Neurological and psychiatric disorders, including substance use disorders, share a range of symptoms, which could be the result of shared genetic background. Many genetic loci have been identified for these disorders using genome-wide association studies, but conclusive evidence about cell types wherein these loci are active is lacking. We aimed to uncover implicated brain cell types in neuropsychiatric traits and to assess consistency in results across RNA datasets and methods. We therefore comprehensively employed cell type enrichment methods by integrating single-cell transcriptomic data from mouse brain regions with an unprecedented dataset of 42 human genome-wide association study results of neuropsychiatric, substance use and behavioral/quantitative brain-related traits \( (n = 12,544,007 \text{ individuals}) \). Single-cell transcriptomic datasets from the Karolinska Institute and 10x Genomics were used. Cell type enrichment was determined using Linkage Disequilibrium Score Regression, Multi-marker Analysis of GenoMic Annotation, and Data-driven Expression Prioritized Integration for Complex Traits. We found the largest degree of consistency across methods for implication of pyramidal cells in schizophrenia and cognitive performance. For other phenotypes, such as bipolar disorder, two methods implicated the same cell types, i.e., medium spiny neurons and pyramidal cells. For autism spectrum disorders and anorexia nervosa, no consistency in implicated cell types was observed across methods. We found no evidence for astrocytes being consistently implicated in neuropsychiatric traits. In conclusion, we provide comprehensive evidence for a subset of neuronal cell types being consistently implicated in several, but not all psychiatric disorders, while non-neuronal cell types seem less implicated.

Molecular Psychiatry (2022) 27:947–955; https://doi.org/10.1038/s41380-021-01324-6

INTRODUCTION

It is well documented that several neuropsychiatric disorders, including substance use disorders (SUDs), share symptoms, which could be the result of shared genetic underpinnings \cite{1, 2}. Much of the heritability \( (h^2) \) of genetically complex or polygenic brain disorders—e.g., schizophrenia (SCZ), Parkinson’s disease, and alcohol use disorder—is due to common genetic variation \cite{3}. In addition, genome-wide association studies (GWASs) have deepened the understanding of such disorders, unraveling thousands of associated loci \cite{4, 5}. However, elucidating disease mechanisms has remained challenging. One reason is missing heritability, meaning the gap between twin-based and single-nucleotide polymorphism (SNP)-based \( h^2 \) estimates, which may result from limited statistical power, phenotypic heterogeneity, clinical misclassification, GWASs not probing associations with rare variants, epigenetics, genomic interactions, and structural genomic alterations \cite{6}. Due to missing heritability, underlying causal genetic contributors might remain undetected. This impedes the translation of GWAS associations into their functional effects. Another reason explaining why elucidating disease mechanisms in neuropsychiatric disorders has remained challenging is that over 90% of identified variants are located within non-coding regions of the genome, indicating that regulatory elements—e.g., promoters, enhancers, and insulators—may explain part of the underlying genetic mechanisms in some polygenic disorders \cite{4, 7}. Due to extensive linkage disequilibrium (LD), it is also challenging to identify a causal variant within a given associated locus \cite{4}.

To overcome gaps between associated and causal genetic loci, their functional effects, and ultimately the biological pathways, extensive research has been performed to identify brain tissues having a role in neuropsychiatric disease. Functional genomic studies using macroscopic brain samples point to enrichment in phylogenetically conserved areas of the brain in psychiatric disorders and brain-related behavioral phenotypes, whereas typically fewer brain regions are found to be enriched in neurological disorders \cite{3}. However, identification of specific cell types within brain tissues is considerably less well studied. Specific cell types that are associated with SCZ and anorexia nervosa (AN)
have previously been identified by integrating GWAS findings with mouse single-cell RNA (scRNA) brain data: while medium spiny neurons (MSNs), cortical interneurons, hippocampal CA1 pyramidal cells (pyramidal CA1), and pyramidal cells from the somatosensory cortex (pyramidal SS) seem implicated in SCZ [8], suggestive findings were reported for the enrichment of MSNs and pyramidal cells (CA1) in AN [9]. Recently, more extensive cell type enrichment analysis was performed for 28 phenotypes using mouse gene expression from the entire central nervous system (CNS) [10]. In psychiatric disorders, enrichment was found for MSNs, cortical interneurons, striatal interneurons, neuroblasts, pyramidal cells (CA1 and SS) [10]. In neurological disorders, fewer cell types were identified and these were dissimilar across disorders [10]. These cell type enrichment analyses have mainly been performed using Linkage Disequilibrium Score Regression (LDSC) and/or Multi-marker Analysis of GenoMic Annotation (MAGMA). However, in this landmark and other studies, MAGMA versions <1.08 have been employed [8–10], of which it was recently reported that its SNP-level P value aggregation into gene-level P values might result in type-I errors [11]. In addition, cell type enrichment in SUDs and several other disorders, such as anxiety disorders, has to the best of our knowledge not been studied.

