Human hippocampus niche and circuit neurotypical and major depression profiles

Madeline B. Mariani
Columbia University - NYSPI/RFMH
Cheick A. Sissoko
Columbia University - NYSPI/RFMH
Alexandra M Wamalwa
Columbia University - NYSPI/RFMH
Anthony Ramnauth
Columbia University - NYSPI/RFMH
Yang Xiao
Columbia University - NYSPI/RFMH
, Alexandria N. Tartt
Columbia University - NYSPI/RFMH
Camille Fulmore
Columbia University - NYSPI/RFMH
Ying Liu
Columbia University - NYSPI/RFMH
Wenpin Hou
Columbia University - NYSPI/RFMH
Yung-yu Huang
Columbia University - NYSPI/RFMH
Andrew J. Dwork
Columbia University - NYSPI/RFMH
Gorazd B. Rosoklija
Columbia University - NYSPI/RFMH
Shujuan T. McDonald
Columbia University - NYSPI/RFMH
Victoria Arango
Columbia University - NYSPI/RFMH
J. John Mann
Columbia University - NYSPI/RFMH
René Hen
Columbia University - NYSPI/RFMH
Lewis M. Brown

Columbia University
Hanga Galfalvy
Columbia University - NYSPI/RFMH
Maura B. Dupont
mb928@columbia.edu

Columbia University - NYSPI/RFMH https://orcid.org/0000-0003-3574-7641

Research Article

Keywords:

Posted Date: January 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8593764/v1

Additional Declarations: The authors declare potential competing interests as follows: HG and her family own stocks in Illumina, Inc.; JJM receives royalties from Columbia University for the Columbia Pathways App. and from the Research Foundation of Mental Hygiene for commercial use of the C-SSRS.

Abstract

Hippocampal neuroplasticity regulates memory and emotional responses, which are altered in major depression, a leading cause of disability worldwide[1]. Molecular underpinnings of reduced hippocampal neuroplasticity[2], fewer neurons, and smaller volume[3,4] in MDD are unknown and may involve blunted adult hippocampal neurogenesis (AHN), a debated phenomenon in human brain[5–8]. Dynamic changes in gene expression (GEX) are mediated, in part, by cis-regulatory elements (CREs), in response to developmental cues, environmental signals, and cellular contexts. However, the accessibility of CREs and their relation to GEX in specific hippocampus cell types in MDD is unknown. To this end, we simultaneously profiled chromatin accessibility and GEX in 349,847 human hippocampal nuclei, used machine learning[9,10], RNA Velocity[11] and pseudotime trajectory inference[12] to investigate the presence of progenitor cells and their progeny, and spatial transcriptomics in intact hippocampus sections to sequence cytoplasm and neuronal processes providing anatomical transcriptome mapping. We identified Type I, II and III neural progenitors and immature granule neurons. We find hyperactivated inflammation and apoptosis in immature and mature granule neurons, possibly affecting AHN and cell survival in MDD. Proteomics and transcriptomics convergent dysregulations in MDD, point to lipid metabolism, neurogenesis, synaptogenesis, cell adhesion, and plasticity. Identified molecular markers may drive biomarker and drug discovery for MDD and hippocampus resilience.

Main

Major depressive disorder (MDD) is recurrent and episodic, often treatment resistant[2], a major cause of disability worldwide, and the most frequent diagnosis in those who die by suicide[1]. Human brain postmortem studies found smaller hippocampus dentate gyrus (DG) and fewer granule neurons (GNs)[3,13] in MDD, which may reflect impaired neuroplasticity resulting in MDD symptoms like impaired emotional dysregulation and memory[2]. Moreover, hippocampus neuronal and neuropil loss were correlated with MDD severity and illness duration[3,13]. *In vivo* brain imaging confirmed smaller hippocampal volume in MDD compared to controls[14]. It remains unknown whether smaller hippocampal volume and diminished neuroplasticity in MDD are the result of increased apoptosis, decreased AHN, or other cellular and molecular changes.

Deficits in AHN have been linked to depressive-like behavior in rodents[15]. Conversely, adult born GNs support stress recovery and emotional regulation[16], enhance hippocampal plasticity, and circuit rewiring[17]. Transcriptomics studies agreed that rodent markers are inadequate to identify progenitor cells and their trajectory in humans and non-human primates[5,7,8]. In humans, early studies showed proliferating cells becoming GNs[18] and estimated neuron birth date based on ^{14}C decay indicated substantial AHN in people who had been subjected to radioactive exposure[19]. Nevertheless, human AHN remains debated[20,21], as immunofluorescence (IF), immunohistochemistry (IHC)[22-24], *in situ* hybridization (ISH)[23,25], and transcriptome sequencing studies[5-8] led to opposite conclusions, likely

driven by tissue processing methods, subjects' medical history, mode of death, and psychotropic drugs exposure[26].

To resolve this debate, provide new insights into the molecular characteristics of human hippocampus cell types, and discover the molecular basis of cellular disruption in MDD, we assessed 106 subjects and for the first time profiled almost 350,000 cells in untreated MDD subjects (MDDs) and neurotypical controls (CTRL) with multi-omic single-nucleus RNA sequencing (snRNA-seq) and assay for transposase-accessible chromatin with sequencing (snATAC-seq), spatial RNA-seq, and extensive molecular marker validation (Fig. 1a). Here we present results from the largest and most deeply sequenced single-nucleus multiome and spatial transcriptome resource dataset of neurotypical and MDD human hippocampus cell types to date[27].

Single-nucleus multiome and spatial transcriptome profiling of adult human hippocampus We simultaneously generated snRNA-seq and snATAC-seq data on 349,847 nuclei from deeply clinically phenotyped sudden death subjects with high-quality brain tissue (52 samples from 18 subjects, 9 male MDDs and 9 male CTRL, with 53% of nuclei from MDDs; Fig. 1a-d, Supplementary Tables 1-3; Methods). Number of samples per donor, cells profiled per donor, RNA integrity number, post-mortem interval, and age did not differ between MDDs and CTRL (Extended Data Fig. 1a-e). The medium number of unique molecular indices, chromatin regions, and genes per donor did not differ between MDDs and CTRL (Extended Data Fig. 1f-i). We used Harmony[28] to correct for sample, donor, and batch in RNA and ATAC data separately (Extended Data Fig. 2).

To integrate snRNA-seq and snATAC-seq data, we used weighted nearest neighbor (WNN) analysis[29] (Fig. 1b-d), an unsupervised method in Seurat (v5) to obtain a unified representation of snRNA-seq and snATAC-seq data. We identified 31 unsupervised clusters, representing 12 broader cell types (Fig. 1b-c), that were not driven by batch or donor effects (Fig. 1b-d; Extended Data Fig. 3a-d). In each cluster and cell type, average GEX and predicted GEX inferred by accessible chromatin (or gene activity) data were strongly correlated with each other (Fig. 1e; Extended Data Fig. 4e) and used to annotate cell types based on their top-expressed genes (Fig. 1f-g; Extended Data Fig. 4c-d; Methods; Supplementary Table 4). Clusters included oligodendrocyte progenitor cells (OPC), oligodendrocytes (Olig), endothelial cells (Endo), pericytes (Peri), vascular leptomeningeal cells (VLMC), microglia (Micro), macrophages (Macro), T-cells (Tcell), choroid plexus cells (Ch.Pl), subventricular zone (SVZ) ependymal cells (Epe), Cajal Retzius cells (CR), three astrocyte (Astro1-3), nine inhibitory neuron (InN.SST, InN.LHX6, InN.PVALB, InN.VIP, InN.LAMP5, InN.SLC17A8, InN.PENK, InN.TUBB3, InN.PROX1), six non-granule excitatory neuron (ExN1-6), and two GN (GN1, GN2) clusters (Fig. 1d, Extended Data Fig. 4a-d). These 31 cluster annotations aligned with those from recent human hippocampus snRNA-seq studies that did not include ATAC sequencing[6-8] (Extended Data Fig. 5; Supplementary Table 5). The 12 broader cell types included: OPC, Olig, Ch.Pl, Epe, and CR (identical to the clusters), and merged clusters of vasculature cells (Vasc; including Endo, Peri, and VLMC), immune system cells (Immune; including Micro, Macro, and Tcell), astrocytes (Astro; including all three astrocyte clusters), granule neurons (GN, including both GN

clusters), inhibitory neurons (InN, including all nine InN clusters), CA excitatory neurons (CA.ExN, including ExN1, ExN 3) and other excitatory neurons (O.ExN, including ExN2, ExN4-6 that did not map on CA regions, Fig. 2e-f; Supplementary Table 6).

Spatial GEX is concordant with single-nucleus cell type annotations

To investigate GEX in whole tissue, including cytoplasm and neuronal processes, and provide anatomical mapping of GEX onto hippocampal subfields, we used Visium (10X Genomics) on intact hippocampus proper tissue sections (6 subjects, 2 samples each; 3 male MDDs and 3 male CTRL; Supplementary Table 7). First, we used histology (Fig. 2a) to guide hippocampus manual anatomical annotation into granule cell layer (GCL), subgranular zone (SGZ), molecular layer (ML), CA1-4, and white matter (WM, Fig. 2b). Then we performed unsupervised clustering of Visium spots (55x55 μm in size) which provided a more detailed spot-type annotation than manual annotation of hippocampal subfields and delivered 13 clusters that expressed genes concordant with expected cell types based on spot location in the respective subfields (Fig. 2c-d; Supplementary Table 8). Unsupervised spot clusters included (Methods): 2 hilus clusters (hilus1, hilus2), CA3 stratum pyramidales (ca3), stratum oriens-stratum lucidum (so-slu), so-slu of the CA3 (ca3-so-slu), stratum radiatum (sr), stratum oriens-stratum lacunosum (so-sla), stratum moleculare (ml), stratum granulosum (gcl), SGZ-polymorphic layer (sgz-pl), stratum moleculare (sgz-ml).

We further used Visium for validation of single-nucleus cell type annotations, performing anchor-based integration of sn-multiome-based cell types (reference) with whole-transcriptome spatial GEX (query), and found that Olig and Astro cell types mapped in WM, GN in GCL (Fig. 1h), and CA.ExN in CA regions, while O.ExN had low prediction score in CA regions, and vasculature was spread across the tissue (Fig. 2e). Different ExN clusters mapped on specific subfields, with ExN1 mapping on CA1, ExN3 on CA3, while Ch.Pl and Epe correctly mapped on choroid plexus and SVZ (Extended Data Fig. 4f).

We then used Xenium multiplex ISH (10X Genomics, 266 probes, 10.45x22.45 mm capture area) for additional anchor-based integration of single-nuclei cell types. While CA.ExN mapped on CA1-4, O.ExN mapped on the subiculum (Fig. 2e-f), which was not possible to verify with Visium because the smaller capture area could not include the subiculum. Xenium confirmed the expected location of Astro, Olig, GN (Fig. 1i), and Vasc (Fig. 2f).

Spatial RNA Velocity identifies a cell trajectory from the SGZ to the GCL

To infer cell trajectories based on proportions of spliced and unspliced RNAs, we applied RNA Velocity analysis to spatial RNA-seq rather than to snRNA-seq data because the method relies on both cytoplasmic and nuclear RNAs to infer transcription and degradation rates[11]. Since we hypothesized that, if a neurogenic trajectory exists, it would be in the DG, to infer RNA Velocity, we selected spots located in the SGZ, GCL, and ML (Fig. 2g; Extended Data Fig. 6; Methods). The partition-based graph abstraction (PAGA) analysis identified a lineage from spatial cluster sgz.ml to cluster sgz.pl and then to gcl (Fig. 2h), indicating that sgz.ml could have the most immature cells with cell bodies in the SGZ and

apical dendrites in ML, likely radial-glia-like quiescent neural stem cells[30]. The next cluster on the trajectory was sgz.pl likely containing intermediate progenitors located in SGZ. Ultimately the gcl spatial cluster would comprise GN at different dynamic functional states. Latent time (Fig. 2i) and pseudotime calculated from RNA Velocity (Fig. 2j) agreed with time zero corresponding to cluster sgz.ml and increasing time latency going towards clusters sgz.ml and then gcl. GEX-by-latent-time analysis showed high expression of *PCNA*, *HESX1*, *MKI67*, and *ENOX1* at the earliest stage (sgz.ml spatial cluster), *DCX*, *CALB2*, *SOX11*, *GAD1*, *GAD2*, and *STMN2* at the intermediate stage (sgz.pl spatial cluster), and *CAMK2A*, *NEUROG3*, *CAMK2A*, *SNAP25*, *PPFIA2*, and *RBFOX3* at the final stage of maturation (gcl spatial cluster, Fig. 2k).

