Structural and cellular transcriptome foundations of human brain disease

Yashar Zeighami[1,*], Trygve Bakken[2], Thomas Nickl-Jockschat[3], Jeremy Miller[2], Alan C. Evans[1], Ed Lein[1], Michael Hawrylycz[2,4 *]

[1]Montreal Neurological Institute, McGill University, Canada, Ludmer Centre for Neuroinformatics and Mental Health, McGill University, Canada, [2]Allen Institute for Brain Science, Seattle, WA[3] Department of Psychiatry, Neuroscience and Pharmacology, Iowa Neuroscience Institute, University of Iowa, Iowa City, IA [4]University of Washington, Department of Genome Sciences, Seattle, WA

[*]e-mail: yashar.zeighami@mail.mcgill.ca, mikeh@alleninstitute.org

Abstract

The intersection of human genetics and brain transcriptomics promises to reveal the structural and cellular locus of brain diseases via selective co-expression of risk genes. We find that adult brain-wide transcriptomic profiles of 40 human brain diseases identify four major transcriptional patterns, represented by tumor-related, neurodegenerative, psychiatric and substance abuse, and a mixed group of diseases, with some unexpected disease associations. Based on differential co-expression using bulk transcriptomics, the majority of brain diseases exhibit unique regional transcriptomic signatures that strongly reflect neuronal versus non-neuronal divisions and variation in excitatory and inhibitory neurons across the brain. Single cell transcriptomic data confirms and refines the relationship of different diseases to specific neuronal and non-neuronal subclasses. Disease signatures are largely conserved between mouse and human at the higher cell class level, but with significant species differences emerging at the finer subclass level that may help explain human-specific disease susceptibility.

Introduction

Brain diseases are increasingly recognized as major causes of death and disability worldwide (1-3). These diverse and multifactorial diseases may be largely grouped into cerebrovascular, neurodegenerative, movement related, psychiatric disorders, developmental and congenital disorders, substance abuse disorders, brain tumors, and a set of other brain-related diseases (*Institute for Health Metrics* (IHME), healthdata.org). Phenotypic classification is challenging however and does not uniquely partition characteristics of disease manifestation, genetic risk, and treatment, and economic considerations differ substantially between disease classes, as reflected in the comprehensive *Global Burden of Disease Study* (4) (referred to as **GBD** classes in this study, **Suppl. Figure 1**).

The etiology of brain-related diseases and their genetics is complex and widely studied (5-7). With the exception of Mendelian diseases arising from single gene mutations, most brain disorders present as a complex interplay between genetics and environment through interaction of the brain transcriptome and its regulatory network. Transcriptomic based analysis of brain disease, through profiling of tissues, cells, and more recently at the resolution of single nucleus (8) addresses this complexity providing a means for population scale sampling to disentangle basic molecular and clinical relationships (9, 10). The approach is based on the hypothesis of regional vulnerability that postulates spatial and temporal co-expression across disease genes as indicative of a potential interaction between these genes (11, 12). Studying brain samples from patient populations characterizing major transcriptomic relationships and expression patterns of disease-related genes in neurotypical brains at multiple scales promises important insight in developing future investigations in pathophysiology of brain disorders.

Large transcriptome profiling of the human brain has produced useful resources for exploring the genetics of neurotypical and disease states (13-16) and for describing the larger scale relationship of brain diseases and the neuroanatomy of transcriptomic patterning (14, 17). Here we investigate large-scale anatomic patterning of the transcriptional patterns for 40 brain-related disorders. Transcriptomic relationships at a mesoscale, intermediate between the largest brain structures (e.g. cortex, hypothalamus) and cellular resolution, provide a natural framework for classifying broad disease associations in comparison with common phenotypic grouping. Comparing disease gene expression from 104 structures from cortex, hippocampus, amygdala, basal ganglia, epithalamus, thalamus, ventral thalamus, hypothalamus, mesencephalon, cerebellum, pons, pontine nuclei, myelencephalon, ventricles, and white matter, from the *Allen Human Brain Atlas* (human.brain-map.org, (13, 14), with cell type data from the human and

mouse celltypes.brain-map.org (15, 18) provides a strategy for exploring the intersection of brain structure and cell type, and indicates directions for further analysis as brain wide cellular data becomes available.

Brain disorders and associated genes

The diseases selected are representative of the seven major *GBD* classes above, excluding cerebrovascular due to limitations of representative endothelial or pericyte cell types in data sources. **Suppl. Table 1** contains definitions, gene sets, and metadata identifying each disease (**Methods**). To identify gene-disease associations, we used the *DisGeNET* database (www.disgenet.org) (19-21) a platform aggregated from multiple sources including curated repositories, GWAS catalogues, animal models and the scientific literature. From an initial survey of the OMIM (www.omim.org) repository we identified 549 brain-related diseases (14) which were intersected with the DisGeNET repository. We required reported gene-disease associations be present in at least one confirmed curated source (See https://www.disgenet.org/dbinfo), and with a minimum of 10 genes per disease. For each disease, the main variant of the disease was selected with rare familial and genetic forms not included in the analysis. Conservative selection resulted in 40 major brain disorders with 1646 unique associated genes (**Suppl. Table 1**). Additionally, ten disorders of public health concern with limited associated genes (n > 5) including opioid-related disorders, sleep disorders, and post-partum depression were also compared. The identified gene sets associated with brain disease range vary widely in size (range: opioid-abuse (g=5), schizophrenia (g=733) and distribute across GBD classes as (number, % unique to GBD class) psychiatric (1107, 0.723), neurodegenerative (257, 0.513), substance abuse (212, 0.320), brain tumors (168, 0.667), developmental disorders (139, 0.676), movement related (136, 0.272) and other brain-related (123, 0.414). By ontology analysis (22), using only those genes *unique* to each GBD class, known defining biological processes and molecular pathways of GBD disease classes emerge (**Suppl. Fig. 2**), and a detailed gene ontological analysis (www.toppgene.org) of individual 40 disease gene sets confirms specificity for each disease and related biological, cellular, and molecular processes (**Suppl. Table 2**).

The proportion of shared genes between diseases is known to be correlated with phenotypic similarity ($\rho = 0.40, p = 6.0 \times 10^{-3}$) based on clinical manifestations (23, 24). The comorbidity of psychiatric diseases and substance abuse is well established (25) and **Suppl. Fig. 3A** shows the largest gene set intersection (g=132) between GBD classes psychiatric and substance abuse, with 62% of substance abuse genes also associated with psychiatric disorders. Movement disorders are also commonly found in

neurodegenerative diseases (26), with GBD neurodegenerative sharing 30% (g=41) of movement related genes, while GBD tumor based and development sharing the least with other classes (2.5%, 2.6% respectively). The 40 diseases and disorders clustered on relative pairwise gene set intersection (Jacard) shows variable agreement with GBD phenotypic groupings (**Supp. Figure 3B),** with the highest mean percentage of shared genes among psychiatric disorders 7.64% (p=1.55×10^{-4}), followed by substance abuse 6.33% (p=2.82×10^{-4}), and brain tumors 5.43% (p=8.35×10^{-3}).   (Significance measures represent likelihood of observed percentage corrected for GBD class size.)

Structural transcriptomic profile of brain diseases

Expression profiles from the *Allen Human Brain Atlas* (AHBA, https://human.brain-map.org) for genes associated with the 40 diseases were obtained from 6 neurotypical donor brains and summarized in an ontology of 104 structures (**Suppl. Table 3**) including cortex (CTX, 8 structures), hippocampus (HIP,7), amygdala (AMG,6), basal ganglia (BG,12), epithalamus (ET,3), thalamus (TH,12), hypothalamus (HY,16), mesencephalon (MES, 11), cerebellum and cerebellar nuclei (CB,4), pons and pontine nuclei (P,10) myelencephalon (MY, 12), ventricles (V,1), white matter (WM,2). We obtain a mean transcriptomic disease profile by averaging expression for genes associated with each of 40 diseases across the 104 structures and z-score normalizing (**Figure 1, Suppl. Table 4)**. Hierarchical clustering with Ward linkage is subsequently performed using Pearson correlation and *removing* genes common to each pair of diseases so that similarities are driven by independent sets of genes (**Methods**). While anatomic signatures are complex, this organization results in four primary anatomic disease groups (**ADG 1-ADG 4**) generally interpretable with respect to GBD classification (**Fig. 1A**, left color bar) as tumor related (**ADG 1**), neurodegenerative (**ADG 2**), psychiatric, substance abuse, and movement disorders (**ADG 3**), and a mixed group without tumor diseases (**ADG 4**). By averaging gene expression within ADG groups across the 104 brain structures we obtain an anatomic representation of this transcriptomic patterning (**Fig. 1B**) whose major anatomy (**Suppl. Fig. 4**) may be described as **ADG 1:** thalamus, brain stem, ventricles, white matter, **ADG 2:** cortico-thalamic, brain stem, white matter, **ADG 3:** (telencephalon) cortex, thalamus, hippocampus, amygdala, basal ganglia, and **ADG 4:** (diencephalon) thalamus, ventral thalamus, hypothalamus, basal ganglia, and mesencephalon, brain stem. ADG transcriptome signatures are consistent across subjects as individual brain holdout analysis (**Suppl. Figs. 5,6**, **Methods)** finds that both the correlation of expression relationships across structures and differential relationships between ADG groups at a fixed structure are preserved between the AHBA brains.