Here, we systematically investigated cell type enrichment in an extensive set of brain-related phenotypes by integrating mouse scRNA brain data from the Karolinska Institute (KI) and 10x Genomics with summary statistics from 42 phenotypes related to neuropsychiatry, SUDs, and brain-related behavior. Our goals were to perform cell type enrichment on a more comprehensive set of brain-related traits than previously studied and to assess consistency in results across a wider array of methods. We went beyond previous studies by systematically performing cell type enrichment analyses using the most recent releases of different methods that rely on different assumptions and algorithms, i.e., LDSC, MAGMA v1.08, Data-driven Expression Prioritized Integration for Complex Traits (DEPICT), and Functional Mapping and Annotation (FUMA). We found evidence for a subset of neuronal cell types being consistently implicated in several, but not all psychiatric disorders, while non-neuronal cell types seem less implicated.

METHODS

GWAS summary statistics

Our goals were to identify salient cell types that are implicated in more brain-related traits than previously studied and to assess consistency in results across methods. Brain-related GWAS summary statistics from predominantly European samples were obtained from publicly available sources. A total of 41 summary statistics from brain-related phenotypes (Table 1, Table S1) were obtained, among which 11 psychiatric disorders (486,142 cases and 1,002,695 controls), 11 neurological disorders (186,171 cases and 2,278,970 controls), and 8 substance use disorders (case/control: 11,569 cases and 34,999; cohorts with continuous substance use data: n = 3,683,037). All psychiatric and neurological disorders for which summary statistics were available had also been included in the Brainstorm project [3]. Substance use disorders were added because of the high comorbidity [12,13] and genetic covariance [14] with psychiatric traits. Eleven well-powered (n = 250,000) brain-related behavioral/quantitative phenotypes (n = 4,166,895) were additionally selected. Because of the association between BMI and brain structure, we considered BMI a brain-related trait than previously studied and to assess consistency in results across a wider array of methods. We went beyond previous studies by systematically performing cell type enrichment analyses using the most recent releases of different methods that rely on different assumptions and algorithms, i.e., LDSC, MAGMA v1.08, Data-driven Expression Prioritized Integration for Complex Traits (DEPICT), and Functional Mapping and Annotation (FUMA). We found evidence for a subset of neuronal cell types being consistently implicated in several, but not all psychiatric disorders, while non-neuronal cell types seem less implicated.

Single-cell RNA sequencing datasets

All cell type enrichment analyses were conducted using the KI dataset [8,16–18] and the 10x Genomics dataset [19]. These datasets were selected because they cover brain regions that are generally accepted to be involved in the pathogenesis of brain-related disorders [20]. Additionally, their high coverage may enable the identification of different cell types. Detailed information about the 10x Genomics dataset, quality control, necessity of a randomized representative subset of cells, and cell type identification are reported in the Supplementary Methods. The quality control of the KI dataset is described elsewhere [16].

Overview of cell type enrichment analyses

To identify cell types underlying various phenotypes, we employed four methods (Fig. 1). LDSC (version 1.0.1) [21,22] was first used to estimate SNP-h2 and bivariate genetic correlations across all traits. We then constructed the specificity metric Sg,c from the 10x Genomics dataset, denoting the specificity of a gene g for cell type c, which was used as input for LDSC, MAGMA and DEPICT (Supplementary Methods). The specificity metric of the KI dataset was previously constructed [8]. Next, LDSC (version 1.0.1) [21,23], MAGMA (version 1.08) [8,24] and DEPICT (version 1, release 194) [25] were employed to test cell type enrichment using the 10x Genomics and KI datasets. We utilized the diversity in statistical approaches of these methods to strengthen the confidence in consistently enriched cell types across methods. In brief, with LDSC we investigated the top 10% cell type-specific genes for enrichment of SNP-h2, with MAGMA we tested whether trait associations linearly increased with cell type-specificity or whether the top 10% cell type-specific genes were associated with gene-level association to traits; with DEPICT we evaluated the enrichment of cell type with gene-level association to traits.