Supervised machine learning characterization of neurogenic cell trajectory

To determine the presence of a human hippocampus neurogenic niche at single-nucleus resolution, we used a supervised machine learning approach. Supervised models have previously been used to distinguish transcriptionally ambiguous cell subtypes[9,10], including those of the neurogenesis trajectory[8]. We scored all 349,847 high quality nuclei using normalized counts of genes known to be expressed in mammalian hippocampus quiescent neural stem cells (Type I), amplifying neural progenitors (Type II), intermediate neural progenitors (Type III), and immature GN (ImGN; Fig. 3a) as per rodent and human literature[5,8,20,30,31]. A percentage of high confidence cells were selected (Methods). High confidence cells with high expression of required marker genes and low expression of genes typically not expressed in that cell type[5,7,8,20,30] were used to train the models: for Type I we used *GFAP*, *SOX2*, *PAX6*, *NES*, *FABP7*, *HES5*, and *ETNPPL*; Type II had *ASCL1*, *FOXO3*, *NR2E1*, *EOMES*, *NEUROG2*, *NES*, *MCM2*, *PCNA*, *MKI67*, *TOP2A*, and not *MBP* or *MOBP*, to distinguish them from oligodendrocyte progenitors; for Type III we used *MYT1L*, *PROX1*, *DCX*, *CALB2*, *RELN*, *ST8SI2*, *ST8SIA4*, and not mature neuron marker *RBFOX3*, or inhibitory neuron marker *VIP*; for ImGN, *RBFOX3*, *PROX1*, *NEUROD1*, *DCX*, *BHLHE22*, *COL25A1*, and *POSTN*. Cells with 85% prediction score or higher were annotated as TypeI (4,413 cells, 7.47% of GN) TypeII (1,680 cells, 2.83% of GN), TypeIII (4,423 cells, 7.49% of GN) and ImGN (5,611 cells, 9.50% of GN) and clustered separately (Fig. 3b). The LASSO approach extracted these cells from the original clusters (Supplementary Table 9). TypeI derived from Astro, as radial-glia-like progenitors are most like astrocytes[30]. TypeII derived mostly from OPC, Astro, and Endo, as those clusters contain proliferating cells. TypeIII derived from InN and CR clusters. CR cells have a prominent role in brain development[32] and express CALB2 and TUBB3 and not SYT1 (Extended Data Fig. 4c-d), not typical of adult excitatory CR. InN clusters expressed immature neuron markers: *CALB2* (InN.VIP, InN.PROX1), *TUBB3* (InN.TUBB3), *RELN* (InN.LAMP5), *DCX* (all InN subtypes), and *PROX1* (InN.PROX1, InN.VIP, InN.LAMP5; Fig. 1f-g, Extended Data Fig. 4c,d). Moreover, *GAD1* and *GAD2*, classically used as InN markers, have transcript variants expressed during brain development until neurons fully differentiate[33]. Using Xenium and RNAscope (ACDbio), we examined the expression of *GAD1* in the DG, finding *GAD1* co-expressed with *PROX1* and *DCX* in the GCL (Fig. 3h). Finally, ImGN were pulled mostly from GN, in line with a recent publication[8]. To validate spatial location of ImGN, we used anchor-based integration of single-nuclei GEX (reference) and Xenium ISH (query) and found that ImGN are in SGZ and GCL (Fig. 3g).

The maturity stages of the machine learning-identified cell types were reflected by gene ontology terms showing TypeI enriched for cell fate commitment, TypeII for stem cell differentiation and neural precursor cell proliferation, TypeIII for neuron migration and axon development, and ImGN for cell growth and learning or memory (Supplementary Table 10) .

In a larger sample (Supplementary Table 1), using RNAscope, we validated expression of DCX in human SGZ and GCL as previously shown[21], and demonstrated DCX co-localizing with NEUROD1, PROX1, TUBB3 and RBFOX3 (Extended Data. Fig. 7a-g). Using IHC, we confirmed expression of Nestin and Ki67 proteins in human SGZ, DCX protein in human SGZ and GCL, and NeuN protein in GCL, as previously reported[18]. Using double IF, we replicated co-localization of DCX with PSA-NCAM[18], and showed DCX co-labeling with neurofilament, calretinin, calbindin, and NeuN, and no DCX co-localization with glial markers GFAP and IBA1 (Extended Data Fig. 7-8).

Single-nucleus pseudotime trajectory inference supports a human neurogenic lineage

To test if machine learning-identified neurogenic cell types TypeI, TypeII, TypeIII, and ImGN belonged to a differentiation trajectory, we used Palantir[12] to model differentiation capturing the continuity in cell states and cell fate determination (Fig. 3c-e). We included TypeI, TypeII, TypeIII, ImGN, and the cells remaining in the GN cell type after applying the LASSO model (Fig. 3b). Palantir assigned pseudotime in concordance with our hypothesis and ordered cell types from TypeI, TypeII, TypeIII to ImGN and GN (Fig. 3c). GEX along the neurogenic lineage included *PAX6*, *ETNPPL*, and *SOX2* in TypeI quiescent progenitors; *MKI67*, *TOP2A*, *PCNA* in TypeII proliferating progenitors; *DCX*, *GAD1*, *GAD2*, *EOMES*, *TUBB3*, *PROX1*, *STMN1* in TypeIII neuroblasts and ImGN; and finally, *SCL17A7*, *CALB1*, *CAMK2A*, *SYT1*, and *RBFOX3* in mature GN, also corresponding to open chromatin at those genes (Fig. 3d-f).

Moreover, Palantir lineage cell GEX is concordant with progressive GEX in Visium spots ordered by latent time based on RNA Velocity (Fig. 2k), and they both suggest a phase in which *GAD1* and *GAD2* are expressed in TypeIII. In fact, all InN expressed *DCX*; InN.LAMP5 expressed *RELN*; InN.VIP and InN.PROX1 expressed *CALB2*; InN.TUBB3 expressed *TUBB3*; and InN.PROX1, InN.VIP, and InN.LAMP5 expressed *PROX1* (Fig. 1f; Extended Data Fig. 4c).

Cell-type specific gene expression and chromatin accessibility are altered in MDD

To understand molecular underpinnings of hippocampus anatomical and functional disruption in MDD[2,3,13,14], we analyzed cell-type specific differentially expressed genes (DEGs), differentially accessible chromatin regions (DARs), and predicted differentially active genes (DAGs) both in the 31 clusters and the 16 cell types (OPC, Olig, Ch.Pl, Epe, CR, Vasc, Immune, Astro, TypeI, TypeII, TypeIII, ImGN, GN, InN, CA.ExN, O.ExN, Supplementary Tables 11-13).

The clusters with the highest number of DEGs were GN2, GN1, ExN2, Epe, OPC, and Oligo, (Fig. 4a-b). GN1 and GN2 had 75 and 43 DEGs, contributing to 13.2% and 7.6% of all DEGs respectively. Epe and ExN2 had 182 DEGs, both contributing to 32% of DEGs. Regarding DARs, ExN1 had the most with 497

hits (87.5% of all), Oligo and GN1 had 39 and 36, contributing to 6.9% and 6.3% of DARs respectively (Fig. 4c-d). As for predicted DAGs, 97.6% were identified in OPCs, and GN1, GN2, and Oligo had 15, 13, and 13 respectively (Fig. 4e-f).

When analyzing cell types, GN had 101 DEGs, O.ExN had 38, ImGN had 20, TypeI and TypeIII each had one DEG: *LINC02315* (upregulated) and *ZNF257* (upregulated) respectively (Fig. 4a-b). DARs were the highest in Immune (14227 DARs), followed by GN (41 DARs; Fig. 4c-d). DAGs were 60 in GN, 41 in O.ExN, 2 in ImGN, and one in TypeI and TypeIII (*ACOT1,* upregulated in MDDs, Fig. 4e-f). Taken together, GN had the most consistently dysregulated genes across modalities.

PsyGeNET Analysis

To explore which of the identified dysregulated genes had previously been associated with MDD, we used PsyGeNET text mining database[34]. In all clusters, we found 43 genes with DEGs and DAGs overlapping with psychiatric disease-associated genes in PsyGeNET, and significant for depressive disorders (p=0.033017458) with 15 overlapping genes (*ADCY8*, *ATF3*, *DISC1*, *EMP1*, *ENPEP*, *FOLH1*, *FOS*, *GRIN2A*, *GRK5*, *IL18*, *IL19*, *IL6R*, *SOD2*, *SYNDIG1*), and for alcohol use disorder (pl=0.009390476), with 83 overlapping genes (Fig. 5a-b, Supplementary Table 14).

When analyzing cell types, GN had 7 DEGs previously linked to psychiatric disorders, including depression, schizophrenia, and mood disorders (*ADCY8*, *CCND2*, *DNAH8*, *GABBR1*, *HLA-DQB1*, *PRKCB*, *THEMIS*; Fig. 5c), without any significant disease-specific associations.

Biological pathways of cell-type specific DEGs in MDD

To infer pathways associated with cell-type specific multi-omics changes in MDD, we performed pre-ranked gene set enrichment analysis (GSEA, Supplementary Table 15). We focused on GN that had the most dysregulated genes across modalities, and ImGN to understand why GN are fewer in MDD[3,4,13]. In ImGN, pathways related to interferon signaling, cellular stress, Golgi, MTORC1, sex hormone response and apoptosis were enriched, while lipid homeostasis and glycosylation were negatively regulated in MDD (Fig. 5d-e). In GN, pathways related to PTEN, MTORC1, apoptosis, sex hormone and immune response, and thyroid, Parkinson, Huntington and Alzheimer's diseases were enriched, and lipid homeostasis was negatively regulated in MDD (Fig. 5f-g). Notably, in both ImGN and GN, apoptosis was enriched in MDD with top genes *CCND2*, *JUN*, *MADD*, *BCAP31*, *DNAJA1*, *NEFH*, *PEA15*, *DAP*, *ENO2*, *BMP2* (Supplementary Table 15).

Weighted gene co-expression networks dysregulation in MDD

While differential GEX analysis provides information on single-gene expression differences, MDD is highly polygenic[35] and disease state may be driven by gene network changes rather than single-gene alterations[36]. Therefore, we used pseudo-bulk GEX in GN and ImGN to perform weighted-gene co-expression network analysis (WGCNA)[37] and determine clusters of highly correlated genes that have an

association with MDD. Each cluster was represented by a color-coded module eigengene (ME), the first principal component of the module representing the overall expression level of the module.

In ImGN, six out of seven modules had correlations with MDD (Fig. 5h-j; Extended Data Fig. 9a-d). Four ME were positively associated and two were negatively associated with MDD. MEblue, enriched for RNA metabolism, splicing, translation and response to cellular stress, was positively correlated with MDD (correlation 0.99, $p=1\times 10^{-40}$), in line with GSEA findings described above. The MEturquoise module, enriched for lipid modification and metabolism, was negatively correlated with MDD (correlation -0.99, $p=2\times 10^{-47}$) underscoring GSEA findings of lipid homeostasis being negatively regulated in MDD ImGN and GN (Fig. 5e-g).

In GN, 8 modules out of 9 had correlations with MDD (Fig. 5k-m; Extended Data Fig. 9e-h). Three ME were positively associated and five were negatively associated with MDD. The MEturquoise module, enriched for lipid modification and metabolism, was positively correlated with MDD (correlation 0.99, $p=2\times 10^{-43}$) and showed significant overlap ($p=2\times 10^{-04}$) with upregulated DEGs in GN, and with GSEA findings (Fig. 5e-g). The MEblue module, enriched for development and cellular morphogenesis, was negatively correlated with MDD (correlation -0.99, $p=1\times 10^{-50}$) and showed significant overlap ($p=7\times 10^{-06}$) with downregulated DEGs in GN, and with GSEA findings (Fig. 5e-g).

Quantification of target marker RNA and protein expression in MDD

In separate and enlarged samples, using unbiased stereology, we quantified number of cells expressing neurogenic lineage RNAs and proteins labeled by multiplex ISH and IF, and single IHC, and used quantitative real-time polymerase chain reaction (qPCR) to quantify RNA in bulk tissue (Methods; Supplementary Table 16, Extended Data Fig. 7-8).

RNAscope DCX+/TUBB3+ cells were fewer in MDDs (N=11) *vs.* CTRL (N=11) in GCL (t=2.294; p=0.04) but not SGZ (Extended Data Fig. 7n). DCX RNA measured by qPCR was lower in MDDs (N=5) *vs.* CTRL (N=5) in anterior (t=3.371, p=0.0145) but not posterior DG (Extended Data Fig. 7o).

Single-IHC Ki67+ cells were fewer in MDDs (N=17) *vs.* CTRL (N=23) in anterior (t=2.780; p=0.0084) but not mid or posterior DG (Extended Data Fig 8p). MDDs had fewer Nestin+ cells in anterior (34 MDDs, 38 CTRL, t=4.360; p<0.0001) and posterior DG (31 MDD, 28 CTRL, t=3.654; pval=0.0008) but not in mid DG (Extended Data Fig. 7q). Nissl+ glial cells were not different between MDDs and CTRL in any region (Extended Data Fig. 7r). Single-IHC DCX+ cells were fewer in MDDs (N=22) *vs.* CTRL (N=36) in anterior (t=3.547; p=0.0008), but not in mid or posterior DG (Extended Data Fig. 7s). MDD had fewer single-IHC NeuN+ GNs in anterior (35 MDDs, 32 CTRL, t=2.951; p=0.004) and mid DG (28 MDDs, 26 CTRL, t=2.410; p=0.01), but not in posterior DG (Extended Data Fig. 7t).

