standard approach, adhering closely to the established GTEX eQTL pipeline. We identified 2,542,908 eQTLs (2,097,741 eQTLs after LD pruning) and 32944 e-genes (including non-coding ones) in DLPFC. There are 1,341,182 (25%) unique snps involved in the eQTLs from the 5,297,875 potential snps (within 1 Mb of gene TSS). This conservative estimate is a substantially larger number of eQTLs and eGenes than previous brain eQTL studies such as CommonMind and reflects a large sample size and statistical power in psychencode (Supplement). We believe the eQTL number is close to saturation, in terms of associating almost every variant with some expression modulating characteristic. We also applied the same QTL calculation pipeline to splicing and identified 157,592 sQTLs.

For the cQTLs, the situation is more complicated. There are no established standard methods for calculating these on a large scale, though previous efforts have detected QTLs associated with various chromatin activities on non-brain context \cite{25799442, 26300125} To properly identify them, we focused on a reference set of enhancers to define the region associated with chromatin activity and then looked at how this activity varies in these enhancers across 292 individuals (See methods). Overall, we were able to identify ~2000 cQTLs in addition to the 6200 cQTLs identified using individuals from CommonMind \cite{ https://doi.org/10.1101/141986}. [[{{why mention commonmind?}]]

Next, we were interested to see if any SNVs were associated with changes in the relative fractions of cell types across individuals; i.e., cell fraction QTLs (fQTLs). In total, we identified the 3720 distinct SNPs constituting 4186 different fQTLs between different cell types. Of these the fractions of microglial and a particular type of excitatory neuron (Ex8) were associated with the most variants. Next, after factoring out these cell type differences, we also identified 260,280 SNPs significantly associated with the gene expression changes across individual tissues - these "residual trans-eQTLs" represent SNP-expression associations unexplained by variation in cell types.

To further dissect the associations between genomic elements and the QTLs, we intersected our QTL lists with each other and also with a comprehensive set of genomic annotations. (Figure 4). For the overlap among different QTLs, we originally expected that many of the other QTLs would be a subset of the very large number of eQTLs but in fact an appreciable number of sQTLs, fQTLs and cQTLs were actually not overlapped with any eQTLs as shown in (Figure 4C). {{[[chjeck]]} As expected, we found that eQTLs tended to enriched on

promoter regions and cQTLs were mostly enriched on enhancer regions and fQTLs tended to occur broadly in many different elements (Figure 4D).

Regulatory networks (421)

In this section, we provided an integrative analysis at the gene regulatory level for the data and genomic elements described above. We created a putative regulatory network revealing how the genotype and regulators control target gene expression in adult brain. To this end, we first processed a Hi-C dataset for adult brain in the same reference samples used for enhancer identification, providing potential physical boundaries of interactions between enhancers and promoters (Figure 5A). \cite{27760116} (Supplement). In total, we identified 2,735 topologically associating domains (TADs) which set potential physical boundaries of enhancer-promoter interactions. This HiC dataset is substantially different than the fetal brain HiC data set highlighting the importance of the developmental stage in chromatin topology (see suppl).

As expected, we found that ~75% of enhancer-promoter interactions occur in the same TADs (Figures Sxx, 5B), suggesting that the majority of cis-regulation occurs within TADs. Also, as expected, the genes that have more potentially regulatory enhancers interacting with their promoters tend to express higher (Figure 5B). We next integrated the Hi-C dataset with eQTLs to assess how much of the common variation-associated gene regulation is mediated by chromatin interactions. Interestingly, 32% of eGenes show evidence of chromatin interactions, accounting for 239,837 eQTLs and 3,235 sQTLs (Figure 5C). To our surprise, enhancer e/sQTLs supported by Hi-C evidence showed stronger associations than exonic and promoter e/sQTLs.

Using the HiC TADs, we then tried to impute a full regulatory network, linking all possible regulatory elements such as enhancers, TFs, [[{weak}]] eQTLs and target genes. In particular, we first found all possible enhancer-promoter pairs in TADs and TFs having binding motifs on these enhancers and promoters (Supplement), providing a reference wiring network for at large potential TF-target regulatory linkages. Second, based on these "wiring" relationships, we used elastic-network regression to find the TFs-target relationships likely control target gene expression (ie associated with a high coefficient for predicting the target gene expression from TF activity). We model them as simple linear relationships but regularize to minimize the number of connections (Methods). Overall, we found this model could successfully predict expression of >4.8k genes with the minimum mean square errors < 0.05 (Figure Sxx). For example, the expression of schizophrenia risk gene, DGCR2 can be predicted by its TFs expression with MSE<0.03 based on our model.

Finally, to complete building gene regulatory network, we used the imputed TF-gene linkages to filter potential enhancer-target gene linkages, finding a robust set of enhancer-target gene linkages. In total, the network is summarized in Fig XXX. It is scale-free and has a hierarchical structure (Figure Sxx).

Associating GWAS variants with genes (285)

Next, we tried to use the regulatory network to link GWAS variants to potential target genes. First, we found significant associations between our eQTLs and GWAS associated with brain diseases. In particular, we calculated the overlap enrichment of our brain cis-eQTL SNPs and GWAS SNPs for schizophrenia, bipolar disorder and parkinson's disease and compared this to that for non-brain related disorders (CAD, asthma and type 2 diabetes). [[{correct ?}]] As

expected, enrichment for GWAS SNPs for brain disorders are more significant, with schizophrenia GWAS SNPs having the highest enrichment (Figure 4E).