The association of diseases within ADG groups often differs from their GBD phenotype classification. Examples include multiple sclerosis, migraine disorders, and dementia grouped with tumor-based glioma diseases in ADG 1, and developmental Williams syndrome and spastic paraplegia, hereditary grouped with the neurodegenerative diseases Alzheimer's and ALS in ADG 2, while other neurodegenerative diseases Huntington's disease and amnesia of ADG 4 having a distinct structural profile. Although ADGs reflect the broadest transcriptome patterning there is considerable variation within the groups. For example, substantia nigra (SNC, SNR) and ventral tegmental area (VTA) enrichment is seen in genes associated with dementia, Parkinson's and related disorders (27), basal ganglia and forebrain involvement in movement disorders and dystonia (ADG 4) (28), ventral thalamus expression in neurodegenerative diseases (29), mamillary body (MB) of the hypothalamus expression in (ADG 2) neurodegenerative altered metabolism(30), and low expression of genes in epithalamus for all but ADG 1 tumor-based diseases. The distinction between ADG 1 and ADG 2 here is subtle with significant variation (**Fig 1C**, row 1-2) only in CTX (4.96×10^{-3}), hippocampus (HIP, 1.49×10^{-3}), and epithalamus (ET, 4.79×10^{-3}) and further analysis without removing common genes (**Suppl. Fig. 7**) and leveraging cell type data (**Fig. 3**) will distinguish these groups.

To compare the ADG signatures of **Fig. 1A** we combine ANOVA for mean differences in expression across ADG groups at each fixed structure (BH corrected p-values, top panel, **Fig. 1**.), with paired corrected t-tests (**Fig. 1C**) between groups showing the most significant expression differences. Particularly striking in **Fig. 1A** is the white matter signature of tumor and neurodegenerative diseases (ADG 1-2), effectively absent in psychiatric disorders and diseases of addiction (ADG 3-4) (31), and substantial enrichment of telencephalon expression (CTX, HIP, AMG, BG) in ADG 3 (32, 33). There is significant variation across ADG groups in (BG) globus pallidus (GP, $p < 7.88 \times 10^{-10}$), cerebellar nuclei (FAS, 2.74×10^{-7}), substantia nigra, pars compacta (SNC, 2.87×10^{-7}), basal forebrain olfactory tubercle (MES) (Olft, 3.07×10^{-7}), and (WM) cingulum bundle (cgb, 4.01×10^{-7}), whereas variation is not significant in the thalamus (TH, 0.417), myelencephalon (0.271), and cerebellum (CB, 0.146). Comparing individual pairs of ADG groups (**Fig. 1C**), the most significant differences are telencephalic between tumor-based (ADG 1) and psychiatric and substance groups (ADG 3): hippocampus (HIP, 1.29×10^{-8}), amygdala (AMG, 7.51×10^{-8}), and cortex (CTX, 8.93×10^{-6}), and between neurodegenerative-based ADG 2 and ADG 3 groups (HIP, 1.30×10^{-6}), ventral thalamus (VT, 1.60×10^{-6}), (AMG, 1.30×10^{-6}), and in white matter (WM, 8.56×10^{-5}).
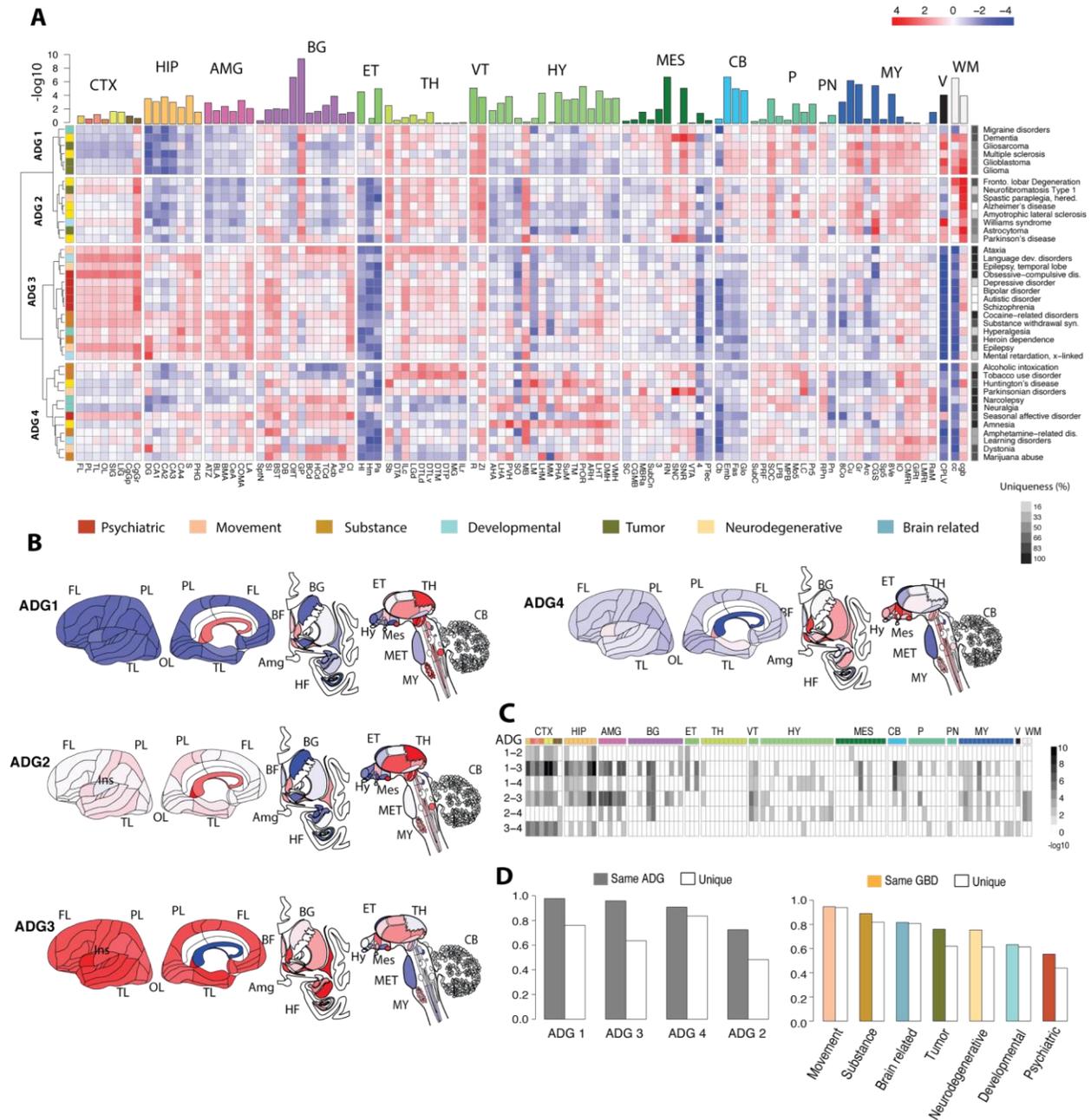
Figure 1. Transcriptome patterning of major brain diseases. A) Mean gene expression profiles for genes associated with 40 major brain diseases and disorders profiled over 104 anatomic structures from 15 major regions cortex (CTX), hippocampus (HIP), amygdala (AMG), basal ganglia (BG), epithalamus (EPI), thalamus (TH), ventral thalamus (VT), hypothalamus (HY), mesencephalon (MES), cerebellum (CB), pons (P), pontine nuclei (PN), myelincephalon (MY), ventricles (V), white matter (WM) and z-score normalized. Hierarchical clustering based on disjoint gene sets between pairs and z-score normalized yields four primary groups ADG 1 – ADG 4. Row annotation left, GBD classification. Column annotation is ANOVA for expression values at a fixed structure grouped by ADG. Row

annotation right, percentage exact match in other AHBA brains (**Methods**.) B) Anatomic patterning of classes ADG 1-4. C) Corrected pairwise t-tests for ADG differences across brain structures. D) Reproducibility of disease transcriptomic profiles. Left: (solid) frequency over six brains that disease in a given ADG group is most closely correlated with profile of the same ADG in other AHBA subjects (decreasing frequency.) (open) Frequency exact disease is identified in other subjects. Right: similar for diseases by GBD group.