IF-stained DCX+/NeuN- cells were fewer in MDDs (N=8) *vs.* CTRL (N=8) in SGZ (welch-corrected t [t]= 2.318; p=0.04) but not GCL, and DCX+/NeuN+ cells did not differ between groups (Extended Data Fig. 8l-o).

Proteomics

We further used high resolution shotgun proteomics liquid chromatography with tandem mass spectrometry (LC/MS/MS)[38], to quantify expression of proteins corresponding to DEGs in MDDs *vs.* CTRL as per snRNA-seq data (Methods; Extended Data Fig. 9i-m; Supplementary Table 17).

After running proteomics as we described[38], Elucidator (Methods) detected 80,504 peptides across 48 runs and identified 1811 proteins with protein score at a 4% false discovery rate (Extended Data Fig. 9i-m; Supplementary Table 17). The 1131 proteins with two or more peptides and protein scores greater than 250 were used for the analyses.

We found 308 differentially expressed proteins between MDDs and CTRL, of which four corresponded to detected DEGs: CPNE8, HSDL2, NCH1, and RAC1 (Extended Data Fig. 9n).

HSDL2, a key factor of fatty acid regulation in lipid metabolism[39], was upregulated at both the RNA level in GN and the protein level in bulk tissue.

RAC1, important in neurodevelopmental function including neurogenesis, synaptogenesis, and plasticity and expressed in migrating cortical neurons and adult neurons[40], was lower at the protein level in MDDs *vs.* CTRL. RAC1 is negatively regulated by ARHGAP15[41], whose genomic sequence overlaps with long non-coding RNA *AC079793.2*, also downregulated in ImGN in MDD.

Conversely, CPNE8 regulating cell adhesion[42] showed downregulated RNA expression in MDD GN and upregulated protein expression in bulk MDD tissue.

Similarly, NCEH, a protector from α-synuclein neurotoxicity[43], was downregulated at the RNA level in Epe but the protein was upregulated in bulk hippocampus tissue in MDD.

Discussion

Employing the first single-nuclei multiome and spatial transcriptome profiling of human anterior hippocampus cells, combined with shotgun proteomics in anterior hippocampus tissue, we compared sudden death neurotypical adult subjects and unmedicated patients with MDD.

Implementing a machine learning approach for rare cells identification, we detected quiescent (Type I), amplifying (Type II), and intermediate (Type III) progenitor cells, and replicated the identification of immature neurons[8]. Our integrated analysis identified distinct clusters representing all hippocampal cell types, in agreement with other publications[6–8], identified a neurogenic lineage from early progenitors to immature granule neurons, and defined cell type-specific gene expression and chromatin accessibility dysregulation in MDD. Importantly, the annotation of these cell types, including those of the neurogenesis trajectory, was validated by spatial GEX profiling using Visium, RNAScope, and Xenium,

and by protein marker expression identification by double IF and IHC, providing robust evidence of cell identity and relative cellular and spatial distribution within the hippocampus.

Cell trajectory inference using both single-nuclei and spatial transcriptome sequencing, supported the neurogenic cell-lineage trajectory, and gene expression corresponding to maturational stages described in mice[5,7].

Findings suggest transcriptional and epigenetic similarities of astroglia-like (TypeI) progenitors with astrocytes, amplifying progenitors (TypeII) with other dividing cells (OPC, Astro and Endo).

These data do not demonstrate when cells of the neurogenic trajectory were generated but the expression of proliferation markers (*PCNA*, *MKI67*, *TOP2A*; Fig. 2k, 3e) in TypeII suggests the presence of mitosis.

TypeIII origination from GAD + neurons initially classified as InN, is largely due to DCX expression in GAD + cells, and warrants further investigation on transitional GAD expression in maturing cells.

We demonstrated for the first time *GAD1* expression in TypeIII, colocalizing with *DCX*, *TUBB3* and *PROX1*, and *DCX* mapping onto SGZ and GCL, in line with findings that GAD1 alternative transcripts are expressed in human hippocampus during neuronal development[33,44]. Additionally, GABAergic basket cells could be generated in human DG as demonstrated in adult rats[45].

ImGN coming from the GN cluster has been previously observed. Notably the number of ImGN is similar to what previously estimated (close to 10% of all GNs), suggesting cells might remain in this immature status for a long time, which has profound implications for the functional significance of AHN in the human brain.

The presence of immaturity markers suggests a regenerative potential of the human SGZ. It remains unknown how prolonged the immature state is and how long it takes to make a human neuron. Different sequencing technologies, including epigenome analysis and deeper sequencing, may explain our finding the full neurogenic lineage when others have not[6–8].

This validity of our trajectory reconstruction is strengthened by RNA Velocity and pseudotime analysis inferring of cell trajectory, and its identified molecular signatures that characterize neural stem cell activation, proliferation, differentiation, and ultimately, neuronal maturation, matching those found in the literature[20,26,30]. These findings provide a valuable resource for future studies investigating the regulation of AHN and its potential modulation for therapeutic purposes.

In MDD, elevated inflammation, response to cellular stress and apoptosis pathways in ImGN and GN suggests the molecular basis of smaller *in vivo* hippocampal volume[14] and postmortem findings of fewer GNs[4,13] and agree with finding in MDD PFC[27] and patient blood[46–48]. Overlap between DEGs and PsyGeNET MDD and alcohol use disorder genes, may explain the comorbidity between the two.

Upregulation of *LINC02315* in TypeI and OPC in MDD could prevent proliferation or transition to the amplifying stage because this long-intergenic non-coding RNA is protective in lung squamous cell carcinoma by blocking proliferation[49]. In TypeIII, InN and Olig, upregulation of *ZNF257* which has over 1,000 binding sites in gene promoters and is believed to regulate transcription in the developing brain[50], may contribute to increased nucleic acid binding, transcription and Vitamin D receptor pathway activity in MDD affecting developmental processes. Dysregulation of MTORC1 in ImGN and GN supports altered cell growth and metabolism[2].

Using both proteomics and transcriptomics, we find genes dysregulated at both levels. Upregulated HSDL2, a key factor of fatty acid regulation in lipid metabolism[39], at the RNA level in GN and protein level in bulk tissue suggests lipid metabolism dysregulation in MDD. Downregulation of both long non-coding RNA *AC079793.2* in ImGN, overlapping in genomic sequence with *ARHGAP15*, and possibly negatively regulating it, and its target protein RAC1 in bulk tissue, suggest reduced neurogenesis, synaptogenesis, and plasticity[40]. The fact that *CPNE8*, regulating cell adhesion[42], and NCEH, protecting from α-synuclein neurotoxicity[43], have downregulated RNA expression in GN and Epe respectively, while the proteins are upregulated in bulk tissue, may be due to tissue-specific transcriptomic differences involving DG and Epe, or compensatory post-transcriptional mechanisms.

Although this is the first study combining single-nucleus multi-omic (RNA and ATAC) profiling, with spatial transcriptomic and bulk proteomics in human hippocampus, and comparing sudded death neurotypical controls with unmedicated depressed subjects that had MDD, limitations include small sample, not including women, possible clinical variability, and few subjects having received benzodiazepines prior to death – although we excluded comorbidity and patients on antidepressants or antipsychotics. Results should be replicated in larger samples that include women and low testing effects of clinical and sociodemographic variables, because in animal models stress greatly affects neurogenesis.

Overall, this comprehensive single-nuclei multi-omic, spatial transcriptomic, and proteomic profiling of brain tissue supports the presence of a neurogenic niche and suggests molecular underpinnings of anterior hippocampus atrophy in MDD. Dysregulated markers will drive future mechanistic studies investigating their cellular, molecular, and behavioral effects, and will lead blood and brain imaging biomarker discovery and drug development for MDD and hippocampus trophic support.

Declarations

Data availability

The single cell and spatial genomics data will be available at the GEX Omnibus database upon publication. Specimen information and sequencing statistics are described in Supplementary Tables 1-3

and 7. Sources of the published sequencing datasets used in this study are described in Supplementary Table 5.

Raw mass spectroscopy data are available at the MassIVE repository entry under dataset identifier MSV000094862 (reviewer password: reviewer052424).

The final datasets for this study, which will be archived in our servers, will include demographic data, psychiatric assessment data, psychological ratings, and postmortem studies data. Because the data we collect are sensitive in nature, we carefully ensure that appropriate privacy safeguards are in place. All data will be stripped of identifiers prior to release for sharing.

In addition, we will provide relevant protocols and published genetic and phenotypic data upon request. Material transfers will be made with no more restrictive terms than in the Simple Letter Agreement (SLA) or the Uniform Biological Materials Transfer Agreement (UBMTA) and without reach through requirements.

Should any intellectual property arise which requires a patent, we will ensure that the technology (materials and data) remains widely available to the research community in accordance with the NIH Principles and Guidelines.

Code availability

The computational code used in this study is available at GitHub or upon request.

Acknowledgements

We thank the families for donating the brain tissue used in this study. This work is supported by NIH grants AI164769, AG076949, MH133561, AG080790. We thank Stefano Marenco and the Human Brain Collection Core at the National Institute of Mental Health, and Mark Underwood at the New York State Psychiatric Institute for providing some of the postmortem human brain tissue samples. We thank Erin Bush and Peter Sims and the Columbia Single Cell Analysis Core and JP Sulzberger Columbia Genome Center for their services. We thank Jeremy Worley and the Columbia Genome Engineering Core for Xenium (10X Genomics) experimentation and analysis services. We thank Vinod Gupta and the Columbia Center for Computational Biology and Bioinformatics for services in data storage and high-performance computing infrastructure usage. We thank Tao Su and the Molecular Pathology Shared Resource at Columbia for RNA extraction and quantification services. This study used the Confocal and Specialized Microscopy Shared Resource of the Herbert Irving Comprehensive Cancer Center at Columbia University, funded in part through the NIH/NCI Cancer Center Support Grant P30CA013696. Some schematic illustrations were created or modified using BioRender.com.

Contributions

MBM and CAS performed sequencing experiments, single-nuclei and spatial library generation, and bioinformatics. AMW performed WGCNA analysis. AR worked on spatial data analysis and data interpretation. YX advised on bioinformatics. ANT, CF, and AMW completed IHC, IF and ISH experiments. YL designed the machine learning approach. WH advised on spatial transcriptomics approaches. YYH worked on nuclei isolation. MBD, GR, JJM, VA and AD provided the human postmortem specimens and clinical characterization of the subjects. STMD and LB worked on proteomics. RH helped with the interpretation of the neurogenesis findings. HG designed the biostatistics approach. MBD secured funding for this project, guided experimental design and approaches, worked on result interpretation and reporting strategy. MBM, and MBD wrote the first draft of the manuscript. All authors reviewed the manuscript.

Competing Interests

HG and her family own stocks in Illumina, Inc.; JJM receives royalties from Columbia University for the Columbia Pathways App. and from the Research Foundation of Mental Hygiene for commercial use of the C-SSRS.

Corresponding author

Correspondence to mb928@cumc.columbia.edu

References

1. James, S. L. *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **392**, 1789-1858 (2018).
2. Tartt, A. N., Mariani, M. B., Hen, R., Mann, J. J. & Boldrini, M. Dysregulation of adult hippocampal neuroplasticity in major depression: pathogenesis and therapeutic implications. *Mol Psychiatry* (2022). https://doi.org:10.1038/s41380-022-01520-y
3. Cobb, J. A. *et al.* Hippocampal volume and total cell numbers in major depressive disorder. *J Psychiatr Res* **47**, 299-306 (2013). https://doi.org:10.1016/j.jpsychires.2012.10.020
4. Boldrini, M. *et al.* Resilience Is Associated With Larger Dentate Gyrus, While Suicide Decedents With Major Depressive Disorder Have Fewer Granule Neurons. *Biological Psychiatry* **85**, 850-862 (2019). https://doi.org:https://doi.org/10.1016/j.biopsych.2018.12.022
5. Wang, W. *et al.* Transcriptome dynamics of hippocampal neurogenesis in macaques across the lifespan and aged humans. *Cell Research* **32**, 729-743 (2022). https://doi.org:10.1038/s41422-022-00678-y