To further predict the target genes of GWAS variants we exploited the combined Hi-C, eQTL-eGene and enhancer-gene linkages. For example, for the newly identified 142 schizophrenia GWAS loci \cite{27869829}, in total, we identified a set of 488 putative schizophrenia-associated genes, hereby referred as "SCZ genes", and 99 genes that show evidence both at the level of Hi-C and eQTLs, providing a high-confidence subset (Figure 5xx). This is a huge increase from the previously annotated 22 genes across 19 loci based on CMC brain eQTLs \cite{27869829, 27668389}. The majority of SCZ genes (288 genes[[{{488 or 288?]}]], ~59%) were not in linkage disequilibrium (LD, r_2>0.6) with index SNPs (Figure 5E), consistent with the previous observations that regulatory relationships often do not follow linear genome organization {{[[ref ?}]]}

We then looked at the characteristics of the 488 SCZ genes. As expected, these genes shared many of the characteristics of known schizophrenia-associated genes. In particular, they are enriched in genes intolerant to loss-of-function mutations \cite{27869829}, translational regulators, cholinergic receptors, calcium channels, and synaptic genes and genes known to be differentially expressed in schizophrenia (Figure Sxx). Next, we further integrated this gene list with the single-cell profiles and found, interestingly, they were enriched in a variety of different neural types.

Integrative deep-learning model (830)

The interaction between genotype and phenotype involves multiple levels; in this section, we perform another level of integrative analysis by embedding our regulatory network derived above into a larger multi-level model. For this purpose, we introduce an interpretable deep-learning framework, the Deep Structured Phenotype Network (***DSPN, Figure 6, Supplement Sec. 7). This model combines a Deep Boltzmann Machine architecture with conditional and lateral connections derived from our elastic net regression (which, in turn, incorporates the QTLs and Hi-C regulatory connections). As shown (Figure 6a), traditional classification methods such as logistic regression predict phenotype directly from genotype, without inferring intermediates such as the transcriptome. In contrast, DSPN is constructed via a series of intermediate models adding layers of structure, including a layer for intermediate molecular phenotypes such as gene expression and chromatin state, multiple higher layers for functional modules and other intermediate phenotypes[[{resusing this term}]] which may be inferred as hidden nodes in the network, and a top-level layer for observed phenotypes such as psychiatric disorders. Finally, we use special forms of connectivity (including sparsity and lateral intra-level connections) to integrate our knowledge of QTLs, regulatory network structure, and co-expression modules from earlier sections of the paper (Figure 6b). By using a generative architecture, we ensure that the model is able to impute intermediate phenotypes when needed, as well as providing a forward predictive model for observed phenotypes.

Using the full model with genome and transcriptome data provided, we show that adding the extra layers of structure in the DSPN allows us to achieve substantially better prediction of disease and other observed traits than without -- ie just using a traditional genotype to phenotype model (Figure 6d). In particular, we achieve a prediction of 73.6% with the extra layers for Schizophrenia vs 63% without (Figure 6d, DSPN vs LR-trans). Further, comparison with a simple logistic predictor from the genome alone shows that the transcriptome carries significant further trait relevant information, which the DSPN is able to extract. For instance, a logistic predictor is able to gain a 2.8X improvement when using the transcriptome versus the

genome for Schizophrenia prediction (+13% vs. +4.6% from 50% chance), while the DSPN is able to gain a 5X improvement (+23% vs. 4.6%); this may reflect the need to incorporate [[{rewrite & expl better}]] non-linear interactions between intermediate phenotypes at multiple layers as in the DSPN. Moreover, DSPN also, allows us to perform joint inference and imputation of the intermediate phenotypes (i.e. transcriptome and epigenome) and observed traits from the genotype alone. Using imputation, we achieved 63.2% accuracy for disease trait prediction averaged across conditions, which is better than direct prediction with a logistic model (53.8% or ~3.5X gain in accuracy above chance (+13.2% vs. +3.8%, see Figure 6d, DSPN-Imput vs LR-gen)[[{rewrite}]]. These results demonstrate the usefulness of even a limited amount of functional genomics information for unraveling gene-disease relationships and show that the structure learned from such data can be used to make more accurate predictions of high-level traits even when absent.

We transform the results above to the liability scale in order to compare with heritability (using GCTA, see methods). Using the PsychENCODE cohort, we estimate that common SNPs and eSNPs explain x% and x% of the liability variance for Schizophrenia respectively, comparable to previous estimates. The imputation-based DSPN model explains a comparable level of variance to eSNPs (4.5%, Fig. 6d). For the full DSPN model, the variance in liability explained by the transcriptome is ~32.8%.[[{no decimal for precision}]] Although we expect that a large portion of this overlaps with the common SNP based additive variance (which has previously been estimated as 25.6%) and genetically determined epistatic interactions, it may also include environmentally determined and trait-influenced contributions (Supp Fig. A1), meaning that it is an upper-bound on the genetically determined variance in liability modeled by DSPN. Similar estimates of the variance in liability modeled by the DSPN for Bipolar and ASD are given (imputation and full models) (Figure 6d).

We examined the specific connections learned by the DSPN between intermediate and high-level phenotypes for potentially relevant biological interactions. We included known co-expression modules in the model and examined which of these DSPN prioritized as well as new sets of genes associated with the DSPN latent nodes that were uncovered at each hidden layer using a multilevel prioritization scheme (Supplement Sec. 7.5 and Supp Fig. A2). Overall, we can show that the modules prioritized by the DSPN are strongly enriched for known GWAS variants (Supplement). We provide a full summary of the functional enrichment analysis for the prioritized modules associated in supplement (Supp section xx) and highlight some of the associations found using the Schizophrenia model (Figure 6c). Among the highest SCZ prioritized modules, we found modules associated with the glutamatergic-synapse and calcium-signaling pathways, with Astrocyte-marker genes, and with complement cascade pathway -- all of which confirms and extends previous analyses \cite{26814963}.[[{correct??}]] Further, we found enriched in the prioritized modules for aging the marker genes for Ex4 cell-type and the gene NRGN (in a module associated with synaptic and longevity functions), consistent with our differential expression analysis (Fig XXXX).