While the expression of disease genes may vary considerably in a population (34, 35), the anatomic expression signature of each disease in an individual brain is most closely correlated with a disease in the same ADG group (ADG 1-4: 97.7, 72.5, 95.9, 90.8 %), and typically identifies the exact disease in other subjects (**Fig. 1D, Methods**. In particular, the expression pattern defining each of the ADG 3 group diseases ataxia, language development disorders, temporal lobe epilepsy, obsessive compulsive disorder, and cocaine-related disorder most closely correlates with these same diseases in each of the AHBA subjects. Similarly, ADG 4 diseases Parkinsonian disorders, amnesia, narcolepsy, neuralgia, and tobacco use disorder exhibit highly unique profiles due to consistent, complex and differentiated expression in the basal ganglia, hypothalamus, and myelencephalon. Conversely, ADG 2 Alzheimer's disease and amyotrophic lateral sclerosis, and ADG 3 bipolar disorder, autistic disorder, and schizophrenia while having consistent ADG profiles are less unique to that disease. This identifiability shows a disease dependent level of specificity in transcriptomic signature beyond ADG grouping. With respect to phenotypic GBD classification, diseases of movement and substance abuse have the most identifiable signatures across AHBA subjects (94.6,89.0%) respectively, while psychiatric diseases map within group least (55.5%). Ten additional diseases with limited gene sets (5-10) are presented in **Suppl. Fig. 8**, with their relationship to ADG groups.

Reproducible transcription patterns of brain diseases

The finer anatomic structure of transcription patterns of brain diseases can be directly studied through conserved differential expression relationships providing a bridge to implicated cell types. *Differential stability* (DS), quantified as the mean Pearson correlation ρ of expression in a fixed set of anatomic regions between pairs of brain specimens (14) measures the fraction of preserved differential relationships between anatomic regions over subjects, with a bias for stronger differentials and structural markers. The gene GABBR2 identified in temporal lobe epilepsy (ADG 3) (**Figure 2A)** has a highly reproducible brain wide expression profile across AHBA subjects (DS=0.904), with consistent expression in telencephalon and

highest in thalamus. Abnormal gamma-aminobutyric acid B receptors (GABBR1, GABBR2) can provoke seizures in mesial temporal lobe epilepsy (36).

Although there is no significant difference in their expression levels (**Suppl. Figure 9**), and DS for disease genes is uncorrelated with the number of associated genes (ρ=0.035), GBD classes exhibit a significantly higher percentage (**Figure 2B**) of differentially stable genes, particularly for substance abuse, (mean DS=0.702, p<4.7×10^{-21}), psychiatric (DS=0.675, p<3.17×10^{-17}), and movement disorders (DS=0.635, p<9.1×10^{-17}In particular, the stability of ADG 3 (median 0.683, p<4.24×10^{-10}), and ADG 4 (0.698, 6.91×10^{-10}) genes are markedly higher than ADG 1 (0.589, p<0.0028) and ADG 2 (0.57,p<0.0087),). indicating a higher percentage of transcriptomic structural markers for diseases in these groups. **Figure 2C** shows the distribution of DS for the 40 brain diseases with minimum and maximum DS genes and details the consistency ranking of anatomic signatures described in **Fig.1.**
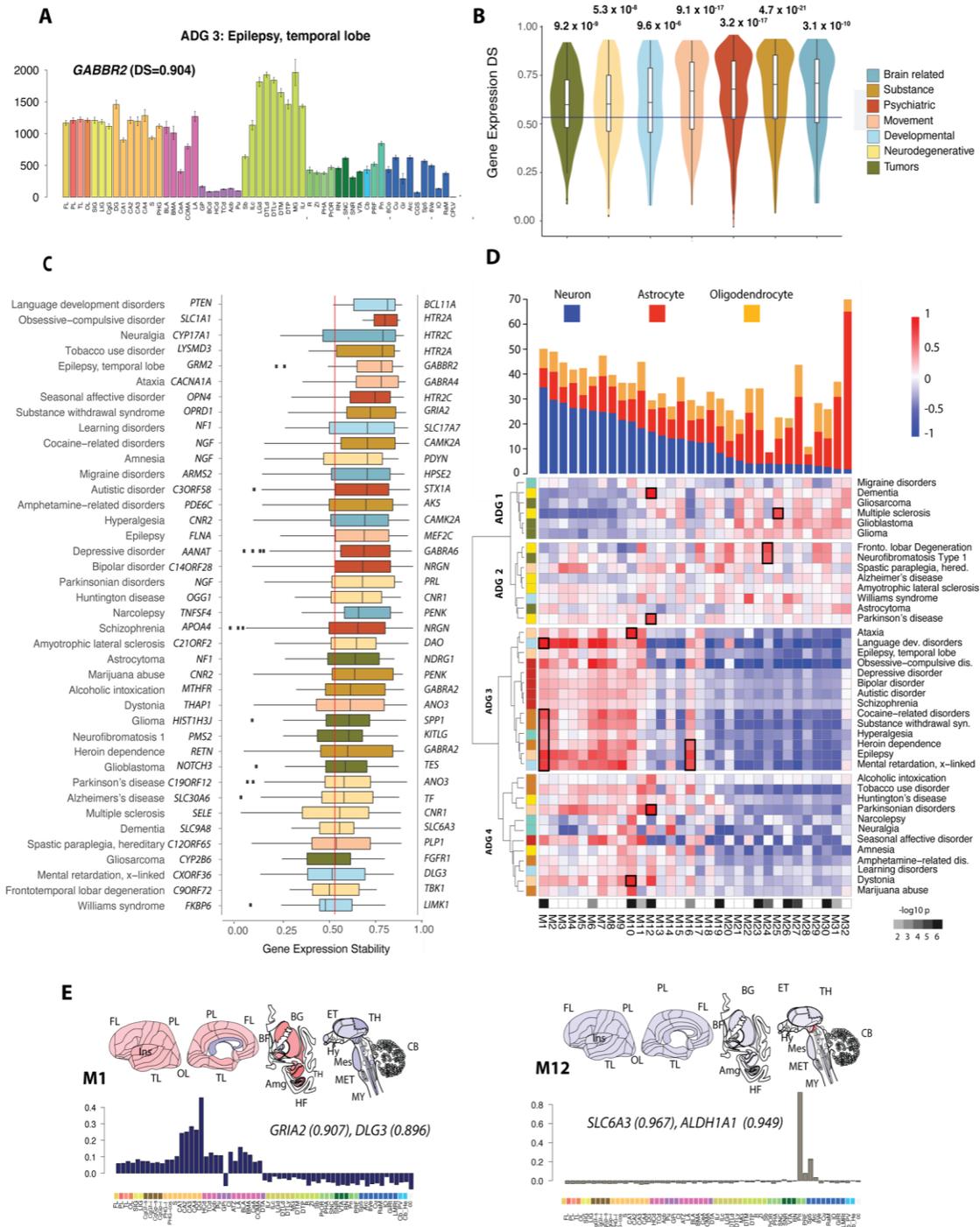
Figure 2. Reproducible transcription patterns in human brain diseases. A) Expression profile for gene GABBR2 with error bars (human.brain-map.org.) Differential stability measures reproducible expression patterns from the Allen Human Brain Atlas (AHBA). B) Distribution of DS by major *Global Burden of Disease* classes. Horizontal bar mean ρ=0.521 of 17,348 genes, with p-values shows significance (corrected for class size) of GBD mean differing from

global mean. C) Disease gene stability sorted by median DS; colors are GBD classification. Minimum and maximum stable genes for each disease are shown. GO analysis (toppgene.cchmc.org) reveals that higher DS genes are orders of magnitude more significant in biological process (e.g. anterograde trans-synaptic signaling, low 4.93×10^{-4}, high 1.80×10^{-12}) molecular function (ion channel activity, 0.59, 1.35×10^{-6}), drug targets (dopamine, 2.02×10^{-3}, 1.80×10^{-12}), and pathways. D) Map of canonical expression modules **M1**-**M32** to diseases. Associated genes for each disease are correlated with each module eigengene independently (**Methods.**) Row ordering is as identified in **Fig. 1** clustering. Modules are ordered based on their neuronal, astrocyte, oligodendrocyte content. Bar below heatmap highlights modules that occur with unexpected frequency compared with genome wide distribution (**Suppl. Fig. 14)**   E) Canonical eigengenes over 62 anatomic regions shown for **M1** telencephalic (language development, epilepsy) and **M1**2 substantia nigra (Parkinson's disease, dementia), with mean Pearson correlation for representative genes. Boxes indicate representative modules of **Suppl. Fig 11**.