LDSC computes an LD score by summarizing the correlations of a given SNP with all neighboring SNPs within 100 kb flanks. Then, the GWAS test statistic χ2 is regressed against the LD score, of which the slope is rescaled into an estimate of SNP-h2, explained by all SNPs included in the LD score. Based on Sg,c, we binned genes in specificity deciles and used LDSC to test for enrichment of SNP-h2 in the top 10% most associated genes. MAGMA aggregates χ2 association statistics within a 10 kb upstream and 1.5 kb downstream window into a gene-level P value (Supplementary Methods). To compute gene-level P values, Brown’s method [26] was updated into lmhoF’s approach [27] in version 1.08 [28]. DEPICT maps genes to loci by first selecting all significant SNPs and preserving lead SNPs as the most significant SNP out of possible SNP-pairs in LD (r2 > 0.1) and/or within < 1 MB distance of each other (Supplementary Methods). Then, boundaries of the most distal SNP on either side around lead SNPs (r2 > 0.5) were used as criteria to list genes from SNPs, from which potential enrichment of these genes to specific cell types could later be assessed. By comparing cell type enrichment results of LDSC, DEPICT, and MAGMA, we evaluated the relative stringency of each method. This was accomplished by comparing the P values, denoting the strength of association of a given cell type with a given phenotype, of any two methods with one another.

Finally, additional human and mouse scRNA datasets (Table S2) were used to conduct additional cell type enrichment analysis, using FUMA (version 1.3.6a) [29] (Supplementary Methods). FUMA builds on MAGMA (version 1.08). However, FUMA includes averaged expression per gene across cell types and the sensitive enrichment analysis with FUMA was considered here to be a good addition to the other methods.

Cell type enrichment using LDSC

Human orthologs were obtained using the One2One R package that is incorporated in the MAGMA_Celotyping R package [8]. SNPs were annotated to the human genome (hg19), version 33 of the GENCODE project [30]. Binary annotations files were created for each cell type, containing 11 sub-annotations. The first sub-annotation contained SNPs that mapped to genes without a human ortholog (1–SNP belongs to a sub-annotation). The other ten sub-annotations represented the SNPs in specific loci, defined by a particular cell type in increasing order. These specificity deciles were obtained by restructoring the specificity metric Sg,c described in the Methods section “Overview of cell type enrichment analyses” using the preparequantile function in the MAGMA_Celotyping package [8]. LD scores were then...
| Phenotype descriptions | Phenotype Cases | Controls | Total number of participants | Source | Ancestry |
|------------------------|----------------|----------|-----------------------------|--------|----------|
| **Psychiatric disorders** |                |          |                             |        |          |
| Attention-deficit/hyperactivity disorder (ADHD) | 19,099 | 34,194 | 53,293 | PGC | European |
| Anorexia nervosa (AN) | 16,992 | 55,525 | 72,517 | PGC | European |
| Anxiety disorders | 7016 | 14,745 | 21,761 | ANGST | European |
| Autism spectrum disorder (ASD) | 18,382 | 27,969 | 46,351 | PGC | European |
| Bipolar disorder (BIP) | 20,352 | 31,358 | 51,710 | PGC | European |
| Cross disorders | 162,151 | 276,846 | 438,997 | PGC | European |
| Major depressive disorder (MDD) | 170,756 | 329,443 | 500,199 | PGC | European |
| Obsessive-compulsive disorder (OCD) | 2688 | 7037 | 9725 | PGC | European |
| Post-traumatic stress disorder | 23,212 | 151,447 | 174,659 | PGC | European |
| Schizophrenia (SCZ) | 40,675 | 64,643 | 105,318 | PGC | European |
| Tourette syndrome | 4819 | 9488 | 14,307 | PGC | European |
| **Substance use disorders** |                |          |                             |        |          |
| Alcohol use | N/A | N/A | 121,604 | PGC | European |
| Alcohol dependence | 11,569 | 34,999 | 46,568 | PGC | European |
| Drinks per week | N/A | N/A | 941,280 | GSCAN | European |
| Cannabis use | N/A | N/A | 162,082 | PGC | European |
| Age smoking initiation | N/A | N/A | 341,427 | GSCAN | European |
| Ever smoked regularly | N/A | N/A | 1,232,091 | GSCAN | European |
| Cigarettes per day | N/A | N/A | 337,334 | GSCAN | European |
| Smoking cessation | N/A | N/A | 547,219 | GSCAN | European |
| **Neurological disorders** |                |          |                             |        |          |
| Amyotrophic lateral sclerosis | 20,806 | 59,804 | 80,610 | AVS | European |
| Alzheimer’s disease | 71,880 | 383,378 | 455,258 | PGC | European |
| Epilepsy | 15,212 | 29,677 | 44,889 | ILAE | Multi-ancestry |
| Generalized epilepsy | 3769 | 29,677 | 33,446 | ILAE | Multi-ancestry |
| Focal epilepsy | 9671 | 29,677 | 39,348 | ILAE | Multi-ancestry |
| Stroke | 40,585 | 406,111 | 446,696 | MEGASTROKE | European |
| Ischemic stroke | 34,217 | 406,111 | 440,328 | MEGASTROKE | European |
| Large artery stroke | 4373 | 406,111 | 410,484 | MEGASTROKE | European |
| Cardioembolic stroke | 7193 | 406,111 | 413,304 | MEGASTROKE | European |
| Small vessel stroke | 5386 | 406,111 | 411,497 | MEGASTROKE | European |
| Parkinson’s disease | 37,688 | 1,400,000 | 1,437,688 | IPDGC-PDWBS- SGPD | European |
| **Behavioral/quantitative** |                |          |                             |        |          |
| Body mass index (BMI) | N/A | N/A | 681,275 | GIANT | European |
| Chronotype | N/A | N/A | 449,734 | SDKP | European |
| Excessive daytime sleepiness | N/A | N/A | 452,071 | SDKP | European |
| Sleep duration | N/A | N/A | 446,118 | SDKP | European |
| Short sleep duration | 106,192 | 305,742 | 411,934 | SDKP | European |
| Long sleep duration | 34,184 | 305,742 | 339,926 | SDKP | European |
| Insomnia | N/A | N/A | 453,379 | SDKP | European |
| Intelligence | N/A | N/A | 269,867 | CTGLAB | European |
| Educational attainment | N/A | N/A | 766,345 | SSGAC | European |
| Cognitive performance | N/A | N/A | 257,828 | SSGAC | European |
| Neuroticism | N/A | N/A | 390,278 | CTGLAB | European |
| **Non-brain-related control** |                |          |                             |        |          |
| Height | N/A | N/A | 693,529 | GIANT | European |
| Total | **683,882** | **3,316,664** | **12,544,007** | | |