Discussion (206)

We integrated PsychENCODE datasets with other resources and developed a comprehensive resource of various functional genomic elements for the adult brain. Overall, our study has identified a large-scale set of eQTLs for adult brain, several folds more than previous studies, almost achieving saturation for protein coding genes. We suspect that larger population studies will not significantly expand on these. However, there exist other aspects of brain QTLs that can be extended in the future. For example, increasing individual sample size of chromatin data size

and quality such as identifying enhancers using STARR-seq will help discover more accurate cQTLs.

Another area of future development is single-cell analysis. In this study, we found that a set of basic and known cells could explain large expression variations across tissues. However, there still exist the gene expression heterogeneities among these cell types (Figure Sxx), implying subdivision into potential additional types in the future. Also, current techniques suffer from low capture efficiency, and so it remains challenging to reliably quantify low-abundant transcripts/genes \cite{26949524, 25053837}, and particularly have problems for some specific cell types such as axons and dendrites; e.g., very noisy expression measurements \cite{25053837}. Thus, we may see the novel techniques such as single-nucleus RNA sequencing techniques that emerge to address these problems.

X. Figures

Figure 1. Comprehensive data resource of functional genomics in adult brain. As shown on data cubic, this resource includes multi-dimensional functional genomic data for adult brain generated from PsychENCODE and other related consortia such as GTEx, Epigenomics Roadmap, ENCODE. The resource data has various types including genotypes (N=xx), RNA-seq (N=xx), ChIP-seq (N=xx), ATAC-seq (N=xx), Hi-C (N=xx), single-cell transcriptomic data (N=xx), and in total ~6k samples from 1931 individuals. These individuals represent major brain phenotypes, gender and psychiatric disorders: Health (n=1445), SCZ (n=270), BP (n=160), ASD (n=65), AFF (n=8), Male (n=1244), Female (n=700), Developmental stages, Age, etc. (details in Supplement). In addition, these datasets cover major brain regions relating to brain disorders: DLPFC (n=xxx), Temporal cortex (n=xx), Cerebellum (n=xx).

Figure 2. Deconvolution analysis of Bulk and single cell transcriptomics reveals cell fraction changes across tissue phenotypes and disorders. **(A)** The genes have significantly higher expression variability across single cells than tissue samples. Left: dopamine gene, DRD3; Right: Age gene, xxx. **(B)** Top: the bulk tissue gene expression matrix (B, genes by individuals) can be decomposed by non-negative matrix factorization (NMF) to the product of two matrices: NMF component matrix (V, genes by top NMF components; i.e., NMF-TCs) and component fraction matrix (H, top NMF components by individuals); i.e., B~=V*H. Bottom: the bulk tissue gene expression matrix B can be also deconvolved by the single cell gene expression matrix (C, genes by cell types) to estimate the cell fractions across individuals (the matrix, W); i.e., B~=C*W. Three major cell types are neuronal cells (blue), non-neuronal cells (red), developmental (dev) cells (green), as highlighted by columns groups in C (also row groups in W). **(C)** The heatmap shows the Pearson correlation coefficients of gene expression between the NMF TCs and single cell types for the biomarker genes (N=xxx). For example, NMF-15 highly correlates with Ex3 cell type (r=xxx). **(D)** The estimate cell fractions contribute >85% bulk tissue expression variations; i.e., 1-||B-C*W||/||B||>0.85. **(E)** The cell fractions change across brain phenotypes and disorders. The neuronal cell types (e.g., Ex3 and In6) have significantly higher fractions in Male than Female samples (p<xxx, and p<xxx). Also, they and Oligodendrocytes have lower fractions in ASD than others. In particular, Ex3 cell fractions increase with aging (trend analysis p<6.3e-10).

Figure 3. Comparative analysis for transcriptomics and epigenomics between brain and other tissues. **(A)** Chromatin features of the reference brain are used to identify active enhancers. Enhancers are located in open chromatin region (high ATAC-seq signal), with strong H3K27ac signal and lack H3K4me3 signal. Enhancer activity varies among individuals, as

indicated by the varying H3K27ac peak signal in the cohort. Each row corresponds to an individual, with shallow to dark blue indicating low to high signal peak value. **(B)** The number of brain active enhancers achieves a saturation to more than 70K with 20 samples. **(C)** The coefficients (PC1 vs. PC2) of RCA analysis for gene expression data of PsychENCODE samples (dark green), Other external brain samples (light green) including GTEx and Other tissue samples (magenta). **(D)** The center (cross) and ranges of different tissue clusters (dashed ellipse) on RCA scatterplot of (C). **(E)** The tissue clusters of RCA coefficients (PC1 vs. PC2) for chromatin data of active brain enhancers. The brain cluster (green ellipse) includes the PsychENCODE samples and other tissue clusters (magenta ellipses) consist of Epigeomics Roadmap samples. **(F)** The transcriptional diversity on coding (circle) and non-coding (triangle) regions among the tissue samples (inter-sample on x-axis) vs. on cumulative tissue samples (y-axis) for select major tissue types including Cerebellum (lightblue), Cortex (xxx), Lung (xxx), Ovary (xxx), Skin (xx) and Testes (xx).