A characterization of the reproducible gene co-expression patterns (14) in *Allen Human Brain Atlas* using the top half of DS genes (DS > 0.5284, g = 8,674) previously identified 32 primary transcriptional patterns, or *modules*, each represented by a characteristic expression pattern (i.e. eigengene) across brain structures. **Fig 2E** illustrates two representative modules **M1** and **M12**: Module **M1** is telencephalic having strong expression in the hippocampus, in particular dentate gyrus, and representative genes include *GRIA2* (correlation to eigengene, ρ=0.907)*,* and *DLG3* (ρ=0.896)*.* Alterations in glutamatergic neurotransmission have known associations with psychiatric and neurodevelopmental disorders and only mutations in *GRIA2* and *GRIA3* have been related with these disorders (37). Analysis of the 157 **M1** correlated genes of the present study **(Methods)** reveals neurotransmitter gated ion channel (q-value, 7.33×10^{-18}), glutamatergic synapse (4.14×10^{-35}) activity and strong association with **ADG 3** diseases: bipolar disorder (q-value, 6.92×10^{-30}), schizophrenia (1.23×10^{-23}), and autism (1.26×10^{-15}). Module **M12** exhibits unique expression in the substantia nigra, *pars compacta* and *pars reticulata*, and ventral tegmental area, the source of dopamine generation and dysregulation in Parkinson's disease and Parkinsonian disorders (2.07×10^{-6}) (37). Dopamine transporter gene *SLC6A3* (ρ=0.967), a strong marker for this pattern, is considered as a candidate risk factor for dopamine or other toxins in the dopamine neurons (38, 39)   Another well-known marker, aldehyde dehydrogenase-1 (*ALDH1A1)*, whose polymorphisms are implicated in alcohol use disorders, is strongly associated with **M12** (ρ=0.949) (40). Other key disease related modules are shown in **Suppl. Figure 10**.

We map brain-related diseases to the canonical patterns by finding the closest correlated module eigengene for each disease gene (**Supp. Table 5**, **Methods**). **Figure 2D s**hows the mean correlation of the 40 disease associated gene sets with the module **M1-M32** eigengenes ordered by **ADG 1-4** as in **Fig.1**. The basic cell class composition neuronal, oligodendrocyte, astrocyte of AHBA tissue samples was determined by earlier single cell studies (14) and the modules **M1-M32** ordered by decreasing proportion of neuron-enriched membership. **Fig. 2D** clarifies the cell type composition and distinction between ADG groups in **Fig. 1**. In particular, primarily tumor-based **ADG 1** maps to modules **M21-M32** having enriched glial content (p<2.413$\times$ 10^{-15}), while **ADG 3** psychiatric and substance abuse related diseases map to neuronal enriched patterned modules **M1-M10** (2.2$\times$ 10^{-16}). By contrast the neurodegenerative disorders of **ADG 2** including Alzheimer's, Parkinson's, ALS, and frontotemporal lobe degeneration show more uniform distribution across the modules, clearly separating this group from **ADG 1** (1.55$\times$ 10^{-15}). **ADG 4** has lower expression in neuronal modules **M1-12** than **ADG 3** (1-sided, p<3.84$\times$ 10^{-13}) and is enriched in isolated structural markers **M10** (striatum, narcolepsy, marijuana abuse), **M11** (thalamus, Parkinsonian disorders) **M12** (substantia nigra, dystonia), and **M16** (dentate gyrus, amnesia, tobacco use). The canonical module distribution of **Figure 2D** further validates the clustering of **Fig. 1B,** improves the unique disease identification for ADG and GBD groupings (**Suppl. Figs. 11,12**) and provides a bridge to understanding their cell type content.

Amongst disease genes several modules were found represented at unexpected proportions (**Suppl. Figure 13**) including **M1** (**Fig. 3A**, telencephalon, higher frequency, psychiatric) functionally connected with synaptic transmission and plasticity $(p < 1.04 \, x \, 10^{-10})$, **M10** (striatum, higher frequency, movement disorders, addiction related) with GPCRs and dopamine receptor signaling $(7.27 \, x \, 10^{-5})$, primarily glial **M19 (**thalamus, cerebellar nuclei, brainstem, lower, Alzheimer's) in mitochondrial pathway dysfunction $(5.50 \, x \, 10^{-82})$ and **M30** (white matter, ventral thalamus, globus pallidus lower, neurodegenerative) for myelination and axon ensheathment $(5.63 \, x \, 10^{-9})$. Disease genes are correlated with neuronal-based modules more frequently than expected, and with glial modules less than expected (p=0.0003).

Disease genes and cell types of middle temporal gyrus

Brain structure-level analysis describes global anatomic relationships of transcriptomic patterns, however it is limited in its ability to implicate specific cell types in diseases ((11, 12). We used human single nucleus RNA-seq data from eight donor brains (15,928 nuclei) from the middle temporal gyrus (MTG) (15) (**Figure 3A**) where 75 transcriptomically distinct cell types were previously identified, including 45 inhibitory neuron types and 24 excitatory types. All inhibitory types express the GABAergic interneuron maker *GAD1* implicated in bipolar disorder (41) depression (42), and schizophrenia (43) and excitatory types express the vesicular glutamate transporter *SLC17A7,* associated with learning and memory disorders (44) (as well as psychiatric diseases), and 6 non-neuronal cell types that express the glutamate transporter *SLC1A3*; autism (45) and epilepsy (46). A set of 142 genes are used to differentially characterize the MTG clusters (15). Structurally these form a very high differentially stable group (DS=0.734, p<8.66E-07, all disease genes, DS=0.630), 30 of which are among the disease genes, and several are uniquely associated with a disease (**Suppl. Table 6**).
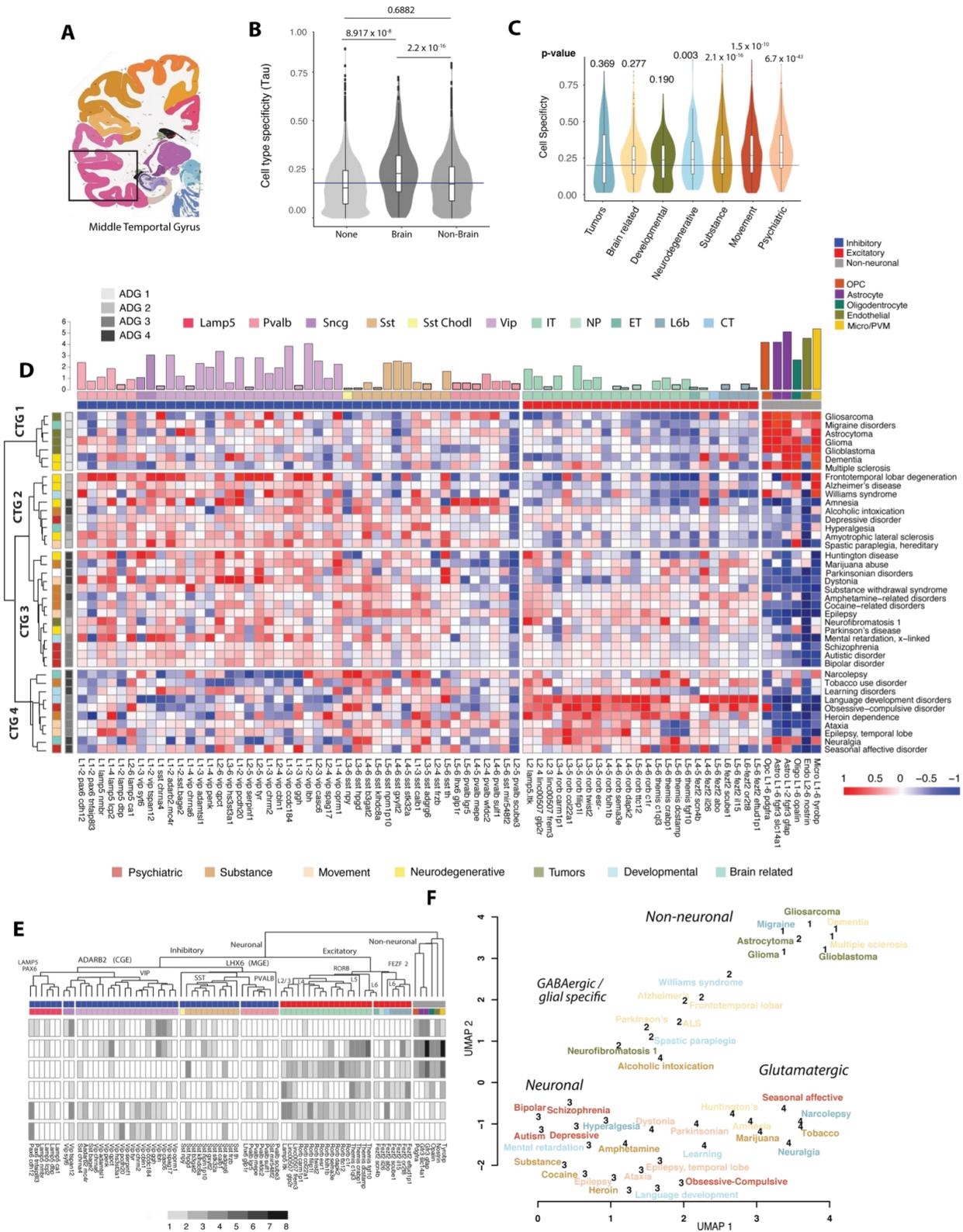
Figure 3. Disease genes and cell types of middle temporal gyrus. A) Reference plate from the Allen Human Brain Atlas (http://human.brain-map.org.) containing middle temporal gyrus region B) Cell type specificity measure τ (tau) derived as the concentration of expression among 75 cell types from (15)(**Methods**). Distribution of τ for brain disease associated, non-brain disease, and unassociated genes (14). C) Specificity measure τ averaged over disease (**Suppl. Fig. 14**) and then by GBD class. D) Mean cell type expression (fpkm) of 40 brain diseases over 75 types identified in (Hodge et al., 2019), clustered removing common genes with fixed taxonomic order (Hodge et al., 2019). Top annotation: major cell class; excitatory, inhibitory, non-neuronal, and ANOVA at fixed cell type colored by *subclass* level; inhibitory (Lamp5, Sncg, Vip, Sst Chodl, Sst, Pvalb), excitatory (L2/3 IT, L4 IT, 5 IT, L6 IT, L6 IT Car3, L5 ET, L5/6 NP, L6 CT, L6b), non-neuronal (OPC, Astrocyte, Oligodendrocyte, Endothelial, Micro-glial/perivascular macrophages.) Hierarchical clustering resolves into four groups **CTG 1-4.** Left annotation: **ADG** grouping consistent with **Fig. 2**, and GBD classification. E) Corrected t-tests between **CTG 1-4** at fixed cell type with subclass and layer annotation. F) Consensus UMAP combining cell type and structural disease relationships. Numbers indicate **ADG** group (**Methods**).