Detailed descriptions, including references, are listed in Table S1.
PGC Psychiatric Genomics Consortium; ANGST Anxiety Neuro Genetics STudy, GSCAN GWAS and Sequencing Consortium of Alcohol and Nicotine use, AVS ALS Variant Server, ILAE International League Against Epilepsy, IPDGC International Parkinson’s Disease Genomics Consortium, SGPD Systems genomics of Parkinson’s disease consortium, PDWBS Parkinson’s disease web based study, SDKP Sleep Disorder Knowledge Portal, CTGLAB Complex Traits Genetics Lab, SSGAC Social Science Genetic Association Consortium.

*May include sample overlap with AN, ADHD, ASD, BIP, MDD, OCD, SCZ, and Tourette syndrome.

*Sample count of a phenotype that is part of larger group.
Brain transcriptomics datasets
- **Data format:**
  - \( S \) = Expression of gene \( g \) in cell type \( c \) relative to all cell types

GWAS datasets
- **Phenotypes** (n=42):
  - Psychiatric (n=11)
  - Neurological (n=11)
  - Substance use (n=8)
  - Behavioral / Quantitative (n=11)
  - Non-brain control (n=1)

Combined enrichment methods
- DEPICT
- MAGMA
- LDSC

Enriched brain cell types
- **Comparison with external cell types**
  - SNP-heritability
  - Genetic correlations

| **External datasets (n=18)** |
|-----------------------------|
|  - Human brain transcriptomics (n=7) |
|  - Mouse brain transcriptomics (n=11) |

**10x Genomics (cells=100,436)**
- **Gene expression patterns**
- **Cell type identification (n=16)**

**KI (cells=9,970)**
- **Annotated cell types**
  - First level of 24 cell types, second level of 149 cell types (sub-types)

**Combined enrichment methods**
- DEPICT
- MAGMA
- LDSC

Fig. 1 Overview of the approach of dataset integration as inputs for enrichment methods, in order to detect implicated brain cell types for various phenotypes. Two mouse brain transcriptomic datasets (10x Genomics, KI) have the data format \( S_{gc} \) of cell type specificity for genes, which was calculated by dividing expression of gene \( g \) in cell type \( c \) by expression of \( g \) in all cell types of a given dataset. Custom cell type identification was performed for 10x Genomics (16 detected cell types), while existing annotation was re-used for KI (first level of 24 cell types and second level of 149 cells (sub-types)). The datasets were integrated with genome-wide association study (GWAS) data, and these were the input for cell type enrichment methods DEPICT, MAGMA and LDSC. External human and mouse brain transcriptomics data were used in cell type enrichment method FUMA, so that enriched cell types from any of the other three methods could be compared to FUMA-enriched cell types. Finally, LDSC was also used to estimate SNP-based heritability for each GWAS phenotype and to calculate genetic correlations across all phenotypes.

**RESULTS**

**Cell type-specific gene expression in the 10x Genomics dataset**
- In the quality control of a randomized representative subset (n = 108,844) of the 10x Genomics dataset, 8408 cells and 6419 non-expressed genes were removed from further analyses (Figs. S1–S3).
- Altogether, the subset consisted of a matrix with 21,579 genes and 100,436 cells with 16 cell clusters (Tables S3, S4). All cell clusters were subsequently mapped to brain cell types by specifically expressed marker genes (Figure S4).