Figure 4. Summary of QTLs of human adult brain DLPFC. (**A**) Example of H3K27ac signal of individual brains in a representative genomic region showing largely congruent identification of regions of open chromatin. Region in dashed frame represents a chromatin QTL, the signal magnitudes of individuals with G/G or G/T genotype are lower than the ones with T/T genotype. (**B**) Numbers of genes with at least one eQTL (eGenes) vs. sample size in different studies. Numbers of eGenes increase as the sample size increases. The eGene of PsychENCODE is close to saturation. (**C**) Overlap of eQTL, sQTL, fQTL and cQTL snps. 80% of eQTLs were overlapped with other QTLs. The percentage of overlapped snps of eQTL with other QTLs was the highest among all QTLs. 31% of fQTL snps overlapped with other QTLs which was the lowest among all QTLs. 36% of sQTL and cQTL snps overlap with other QTL snps. fQTL overlap more with sQTL(17%) than eQTL(9%) (p<1e-15, odds ratio:2.07). (**D**) Enrichment of genomic regions annotations of QTLs. (E) Enrichment of GWAS snps on cis-eQTL snps. Enrichment for GWAS SNPs of brain disorders on cis-eQTLs SNPs are more significant than the ones of non-brain disorder GWAS SNPs. Schizophrenia GWAS SNPs have the highest enrichment on cis-eQTLs SNPs among those three brain disorders.

Figure 5. Data integration and modeling predicts gene regulatory network, revealing additional GWAS genes for psychiatric disorders. (**A**) The full Hi-C data for adult brain found a variety of TADs (contact heatmap) which provide the genomic regions for potential enhancer-promoter interactions. The example highlights the gene XXX where its promoter, enhancers, Hi-C, eQTL, and TADs all point to the same regulatory relationship. **(B)** The number of genes (left y-axis) and the normalized gene expression levels (right y-axis) vs. the number of enhancers that interact with the gene promoters. **(C)** eQTLs and sQTLs that are supported by Hi-C evidence show more significant P-values than those that are not. **(D)** A full gene regulatory network of TFs (green), eQTLs (red), enhancers (blue) and target genes (xxx). The TF-target relationships are predicted using their gene expression data across individual tissues and Elastic net regression. **(E)** Schematic showing how to identify putative target genes of schizophrenia GWAS loci using Hi-C and eQTL. Hi-C identified genes and eQTL identified genes are highly overlapping. The SCZ genes show higher gene expression levels on neuronal cell types than others. **(F)** Evidence shows that GWAS SNPs overlap eQTLs and Hi-C point to the same gene; e.g., TSNARE1.

Figure 6. Deep-learning model predicts genotype-phenotype and reveals intermediate molecular mechanisms. (**A**) The schematic outlines the model structures for Logistic Regression (LR), conditional Restricted Boltzmann Machine (cRBM), conditional Deep

Boltzmann Machine (cDBM) and Deep Structured Phenotype Network (DSPN) models. The hierarchical layers from genotype to phenotype, representing various multi-scale functional elements in adult brain, include genotype (yellow), gene expression, enhancer activity and other intermediate phenotypes (green), function and module (purple) and high-level phenotype (red). The model also allows the gene regulatory linkages (green) on the same layer. **(B)** shows the DSPN structure in further detail, with different types of regulatory linkages highlighted. **(C)** shows examples of associations found: the DSPN prioritized functional modules associated with Dopaminergic and Glutamatergic synapse, and calcium signaling pathways, as well as ones associated with Oligodendrocyte markers and the complement cascade pathway for Schizophrenia. **(D)** summarizes the performance of different models, comparing performance as layers of structure are added to the model; for different predictors (genotype/transcriptome/all intermediate phenotypes); and with or without imputation. Performance accuracy on a balanced sample is shown first, with variance explained on the liability scale shown in brackets.

Page 1: [1] Formatted	Brain	3/14/18 1:50:00 PM

Header

Page 1: [2] Formatted	Brain	3/14/18 1:50:00 PM

Footer, Left

Page 1: [3] Style Definition	Brain	3/14/18 1:50:00 PM

Subtitle

Page 1: [3] Style Definition	Brain	3/14/18 1:50:00 PM

Subtitle

Page 1: [3] Style Definition	Brain	3/14/18 1:50:00 PM

Subtitle

Page 1: [3] Style Definition	Brain	3/14/18 1:50:00 PM

Subtitle

Page 1: [3] Style Definition	Brain	3/14/18 1:50:00 PM

Subtitle

Page 1: [3] Style Definition	Brain	3/14/18 1:50:00 PM

Subtitle

Page 1: [3] Style Definition	Brain	3/14/18 1:50:00 PM

Subtitle

Page 1: [3] Style Definition	Brain	3/14/18 1:50:00 PM

Subtitle

Page 1: [3] Style Definition	Brain	3/14/18 1:50:00 PM

Subtitle

Page 1: [4] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, Font color: Black

Page 1: [5] Formatted	Brain	3/14/18 1:50:00 PM

Header distance from edge: 0.5"

Page 1: [6] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, Font color: Black

Page 1: [7] Formatted	Brain	3/14/18 1:50:00 PM

Font: Times New Roman, 12 pt, Not Bold

Page 1: [8] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, Font color: Black

Page 1: [9] Formatted	Brain	3/14/18 1:50:00 PM

Level 1

Page 1: [10] Formatted	Brain	3/14/18 1:50:00 PM

Font: Times New Roman, 12 pt, Not Bold

Page 1: [11] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [11] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [12] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [12] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [13] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [14] Deleted	Brain	3/14/18 1:50:00 PM