To assess the tendency for disease gene co-expression to concentrate in a specific cell type, we used the *Tau*-score (τ) defined in (47) (**Methods**). While the distinction in *fpkm* values between brain and non-brain disease genes is not significant (p=0.021) (**Supp. Fig 14)**, the τ specificity difference between brain disease genes and no disease is significant ($8.92 \, x \, 10^{-8}$) and more remarkably for brain disease and non-brain disease genes ($2.2 \, x \, 10^{-16}$) indicating potential cell type specific involvement in genes associated with brain diseases (**Figure 3B**). Averaging *tau* over sets of genes representing a given disease, we obtain a measure of cell type specificity of each disease within MTG (**Supp. Fig 15**). Pooling to the 7 GBD categories, the classes psychiatric ($6.7 \, x \, 10^{-43}$), movement ($1.5 \, x \, 10^{-10}$), and substance abuse disorders ($2.1 \, x \, 10^{-16}$) show high cell type specificity, while tumors, developmental disorders and neurodegenerative diseases show less. (**Figure 3C**). *GBD* brain-related category is a mismatch with the structural approach likely due to non-MTG related cell type expression.

A cell type disease-based clustering was performed over the 75 MTG cell types, again removing common genes similarly to **Fig. 1** presents the clustering of the mean expression profiles across the 40 studied brain diseases. This approach also identifies four major groups: (**CTG 1**) diseases with strong non-neuronal enrichment, mainly tumors (**ADG 1**), (**CTG 2**) predominant enrichment in inhibitory neurons and specialized non-neuronal enrichment, associated with neurodegenerative diseases (**ADG 2),** (**CTG 3**) strong enrichment in both inhibitory and excitatory neurons for major psychiatric (48) and substance

abuse disorders (49) (**ADG 3,4**), and (**CTG 4**) primarily neuronal and strongly in excitatory cell types, representing movement, substance and developmental disorders (**ADG 3,4**). Grouping of cell types for disease genes is seen both at the major class (inhibitory, excitatory, non-neuronal) as well the *subclass* level of the ontology (**Supp. Fig 15**). There is notable clustering of diseases within GBD classes including psychiatric (e.g. schizophrenia, bipolar, autistic disorders) and substance abuse (e.g. amphetamine and cocaine related, substance withdrawal), and neurodegenerative (e.g. Alzheimer's, ALS, amnesia, frontotemporal dementia).

Analysis of variance at fixed cell types (**Fig. 3D,** annotation) shows that the highest variation between CTG groups occurs for non-neuronal, followed by *VIP*, *SNCG* interneurons, and groupwise CTG comparisons (**Fig. 3E**) highlight significant variation between neuronal subtypes, in particular excitatory (L2/3 - 5 IT, L6 CT). There is surprising consistency between the structural and cell type analysis in transcriptomic proximity between diseases and their grouping by GBD class, despite data being limited to nuclei from a single cortical area (**Suppl. Fig. 16**). Comparing structural and single cell data connects these scales and provides a more detailed molecular map of brain diseases and associated cell types. By combining the correlation matrices for these approaches (**Methods**) a UMAP can be derived illustrating a consensus disease map, highlighting ADG-CTG consistency, its relationship to GBD phenotypic groupings and major cell types **(Fig. 3F, Suppl. Fig. 16B**).

Brain diseases in mouse and human cell types
Single cell profiling now enables the alignment of cell type taxonomies between species, analogously to homology alignment of genomes between species. To examine conservation of disease based cellular architecture between mouse and human, we used the alignment of transcriptomic cell types obtained in (15) of human MTG to two distinct mouse cortical areas: primary visual cortex (V1) and a premotor area, the anterior lateral motor cortex (ALM) (18). This homologous cell type taxonomy is based on expression covariation and alignment shows a largely conserved cellular architecture between cortical areas and species identifying 20 interneuron, 12 excitatory, and 5 non-neuronal types (**Figure 4A**.) We use this alignment to study species specific cell type distribution over the 40 disease groups and at two levels of resolution: broad cell type class (N=7, e.g. excitatory), and subclasses (N= 20, as in **Fig. 3**) where non-neuronal cell types were common between the two levels of analysis.

To identify cell types showing enriched gene expression associated with brain disorders in mouse and human cell types we used expression-weighted cell-type enrichment (EWCE) analysis (48). Briefly, EWCE compares expression levels of a set of genes associated with a given disease to the genomic background, excluding disease-related genes, by performing permutation analysis (**Methods**) thereby defining the probability for the observed expression level of the given gene set compared with random sets of genes. In contrast to the *tau* measure (**Fig. 3B,C**) which examines each gene separately and measures only divergence of cell type expression from uniformity, EWCE evaluates all genes in a disease simultaneously and identifies concentration of cell-type expression.. The correlation of EWCE values aligned between mouse and human is ρ=0.621, p<2.2×10^{-16}, reflective of broadly conserved expression patterns (14, 15) with no significant difference in EWCE distribution (**Suppl. Fig. 17**, K-S test, p=0.126). Further, simultaneous clustering of EWCE mouse and human aligned cell types (**Figure 4C,** orange (mouse), blue(human)) demonstrates a highly conserved cell type signature at the subclass level across species for many diseases (median ρ=0.645).