**LDSC, MAGMA, and DEPICT sensitivity analyses and quality control**
- For cell type enrichment analyses using LDSC, we initially adopted the same parameters that were previously described [8]. Additionally, to optimize the cell type enrichment pipeline we tested various settings (Fig. S5, Table S5). The parameters described in the Methods section “Cell type enrichment using LDSC” provided consistent cell type enrichment results that were most consistent with DEPICT and MAGMA. For MAGMA, we found that restricting to HapMap3 SNPs and excluding the MHC region increased statistical power to identify associated cell types, whilst not inflating cell types that were not associated (Fig. S6, Table S6). In addition to MAGMA version 1.08, we also performed cell type enrichment analyses using MAGMA version 1.07b (Table S7). We found that, although cell type associations follow similar patterns using both versions, the updated SNP-wise mean gene analysis model modestly exerts effects on cell type enrichment results, resulting in differently associated cell types (Fig. S7). However, no consistent unidirectional differences in cell type enrichment results were observed. Finally, for DEPICT, we found that not restricting to HapMap3 SNPs increased statistical power to identify associated cell types, without an upwards bias for non-associated cell types (Fig. S8, Table S8).

**Cell type enrichment analyses using the 10x Genomics dataset**
- Consistent with SNP-h2 estimate patterns (Table S9, Figs. S9, S10), Supplementary Methods, Supplementary Results) and genetic correlations (Table S10, Figs. S11, S12, Supplementary Methods, Supplementary Results), we found that cell type association patterns of neurological disorders were distinct from psychiatric, substance use and behavioral/quantitative association patterns by...
hierarchical clustering (Figs. S13, S14). For brain-related phenotypes, there were no cell types in the 10x Genomics dataset to which implicated genomic loci consistently mapped using all three methods (Figs. S13–S15, Table S11). By comparing the P values of each cell type computed by LDSC, MAGMA, and DEPICT against one another, we found that MAGMA linear mode was considerably more lenient (Fig. S16) and susceptible to bias due to GWAS sample size (Fig. S17). Therefore, MAGMA linear mode was excluded from our main analyses.

Two methods provided evidence for certain neurons to be implicated in cross-disorders (eight psychiatric disorders jointly studied) [5]. A deeper cellular determination was not possible due to low sequencing depth. The same neuronal cells were also associated with educational attainment along with certain interneurons according to two methods. Additionally, these neurons were associated with intelligence. We also found evidence by two methods that implicated genomic loci of cognitive performance specifically mapped to certain neuroblasts. Suggestive findings are reported in the Supplementary Results.

**Cell type enrichment analyses using the KI dataset**

After integrating GWAS findings with the 10x Genomics dataset, we leveraged cell type enrichment analyses by using the KI dataset, which is distinct from the 10x Genomics dataset because of its better coverage, higher resolution, and higher specificity to identify cell type-specific gene expression markers. Therefore, considerably more cell types were identified on a deeper cellular level with the KI dataset. The largest degree of consistency across methods in brain-related traits was found for SCZ and cognitive performance (Figs. 2 and 3, Figs. S18, S19, Table S12). Genetic loci that are associated with SCZ consistently mapped to excitatory pyramidal cells (CA1 and SS), while those associated with cognitive performance only mapped to pyramidal cells (SS). For SCZ, we found evidence by two methods that MSNs were the main implicated inhibitory neurons. MSNs and both types of pyramidal cells were found to be associated with cross-disorders by two methods, while only pyramidal cells were associated with educational attainment, and MSNs and pyramidal cells (CA1) were implicated in bipolar disorder. Suggestive findings are reported in the Supplementary Results.

To identify more differentiated cell types, the analysis was expanded by using the KI level 2 dataset (Figs. S20–S22, Table S13), which includes 149 cell types that were subtypes of the cell types identified in the level 1 dataset. Using both the KI level 1 and level 2 datasets, we again found that MAGMA linear mode, compared to strength of association estimates computed by LDSC, MAGMA top 10% mode and DEPICT, provided disproportionately large estimates (Figs. S23, S24) and was prone to inflated results due to GWAS sample size (Fig. S25).