, including genotyping, RNA-seq, ChIP-seq, ATAC-seq, Hi-C and single-cell data on healthy and diseased tissues of thousands of individuals

Page 1: [15] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [16] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [17] Deleted	Brain	3/14/18 1:50:00 PM

 including ~2 million QTLs for expression and chromatin, and ~80k active enhancers. Leveraging single cell data explaining >85% tissue expression variation, we

Page 1: [18] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [18] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [19] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [19] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [20] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [21] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [22] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [23] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [24] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [25] Deleted	Brain	3/14/18 1:50:00 PM

has unique expression and greater non-coding transcription than most

Page 1: [26] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [27] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [28] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [29] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [30] Deleted	Brain	3/14/18 1:50:00 PM

 linking all possible functional genomic elements, identifying

Page 1: [31] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [32] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [33] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [34] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [34] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [35] Deleted	Brain	3/14/18 1:50:00 PM

achieved by using functional genomics to learn to impute intermediate phenotypes first compared to directly predicting high-level phenotypes from the genotype.

Page 1: [36] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [37] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [38] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [39] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [40] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [41] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, Font color: Black

Page 1: [41] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, Font color: Black

Page 1: [42] Formatted	Brain	3/14/18 1:50:00 PM

Level 1

Page 1: [43] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, Font color: Black

Page 1: [43] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, Font color: Black

Page 1: [44] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [45] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [46] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [47] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [48] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [49] Formatted	Brain	3/14/18 1:50:00 PM

Font: Times New Roman

Page 1: [50] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [50] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [51] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [52] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [53] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [53] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [54] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [55] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 1: [56] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 1: [56] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 1: [57] Deleted	Brain	3/14/18 1:50:00 PM
identify brain eQTLs using bulk tissue RNA-Seq data		

Page 1: [58] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 1: [59] Deleted	Brain	3/14/18 1:50:00 PM
such as missing QTLs with chromatin activities and single cell fractions. Thus, comprehensive data integration for functional genomic data is essential to fully understand the molecular mechanisms in adult brain. In particular, these genomic elements relating to adult brain from previous genomic studies suggest that the gene regulation is a key molecular mechanism revealing how genomic variants affect gene expression in brain, so		

Page 1: [60] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 1: [60] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 1: [61] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 1: [62] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 1: [63] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 1: [64] Formatted	Brain	3/14/18 1:50:00 PM
Header		

Page 1: [65] Formatted	Brain	3/14/18 1:50:00 PM
Footer, Left		

Page 2: [66] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 2: [67] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 2: [67] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 2: [68] Formatted	Brain	3/14/18 1:50:00 PM

Font: 12 pt

Page 2: [69] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [70] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [70] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [71] Deleted	Brain	3/14/18 1:50:00 PM

to predict the molecular relationships from genotype to phenotype with high accuracy and impute missing data. The results obtained from this model are then studied in relation to specific brain phenotypes and psychiatric disorders.

Page 2: [72] Formatted	Brain	3/14/18 1:50:00 PM

Font: Times New Roman, 12 pt

Page 2: [73] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, Font color: Black

Page 2: [74] Formatted	Brain	3/14/18 1:50:00 PM

Normal, Level 1, Space After: 0 pt

Page 2: [75] Formatted	Brain	3/14/18 1:50:00 PM

Font: Times New Roman, 12 pt, Not Bold

Page 2: [76] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [77] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [78] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [79] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [80] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [81] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [81] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [81] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [82] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [83] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [84] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [84] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [85] Deleted	Brain	3/14/18 1:50:00 PM

 The major data types include RNA-seq, ChIP-seq, ATAC-seq, HiC, single-cell data, and genotyping. (The later required large-scale imputation for all the PsychENCODE datasets, and we make full genotype sets available).

Page 2: [86] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [86] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [87] Formatted	Brain	3/14/18 1:50:00 PM

Level 1

Page 2: [88] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, Font color: Black

Page 2: [89] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, Font color: Black

Page 2: [90] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, Font color: Black

Page 2: [91] Deleted	Brain	3/14/18 1:50:00 PM

: deconvolution explains gene expression in terms of cell fraction changes

Page 2: [92] Formatted	Brain	3/14/18 1:50:00 PM

Font: Times New Roman, 12 pt, Not Bold

Page 2: [93] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [94] Formatted	Brain	3/14/18 1:50:00 PM

Space After: 4 pt

Page 2: [95] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [96] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt

Page 2: [96] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt

Page 2: [97] Deleted	Brain	3/14/18 1:50:00 PM

characterizing various brain regions, phenotypes and disorders

Page 2: [98] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [99] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [100] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [101] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [102] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [102] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [103] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [103] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [104] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [105] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [106] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [107] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [107] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [108] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [109] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [110] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [111] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [112] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [113] Deleted	Brain	3/14/18 1:50:00 PM

in our brain tissue samples is reliably determining whether the changes

Page 2: [114] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [114] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [115] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [116] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 2: [116] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [117] Formatted	Brain	3/14/18 1:50:00 PM

Header

Page 1: [118] Formatted	Brain	3/14/18 1:50:00 PM

Footer, Left

Page 3: [119] Deleted	Brain	3/14/18 1:50:00 PM

To address this, we integrated single cell transcriptome data to discover how the gene expression from cell types contribute to bulk gene expression.