Despite distributional similarities there are notable species differences in cell type enrichments. The EWCE outliers of **Fig. 4A** (pie charts) are dominated by **ADG 4** GBD substance abuse (cocaine, tobacco abuse) and brain-related (neuralgia, narcolepsy) categories. **Figure 4D** presents a more detailed view of mouse and human cell type enrichments. While there are no significant neuronal enrichments in either species for tumor based **CTG 1** (**ADG 1**), non-neuronal include known associations of endothelial cells (50), human specific oligodendrocyte precursor cells (OPC) identified as a potential origin of glioblastoma (p<0.003) (51), and migraine disorder where genetic malfunction of astrocytes is implicated (52)(p<0.0009). Unique human microglia/PVM enrichment for Alzheimer's disease (p<0.0012), and oligodendrocytes in hereditary spastic paraplegia (p<0.0008), (53) are seen and there are no significant non-neuronal enrichments among **ADG 2-4.** GBD classes psychiatric and substance abuse (**ADG 3-4**) dominate cell type enrichments for both inhibitory (64%) and excitatory (70%) types. However, at the subclass level these are primarily inhibitory types in both mouse and human (**Fig 4D**, top), in particular *Snc*g, *Vip*, and *Sst Chodl*, and significant excitatory types are found in psychiatric disorders, e.g. depressive disorder *L2/3 IT* (p<0.0015), *L5 ET* (p<0.00153), and bipolar disorder *L6 IT Car* (p<0.004). In summary, although distribution of disease implicated cell types is conserved **Fig. 4D** identifies a number of species-specific differences.
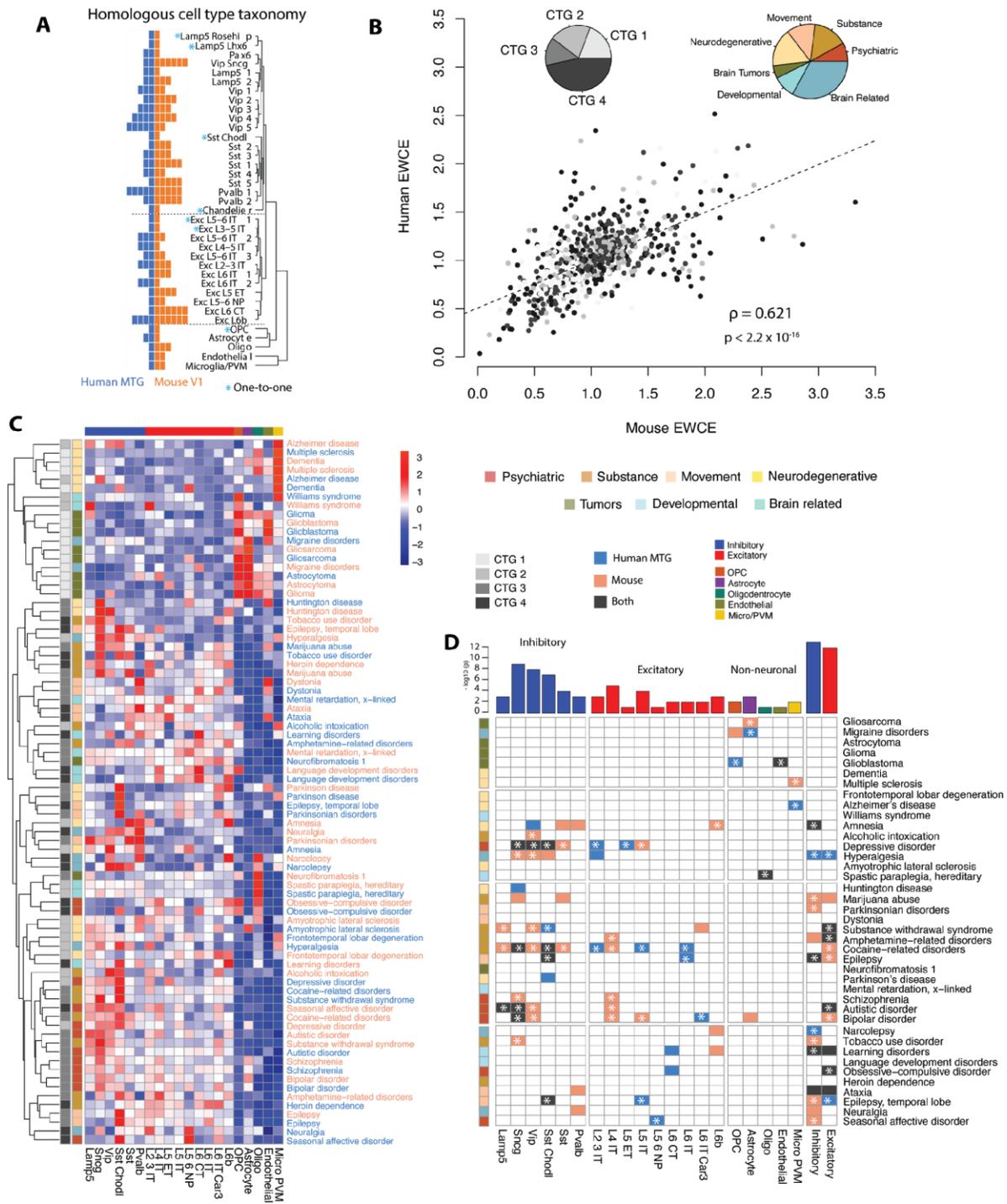
Figure 4. Disease based cell type expression in mouse and human. A) Aligned transcriptomic taxonomy of cell types in human MTG to two distinct mouse cortical areas, primary visual cortex (V1) and a premotor area, the anterior lateral motor cortex (ALM) from (15) allows comparison of cell type enrichments between species. B) Scatterplot of disease-subclass EWCE values for mouse and human colored by **CTG 1-4.** Pie chart insets show percentages of CTG

and GBD classes of top 10% outliers from the regression line, represent most significant EWCE differences.   C) Simultaneous clustering of mouse and human EWCE disease signatures at subclass level (orange: mouse, blue:human). Annotation top major cell classes, side disease GBD and CTY membership.   D) Significant species EWCE by disease and cell type with similar color code, black: both species. Disease clustering is as in **Fig. 3** with GBD annotations. Top barplot: number of cell type enrichments. * indicates surviving multiple hypothesis testing B-H q<0.01.

Discussion

We presented a brain-wide molecular characterization of common brain diseases from the perspective of neuroanatomic structure and cell type specificity, striving to describe relationships between multiscale approaches and key associations. Combining structural patterns with cell type-specific analyses allows an identification of major disease groups and brain-wide and cell type specific transcriptomic signatures of common brain diseases. Such analysis may potentially indication clinically actionable strategies, as shared cellular and mesoscopic expression patterns hint at potential interactions between genes across regions of the brain. Precise phenotypic classification of diseases is challenging due to variations in manifestation, severity of symptoms, and comorbidities (54). We chose for a high-level phenotypic categorization from the *Global Burden of Disease* study from the Institute for Health Metrics and Evaluation as this is a continuously updated, comprehensive, and data-driven resource. For associated genes, *DisGeNet* is one of the largest resources integrating human disease genes and variants from curated repositories and provided a standard approach to select genes for the study. Determining implicated genes in disease states presents considerable uncertainty, and any study is likely to miss important associations. In particular, notably absent from our analysis are cerebrovascular diseases that account for the largest global burden of disability (55), and this limitation is due to relative under-sampling of rare vascular cell types in the *Allen Human Brain Atlas.* However, the approach presented is flexible and data driven and can be immediately updated with other diseases of interest or associated genes.

Transcriptomic patterns of brain diseases cluster structurally into four major disease groups (**ADG 1-4**), largely recapitulated using cell type data from a single cortical area. **ADG 1** consists primarily of glial diseases including most brain tumors, multiple sclerosis, migraine and certain dementias and are transcriptomically distinct (**ADG 1**). Most neurodegenerative diseases (**ADG 2**) involve common neuronal (particularly cortex and hippocampus) and strong glial patterning (**Fig. 2D** and **3D**) effectively

distinguishing them from largely glial based **ADG 1**. **ADG 3** shows the strongest neuronal patterning of all four disease groups with pronounced expression within the telencephalon, while glial content is minimal. This group mainly consists of psychiatric and substance abuse disorders, and epilepsies, recapitulating the known close genetic relationship between these disease groups (56). Finally, **ADG 4** represents a combination of GBD diseases with modest cortical expression, enriched in glutamatergic cell types, and with a larger number (26%) of anatomic structural markers in basal ganglia, hypothalamus, and lower brain structures. The general association of these disease groups is reproduced both in cell type specific analysis and corroborated by association in the mouse.

Diseases with diverse phenotypes and clinical presentations have shared anatomic expression patterns, and this may provide insight into disease mechanisms and frequency of comorbidity. For example, autism, OCD and epilepsy are phenotypically diverse, yet all belong to **ADG 3,** and cell type analysis in **Fig. 3D** illustrates a correlated cell type signature with strong IT neuron subclass expression. While these are broad categorizations, there is highly reproducible structure to anatomic disease profiles illustrated through differential expression stability analysis and through correspondence between mouse and human cell type profiles (**Fig. 4C**). This potentially provides a novel approach to utilize brain wide relationships to study and characterize brain disorders, especially as large-scale cell type studies become available, supplying spatially localized expression of disease genes and details of coregulation. The general correspondence of structural and cell type approaches even restricted to a single cortical area, the middle temporal gyrus, while an imperfect association indicates a consensus organization and encourages examining these results in other areas using cell type data.

This work is complementary to studies on shared genetic heritability of common disorders of the brain. The *Brainstorm* consortium studied a large cohort GWAS meta-analysis demonstrating that common genetic variation contributes to the heritability of brain disorders, and showing that psychiatric disorders share common variant risk, while other neurological disorders appearing more distinct from one another and psychiatric disorders (57). This result is also seen in the present study, with a lower transcriptional variance in both structural and cell type profiling between schizophrenia, bipolar and autistic disorders compared with a wider range of anatomic patterning in other neurological disorders (**ADG 4.)** A striking finding is the enrichment of inhibitory neurons in psychiatric diseases and substance use disorders, and notable species-specific expression differences in these classes. For example, microglial involvement in Alzeimer's disease is well established and seen in **Fig 3D** and uniquely human enriched (**Fig. 4A**). While

there are several lines of evidence that inhibitory cell types are impaired in psychiatric disorders (58) (e.g. depression, bipolar disorder and schizophrenia, (59), there are, however, limitations to a cell type enrichment approach. In particular, some diseases may involve gene pathways shared across cells rather than involvement of subsets of cell types or brain regions, and we find that cell type enrichment of disease genes here does not necessarily match cell types with expression differences in disease vs. control tissue. Exploring the transcriptomic architecture of these disorders is a fully new field that has been underexplored and these findings support the transcriptomic hypothesis of vulnerability that in polygenic disorders, genes that are co-expressed in a certain brain region or cell type are much more likely to interact with each other than those that do not follow such a pattern (11, 12).

While previous work has shown conservation of neuronal enriched expression patterning between the mouse and human (14), a recent novel alignment of mouse and human cell types in middle temporal gyrus (15) now allowed for a more specific analysis. Here we show that the mouse appears to be evolutionarily sufficiently close to identify potentially relevant cell types and a striking conserved signature across subclass cell types for many diseases. While homology alignment of cell types between mouse and human may provide insight into convergent mechanisms based on species-specific differences, further human data is needed to implicate disease genes with cell function. We are aware of the different sampling protocols and numerous complications of cross-species comparison and our work addresses broad trends across diseases and cell types as opposed to differential expression of specific genes. Our results describe the structural and cellular transcriptomic landscape of common brain diseases in the adult brain providing an approach to characterizing the cellular basis of disorders as brain-wide cell type studies become available.