6. Ayhan, F. *et al.* Resolving cellular and molecular diversity along the hippocampal anterior-to-posterior axis in humans. *Neuron* **109**, 2091-2105.e2096 (2021). https://doi.org:10.1016/j.neuron.2021.05.003
7. Franjic, D. *et al.* Transcriptomic taxonomy and neurogenic trajectories of adult human, macaque, and pig hippocampal and entorhinal cells. *Neuron* **110**, 452-469. e414 (2022).
8. Zhou, Y. *et al.* Molecular landscapes of human hippocampal immature neurons across lifespan. *Nature* **607**, 527-533 (2022). https://doi.org:10.1038/s41586-022-04912-w
9. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nature methods* **14**, 955-958 (2017).
10. Shekhar, K. *et al.* Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. *Cell* **166**, 1308-1323 e1330 (2016). https://doi.org:10.1016/j.cell.2016.07.054
11. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494-498 (2018). https://doi.org:10.1038/s41586-018-0414-6
12. Setty, M. *et al.* Characterization of cell fate probabilities in single-cell data with Palantir. *Nat Biotechnol* **37**, 451-460 (2019). https://doi.org:10.1038/s41587-019-0068-4
13. Boldrini, M. *et al.* Hippocampal Granule Neuron Number and Dentate Gyrus Volume in Antidepressant-Treated and Untreated Major Depression. *Neuropsychopharmacology* **38**, 1068-1077 (2013). https://doi.org:10.1038/npp.2013.5
14. Nolan, M. *et al.* Hippocampal and Amygdalar Volume Changes in Major Depressive Disorder: A Targeted Review and Focus on Stress. *Chronic Stress* **4**, 2470547020944553 (2020). https://doi.org:10.1177/2470547020944553
15. Surget, A. *et al.* Antidepressants recruit new neurons to improve stress response regulation. *Mol.Psychiatry* (2011).
16. Anacker, C. *et al.* Hippocampal neurogenesis confers stress resilience by inhibiting the ventral dentate gyrus. *Nature* **559** (2018). https://doi.org:https://doi.org/10.1038/s41586-018-0262-4
17. Temprana, S. G. *et al.* Delayed coupling to feedback inhibition during a critical period for the integration of adult-born granule cells. *Neuron* **85**, 116-130 (2015). https://doi.org:10.1016/j.neuron.2014.11.023
18. Eriksson, P. S. *et al.* Neurogenesis in the adult human hippocampus. *Nat.Med.* **4**, 1313-1317 (1998).
19. Spalding, K. L. *et al.* Dynamics of hippocampal neurogenesis in adult humans. *Cell* **153**, 1219-1227 (2013). https://doi.org:10.1016/j.cell.2013.05.002
20. Gage, F. H. Adult neurogenesis in mammals. *Science* **364**, 827-828 (2019).
21. Tosoni, G. *et al.* Mapping human adult hippocampal neurogenesis with single-cell transcriptomics: Reconciling controversy or fueling the debate? *Neuron* **111**, 1714-1731. e1713 (2023).
22. Boldrini, M. *et al.* Human hippocampal neurogenesis persists throughout aging. *Cell stem cell* **22**, 589-599. e585 (2018).

23. Sorrells, S. F. *et al.* Positive controls in adults and children support that very few, if any, new neurons are born in the adult human hippocampus. *Journal of Neuroscience* **41**, 2554-2565 (2021).
24. Moreno-Jiménez, E. P., Terreros-Roncal, J., Flor-García, M., Rábano, A. & Llorens-Martín, M. Evidences for adult hippocampal neurogenesis in humans. *Journal of Neuroscience* **41**, 2541-2553 (2021).
25. Tartt, A. N. *et al.* Considerations for Assessing the Extent of Hippocampal Neurogenesis in the Adult and Aging Human Brain. *Cell Stem Cell* **23**, 782-783 (2018). https://doi.org:10.1016/j.stem.2018.10.025
26. Kempermann, G. *et al.* Human adult neurogenesis: evidence and remaining questions. *Cell stem cell* **23**, 25-30 (2018).
27. Maitra, M. *et al.* Cell type specific transcriptomic differences in depression show similar patterns between males and females but implicate distinct cell types and genes. *Nat Commun* **14**, 2912 (2023). https://doi.org:10.1038/s41467-023-38530-5
28. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature methods* **16**, 1289-1296 (2019).
29. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e3529 (2021). https://doi.org:https://doi.org/10.1016/j.cell.2021.04.048
30. Ming, G.-l. & Song, H. Adult neurogenesis in the mammalian brain: significant answers and significant questions. *Neuron* **70**, 687-702 (2011).
31. Abbott, L. C. & Nigussie, F. Adult neurogenesis in the mammalian dentate gyrus. *Anatomia, Histologia, Embryologia* **49**, 3-16 (2020). https://doi.org:https://doi.org/10.1111/ahe.12496
32. Gil-Sanz, C. *et al.* Cajal-Retzius cells instruct neuronal migration by coincidence signaling between secreted and contact-dependent guidance cues. *Neuron* **79**, 461-477 (2013). https://doi.org:10.1016/j.neuron.2013.06.040
33. Tao, R. *et al.* GAD1 alternative transcripts and DNA methylation in human prefrontal cortex and hippocampus in brain development, schizophrenia. *Mol Psychiatry* **23**, 1496-1505 (2018). https://doi.org:10.1038/mp.2017.105
34. Gutiérrez-Sacristán, A. *et al.* Text mining and expert curation to develop a database on psychiatric diseases and their genes. *Database (Oxford)* **2017** (2017). https://doi.org:10.1093/database/bax043
35. Kendall, K. M. *et al.* The genetic basis of major depression. *Psychol Med* **51**, 2217-2230 (2021). https://doi.org:10.1017/s0033291721000441
36. Perez, J. M. *et al.* Hippocampal subfield transcriptome analysis in schizophrenia psychosis. *Molecular Psychiatry* **26**, 2577-2589 (2021). https://doi.org:10.1038/s41380-020-0696-6
37. Li, J. *et al.* Application of Weighted Gene Co-expression Network Analysis for Data from Paired Design. *Scientific reports* **8**, 622 (2018). https://doi.org:10.1038/s41598-017-18705-z
38. Werth, E. G., Rajbhandari, P., Stockwell, B. R. & Brown, L. M. Time Course of Changes in Sorafenib-Treated Hepatocellular Carcinoma Cells Suggests Involvement of Phospho-Regulated Signaling in

Ferroptosis Induction. *Proteomics* **20**, e2000006 (2020). https://doi.org:10.1002/pmic.202000006

39. Yang, Y. *et al.* Lipid metabolism regulator human hydroxysteroid dehydrogenase-like 2 (HSDL2) modulates cervical cancer cell proliferation and metastasis. *J Cell Mol Med* **25**, 4846-4859 (2021). https://doi.org:10.1111/jcmm.16461
40. Liaci, C. *et al.* Loss of ARHGAP15 affects the directional control of migrating interneurons in the embryonic cortex and increases susceptibility to epilepsy. *Front Cell Dev Biol* **10**, 875468 (2022). https://doi.org:10.3389/fcell.2022.875468
41. Zamboni, V. *et al.* Disruption of ArhGAP15 results in hyperactive Rac1, affects the architecture and function of hippocampal inhibitory neurons and causes cognitive deficits. *Scientific Reports* **6**, 34877 (2016). https://doi.org:10.1038/srep34877
42. Zhang, P. *et al.* CPNE8 Promotes Gastric Cancer Metastasis by Modulating Focal Adhesion Pathway and Tumor Microenvironment. *Int J Biol Sci* **18**, 4932-4949 (2022). https://doi.org:10.7150/ijbs.76425
43. Zhang, S., Glukhova, S. A., Caldwell, K. A. & Caldwell, G. A. NCEH-1 modulates cholesterol metabolism and protects against α-synuclein toxicity in a C. elegans model of Parkinson's disease. *Hum Mol Genet* **26**, 3823-3836 (2017). https://doi.org:10.1093/hmg/ddx269
44. Ramnauth, A. D. *et al.* Spatially-resolved transcriptomics of human dentate gyrus across postnatal lifespan reveals heterogeneity in markers for proliferation, extracellular matrix, and neuroinflammation. *bioRxiv* (2023). https://doi.org:10.1101/2023.11.20.567883
45. Liu, S. *et al.* Generation of functional inhibitory neurons in the adult rat hippocampus. *Journal of Neuroscience* **23**, 732-736 (2003).
46. Felger, J. C. *et al.* What does plasma CRP tell us about peripheral and central inflammation in depression? *Molecular Psychiatry* **25**, 1301-1311 (2020). https://doi.org:10.1038/s41380-018-0096-3
47. Enache, D., Pariante, C. M. & Mondelli, V. Markers of central inflammation in major depressive disorder: A systematic review and meta-analysis of studies examining cerebrospinal fluid, positron emission tomography and post-mortem brain tissue. *Brain, Behavior, and Immunity* **81**, 24-40 (2019). https://doi.org:https://doi.org/10.1016/j.bbi.2019.06.015
48. Zhan, Y. *et al.* Alterations of multiple peripheral inflammatory cytokine levels after repeated ketamine infusions in major depressive disorder. *Translational Psychiatry* **10**, 246 (2020). https://doi.org:10.1038/s41398-020-00933-z
49. Liu, J., Yao, Y., Hu, Z., Zhou, H. & Zhong, M. Transcriptional profiling of long-intergenic noncoding RNAs in lung squamous cell carcinoma and its value in diagnosis and prognosis. *Molecular genetics & genomic medicine* **7**, e994 (2019).
50. Farmiloe, G., Lodewijk, G. A., Robben, S. F., van Bree, E. J. & Jacobs, F. M. J. Widespread correlation of KRAB zinc finger protein binding with brain-developmental gene expression patterns. *Philos Trans R Soc Lond B Biol Sci* **375**, 20190333 (2020). https://doi.org:10.1098/rstb.2019.0333

Methods

Brain Collection

Brain tissue was obtained from the Brain Collection of the New York State Psychiatric Institute (NYSPI) at Columbia University, which includes brain samples from the Republic of Macedonia[1], and the Human Brain Collection Core (HBCC) of the National Institute of Mental Health (NIMH). Brain tissue collection was conducted with IRB approval and informed consent obtained from next of kin who agreed to donate the brains and participate in psychological autopsy interviews[2] which provide diagnoses according to the Diagnostic Statistical Manual of Mental Disorders (DSM) through structured proxy-based interviews comparable in reliability to the Structured Clinical Interview for DSM diagnoses (SCID) in living patients, and has been extensively used in postmortem studies[3-7].

Subjects

We included 106 subjects with sudden death, no neuropathology including microvascular disease, short agonal state, no resuscitation with prolonged (>10 min) hypoxia, no alcohol or drug use disorder, or chronic illness that may affect CNS function (e.g. epilepsy, renal failure, metastatic malignancy, HIV, Supplementary Table 1).

All 51 subjects with Major Depressive Disorder (MDD) included in this study had a diagnosis of MDD, a major depressive episode within 4 months of death, were not in remission at death, had no history of treatment during the three months before death, and had negative toxicology for psychoactive drugs, except for 15 subjects with positive toxicology for benzodiazepines that could have been given at the time of death in the emergency room, or previously prescribed.

All 55 controls included in this study had no neurological or psychiatric diagnoses, no pathological or traumatic condition affecting the brain, clear toxicology reports, and had no history of psychiatric treatments or suicide attempts.

Brain tissue processing

Right hemispheres were sliced at 2-cm intervals, flash frozen and stored at -80ºC at the time of death. Routine neuropathological examination and toxicology exams were performed. If the subject was 45 years old or older, hippocampal formation and neocortex are stained for senile plaques and neurofibrillary tangles.

The whole hippocampal formation was dissected from two or three consecutive 2 cm-thick frozen coronal blocks. Frozen adjacent sections of 50 m thickness were collected in tubes for snRNA/ATAC-seq, and two adjacent sections of 14 m were mounted on barcoded slides for Visium spatial transcriptomics (10X Genomics) per donor. Frozen tissue sections (16 m) were also used for qPCR, RNAscope® Duplex, Multiplex and Hiplex in situ hybridization (ACDbio), and for Western Blots (WB). Tissue fixed in 4% paraformaldehyde (length of tissue fixation was 5’ per each mm of block thickness)

was used for immunohistochemistry (IHC) and immunofluorescence (IF). Supplementary Table 1 describes experiments performed in each sample.

Nuclei isolation

We adapted methods from Maitra et al. 2021[8] for isolating high-quality nuclei from human frozen brain tissue. Using a glass douce tissue homogenizer (Wheaton, 357542), approximately 90 mg of tissue were homogenized in 5mL lysis buffer (10mM tris-HCl pH 7.4, 10mM NaCl, 3mM $MgCl_2$ $6H_2O$, 0.05% NP-40) and incubated on ice for 5 minutes. Then, 5mL nuclei wash buffer (NWB: 5% BSA, 0.25% glycerol, 0.001% protector RNase inhibitor (MilliporeSigma, cat. no. 333539901) in 1x PBS) were added to sample to quench the lysis. Homogenate was filtered through a 40 m Corning cell strainer (Millipore Sigma, cat. no. CLS431750). Filtrate was centrifuged three times at 500g for 5 minutes at 4°C. After each incubation, pellet was resuspended in NWB. Final resuspension was ultracentrifuged at 10,000g for 30 minutes at 4°C over iodixanol cushion (Millipore Sigma, cat. no. D1556-250). Nuclei pellet was resuspended in less than 100 L nuclei wash buffer (10X Genomics). Nuclei concentration and cell viability were estimated using trypan blue and DAPI in Countess II Hemocytometer (Applied Biosystems). Nuclei concentrations varied between 2-9 10^6 nuclei/mL.