Page 3: [120] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [120] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [121] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [122] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [123] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [124] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [125] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [126] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [127] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [128] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [129] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [130] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [130] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [131] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [132] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [133] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [134] Deleted	Brain	3/14/18 1:50:00 PM

in the data and the 25 standard gene expression signatures of single cells

Page 3: [135] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [136] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt

Page 3: [136] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt

Page 3: [137] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [138] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [139] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [140] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [141] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [141] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [142] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [142] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [143] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [144] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [144] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [145] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [145] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [146] Deleted	Brain	3/14/18 1:50:00 PM

Previous studies found that the gene expression changes at the tissue level can be significantly associated with various cell types; e.g., the cell type specific gene co-expression modules \cite{18849986, 19829370}, but have not systematically and quantitatively revealed how different cell types contribute to the tissue gene expression changes. Thus, we used the supervised method to estimate the cell fractions for tissue samples of individuals.

Page 3: [147] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [148] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [149] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [149] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [150] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [151] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [152] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 3: [152] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

| Page 3: [152] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 3: [152] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 3: [152] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 3: [153] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 3: [153] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 3: [154] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 3: [154] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 3: [155] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 3: [156] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 3: [157] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 3: [158] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 3: [159] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 3: [160] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 3: [161] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 3: [162] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 3: [163] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 3: [164] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 3: [165] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

Font: Arial, 11 pt, Font color: Black

Font: Arial, 11 pt, Font color: Black

Font: Arial, 11 pt, Font color: Black

Font: Arial, 11 pt, Font color: Black

Font: Arial, 11 pt, Font color: Black

Font: Arial, 11 pt, Font color: Black

Font: Times New Roman

Level 1

Font: Arial, Font color: Black

Font: Arial, Font color: Black

In addition to the transcriptome data, we uniformly processed chromatin data in the resource to give uniform peak calling lists and single tracks for adult brain epigenomics. This includes the H3K27ac ChIP-seq data of a cohort of 150 postmortem samples from normal individuals in three different brain regions, namely, prefrontal cortex (PFC), temporal cortex (TC) and cerebellum (CBC). We pooled together the processed peaks from the cohort respectively in each region, generating a consistent set of 37,761 H3K27ac peaks in PFC, 42,683 peaks in TC, and 26,631 in CBC. Each of the pooled peaks is in more than half of the samples of the corresponding brain region. We also processed the H3K27ac and H3K4me3 ChIP-seq data for the reference brain. Combining the ChIP-seq data and the ATAC-seq data from the same reference brain, we identified a consistent set of ~80k brain active enhancers in the prefrontal cortex, >90% of which overlap with the Epigenomics Roadmap annotation (See details in the Supplement). We examined the overlaps of these enhancer regions with the pooled H3K27ac peaks and found that around 90% of the identified enhancers reside in PFC H3K27ac peaks and TC H3K27ac peaks, despite that temporal cortex has a slightly larger set of pooled peaks. However, only 34% of the PFC enhancers are found in the cerebellum pooled H3K27ac peaks, indicating very different epigenome landscape of cerebellum from prefrontal cortex. This is

consistent with previous findings that prefrontal cortex and temporal cortex shared similar chromatin dysregulation signature in ASD and is largely different from that of cerebellum.

We then looked at the epigenetic signal variations across individuals at these enhancers. For this we examined the H3K27ac peaks across 50 PFC samples (Figure 3A). As expected, not all the identified enhancers have H3K27ac signal in every individual. In fact, on average only about 51K of the identified enhancers are found to be active in one individual person at a time, with an even smaller fraction of enhancers have H3K27ac signals across all 50 PFC samples. Yet the majority (~68%) of the identified enhancers are active in more than half of the population. The variation of enhancer state can be also observed from the cumulative numbers of active enhancers by increasing the sample size. For example, the cumulative number increase dramatically for the first 20 sample examined, but becomes nearly saturated at the 30th sample to more than 75K (Figure 3B). Again, this suggests that enhancer activity varies across individuals, yet the majority of brain enhancers are active in most of the population. We also compared the distribution of the saturation curve on the normal samples (N=50) with the ASD samples (N=43) and found no significant differences in overall enhancer activity (KS-test p>0.1).

V. Consistently comparative analysis reveals the brain related

Page 1: [176] Formatted	Brain	3/14/18 1:50:00 PM
Header		

Page 1: [177] Formatted	Brain	3/14/18 1:50:00 PM
Footer, Left		

Page 5: [178] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 5: [178] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 5: [179] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 5: [179] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 5: [180] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 5: [181] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 5: [182] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 5: [182] Formatted	Brain	3/14/18 1:50:00 PM
Font: Arial, 11 pt, Font color: Black		

Page 5: [183] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [184] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [184] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [184] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [185] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [186] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [187] Formatted	Brain	3/14/18 1:50:00 PM

Font: Times New Roman, 12 pt, Font color: Auto

Page 5: [188] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [188] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [189] Formatted	Brain	3/14/18 1:50:00 PM

Level 1

Page 5: [190] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, Font color: Black

Page 5: [191] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, Font color: Black

Page 5: [191] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, Font color: Black

Page 5: [192] Deleted	Brain	3/14/18 1:50:00 PM

To understand how the genotype affects the transcriptome and epigenome in the adult brain, we

Page 5: [193] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [194] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [195] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [195] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [195] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [196] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [197] Deleted	Brain	3/14/18 1:50:00 PM

 by using matched genotype and gene expression data of 1387 adult individuals.