Methods

Disease genes. To obtain the gene disease associations, we used the DisGeNET database, a discovery platform with aggregated information from multiple sources including curated repositories, GWAS catalogues, animal models and the scientific literature. DisGeNET provides one of the largest gene-disease association collections. The data was obtained from the April 2019 update, the latest update related to the gene-disease association at the time of analysis. An original list of 500 diseases with connection to the brain were intersected with the provided repository at the DisGeNET. For each disease the main variant was selected, and rare familial/genetic forms were not included in the analysis. For this study, we

included genes with gene-disease association reported at least in one confirmed curated (i.e. UNIPROT, CTD™, ORPHANET, CLINGEN, GENOMICS ENGLAND, CGI, PSYGENET) (for details see https://www.disgenet.org/dbinfo). Since the goal of the study is to investigate the similarities and distinctions between brain-related disorders, disorders with less than 10 genes associated with them for the analysis were excluded. Finally, 15 disorders of peripheral nervous system or a second level association to the brain (e.g. retinal degeneration) were removed. This procedure resulted in 40 brain disorders with their corresponding associated genes. Finally, for these 40 disorders, we performed a literature review of the current GWAS studies to add all the missing genes from the DisGeNET dataset. These 40 diseases include brain tumors, substance related, neurodevelopmental, neurodegenerative, movement, and psychiatric disorders (**Supp. Figure 1**).

Dataset description. Anatomic based gene expression data was extracted from 6 post-mortem brains (13). The extracted samples were divided into 132 regions based on the anatomical/histological extraction regions. These 132 regions were further pooled/aggregated into 104 regions including cortex (CTX,8), hippocampus (HIP,7), amygdala (AMG,6), basal ganglia (BG,12), epithalamus (ET,3), thalamus (TH,10), ventral thalamus (VT,2), hypothalamus (HY,16), mesencephalon (MES, 11), cerebellum (CB,4), pons (P,8), pontine nuclei (PN,2), myelencephalon (MY, 12), ventricles (V,1), white matter (WM,2) (**Suppl. Table 3**). The resulting gene by region matrix was further averaged between subjects to produce one representative gene expression by region matrix. Each gene expression profile was further normalized across the brain regions. Cell type data in human is snRNA-seq from middle temporal gyrus (MTG) largely from postmortem brains (15). Nuclei were collected from eight donor brains representing 15,928 nuclei passing quality control, including those from 10,708 excitatory neurons, 4,297 inhibitory neurons and 923 non-neuronal cells. Cell type data from the mouse represents 23, 822 single cells isolated from two cortical areas (VISp, ALM) from the C57GL/6J mouse (18).

Cell-type specificity. Calculated based on the Tau-score defined in (47). This measure has previously been employed using the same dataset (15). Briefly, Cell-type specificity τ is defined as:

$$\tau = \frac{\sum_1^N (1 - x(i))}{(N-1)}$$

where, x(i) is the gene expression level in each cell-type for a given gene normalized by the maximum cell-type expression of that gene, and N is the number of the cell-types in the analysis.

Disease-Disease similarity index. In order to calculate the similarity between each pair of disorders we used the gene expression patterns across 104 brain structures, removing overlapping genes from each pair of disorders during clustering. In presenting heatmaps the full set of genes in each disease are averaged. Distance metric between diseases is 1 - ρ where ρ is Pearson correlation between structure or cell type profile also removing common genes. The procedure for disease similarity using cell-type data used the gene expression pattern across the 75 cell-types (instead of brain regions) in human cells extracted from MTG. For clustering in both cases we used agglomerative hierarchical clustering with Ward linkage algorithm (i.e. Ward.2 in R hclust function, R version 3.6.3)

Gene Expression Differential stability (DS). Gene expression differential stability was calculated for each gene as the similarity of its expression pattern across 6 post-mortem brains. For each pair of brains, the correlation of expression pattern across overlapping brain structures was calculated. The mean correlation over these 15 pairs was used as the differential stability for the given gene. (For more details see (14).

Disease-module association. For each gene the relationship between gene expression for each gene and module is calculated as explained in (14) (i.e. each gene expression pattern is correlated with eigengene pattern from modules across within each of 6 postmortem brains). These correlation values are then normalized using Fisher r-to-z transform and averaged across brains. For each module the gene associations were then standardized (mean=0, sd=1). Finally, these values are averaged across genes associated with each disease to calculate the disease module association.

Disease related gene expression within cell-types. We used expression-weighted cell-type enrichment (EWCE) analysis (https://bioconductor.riken.jp/packages/3.4/bioc/html/EWCE.html; (48) to identify cell types showing enriched gene expression associated with each of the 6 brain disorders. Briefly, EWCE compares the expression levels of the genes associated with a given disease to the background gene expression (all genes, excluding the disease-related genes) by performing permutation analysis and defining the probability for the observed expression level of the given gene set compared against a random set of genes. We used N=100,000 as the permutation parameter and performed the analysis at two cell-type category levels. The two levels included broad cell-types (N=7) and cell-subclasses (N= 20). The non-neuronal cell types were common between the two levels of analysis. These two levels were selected due to the availability of the homologous cell types in mouse and human cell dataset. Finally for

each disease, we used false discovery rate (*FDR*) correction (60) for multiple comparisons for disease-cell type associations.

Consensus Representation. Consensus UMAP was constructed by taking pairwise disjoint gene set correlation matrices for structural and cell type data, normalizing 2D UMAP embedding to maximum distance and averaging coordinates.

Statistical Analysis. All statistical analysis and visualization were conducted in R (www.r-project.org), a Jupyter notebook reproduces all analysis. To examine the differences in mean expression level between ADG groups we performed ANOVA tests. This was followed by direct comparisons between ADG pairs using unpaired t-test. All results were corrected for multiple comparisons using Benjamini-Hochberg correction controlling the false discovery rate. To examine the stability of the gene expression profiles, we repeated our analysis across 6 brains and searched for the matching pattern in other subjects for any given brain across ADG and GBD disease groups. The gene expression DS within each GBD group was compared to the general DS of all other genes in the dataset using independent t-tests.

Data Availability. All data used in this manuscript are publicly available. The gene disease association data can be downloaded from https://www.disgenet.org/. The large-scale anatomic transcriptional patterns can be downloaded from http://human.brain-map.org/ and cell type data is available at http://celltypes.brain-map.org/.

Code Availability

The script alongside a notebook file and all necessary data files for producing the figures are provided at https://github.com/yasharz/human-brain-disease-transcriptomics .

Acknowledgements

References

1.   Carroll LS, Williams NM, Moskvina V, Russell E, Norton N, Williams HJ, et al. Evidence for rare and common genetic risk variants for schizophrenia at protein kinase C, alpha. Mol Psychiatry. 2010;15(11):1101-11.

2.   DiLuca M, Olesen J. The cost of brain diseases: a burden or a challenge? Neuron. 2014;82(6):1205-8.

3.   Olesen J, Leonardi M. The burden of brain diseases in Europe. Eur J Neurol. 2003;10(5):471-7.

4.   Collaborators GBDV. Five insights from the Global Burden of Disease Study 2019. Lancet. 2020;396(10258):1135-59.

5.   Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. 2007;25(11):1251-5.

6.   Tropea D, Harkin A. Editorial: Biology of Brain Disorders. Front Cell Neurosci. 2017;11:366.

7.   Glorioso C, Sibille E. Between destiny and disease: genetics and molecular pathways of human central nervous system aging. Prog Neurobiol. 2011;93(2):165-81.

8.   Ding J, Adiconis X, Simmons SK, Kowalczyk MS, Hession CC, Marjanovic ND, et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. Nat Biotechnol. 2020;38(6):737-46.

9.   McGuire AL, Gabriel S, Tishkoff SA, Wonkam A, Chakravarti A, Furlong EEM, et al. The road ahead in genetics and genomics. Nature Reviews Genetics. 2020;21(10):581-96.

10.   Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nature Communications. 2017;8(1):14049.

11.   Kumar VJ, Grissom NM, McKee SE, Schoch H, Bowman N, Havekes R, et al. Linking spatial gene expression patterns to sex-specific brain structural changes on a mouse model of 16p11.2 hemideletion. Translational Psychiatry. 2018;8(1):109.

12.   Seidlitz J, Nadig A, Liu S, Bethlehem RAI, Vértes PE, Morgan SE, et al. Transcriptomic and cellular decoding of regional brain vulnerability to neurogenetic disorders. Nature Communications. 2020;11(1):3358.

13.   Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature. 2012;489(7416):391-9.

14.   Hawrylycz M, Miller JA, Menon V, Feng D, Dolbeare T, Guillozet-Bongaarts AL, et al. Canonical genetic signatures of the adult human brain. Nat Neurosci. 2015;18(12):1832-44.

15.   Hodge RD, Bakken TE, Miller JA, Smith KA, Barkan ER, Graybuck LT, et al. Conserved cell types with divergent features in human versus mouse cortex. Nature. 2019;573(7772):61-8.

16.   Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A, et al. Transcriptional landscape of the prenatal human brain. Nature. 2014;508(7495):199-206.