Library preparation and sequencing

Approximately 16,000 nuclei for each sample were loaded into the 10X Chromium Controller immediately after quantification. Gene expression (GEX) and Assay for Transposase assessable chromatin (ATAC) libraries were prepared according to Chromium Next GEM Single Cell Multiome ATAC + GEX (CG000338) User Guide. We generated 2-4 libraries per sample. The libraries were sequenced on an Illumina NovaSeq 6000 using v4 chemistry to an average sequencing depth of 80,611 reads per cell for RNA libraries and 37,119 reads per cell for ATAC libraries (Supplementary Tables 3 and 7).

Pre-processing and filtering multiome data

A count matrix and a fragment file were generated by 10X CellRanger (v2.0.0) pipeline with the chemistry parameter set to Single Cell Multiome ATAC + GEX v1 (Supplementary Table 3). The reads were aligned to the GRCh38 human genome. RNA and ATAC data for each sample were preprocessed in parallelly using Seurat[9,10] (V5) and Signac[11] (V1)R packages. The count matrix was filtered for cells that have greater than 500 total number of unique transcripts and 200 genes expressed detected to remove dead or dying cells, less than 10% mitochondrial DNA, less than 150,000 UMIs, and less than 20,000 genes expressed to remove multiplets, an artifact that occurs when two or more cells are encapsulated in the same GEM (Gel beads-in-EMulsion) and are tagged with the same barcode (Extended Data Fig. 3e). Multiplets were then removed by considering canonical markers of expected cell types. Clusters that expressed multiple markers of more than one cell type will be removed. Remaining cells were pre-processed again using the standard Seurat workflow. For all samples, peaks BED (Browser Extensible Data) files output from CellRanger were converted to GenomicRanges[12], an R/Bioconductor package for

manipulating genomic intervals and combined to create a unified peak set representing all samples in the dataset. Peaks longer than 10,000 and shorter than 20 bases were removed. Barcodes for cells that were previously identified to be multiplets, dying, or quality were removed. Cells with transcription start site (TSS) enrichment score less than 1, nucleosome signal greater than 3, ATAC read counts less than 200 and greater than 100,000 were removed (Extended Data Fig. 3e). Cells that passed quality control (QC) in both modalities were retained for further analysis.

Multi-modal integration

In RNA and ATAC modalities, 52 samples were merged. Genes and peaks expressed in less than 0.01% of cells that passed QC were removed. For chromatin accessibility data, we followed the Signac workflow for merging multiple objects. Peaks on nonstandard chromosomes and in blacklist regions from the hg38 genome were removed. Merged gene expression data were processed as follows: to account for different number of counts per gene in each cell, the gene counts for each cell were divided by the sum of counts for that cell, multiplied by 10,000 and natural log transformed using Seurat function NormalizeData with default parameters. The intersection of the top 2,000 variable genes and a list of 93 genes involved in neurogenesis and neural maturation (sum: 2,065 genes) was passed to Seurat function ScaleData for linear transformation, to center mean expression across cells for each gene to zero, and scale expression of each gene, so that highly expressed genes were not given more weight in downstream analysis. Principal component analysis (PCA) was run (RunPCA) with features set to these 2,065 genes, to prevent overfitting. Otherwise, default parameters were used. Fifty PCA cell embeddings were passed to Seurat's wrapper for Harmony[13] with *group.by.vars* set to sample, donor, and batch. The Harmony-corrected PCA embedding was used for UMAP (Uniform Manifold Approximation and Projection) and Louvain clustering (Extended Data Fig. 2). Similarly, accessible chromatin data are processed using Signac workflow. We used latent semantic indexing (LSI) first by term frequency-inverse document frequency normalization (TF-IDF) and then singular value decomposition (SVD) dimension reduction with Signac functions *RunTFIDF* and *RunSVD*. TF-IDF accounts for differences in sequencing depth between cells and gives higher values to rarer peaks, so they are not overlooked in downstream analysis. All features in the TF-IDF matrix are passed to SVD. A total of 50 LSI components, excluding the first, were passed to *RunHarmony* with *group.by.vars* parameter set to sample, donor, and batch. The Harmony-corrected PCA and LSI reductions were passed to Seurat *FindMultiModalNeighbors* function for Weighted Nearest Neighbor (WNN) analysis[9]. The weighted nearest neighbor and weighted shared nearest neighbor graphs were passed to UMAP and Smart Local Moving (SLM) clustering algorithm according to Seurat WNN vignette. Unsupervised clustering implemented with *FindClusters* function in Seurat with resolution set to 1.6 yielded 113 clusters. Clusters with less than 20 cells and clusters driven by one donor were merged with the nearest cluster using *BuildClusterTree* with slot set to 'scale.data' resulting in 31 clusters (Fig. 1d, Extended Data Fig. 3). We then created a gene activity matrix inferred from single nuclei ATAC sequencing (snATAC-seq) data with the *GeneActivity* function in Signac which quantifies ATAC-seq counts in the 2 kb-upstream region and gene body.

Cell cluster identification

The 31 cell clusters were annotated using canonical gene expression in accordance with published single-cell transcriptomics studies of mammalian hippocampus[14-18], markers identified by a Wilcoxon rank-sum test implemented by FindAllMarkers function in Seurat (parameter setting RNA: only.pos = TRUE, min.pct = 0.2, min.diff.pct = 0.1, logFC_threshold = 1; parameter setting ATAC: only.pos = T, min.pct = 0.05, test.use = 'LR', latent.var = 'nCount_ATAC'), and prediction scores from integration with spatial data implemented with TransferData function in Seurat. The following canonical cell types were annotated using the respective recognized markers (ordered as in Fig. 1f-g): oligodendrocyte progenitor cell (OPC): *PDGFRA*[14,16,19-21], *OLIG1*[21]; oligodendrocyte (olig) *MBP*[19,22]; vasculature cells including endothelial cells (Endo): *FLT1*, *PECAM1* (protein name: CD31), *CLDN5*, *NES*; vascular leptomeningeal cells (VLMC): *CEMIP*, *COL1A2*; pericyte (Peri): *PDGFRB*, *ABCC9*; immune cells including microglia (Micro) *APBB1IP*, *PTPRC*[14,20], *P2RY12*[14], macrophage (Macro): *CD163* (alias is macrophage associated protein); T-cell: *CD247* (alias is T-cell surface glycoprotein); ependymal cells of the subventricular zone (Epe): *CFAP54*, *DNAH9*, *FOXJ1*; choroid plexus cells (Ch.Pl): *HTR2C*, *CLIC6*, *TTR*; astrocytes (Astro, 3 clusters): *AQP4*, *SLC1A2*, *ALDH1L1*, *GLUL*, *GFAP*, with subtypes Astro1: *GPC5*, *WIF1*, GFAP; Astro2: *CD44*, *TNC*, *GFAP*; Astro3: *TNC*, *GFAP*; Cajal Retzius cells (CR): *RELN*[14,16]; granule neurons (GN, 2 clusters): *RBFOX3*, *SYT1*, *MYT1L*, *SLC17A7*, *CAMK2A*, *PROX1*, *CALB1*, *TLL1*, *PPFIA2*, *SEMA5A*, *BDNF*, with subtypes *POSTN*$^{-}$ (GN1) and *POSTN*$^{+}$ (GN2); excitatory neurons (ExN, 6 clusters): with *RBFOX3*, *MYT1L*, *CAMK2A*, *SLC17A7*, *SV2B* expressed by all ExN and specific subtype markers for ExN1: *CADPS2*, *NEUROD6*; ExN2: *SATB2*, *SLC22A10*; ExN3: *NEUROD6*, *TRHDE*, *CPNE4*, *HTR7*; ExN4: *SATB2*, *SLC22A10*, *HS3ST2*, *SYNPO2*; ExN5: *HS3ST2*, *NTNG1*; ExN6: *HS3ST2*, *CHRM2*, *PCP4*; inhibitory neurons (InN, 9 clusters): all expressing *RBFOX3*, *MYT1L*, *GAD1*, *GAD2*, *SLC32A1*, *DCX*, with InN subtypes expressing: *VIP*, *SST*, *PVALB*, *LHX6*, *LAMP5*, *SLC17A8*, *TUBB3*, *PROX1*, *PENK*.

Out of the 31 cell clusters (Extended Data Fig. 4), 12 cell types were identified by merging all astrocyte clusters, all inhibitory neuron clusters, all immune clusters, all vasculature clusters, the two GN clusters, ExN1 and ExN3 merged to CA.ExN (*EPHA5*[16], *NEUROD6*[19]), and ExN2 and ExN4-6 merged to Other ExN (O.ExN; expressing *TSHZ2*[23,24]).

Single nuclei integration with other datasets

We integrated our database with publicly available databases to achieve a better characterization of the molecular landscapes defining cell types and states in the human hippocampus in neurotypical controls and MDD. We downloaded the count matrix of datasets generated by recently published studies focusing on the hippocampus[14,16,25,26]. Independently generated datasets were preprocessed using our standard pipeline using Seurat[10] and Harmony[13] in R. Then, following a published approach[14], for the top 2000 variably expressed genes, Pearson's R was calculated for the average expression across cell types (Extended Data Fig. 5).

Tissue and library preparation for Visium spatial gene expression

Two adjacent sections per donor were sectioned at 14 µm at -20C and mounted within capture areas on Visium slides. Slides were stored at -80C, then fixed, stained, and imaged according to 10X User Guide CG000160 Rev C. Spatial gene expression libraries were prepared according to 10X User Guide CG000239 Rev F. The libraries were sequenced on an Illumina NovaSeq 6000 using v4 chemistry to an average sequencing depth of about 67,000 reads/spot and 1965 median genes/spot (Supplementary Table 7). SpaceRanger (v1.3.0) with chemistry set to Spatial 3' v1 used to generate count matrices.

Tissue preparation and loading for Xenium in-situ hybridization

Fresh frozen tissue was sectioned at 10µm onto 10.45mm x 22.45mm sampling area on 10X Genomics Xenium Slide according to 10X Demonstrated Protocol CG000579 Rev C. Tissue was fixed and permeabilized according to 10X Demonstrated Protocol CG000581 Rev C. Probe hybridization, ligation, and amplification were carried out according to 10X User Guide CG000582 Rev C. The output from the Xenium analyzer was loaded into Xenium Explorer 1.3 or RStudio for integration with single-nuclei gene expression data.

Spatial gene expression data pre-processing, integration, and clustering

Visium (10X Genomics) count matrices and spatial data were loaded into Seurat using the Load10X_Spatial command with the default parameters. Feature counts for each spot were divided by total counts for that spot and multiplied by scale factor (10,000) and natural log transformed using function NormalizeData. Features that were outliers to mean variability were found with local polynomial regression of log(variance) and log(mean) which gives the observed mean and expected variance of feature expression implemented using the FindVariableFeatures function. Then, data were scaled and centered using ScaleData function. Spatial data from replicates were merged using the merge function. Dimensionality reduction and clustering were calculated using RunPCA, FindNeighbors, FindClusters, and RunUMAP with dimensions parameter set to 1:30.

Spatial data were integrated with multiome data using FindTransferAnchors and TransferData functions in Seurat[10] (Fig. 1h, Fig. 2e).

Hippocampalsubfields, including granule cell layer (GCL), subgranular zone (SGZ), molecular layer (ML), *Cornu Ammonis* (CA), white matter (WM), vasculature (vasc), chorioid plexus (chor.plex), and subventricular zone (SVZ) were manually annotated in Loupe Browser within the Visium fiducial frame (Fig. 2b). Folds in the tissue were labeled as such and excluded from downstream analysis. Spots with less than 500 unique molecular identifiers were excluded from downstream analysis. Unsupervised clustering of merged spatial gene expression data was driven largely by sample and donor identify, so spatial gene expression data were integrated using canonical correlation analysis (CCA) in Seurat (Extended Data Fig. 6a-b).

Spatial cluster identification

Using *FindClusters* in Seurat, 13 clusters were identified at 0.5 resolution: two hilus clusters, hilus1 comprising mostly excitatory neurons, and hilus2 showing markers of inhibitory neurons, mossy cell markers, serotonin receptor 2c (*HRT2C*) and tryptophan hydroxylases 2 (*TPH2*); CA3 stratum pyramidales (ca3) comprising excitatory neurons; stratum oriens-stratum lucidum (so-slu) representing WM; so-slu of the CA3 (ca3-so-slu) expressing parvalbumin and markers involved in cell-cell adhesion, extracellular matric, methylation, calcium signaling, and trophic factors; stratum radiatum (sr) comprised of axonal markers; stratum oriens-stratum lacunosum (so-sla) made up of dendrites; stratum moleculare (ml) comprised of apical dendrite-associate markers and *HTR2B*; stratum granulosum (gcl) made up mostly of GNs; SGZ-polymorphic layer (sgz-pl) expressing markers involved in neurogenesis, development and differentiation, *TPH1* and *TPH2*; scattered spots in SGZ and stratum moleculare (sgz-ml) expressing stem cell, differentiation and neurogenesis markers and *HTR3A* and *HTR1D*; vasculature (vasc) expressing endothelial and collagen markers, and *HTR2B* and *TPH1*; and choroid plexus (ch.pl) expressing cilia and flagella associated markers, *HTR2C*, and astrocyte markers (Supplementary Table 8).