Page 5: [198] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [198] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [199] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [200] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [201] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [202] Deleted	Brain	3/14/18 1:50:00 PM

. Because of the effect of linkage disequilibrium, many of these eQTL SNPs for the same gene were correlated. We calculated the genotype correlation coefficient (r^2) for all eQTL SNPs and determined pruned eQTLs using $r^2 < 0.6$.

Page 5: [203] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [203] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [204] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [205] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [205] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [206] Formatted	Brain	3/14/18 1:50:00 PM

Font: Times New Roman

Page 5: [207] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [208] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [209] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [210] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [211] Formatted	Brain	3/14/18 1:50:00 PM

Font: Times New Roman

Page 5: [212] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [213] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [214] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [215] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [216] Deleted	Brain	3/14/18 1:50:00 PM

 (all fQTLs were conservatively identified using Bonferroni-corrected p-values to 0.05).

Page 5: [217] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [218] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [219] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [219] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [219] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [219] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [220] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black, Pattern: Clear (Yellow)

Page 5: [220] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black, Pattern: Clear (Yellow)

Page 5: [221] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [221] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [222] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [222] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [223] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [224] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [225] Deleted	Brain	3/14/18 1:50:00 PM

For example, we observed a significant enrichment of eQTLs on ePromoters, TFBS and TSS (xx%,

Page 5: [226] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 5: [227] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [228] Formatted	Brain	3/14/18 1:50:00 PM

Header

Page 1: [229] Formatted	Brain	3/14/18 1:50:00 PM

Footer, Left

Page 6: [230] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [231] Deleted	Brain	3/14/18 1:50:00 PM

enhancers. As expected, there were lots of overlaps among eQTLs,cQTLs, sQTLs and fQTLs

Page 6: [232] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [233] Deleted	Brain	3/14/18 1:50:00 PM

The percentage of overlapped snps of eQTL with other QTLs was the highest among all QTLs. Only 31% of fQTL SNPs overlapped with other QTLs, the lowest among all QTLs.

Page 6: [234] Formatted	Brain	3/14/18 1:50:00 PM

Font: Times New Roman

Page 6: [235] Formatted	Brain	3/14/18 1:50:00 PM

Level 1

Page 6: [236] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, Font color: Black, Pattern: Clear (White), Not Highlight

Page 6: [237] Formatted	Brain	3/14/18 1:50:00 PM

Font: Times New Roman, 12 pt, Not Bold, Not Highlight

Page 6: [238] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [239] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [240] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [241] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [242] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [243] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [244] Deleted	Brain	3/14/18 1:50:00 PM

, which provides direct physical evidence for potential

Page 6: [245] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [246] Deleted	Brain	3/14/18 1:50:00 PM

Specifically, we generated and processed Hi-C data for the same reference adult brain that was used to identify the brain active enhancers

Page 6: [247] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [248] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [249] Deleted	Brain	3/14/18 1:50:00 PM

 and then 149,097 putative enhancer-promoter interactions in adult DLPFC.

Page 6: [250] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [251] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [251] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [252] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [252] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [253] Formatted	Brain	3/14/18 1:50:00 PM

Font: Times New Roman

Page 6: [254] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [255] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [256] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [257] Deleted	Brain	3/14/18 1:50:00 PM

regulations from enhancers to target genes are through interactions to targets' promoters

Page 6: [258] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [259] Deleted	Brain	3/14/18 1:50:00 PM

; e.g., we found ~3.2 million enhancer-target linkages in total

Page 6: [260] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [261] Deleted	Brain	3/14/18 1:50:00 PM

Page 6: [262] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [263] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [264] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [265] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [266] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [266] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [267] Deleted	Brain	3/14/18 1:50:00 PM

not only to e/sQTLs without genomic annotations, but also to

Page 6: [268] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [268] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [269] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [269] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [270] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [271] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [272] Deleted	Brain	3/14/18 1:50:00 PM

we integrated the enhancer-target linkages identified by TADs with other

Page 6: [273] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [274] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [275] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [275] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [276] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [277] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [278] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [279] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [280] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [280] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [281] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [282] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [283] Deleted	Brain	3/14/18 1:50:00 PM

linkages, which include the active regulatory links relating QTLs, enhancers, and transcription factors

Page 6: [284] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [284] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [285] Deleted	Brain	3/14/18 1:50:00 PM

Methods). This network also has a few particular characteristics such as scale-free and hierarchical structures, which have been revealed by previous network analyses (Figure 5D). Given a target gene, we

Page 6: [286] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [287] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [288] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [289] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [290] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [291] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [291] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [292] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [293] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [294] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [295] Deleted	Brain	3/14/18 1:50:00 PM

gene regulatory network comprises ~43k TF-gene linkages (Elastic net coefficient's absolute value>0.2) and ~1.8million eQTL-gene (FDR<0.05) linkages

Page 6: [296] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [297] Formatted	Brain	3/14/18 1:50:00 PM

Font: Times New Roman

Page 6: [298] Deleted	Brain	3/14/18 1:50:00 PM

VIII. GWAS

Using the regulatory map built with Hi-C and eQTLs, we are further interested to predict potential target genes of

Page 6: [299] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 17 pt, Bold, Font color: Black

Page 6: [300] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [300] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [301] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [301] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [302] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [303] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [304] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [305] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [306] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [306] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 6: [306] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 1: [307] Formatted	Brain	3/14/18 1:50:00 PM

Header

Page 1: [308] Formatted	Brain	3/14/18 1:50:00 PM

Footer, Left

Page 7: [309] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [310] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [311] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [312] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

| Page 7: [313] Deleted | Brain | 3/14/18 1:50:00 PM |

. Collectively, these QTLs annotate a larger fraction of GWAS SNPs involving the brain (e.g., 21% in schizophrenia, 18% in bipolar) than previously observed, providing important leads on which genes are affected in disease.