17.   Li M, Santpere G, Imamura Kawasawa Y, Evgrafov OV, Gulden FO, Pochareddy S, et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. Science. 2018;362(6420).

18.   Tasic B, Yao Z, Graybuck LT, Smith KA, Nguyen TN, Bertagnolli D, et al. Shared and distinct transcriptomic cell types across neocortical areas. Nature. 2018;563(7729):72-8.

19.   Pinero J, Ramirez-Anguita JM, Sauch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res. 2020;48(D1):D845-D55.

20.   Queralt-Rosinach N, Pinero J, Bravo A, Sanz F, Furlong LI. DisGeNET-RDF: harnessing the innovative power of the Semantic Web to explore the genetic basis of diseases. Bioinformatics. 2016;32(14):2236-8.

21.   Pinero J, Bravo A, Queralt-Rosinach N, Gutierrez-Sacristan A, Deu-Pons J, Centeno E, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res. 2017;45(D1):D833-D9.

22.   Chen J, Aronow BJ, Jegga AG. Disease candidate gene identification and prioritization using protein interaction networks. BMC Bioinformatics. 2009;10:73.

23.   Chen Y, Zhang X, Zhang GQ, Xu R. Comparative analysis of a novel disease phenotype network based on clinical manifestations. J Biomed Inform. 2015;53:113-20.

24.   Qi M, Fan S, Wang Z, Yang X, Xie Z, Chen K, et al. Identifying Common Genes, Cell Types and Brain Regions Between Diseases of the Nervous System. Front Genet. 2019;10:1202.

25.   Santucci K. Psychiatric disease and drug abuse. Curr Opin Pediatr. 2012;24(2):233-7.

26.   Granholm AC, Boger H, Emborg ME. Mood, memory and movement: an age-related neurodegenerative complex? Curr Aging Sci. 2008;1(2):133-9.

27.   Kelly J, Moyeed R, Carroll C, Albani D, Li X. Gene expression meta-analysis of Parkinson's disease and its relationship with Alzheimer's disease. Mol Brain. 2019;12(1):16.

28.   Pereira JB, Hall S, Jalakas M, Grothe MJ, Strandberg O, Stomrud E, et al. Longitudinal degeneration of the basal forebrain predicts subsequent dementia in Parkinson's disease. Neurobiol Dis. 2020;139:104831.

29.   Henderson JM, Carpenter K, Cartwright H, Halliday GM. Loss of thalamic intralaminar nuclei in progressive supranuclear palsy and Parkinson's disease: clinical and therapeutic implications. Brain. 2000;123 (Pt 7):1410-21.

30.   Vercruysse P, Vieau D, Blum D, Petersen A, Dupuis L. Hypothalamic Alterations in Neurodegenerative Diseases and Their Relation to Abnormal Energy Metabolism. Front Mol Neurosci. 2018;11:2.

31.   Miao J, Liu L, Yan C, Zhu X, Fan M, Yu P, et al. Association between ADORA2A gene polymorphisms and schizophrenia in the North Chinese Han population. Neuropsychiatr Dis Treat. 2019;15:2451-8.

32.   Ramaker RC, Bowling KM, Lasseigne BN, Hagenauer MH, Hardigan AA, Davis NS, et al. Post-mortem molecular profiling of three psychiatric disorders. Genome Med. 2017;9(1):72.

33.   Alheid GF, Heimer L. New perspectives in basal forebrain organization of special relevance for neuropsychiatric disorders: the striatopallidal, amygdaloid, and corticopetal components of substantia innominata. Neuroscience. 1988;27(1):1-39.

34.   Broeckel U, Schork NJ. Identifying genes and genetic variation underlying human diseases and complex phenotypes via recombination mapping. J Physiol. 2004;554(Pt 1):40-5.

35.   Prohaska A, Racimo F, Schork AJ, Sikora M, Stern AJ, Ilardo M, et al. Human Disease Variation in the Light of Population Genomics. Cell. 2019;177(1):115-31.

36.   Wang X, Sun W, Zhu X, Li L, Wu X, Lin H, et al. Association between the gamma-aminobutyric acid type B receptor 1 and 2 gene polymorphisms and mesial temporal lobe epilepsy in a Han Chinese population. Epilepsy Res. 2008;81(2-3):198-203.

37.   Soto D, Altafaj X, Sindreu C, Bayes A. Glutamate receptor mutations in psychiatric and neurodevelopmental disorders. Commun Integr Biol. 2014;7(1):e27887.

38.   Keo A, Mahfouz A, Ingrassia AMT, Meneboo JP, Villenet C, Mutez E, et al. Transcriptomic signatures of brain regional vulnerability to Parkinson's disease. Commun Biol. 2020;3(1):101.

39.   Zhai D, Li S, Zhao Y, Lin Z. SLC6A3 is a risk factor for Parkinson's disease: a meta-analysis of sixteen years' studies. Neurosci Lett. 2014;564:99-104.

40.   Lind PA, Eriksson CJ, Wilhelmsen KC. The role of aldehyde dehydrogenase-1 (ALDH1A1) polymorphisms in harmful alcohol consumption in a Finnish population. Hum Genomics. 2008;3(1):24-35.

41.   Arrue A, Gonzalez-Torres MA, Basterreche N, Arnaiz A, Olivas O, Zamalloa MI, et al. GAD1 gene polymorphisms are associated with bipolar I disorder and with blood homovanillic acid levels but not with plasma GABA levels. Neurochem Int. 2019;124:152-61.

42.   Bountress K, Sheerin C, Amstadter AB, Mandel H, Voltin J, Wang Z. The relation between GAD1 and PTSD symptoms: Shared risk for depressive symptoms. Psychiatry Res. 2017;258:607-8.

43.   Magri C, Giacopuzzi E, La Via L, Bonini D, Ravasio V, Elhussiny MEA, et al. A novel homozygous mutation in GAD1 gene described in a schizophrenic patient impairs activity and dimerization of GAD67 enzyme. Sci Rep. 2018;8(1):15470.

44.   Przybyla-Zawislak BD, Thorn BT, Ali SF, Dennis RA, Amato A, Virmani A, et al. Identification of rat hippocampal mRNAs altered by the mitochondrial toxicant, 3-NPA. Ann N Y Acad Sci. 2005;1053:162-73.

45.   van Amen-Hellebrekers CJ, Jansen S, Pfundt R, Schuurs-Hoeijmakers JH, Koolen DA, Marcelis CL, et al. Duplications of SLC1A3: Associated with ADHD and autism. Eur J Med Genet. 2016;59(8):373-6.

46.   Chivukula AS, Suslova M, Kortzak D, Kovermann P, Fahlke C. Functional consequences of SLC1A3 mutations associated with episodic ataxia 6. Hum Mutat. 2020.

47.   Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics. 2005;21(5):650-9.

48.   Skene NG, Grant SG. Identification of Vulnerable Cell Types in Major Brain Disorders Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment. Front Neurosci. 2016;10:16.

49.   Schall TA, Wright WJ, Dong Y. Nucleus accumbens fast-spiking interneurons in motivational and addictive behaviors. Mol Psychiatry. 2021;26(1):234-46.

50.   Charalambous C, Chen TC, Hofman FM. Characteristics of tumor-associated endothelial cells derived from glioblastoma multiforme. Neurosurg Focus. 2006;20(4):E22.

51.   Zong H, Verhaak RG, Canoll P. The cellular origin for malignant glioma and prospects for clinical advancements. Expert Rev Mol Diagn. 2012;12(4):383-94.

52.   Romanos J, Benke D, Pietrobon D, Zeilhofer HU, Santello M. Astrocyte dysfunction increases cortical dendritic excitability and promotes cranial pain in familial migraine. Sci Adv. 2020;6(23):eaaz1584.

53.   Blackstone C. Cellular pathways of hereditary spastic paraplegia. Annu Rev Neurosci. 2012;35:25-47.

54.   Dozmorov MG. Disease classification: from phenotypic similarity to integrative genomics and beyond. Brief Bioinform. 2019;20(5):1769-80.

55.   Cheon SY, Kim EJ, Kim JM, Koo BN. Cell Type-Specific Mechanisms in the Pathogenesis of Ischemic Stroke: The Role of Apoptosis Signal-Regulating Kinase 1. Oxid Med Cell Longev. 2018;2018:2596043.

56.   Cross-Disorder Group of the Psychiatric Genomics C, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet. 2013;45(9):984-94.

57.   Brainstorm C, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. Analysis of shared heritability in common disorders of the brain. Science. 2018;360(6395).

58.   Schmidt MJ, Mirnics K. Neurodevelopment, GABA system dysfunction, and schizophrenia. Neuropsychopharmacology. 2015;40(1):190-206.

59.   Jansen PR, Watanabe K, Stringer S, Skene N, Bryois J, Hammerschlag AR, et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. Nat Genet. 2019;51(3):394-403.

60.   Benjamini Y, Hochberg, Y. Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. Journal of the Roya Statistical Society Series B: Methodological. 1995(57):289-300.