Integration of spatial and single-nucleus gene expression datasets

We used Seurat v3 anchor-based integration to transfer annotations from single nucleus data to spatial Visium data. Single-nucleus annotations were used as the reference dataset and both whole transcriptome Visium spatial gene expression and Xenium in-situ hybridization were used as query data. Spots were assigned probability scores based on transcriptional congruence with cell types from single nucleus data (Fig. 1h-i, Fig. 2e-f, Extended Data Fig. 4f).

RNA Velocity inference on Visium spatial transcriptomic data

To understand the kinetics of cell state in human hippocampus, we computed RNA velocity using velocyto[27] based on the relative abundance of nascent (unspliced) mRNAs and mature (spliced) mRNAs on Visium spots (55x55 m). Samtools[28] (v1.9) was used to sort bamfiles and *velocyto run*[27] was used to run velocity analysis and output a loom file. Pseudotime lineage analysis was performed by RNA velocity (scVelo Python package[29] , Fig. 2g-k, Extended Data Fig 6). The proportion of spliced/unspliced RNAs did not differ between unsupervised spot clusters, and we did not have power to test differences between MDD and CTRL.

We used unsupervised clusters located in SGZ and GCL (Supplementary Table 8) that expressed proliferation, differentiation, maturation and neurogenesis markers among their top 100, to subset spots hypothesized to be part of neurogenesis trajectory (Extended Data Fig. 6c-h). We selected spatial clusters sgz.ml, sgz.pl, and gcl (Supplementary Table 8) and recalculated UMAP embedding using the top 5000 variably expressed genes and additional genes known to be involved in neurogenesis (total of 5074 genes) and the first 15 PCs for FindNeighbors and RunUMAP. Seurat object was translated to AnnData using SeuratDisk functions SaveH5Seurat and Convert with ‘dest’ set to ‘h5ad.’ The ‘data’ layer from the Seurat object was used for gene expression calculations in scvelo. The AnnData object was

merged with loom files output of velocyto run on the command line. The dynamical model was run with the default parameters.

Identification of neurogenic niche stem cell progeny using machine learning on single nuclei transcriptome data

We apply a supervised learning approach to identify neurogenic cell types, as recently published[16]. Prototypes for each cell type were discovered by first retaining the sparse counts matrix for the top 4000 variable genes and an additional 54 genes involved in neurogenesis and neuronal maturation. The Seurat function *NormalizeData* was rerun on this matrix with the default parameters. For each gene, normalized counts that were above the sum of the third quartile and the product of 1.5 times the inter-quartile range were brought down to this value to reduce the effect of extreme outliers. Prototype scores were assigned to cells for identification of the following cell types: TypeI progenitors (quiescent neural stem cells), TypeII progenitors (intermediate amplifying neural progenitors), TypeIII (differentiating neurons), and ImGN (immature granule neurons, which are maturing and integrating into the DG circuit) using normalized count data. The distribution of summed normalized counts of selected genes expected to be expressed at each maturation stage was used to determine a cutoff for defining prototypical cells, considering that, based on previous studies in rodents and primates, progenitors are expected to be a small proportion of the total number of adult human dentate gyrus cells (<1% of all cells). We summed normalized counts of the following stage-specific genes: for TypeI, we summed normalized counts for *GFAP*, *SOX2*, *PAX6*, *NES*, *FABP7*, *HES5*, and *ETNPPL*; for TypeII, normalized counts for *ASCL1*, *FOXO3*, *NR2E1*, *EOMES*, *NEUROG2*, *NEUROD1*, *NES*, *MCM2*, *PCNA*, *MKI67*, and *TOP2A* were summed, and because oligodendrocyte progenitors can express markers in common to other progenitors, normalized counts for *MBP* and *MOBP* were subtracted so that cells that could be on the oligodendrocyte differentiation trajectory were less likely to be determined to be TypeII prototypes; for TypeIII, we summed normalized counts for *MYT1L*, *PROX1*, *DCX*, *CALB2*, *RELN*, *ST8SIA2*, and *ST8SIA4* and subtracted normalized counts for *RBFOX3* and *VIP* to decrease the likelihood that TypeIII prototypes were at a later stage of maturation or inhibitory neurons; for ImGN prototypes, normalized counts of *RBFOX3*, *PROX1*, *NEUROD1*, *DCX*, *BHLHE22*, *COL25A1*, and *POSTN* were summed. Cells with high prototypes scores were identified as prototypical, or high confidence cells.

We applied the LASSO regression (implemented with *glmnet* in R[30]) for each cell type individually. For each cell type, the prototypes and a randomly sampled quantity of cells required to reach 50,000 were used for training and testing the model. The trained model uses a list of positively and negatively ranked coefficients to rank genes in their ability to predict cell types. This method uses many genes to characterize these cell types (for TypeI, TypeII, TypeIII, and ImGN it used 372, 410, 555, and 422 genes respectively), instead of just a few arbitrarily chosen genes. Few cells (approximately 150 cells) were assigned two identities. For these cells, we manually assigned the identity with the highest predicted probability from *predict.glm* with type set to 'response' and the tuning parameter s was chosen by 10-fold cross validation by cv.glmnet. For final counts of cells in each LASSO identified cell type, see Supplementary Table 9.

Cell trajectory inference based on single nucleus transcriptomic data

In the absence of cytoplasmic mRNA and thus degradation rate in the single-nuclei gene expression data, we used Palantir[31] to characterize the trajectory of cells identified as TypeI, TypeII, TypeIII, ImGN, and GN from LASSO regression. TypeI, TypeII, TypeIII, ImGN, and GN cell types were subset, and re-preprocessed using Seurat and Harmony as described in Multi-modal integration. The first 5 Harmony batch, sample, and donor corrected PCs were used for FindNeighbors and RunUMAP (Fig. 3b). Seurat object was converted to AnnData using SeuratDisk functions SaveH5Seurat and Convert with 'dest' set to 'h5ad.' Palantir function core.run_palantir was used to calculate psueodtime ordering of each cell (Fig. 3c), terminal state probabilities, and entropy (a quantitive measure of the differentiation potential of each cell). Generalized Additive Models[31] were used to determine the gene expression trends along the neurogenesis lineage using the function presults.compute_gene_trends (Fig. 3e).

Differential gene expression analysis in MDD *versus* CTRL

To determine cell-type specific differentially expressed genes (DEGs) between subjects with MDD and neurotypical controls, count, peak, and gene activity matrices were aggregated on the sample and cell type level. Data were normalized using center log ratio transformation implemented using the clr function in the compositions R package. Then, we used the differential expression pipeline from R package limma[32,33]. Samples from the same donor were not aggregated. Instead, we used limma's DuplicateCorrelation function to add a random effect for donor identity because, in some cases, samples from the same donor were processed in different batches. Additionally, since donors had between 2-3 samples, we used voomWithQualityWeights to adjust for differences in the number of samples per donor and number of cells aggregated/sample. We added RIN, PMI, batch, and age to the model as fixed effects. For RIN, PMI, batch, and age, values that were outside the sum of the third quartile and the product of 1.5 times the inter-quartile range from the median were winsorized to reduce the effect of extreme outliers. P-values were adjusted using Benjamini-Hochberg correction. We implemented the Bioconductor scRNA-seq pseudobulk pipeline to analyze DEGs between MDD and neurotypical controls in all clusters[34-38].

Differential chromatin accessibility analysis in MDD *versus* CTRL

To determine cell-type specific differentially accessible regions (DARs) between subjects with MDD and neurotypical controls, the peaks matrix was aggregated and the pipeline for differential gene expression was applied. Additionally, a gene activity matrix was created from snATAC-seq using GeneActivity function in Signac. Chromatin accessibility at transcriptional and promoter regions was used to create the gene activity matrix. Then, the pipeline for predicted differentially active genes (DAGs) was applied to the aggregated gene activity matrices.

Pre-ranked gene set enrichment analysis

We performed pre-ranked Gene Set Enrichment Analysis (GSEA)[39] with FGSEA[40] using the product of log fold change and negative base 10 log of the Benjamini-Hochberg (BH) corrected p-value as the ranking method. We used the Gene Ontology (GO) annotation system to determine molecular function (MF, the molecular activities carried out by gene products), the cellular component (CC, the cellular location where gene products are active and perform their function), and biological process (BP, the processes and activities in which gene products are involved), with gene sets obtained from Molecular Signatures Database (MSigDB)[41]. We used Kyoto Encyclopedia of Genes and Genomes (KEGG)[42] to infer biological pathways related to DEGs, and used the Reactome pathway[43] to understand molecular mechanisms of biological reactions that are interconnected and result in pathways. The following parameters were used for the *fgsea* function: minSize=15, maxSize=400, nperm = 10000. Outputs with BH adjusted p-value less than 0.05 were considered significant.

PsyGeNET analysis

We ran *enrichedP*D from Psygenet2r[44] a R/Bioconductor package for the analysis of psychiatric disease genes, for leveraging PsyGeNET[45] a knowledge resource that collects and validates information of psychiatric diseases and their genes. We used database = 'ALL' to find psychiatric disorders for which DEGs from any cell type showed enrichment. We tested all DEGs from gene expression, all DEGs from gene activity, and the combination, and ran *psygenetGene* with database = 'ALL' and created *geneAttrPlot* for the evidence index. We then ran *psygenetGene* with DEGs from GN and ImGN (Fig. 4c) which were the cell types with most dysregulated genes across multiple modalities: transcript, chromatin accessibility and gene activity.

Weighted gene co-expression network analysis (WGCNA)

Weighted gene co-expression network analysis (WGCNA) was performed to identify co-expression modules using snRNA-seq data[22,46]. First, the aggregated expression for each sample (N = 52; 25 CTRL, 27 MDD) in the GN and ImGN cell types was calculated by summing the counts per gene across all nuclei. To account for known external sample traits, the counts were corrected for age and batch using limma[47]. In addition, lowly expressed genes with total counts of below 5 were removed. A soft thresholding power of 20 and 18, respectively, and a minimum module size of 30 genes, were used for network construction and module detection for GN. Each module was correlated with the phenotype (CTRL *vs.* MDD), and significance was determined using a p-value < 0.05. The online tool Metascape[48] was used to perform GO on the modules that showed a correlation with MDD.

In Situ Hybridization

For each donor, 3 serial sections of 16 m at 80 m intervals were used for Singleplex and Duplex RNAScope assays. For Singleplex RNAscope (ACDbio), the *DCX* probe (RNAscope Probe-Hs-DCX, Cat. No. 489551) was hybridized and amplified. Duplex RNAscope was used to investigate *DCX* co-localization with *NEUROD1*, *PROX1*, *RBFOX3*, or *TUBB3* mRNA in the human DG. Fresh frozen 16 m

tissue sections were fixed in 4% paraformaldehyde for 15 minutes at room temperature (RT). Sections were washed in 1x PBS and then dehydrated in 50%, 70%, and twice in 100% EtOH for 5 minutes. RNAscope was performed according to manufactured instructions. The *DCX* probe (Cat. No. 489551-C2) and *NEUROD1* probe (Cat. No. 437281), *PROX1* probe (Cat. No. 530241), *RBFOX3* probe (Cat. No. 415591), or *TUBB3* probe (Cat No. 318901) were hybridized and amplified in C2 and C1 respectively. Slides were counterstained with 50% hematoxylin (Gil's Hematoxylin No. 1) according to manufacturer's instructions. The slides are then quickly dipped in xylene and mounted with Permount.

Multiplex RNAscope was used to investigate the expression of *GAD1* (Cat No 404031-C2), *PROX1* (Cat No 530241-C3), and *DCX* (Cat No 489551) in the human DG, because single nuclei data showed co-expression of *GAD1* in cell clusters expressing *DCX*. Sections were counterstained with 4', 6-diamidino-2-phenylindole (DAPI).

Gene expression analysis using qPCR

RNA was isolated using RNeasy FFPE Kit (Qiagen, Germantown, MD) according to the manufacturer's instructions. RNA quantity was assessed by evaluation of the A260/280 and A260/230 ratios using a Nanodrop spectrometer (Thermo Scientific 840274100). Superscript III enzyme (Invitrogen) was used to reverse transcribe 200 ng total RNA. Quantitative Real-Time PCR was performed on a StepOnePlus™ Real-Time PCR System using a TaqMan™ Gene Expression Master Mix (Applied Biosystems, Waltham, USA). A ready-made *DCX* TaqMan™ probe (Applied Biosystems, Waltham, USA) was used to amplify a 77bp region spanning an exon boundary within the coding sequence of the target gene. Each sample was assayed in triplicate and *DCX* target gene expression was normalized to the geometric mean of the reference gene beta-actin (*ACTB*). *DCX* gene expression was analyzed in PRISM using a Welch unpaired T-test.

Double Immunofluorescence

To examine expression of cell type specific markers at the protein level in human hippocampus, we performed double immunofluorescence (IF) for protein corresponding to RNAs of interest identified with the single nuclei and spatial transcriptomic analyses.