| Page 7: [314] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Times New Roman

| Page 7: [315] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 7: [315] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 7: [316] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 7: [317] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 7: [318] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 7: [318] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 7: [318] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 7: [319] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 7: [320] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 7: [321] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 7: [322] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 7: [323] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 7: [323] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 7: [323] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Arial, 11 pt, Font color: Black

| Page 7: [324] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Times New Roman

| Page 7: [325] Formatted | Brain | 3/14/18 1:50:00 PM |

Font: Times New Roman

Page 7: [326] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [327] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [328] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [329] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [330] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [331] Deleted	Brain	3/14/18 1:50:00 PM

attributing to schizophrenia and also differentially expressed genes associated with schizophrenia (Figure Sxx). Second, we found that these genes can reveal additional biological mechanisms and functions that GWAS is unable to identify. They were enriched for genes and co-expression modules dysregulated in DLPFC of schizophrenia-affected individuals \cite{27668389}, suggesting that common variation-mediated gene regulation contributes to the gene dysregulation in schizophrenia (Figure Sxxx). This also hints that there likely exists shared genetic etiology between common and structural variation since the SCZ genes are often affected by recurrent CNVs in schizophrenia. We

Page 7: [332] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [333] Formatted	Brain	3/14/18 1:50:00 PM

Font: Times New Roman

Page 7: [334] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black, Pattern: Clear (Yellow)

Page 7: [334] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black, Pattern: Clear (Yellow)

Page 7: [335] Deleted	Brain	3/14/18 1:50:00 PM

IX. Integrative modeling to relate genotype to molecular and high-level phenotypes in the adult brain

Page 7: [336] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [337] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [338] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [339] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [340] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [340] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [341] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [342] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [343] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [344] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [344] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [345] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [346] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [346] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [346] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [346] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [347] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [348] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [349] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [350] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [351] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [351] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [352] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [352] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [353] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [353] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [354] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [355] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [355] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 7: [356] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black, Pattern: Clear (Yellow)

Page 7: [357] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black, Pattern: Clear (Yellow)

Page 7: [358] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black, Pattern: Clear (Yellow)

Page 1: [359] Formatted	Brain	3/14/18 1:50:00 PM

Header

Page 1: [360] Formatted	Brain	3/14/18 1:50:00 PM

Footer, Left

Page 8: [361] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black, Pattern: Clear (Yellow)

Page 8: [361] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black, Pattern: Clear (Yellow)

Page 8: [362] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black, Pattern: Clear (Yellow)

Page 8: [362] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black, Pattern: Clear (Yellow)

Page 8: [363] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [364] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [365] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [366] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [367] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [367] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [368] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [369] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [370] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [371] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [372] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [372] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [373] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [373] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [374] Formatted	Brain	3/14/18 1:50:00 PM

Font: Times New Roman

Page 8: [375] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [376] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [377] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [377] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [377] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [378] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [379] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [380] Deleted	Brain	3/14/18 1:50:00 PM

), although we note that the DSPN may be capturing epistatic interactions not modeled in SNP-based heritability (Figure Sxx). The

Page 8: [381] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [382] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [383] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [384] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [385] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [386] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [387] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [388] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [388] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [389] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [390] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [390] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [391] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [392] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [393] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [393] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [394] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [394] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [395] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [396] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [397] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [398] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [399] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [400] Deleted	Brain	3/14/18 1:50:00 PM

common prioritization scheme (Supplement Sec. 7.5 and Supp Fig. A2). For instance, in Schizophrenia, we found that the highest prioritized module in the DSPN was associated with Dopaminergic and Glutamatergic synapse and calcium signaling pathways, with other modules associated with Oligodendrocyte marker genes, and the Complement cascade pathways, which confirms and extends previous smaller scale analyses \cite{JK's refs}. Further, we found that excitatory neuron marker gene were enriched in the highest prioritized module for age and the gene NRGN occurred in many of the top prioritized modules/submodules, in agreement with the earlier analyses \cite{}. We provide a full summary of the functional enrichment analysis for all disease and high-level traits in supplement. We further show that the modules prioritized by the DSPN are strongly enriched for GWAS variants (Supplement).

Page 8: [401] Formatted	Brain	3/14/18 1:50:00 PM

Font: Times New Roman

Page 8: [402] Deleted	Brain	3/14/18 1:50:00 PM

X.

Page 8: [403] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, Font color: Black

Page 8: [404] Formatted	Brain	3/14/18 1:50:00 PM

Level 1, Space After: 4 pt

Page 8: [405] Formatted	Brain	3/14/18 1:50:00 PM

Font: Times New Roman

Page 8: [406] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [407] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [408] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [409] Deleted	Brain	3/14/18 1:50:00 PM

 including data from 1931 individuals. This resource serves as an important step for linking variants to functional genomic elements in the brain.

Page 8: [410] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [411] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [412] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 8: [413] Formatted	Brain	3/14/18 1:50:00 PM

Font: Arial, 11 pt, Font color: Black

Page 9: [414] Deleted	Brain	3/14/18 1:50:00 PM

it and also complement the gene expression measurement at the whole cell level that may be impacted by various aspects such as intercellular interactions \cite{28729663, 29227469}. Also, our model performance can be potentially improved if we integrate more functional types such as the more accurate QTLs and cell fractions, imaging and medical data. Finally, this model provides a useful compression on the functional genomic datasets; e.g., XXX(TBD) KB for the model representation vs. XXX(TBD) TB for the resource data