**Cell type enrichment analyses using additional scRNA datasets**

Finally, we performed additional analyses with FUMA using additional mouse (n = 11) and human (n = 7) gene expression datasets to compare our findings and to assess consistency between species. We were able to identify at least one implicated cell type in 22 phenotypes (Fig. S26, Table S14). Using human gene expression data, 24 cell types were enriched in at least one phenotype, while using mouse scRNA data 70 cell types were revealed. Consistent with findings from 10x Genomics and KI datasets, pyramidal cells from various mouse brain regions, including pyramidal cells (CA1 and SS), were implicated in SCZ and cognitive performance. Pyramidal cells were also enriched in numerous psychiatric disorders, SUDs, and behavioral/quantitative phenotypes. Along with pyramidal cells, inhibitory GABAergic and MSNs were consistently enriched in psychiatric disorders, SUDs, and behavioral/quantitative phenotypes. Using human datasets, enrichment of pyramidal cells (CA1) was confirmed in SCZ, cognitive performance, intelligence, cross-disorders, and overall sleep duration. Additionally, pyramidal cells (CA1) were enriched in cigarettes smoked per day. However, the strongest consistent evidence was found for enrichment of GABAergic neurons from the prefrontal cortex and midbrain in various psychiatric disorders, SUDs, and behavioral/quantitative phenotypes. Consistent with findings using 10x Genomics and KI data, fewer enriched cell types were identified in neurological disorders. Human microglia were enriched in Alzheimer’s disease and human inhibitory GABAergic neurons from the prefrontal cortex and midbrain in generalized epilepsy. Generalized epilepsy was the only neurological phenotype in which mouse cell types were identified, namely certain pyramidal neurons and certain inhibitory neurons. We thus largely confirmed the main cell type enrichment findings from the 10x Genomics and KI dataset in mouse datasets and in human datasets using FUMA.

**DISCUSSION**

Here, we provide a comprehensive overview of specific brain cell types implicated in a range of brain-related phenotypes using both mouse and human brain scRNA data. We show that results from brain-related GWAS data consistently map to excitatory pyramidal neurons (CA1 and SS) and inhibitory MSNs and less so to glial and embryonic cells. The largest degree of consistency across methods and tissue origins (rodent and human) was found for implication of pyramidal cells in schizophrenia and cognitive performance. In contrast, GWAS data from neurological disorders mapped to fewer cell types and also to cell types that were distinct from psychiatric and substance use disorders.

Our SNP-h² and genetic correlation findings confirm that neurological disorders are genetically distinct from one another and from psychiatric disorders and SUDs, as well as from behavioral/quantitative phenotypes, which is in line with previous evidence [3, 10]. Consistent with these findings, we found that GWAS findings from psychiatric disorders, SUDs, and brain-related behavioral/quantitative phenotypes, but not neurological disorders, consistently map to excitatory hippocampal pyramidal neurons (CA1), excitatory pyramidal neurons (SS) and inhibitory MSNs, and much less to glial and embryonic cells. This concurs with previous lines of evidence pointing to neurological disorders being genetically and functionally distinct from psychiatric disorders, SUDs, and brain-related behavioral/quantitative phenotypes [3, 10]. Alzheimer’s disease was the only malady targeted here that showed evidence of exclusively human glial cells being implicated, underscoring the importance of key transcriptomic differences between human and mouse microglial signatures [33]. We confirmed our main findings with multiple external scRNA datasets using FUMA. This provides further evidence that genetic underpinnings of neurological disorders are distinct from those of psychiatric, SUDs, and behavioral/quantitative phenotypes [8, 10]. Our main findings were based on the identification of cell types by LDSC, DEPICT, and MAGMA top 10% mode. MAGMA linear mode was omitted because its strength of association estimates were consistently deviating substantially from LDSC, DEPICT, and MAGMA top 10% mode. Therefore, it was deemed too lenient and thus prone to type I error inflation. This concurs with previous studies reporting that binned MAGMA analyses in linear mode inflated results since the binned scores can have strong correlations with the average gene expression across cell types [29]. Also in agreement with previous lines of evidence, we confirm that the statistical foundation of the SNP-wise mean gene analysis model MAGMA < 1.07 may result in biased associations of cell types [11].

We envision three broad implications for psychiatry as a medical specialty. First, our results consistently point to specific neuronal cell types being implicated in several psychiatric disorders. Therapeutic targeting of those cells could one day result in innovative treatments. Second, gene sets that are
specifically enriched in those cells (for instance hippocampal pyramidal cells) could be used for risk scoring to stratify patients and tailor therapy. Third, for the disorders with less consistent results, clinicians and researchers should aim to collect more samples and thus ensure future studies may shed light on their implicated cell types.