Human hippocampus tissue sections (50 m thick) were washed and prepared as previously published[49]. The primary antibodies were incubated [(1:1000 guinea pig anti-DCX (Millipore), 1:5000 mouse anti-Glial Fibrillary Associated Protein (GFAP, Sigma-Aldrich), 1:500 rabbit anti-Ionized calcium binding adaptor molecule 1 (Iba1, Fujifilm), 1:50,000 mouse anti-Neurofilament (NF; Sigma-Aldrich); 1:2000 mouse anti-polysialylated neural cell adhesion marker (PSA-NCAM, Miltenyi Biotec); 1:500 rabbit anti-Calretinin (Swant Inc.), 1:500 rabbit anti-Calbindin (Swant Inc.), 1:500 rabbit anti-Neuronal Nuclear Marker (NeuN, Millipore)] in a 25% blocker overnight or for 5-days (Calretinin and Calbinin) at 4°C. The secondary antibodies were incubated overnight at 4°C with Jackson Alexa Fluor 488 goat anti-mouse or anti-rabbit (1:500) and Jackson Alexa Fluor 594 goat anti-guinea pig (1:500). The following day the tissue was washed for 30min (2x15') in 0.05M PBS. Then the tissue was counterstained with 4', 6-diamidino-2-

phenylindole (DAPI) in PBS for 3min before being washed for 15min (3x5') in 0.05M PBS. The tissue was then treated with 0.1% True Black (Biotium) in 70% ethanol in distilled water to quench autofluorescence. Sections were then washed for 1h (6x10') in 0.05M PBS. The tissue was coverslipped using 90% glycerol and PBS.

Immunohistochemistry

Immunohistochemistry (IHC) for DCX was performed as previously described[49] Sections were incubated in guinea pig anti-doublecortin (1:30,000, Millipore) in 10% blocking solution overnight at 4°C on a shaker. Following 2-hour secondary incubation with anti-guinea pig biotin-conjugated antibody (Vector Laboratories) and 1-hour incubation in avidin-biotin-horse radish peroxidase system (Vector Laboratories), diluted 1:200, sections were washed in sodium acetate for 5 min, followed by a PBS wash for 20 min (2x10') before being mounted on glass slides and desiccated. Sections were counterstained with Nuclear Red which involves dehydrating slides in ethanol, clarifying slides in xylenes and cover-slipping with Permount (Sigma-Aldrich).

Sections were processed to identify mature granule neurons (anti-Neuronal Nuclear antigen [NeuN] mouse monoclonal antibody, 1:100,000; Chemicon, Temecula, CA), progenitor cells of Type II or III (anti-nestin mouse monoclonal antibody, 1:8000, Chemicon) and mitotic cells (anti-Ki-67 mouse monoclonal antibody, 1:200, Novocastra Clone-MM1, Newcastle Upon Tyne, UK). Immunohistochemistry and stereology were performed as previously described[50].

Confocal Microscopy

Cell markers co-localization was visualized using an Olympus Fluoview FV1000 Confocal Laser Scanning System (Olympus, Waltham, MA) equipped with an Olympus IX81 Inverted Microscope, with electronic stage controller (Prior Scientific, Rockland, MA), camera interface, and Olympus Fluoview 1000 (v. 1.5) software, using the following objectives: UPLSAPO 10X and 20X air, with numerical aperture (NA) 0.40, 0.75; UPLFLN 40X oil with NA 1.30; and PLAFON 60X oil with NA 1.42 (Olympus).

To quantify co-localization of cell markers (i.e., DCX/NeuN), image stacks of the whole DG were obtained using a confocal scanning microscope (Leica TCS SP8 2-Photon, Leica Microsystems Inc.), and then processed for Stereology using Stereoinvestigator software (MBF, Inc.). Fluorescence from the different fluorophores was detected in the following way: 1) 488-AffiniPure was excited at 488 nm and detected at 505-550 nm, 2) 594-AffiniPure was excited at 552 nm and detected at 600-650 nm, and 3) DAPI was excited at 405 nm and detected at 415-485 nm. For 488- and 552-nm excitation, the beam path included a TD 488/552/638 beamsplitter, while for 405 nm excitation a substrate beamsplitter was used.

Hippocampal sections from each subject were scanned to produce 13 ± 2 μm image stacks, depending on the thickness of each tissue section. All z stacks were imaged with a dry Leica 20X objective (NA 0.70, working distance 0.5 mm), with a field of view of 553.6 × 553.6 μm, a pixel size of 0.54 × 0.54 μm, optical sectioning of 2.36 μm, and a z step of 1 μm. Positive and negative controls were run to ensure

signal specificity and minimal background, creating an optimal signal to noise ratio. The same parameters were applied for every section of every subject. Image stacks were then imported in our MicroBrightfield system (MBF Bioscience, Williston, VT) to perform unbiased stereology.

Cell quantification and statistical analysis

We used an unbiased stereological approach (optical *disector* with fractionator) to estimate numbers of cells expressing DCX (DCX+), DCX/NeuN+ cells, DCX mRNA+ cells and DCX/TUBB3 mRNA+ cells. Olympus BX51-WI microscope was used with electronic stage controller LUDL MAC 6000 XYZ (BioVision Technologies, Exton, PA) and camera interface for Retiga 2000R CCD Scientific Camera (Q-Imaging, Surrey, BC, Canada).

For anatomical alignment along the rostro-caudal axis of the DG, processed sections were anatomically matched with Nissl-stained sections at 1-mm interval using a stereoscope (Leica, Wild M3Z, Heerburg, Switzerland). We defined the anterior DG as the portion from the most rostral appearance of the DG to the start of the lateral geniculate. The mid DG was defined as the portion going from the start of the lateral geniculate to the disappearance of the lateral geniculate, and the posterior DG was the region from the end of the lateral geniculate to the end of the caudal DG. Outlines of the DG including SGZ and GCL were aligned from the most rostral to the most caudal hippocampal sections and used as boundaries for cell counting. To quantify cells, the first hippocampus section to be sampled was the one in which the DG first appeared. Subsequent sections were assayed every 2 mm thereafter in systematic fashion, until the DG disappeared, with an average of 8-10 sections per subject for DCX immunohistochemistry experiments. It should be noted that for DCX/NeuN and RNAscope experiments, an average of 3-5 sections of the anterior hippocampus were used per subject. Using the Stereo Investigator system, DG boundaries were defined at low magnification using the 4x objective. Cells were not counted in the upper and lower 3 µm in the z planes of the tissue sections (the guard zones). The estimated total number of cells was calculated as:

$$N=\sum Q^{-}\times(1/hsf)\times(1/asf)\times(1/ssf)$$

where Σ^{Q-} = total number of cells counted, hsf (height sampling fraction) = h/t, where *t* is the mean section thickness, *h* is the height of the optical *dissector*, *asf* = area sampling fraction (area of the counting frame over the area of x,y step), and *ssf* = section sampling fraction (1/section interval).

The cell counts were analyzed in PRISM using a Welch unpaired T-test.

Proteomics

We used Shotgun proteomics liquid chromatography with tandem mass spectrometry (LC/MS/MS), a strategy with broad applicability, based on digestion of proteins with proteolytic enzymes and analysis of the resulting peptides by mass spectrometry[51].

Fresh-frozen 4% paraformaldehyde post-fixed tissue was dissolved in 0.2% RapiGest SF in 50 mM ammonium bicarbonate, heated to 105 ^{o}C for 30 min, then to 80 ^{o}C for 2 h, digested with trypsin as previously described[52,53], and then 50 fmol of yeast alcohol dehydrogenase were added as internal control.

LC/MS/MS was run for 120 minutes, two chromatograms recorded for each biological replicate, in traveling-wave ion mobility spectrometry (TWIMS) MS^{E} resolution mode [51,53,54]; mass spectra were recorded every 0.6s; every 30s internal “lockmass” spectrum of Glu-1-Fibrinopeptide B was recorded (m/z 785.8426); 11,000 mass spectra were recorded per run; a total of 792,000 mass spectra were analyzed. Of the 1811 proteins detected, 680 were represented by a single peptide or with protein identification quality scores less than 250 and were not included in the analysis.

The two groups (MDDs and CTRL) were compared against each other and differentially expressed proteins were identified with a PeptideTeller predicted error (false positive rate of 1%). P-values used in our report on these experimental results are based on the error function called xdev, which addresses false positives by using an advanced error model as described in previous papers[55,56]. The xdev-derived P-values were generated by the software Rosetta Elucidator Protein Expression Analysis System (V. 4.0.0.2.13, PerkinElmer, Boston, MA,). This software, developed by Merck & Co./Rosetta[57-59] extends the dynamic range of fold-change data, especially for low abundance proteins. This system matches accurate mass and retention time across all LC/MS/MS chromatograms in the experiment, enhancing the effectiveness of the mass spectrometry analyses. The xdev represent the difference of intensities between groups (pairwise comparisons) divided by the error of the difference, and the p-value of differential expression is calculated from an error function of a standard Gaussian distribution. Elucidator ran on Oracle Linux v. 6.5 server with 128 GB RAM and 32 Tb of RAID network attached storage. From within Elucidator, results were searched with Mascot server (Version 2.5.1, Matrix Science Ltd., London, UK). The search was against a SwissProt human canonical protein sequence database with isoforms (UniProtKB/Swiss-Prot Release 2017_01 containing 84,336 sequences; 48,572,808 residues, www.uniprot.org). Our quality control indices included peptide count above two for considering a protein a good candidate, and p-value < 0.05.

Methods References

1 Rosoklija, G., Duma, A. & Dwork, A. J. Psychiatric brain collection in Macedonia: general lessons for scientific collaboration among countries of differing wealth. *Pril (Makedon Akad Nauk Umet Odd Med Nauki)* **34**, 95-98 (2013).

2 Kelly, T. M. & Mann, J. J. Validity of DSM-III-R diagnosis by psychological autopsy: a comparison with clinician ante-mortem diagnosis. *Acta Psychiatr.Scand.* **94**, 337-343 (1996).

3 Zhang, D. *et al.* Spatial epigenome-transcriptome co-profiling of mammalian tissues. *Nature* **616**, 113-122 (2023). https://doi.org:10.1038/s41586-023-05795-1

4 Deng, Y. *et al.* Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature* **609**, 375-383 (2022). https://doi.org:10.1038/s41586-022-05094-1

5 Boldrini, M. *et al.* Resilience Is Associated With Larger Dentate Gyrus, While Suicide Decedents With Major Depressive Disorder Have Fewer Granule Neurons. *Biol Psychiatry* **85**, 850-862 (2019). https://doi.org:10.1016/j.biopsych.2018.12.022

6 Boldrini, M. *et al.* Human Hippocampal Neurogenesis Persists throughout Aging. *Cell Stem Cell* **22**, 589-599 e585 (2018). https://doi.org:10.1016/j.stem.2018.03.015

7 Besnard A, L. T., Levinson S, Chu D, Vicidomini C, Scobie KN, Dwork AJ, Arango V, Rosoklija GB, Mann JJ, Hen R, Leonardo ED, Boldrini M, Sahay A. Targeting Kruppel-like factor 9 (Klf9) in excitatory forebrain 1 neurons protects against chronic stress-induced dendritic spine enlargement and maladaptive fear responses. *Cell Reports* **In Press** (2018).

8 Maitra, M. *et al.* Extraction of nuclei from archived postmortem tissues for single-nucleus sequencing applications. *Nature Protocols* **16**, 2788-2801 (2021).

9 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e3529 (2021). https://doi.org:https://doi.org/10.1016/j.cell.2021.04.048

10 Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology* (2023). https://doi.org:10.1038/s41587-023-01767-y

11 Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nature Methods* **18**, 1333-1341 (2021). https://doi.org:10.1038/s41592-021-01282-5

12 Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLOS Computational Biology* **9**, e1003118 (2013). https://doi.org:10.1371/journal.pcbi.1003118

13 Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature methods* **16**, 1289-1296 (2019).

14 Franjic, D. *et al.* Transcriptomic taxonomy and neurogenic trajectories of adult human, macaque, and pig hippocampal and entorhinal cells. *Neuron* **110**, 452-469.e414 (2022). https://doi.org:https://doi.org/10.1016/j.neuron.2021.10.036

15 Wang, W. *et al.* Transcriptome dynamics of hippocampal neurogenesis in macaques across the lifespan and aged humans. *Cell Research* **32**, 729-743 (2022). https://doi.org:10.1038/s41422-022-00678-y

16 Zhou, Y. *et al.* Molecular landscapes of human hippocampal immature neurons across lifespan. *Nature* **607**, 527-533 (2022). https://doi.org:10.1038/s41586-022-04912-w

17 Hochgerner, H., Zeisel, A., Lönnerberg, P. & Linnarsson, S. Conserved properties of dentate gyrus neurogenesis across postnatal development revealed by single-cell RNA sequencing. *Nature Neuroscience* **21**, 290-299 (2018). https://doi.org:10.1038/s41593-017-0056-2

18 Ayhan, F. *et al.* Resolving cellular and molecular diversity along the hippocampal anterior-to-posterior axis in humans. *Neuron* **109**, 2091-2105.e2096 (2021). https://doi.org:10.1016/j.neuron.2021.05.003

19 Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nature methods* **14**, 955-958 (2017).