The discrepancy between the KI- and 10x Genomics-derived cell types could be a consequence of a lower sequencing depth in the

Fig. 2 Cell type enrichment estimated by DEPICT, LDSC, and MAGMA top 10% mode in selected brain-related phenotypes. Cell type enrichment results are generated using KI data. Bars represent the mean strength of association (-log10(P)) of LDSC, DEPICT, and MAGMA top 10% mode. The red line indicates the Bonferroni threshold $P < 0.05/(24*42)$. The red line is solid if any of the methods identified any cell type as significantly associated, and if none of the methods identified any of the cell types as significantly associated, the red line is dashed. A complete overview of cell type enrichment results using KI data, including MAGMA linear mode is available in the supplementary information (Fig. S18).
between 25,000 and 50,000 mapped reads per cell [34]. This minimum sequencing depth is generally considered to be KI dataset (~1.2 million mapped reads per cell). Notably, the 10x Genomics dataset (~18,500 mapped reads per cell) than in the ADHD; attention de
dataset.

Fig. 3 Overview of enriched cell types of 42 common-variant psychiatric, neurologic, and behavioral/quantitative GWAS results in the KI
dataset. ADH
H; attention deficit hyperactivity disorder, ALS; amyotrophic lateral sclerosis, BMI; body mass index. Analyses from LDSC, DEPICT,
and MAGMA top 10% mode, referred to as “methods” in the graph, show enrichment in MSNs and pyramidal cells (CA1) and pyramidal cells
(SS) across brain-related phenotypes. The largest degree of consistency was found in SCZ and cognitive performance. Phenotypes and cell
types are grouped by hierarchal clustering. Shades of pink are proportional to the mean strength of association (−log10(P)) of all methods. The
color of the frames refers to the number of methods that identified a given cell type as significant in a given phenotype, after Bonferroni
correction (P < 0.05/(24*42)). Gray frames: one method (intelligence, excessive daytime sleepiness, ADHD, drinks per week, ever smoked,
chronotype, overall sleep duration, short sleep duration, MDD). Black frames: two methods (cross-disorders, educational attainment, BIP). Red
frames: all three methods (human height, cognitive performance, SCZ).

Although we provide new insight with the largest and most comprehensive study of cell type enrichment in brain-related
disorders, our results should be interpreted in light of inevitable limitations. First, using MAGMA it is possible to test whether the
genes specific to a phenotype are enriched in genetic associations of that phenotype while controlling for genetic associations of
another phenotype [10]. However, as our main goal was to identify enriched cell types, such conditional analyses are beyond the
scope of this study. Second, we found that microglia associated with age-induced neuroinflammation were exclusively found to be
enriched in Alzheimer’s disease using human scRNA datasets, whereas no enriched glial cells were identified using mouse scRNA
datasets. Therefore, mouse gene expression data from not only a spatial, but also a temporal resolution is warranted for future
research to identify cell types implicated in disease during development. Additionally, improved coverage of brain-related
regions, such as the entire CNS [10], is warranted for future research. Third, a potential limitation of the FUMA model is that
averaging gene expression disregards that low expression levels of certain genes can still be relevant to disease. This caveat
illustrates the challenges of accounting for factors potentially inflating statistical results as well as capturing etiological
mechanisms by examining cell type-specific gene expression [10]. Finally, to identify enriched cell types, we integrated human
genomic findings with mouse scRNA brain expression data. Although considerable differences exist between mice and
humans, we believe our choice is justified because mouse scRNA
datasets cover transcripts that are missed in human single-nuclei
sequencing [8]; cover more brain regions that are believed to be
important in neuropsychiatric disorders than in human, such as
the striatum for which no human scRNA expression data is
available; and reveal key findings consistent with human data
[8, 9]. In addition, gene expression data from rodents are often of
higher quality, as fresh tissue can be more readily obtained and
gene expression data cluster by cell type in different species,
rather than by different cell types in the same species [8].
Although we believe that mouse scRNA data are suitable to apply
to human GWAS data, there are limitations, e.g., less conserved
brain regions might contain cell types that express genes
differently; cell types could be specific to certain species (on the
other hand, it has been shown that gene expression in the brain,
including key gene expression patterns, is well conserved across
species [32]); and cell types could have different functions or could
be connected to and active in different brain circuits.

The identification of a specific subset of brain cell types being
implicated in various brain disorders only marks the beginning of elucidating causal biological pathways. One question future
research should address is what the effects of genetic variants in
the non-coding genome are. One way to address this question is
using an activity-by-contact model [37]. This model allows for the
identification of cell type-specific enhancers and their target genes by leveraging single-cell chromatin accessibility and enhancer activity data. Additional insight could be obtained by performing cell prioritization analyses from human post-mortem brain samples and/or induced pluripotent stem cells from individuals with relevant genetic backgrounds using LDSC, MAGMA, and DEPICT to identify genes that are predicted to be functionally similar to causal genes. Importantly, enriched cell types are not necessarily causal, but might be part of a neural network. To confirm our findings and to elucidate causality, selective chemo- and optogenetic manipulation of identified enriched cell types in rodents might provide additional insight in the role these cells play in the neural circuit underlying brain disorders [38]. Additionally, the recently developed computational toolkit CELL-type Expression-specific integration for Complex Traits (CELLECT) can provide additional insight in cell type enrichment [39]. CELLECT builds upon gene prioritization models, such as LDSC, DEPICT, and MAGMA and subsequently performs cell type prioritization analyses using a continuous representation of cell type expression, rather than binary representation. Finally, statistical power is currently a major challenge in genetic studies. Future studies might benefit from multi-variant analysis of GWAS [40], which is a method for analysis of multiple GWASs, thereby increasing the statistical power of each trait analyzed to identify genetic associations.

In sum, by incorporating different tools that rely on different assumptions and algorithms we provide robust evidence for a subgroup of neuronal cell types consistently implicated in several brain-related phenotypes. We thus provide a framework that furthers the understanding of cell types involved in brain-related phenotypes at a cellular level that can serve as a basis for future, more hypothesis-driven research.

DATA AVAILABILITY
All scRNA datasets used in this study are publicly available. All summary statistics are publicly available and the sources are listed in Table S1. We have made our code publicly available at https://github.com/mitchellolislagers/cell_type_enrichment_pipeline so that with the advent of new GWASs researchers may readily apply our pipeline to new data.

REFERENCES
1. Armstrong TD, Costello EJ. Community studies on adolescent substance use, abuse, or dependence and psychiatric comorbidity. J Consult Clin Psychol. 2002;70:1224–39.
2. Couwenbergh C, van den Brink W, Zwart K, Vreugdenhil C, van Wijngaarden-Cremers P, van der Gaag RJ. Comorbid psychopathology in adolescents and young adults treated for substance use disorders: a review. Eur Child Adolesc Psychiatry. 2006;15:319–28.
3. 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:eaaap8757.
4. Tarn V, Patel N, Turcotte M, Bosse Y, Pare G, Meyne D. Benefits and limitations of genome-wide association studies. Nat Rev Genet. 2019;20:467–84.
5. Cross-Disorder Group of the Psychiatric Genomics C. et al. LD Score regression distinguishes causal from non-causal variants. Nature. 2015;527:73–78.
6. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet. 2015;47:1228–1235.
7. Leeuw CA, Moso MJ, Heskes T, Posthuma D, MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11:e1004219.
8. Pers TH, Karjalainen JM, Chan Y, Westra HJ, Wood AR, Yang J, et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun. 2015;6:5890.
9. Young AI. Solving the missing heritability problem. PLoS Genet. 2015;11:e1004219.
10. Bulik-Sullivan B, Finucane HK, Gusev A, Trynka G, Reshef Y, Loh PR, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet. 2015;47:1228–1235.
11. Frankish A, Diekhans M, Ferrera AM, Johnson R, Jungreis I, Loveland J, et al. GENCODE gene reference annotation for the human and mouse genomes. Nucleic Acids Res. 2019;47:D766–D773.
12. de Leeuw C, Sey NYA, Posthuma D, Won H. A response to Yurko et al. H-MAGMA, inheriting a shaky statistical foundation, yields excess false positives. bioRxiv. 2020.
13. Watanabe K, Umicevic Mirkov M, de Leeuw CA, van den Heuvel MP, Posthuma D. Genetic mapping of cell-type specificity for complex traits. Nat Commun. 2019;10:3222.
14. Shekhar K, Lapan SW, Whitney TE, Tran NM, Macosko EZ, Kowalczyk M, et al. Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. Cell. 2016;166:1308–1323.
15. Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. Nat Genet. 2019;51:1664–1669.
16. Manton TF. Solah VS of mice, men, and microbial opsins: how optogenetics can help hone mouse models of mental illness. Biol Psychiatry. 2016;79:47–52.
ACKNOWLEDGEMENTS
Funding was provided through a personal Rudolf Magnus Talent Fellowship (H150) to Jurjen Luykx and by grants of the Netherlands Organisation for Scientific Research (NWO: 024.004.012, ALWOP.137 and OCENW.KLEIN.071) to Roger Adan. The authors thank Kevin Kenna, Mark Bakker, and Nathan Skene for fruitful input on statistical analyses. The authors thank Albert Batalla Cases for comments and suggestions on the manuscript.

AUTHOR CONTRIBUTIONS
MO, BDL, and JJL conceived and designed the study. MO collected the data. MO and KR performed data analysis and plotted the figures. RAHA provided interpretive advice. BDL and JJL supervised the study. MO and KR wrote the manuscript and RAHA, BDL and JJL revised the manuscript. All authors have read and approved the final version of the paper.

COMPETING INTERESTS
All authors declare they have no conflict of interest.