20 Hao, Z.-Z. *et al.* Single-cell transcriptomics of adult macaque hippocampus reveals neural precursor cell populations. *Nature Neuroscience* **25**, 805-817 (2022). https://doi.org:10.1038/s41593-022-01073-x

21 van Bruggen, D. *et al.* Developmental landscape of human forebrain at a single-cell level identifies early waves of oligodendrogenesis. *Developmental Cell* **57**, 1421-1436.e1425 (2022). https://doi.org:https://doi.org/10.1016/j.devcel.2022.04.016

22 Maitra, M. *et al.* Cell type specific transcriptomic differences in depression show similar patterns between males and females but implicate distinct cell types and genes. *Nature Communications* **14**, 2912 (2023). https://doi.org:10.1038/s41467-023-38530-5

23 Blankvoort, S., Olsen, L. C. & Kentros, C. G. Single Cell Transcriptomic and Chromatin Profiles Suggest Layer Vb Is the Only Layer With Shared Excitatory Cell Types in the Medial and Lateral Entorhinal Cortex. *Front Neural Circuits* **15**, 806154 (2021). https://doi.org:10.3389/fncir.2021.806154

24 Di Bella, D. J. *et al.* Molecular logic of cellular diversification in the mouse cerebral cortex. *Nature* **595**, 554-559 (2021). https://doi.org:10.1038/s41586-021-03670-5

25 Morabito, S. *et al.* Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nature Genetics* **53**, 1143-1155 (2021). https://doi.org:10.1038/s41588-021-00894-z

26 Grubman, A. *et al.* A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nature Neuroscience* **22**, 2087-2097 (2019). https://doi.org:10.1038/s41593-019-0539-4

27 La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494-498 (2018). https://doi.org:10.1038/s41586-018-0414-6

28 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021). https://doi.org:10.1093/gigascience/giab008

29 Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology* **38**, 1408-1414 (2020).

https://doi.org:10.1038/s41587-020-0591-3

30 Friedman, J. H., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* **33**, 1 - 22 (2010). https://doi.org:10.18637/jss.v033.i01

31 Setty, M. *et al.* Characterization of cell fate probabilities in single-cell data with Palantir. *Nature Biotechnology* **37**, 451-460 (2019). https://doi.org:10.1038/s41587-019-0068-4

32 Liu, R. *et al.* Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. *Nucleic Acids Res* **43**, e97 (2015). https://doi.org:10.1093/nar/gkv412

33 Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. K. ROBUST HYPERPARAMETER ESTIMATION PROTECTS AGAINST HYPERVARIABLE GENES AND IMPROVES POWER TO DETECT DIFFERENTIAL EXPRESSION. *Ann Appl Stat* **10**, 946-963 (2016). https://doi.org:10.1214/16-aoas920

34 Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084-2092 (2019).

35 McCarthy, D. J., Campbell, K. R., Lun, A. T. L. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179-1186 (2017). https://doi.org:10.1093/bioinformatics/btw777

36 Amezquita, R. A. *et al.* Orchestrating single-cell analysis with Bioconductor. *Nature Methods* **17**, 137-145 (2020). https://doi.org:10.1038/s41592-019-0654-x

37 Love, M. I., Anders, S. & Huber, W. Analyzing RNA-seq data with DESeq2. *Bioconductor* **2**, 1-63 (2017).

38 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014). https://doi.org:10.1186/s13059-014-0550-8

39 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005). https://doi.org:10.1073/pnas.0506580102

40 Korotkevich, G. *et al.* Fast gene set enrichment analysis. *bioRxiv*, 060012 (2021). https://doi.org:10.1101/060012

41 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015). https://doi.org:10.1016/j.cels.2015.12.004

42 Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**, D353-d361 (2017). https://doi.org:10.1093/nar/gkw1092

43 Milacic, M. *et al.* The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Research* **52**, D672-D678 (2023). https://doi.org:10.1093/nar/gkad1025

44 Gutiérrez-Sacristán, A., Hernández-Ferrer, C., González, J. R. & Furlong, L. I. psygenet2r: a R/Bioconductor package for the analysis of psychiatric disease genes. *Bioinformatics* **33**, 4004-4006 (2017). https://doi.org:10.1093/bioinformatics/btx506

45 Gutiérrez-Sacristán, A. *et al.* PsyGeNET: a knowledge platform on psychiatric disorders and their genes. *Bioinformatics* **31**, 3075-3077 (2015). https://doi.org:10.1093/bioinformatics/btv301

46 Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4**, Article17 (2005). https://doi.org:10.2202/1544-6115.1128

47 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015). https://doi.org:10.1093/nar/gkv007

48 Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523 (2019). https://doi.org:10.1038/s41467-019-09234-6

49 Boldrini, M. *et al.* Human hippocampal neurogenesis persists throughout aging. *Cell stem cell* **22**, 589-599. e585 (2018).

50 Boldrini, M. *et al.* Benzodiazepines and the potential trophic effect of antidepressants on dentate gyrus cells in mood disorders. *Int J Neuropsychopharmacol* **17**, 1923-1933 (2014). https://doi.org:10.1017/S1461145714000844

51 Brown, L. M. Quantitative shotgun proteomics with data-independent acquisition and traveling wave ion mobility spectrometry: a versatile tool in the life sciences. *Adv Exp Med Biol* **806**, 79-91 (2014). https://doi.org:10.1007/978-3-319-06068-2_4

52 Shimada, K. *et al.* Global survey of cell death mechanisms reveals metabolic regulation of ferroptosis. *Nat Chem Biol* **12**, 497-503 (2016).

53 Yang, Wan S. *et al.* Regulation of Ferroptotic Cancer Cell Death by GPX4. *Cell* **156**, 317-331 (2014).

54 Oswald, E. S., Brown, L. M., Bulinski, J. C. & Hung, C. T. Label-free protein profiling of adipose-derived human stem cells under hyperosmotic treatment. *J Proteome Res* **10**, 3050-3059 (2011).

55 Dai, H., Meyer, M., Stepaniants, S., Ziman, M. & Stoughton, R. Use of hybridization kinetics for differentiating specific from non-specific binding to oligonucleotide microarrays. *Nucleic Acids Res* **30**, e86 (2002). https://doi.org:10.1093/nar/gnf085

56 Weng, L. *et al.* Rosetta error model for gene expression analysis. *Bioinformatics* **22**, 1111-1121 (2006).

57 Paweletz, C. P. *et al.* Application of an End-to-End Biomarker Discovery Platform to Identify Target Engagement Markers in Cerebrospinal Fluid by High Resolution Differential Mass Spectrometry. *Journal of Proteome Research* **9**, 1392-1401 (2010).

58 Hendrickson, R. C. *et al.* High Resolution Discovery Proteomics Reveals Candidate Disease Progression Markers of Alzheimer's Disease in Human Cerebrospinal Fluid. *PloS one* **10** (2015). https://doi.org:e0135365

10.1371/journal.pone.0135365

59 McAvoy, T. *et al.* Quantification of tau in cerebrospinal fluid by immunoaffinity enrichment and tandem mass spectrometry. *Clin. Chem.* **60**, 683-689 (2014).

Figures

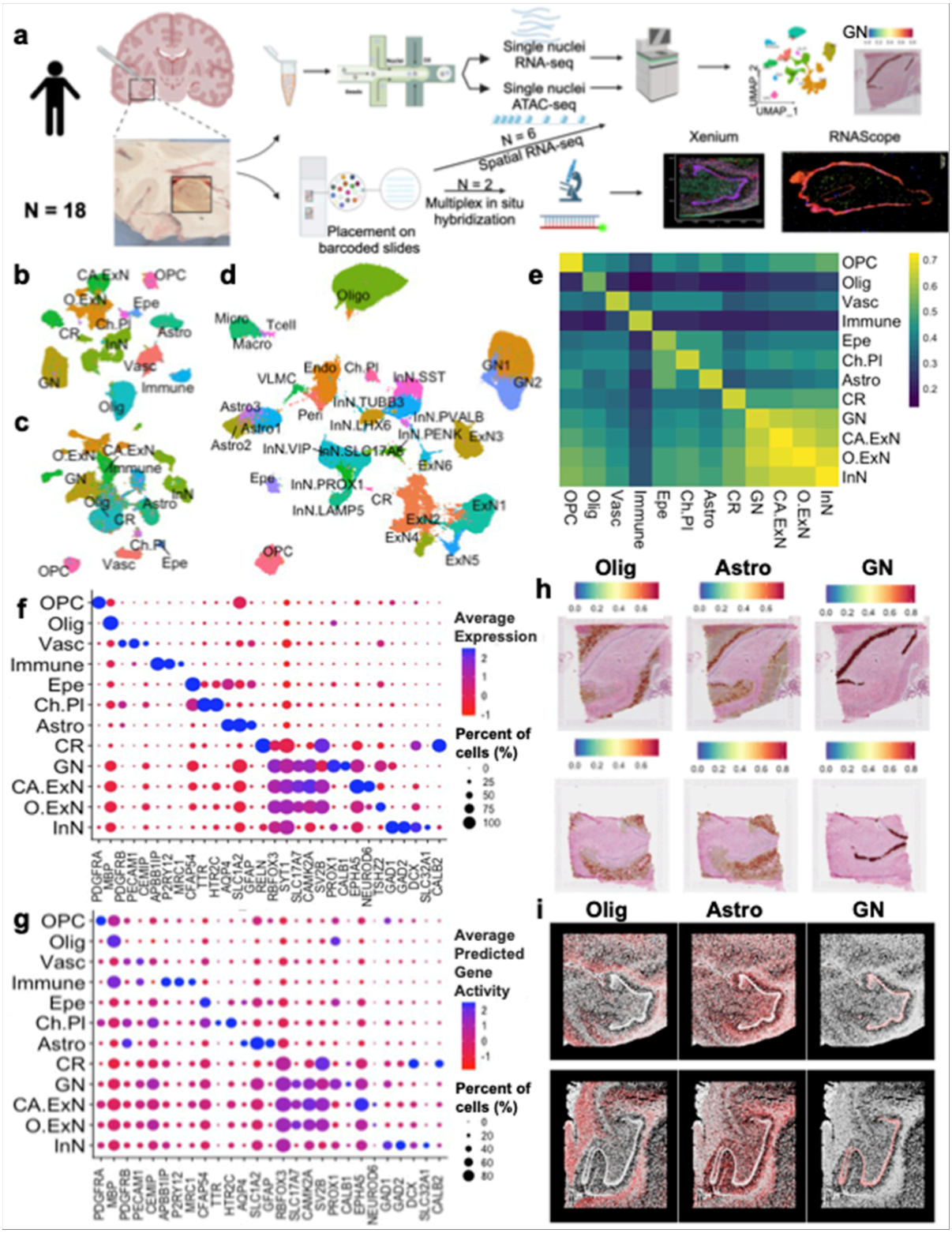

Figure 1

Defining hippocampal cell types characterized by gene expression and accessible chromatin sequencing. a. Schematic of study design. **b-d.** UMAP plots using batch-, sample-, and donor-corrected Harmony components. **b.** UMAP visualization of 349,847 cells derived from gene expression data and colored by cell types. **c.**UMAP visualization of 349,847 cells derived from accessible chromatin data and colored by cell types. **d.** UMAP visualization of 349,847 cells derived from integration of two modalities

using WNN analysis colored by annotations from unsupervised clustering. **e.** Pearson's correlation of natural log1p of the average gene expression and average predicted gene activity calculated from accessible chromatin in each cell type. **f-g.** Dotplot depicting the expression of marker genes (*PDFGRA* – oligodendrocyte progenitor cells (OPC); *MBP* – oligodendrocytes (Olig); *PDGFRB*, *PECAM1*, *CEMIP*– vasculature cells (Vasc); *APBB1IP*, *P2RY12*, *MRC1* – immune cells (Immune); *CFAP54* – ependymal cells (Epe); *TTR*, *HTR2C* – epithelial cells of choroid plexus (Ch.Pl); *AQP4*, *SCL1A2*, *GFAP* – astrocytes (Astro); *RELN* – Cajal Retzius cells (CR); *RBFOX3*, *SYT1*, *SV2B* – neurons; *SLC17A7*, *CAMK2A* – excitatory neurons; *PROX1*, *CALB1* – granule neurons (GN); *EPHA5*, *NEUROD6* – Cornu Ammonis excitatory neurons (CA.ExN); *TSHZ2* – other excitatory neurons (O.ExN); *GAD1*, *GAD2*, *DCX*, *SCL32A1* – inhibitory neurons (InN). (For detailed dotplot with 31 clusters and specific marker genes see Extended Data Fig. 4c). **f.** Dotplot depicting expression of marker RNA transcripts. **g.** Dotplot depicting predicted gene activity based on chromatin accessibility. **h-i.** Anchor-based integration of single-nuclei gene expression data with **(h)** whole-transcriptome spatial gene expression and **(i)** single-cell in-situ hybridization with 266 probes.

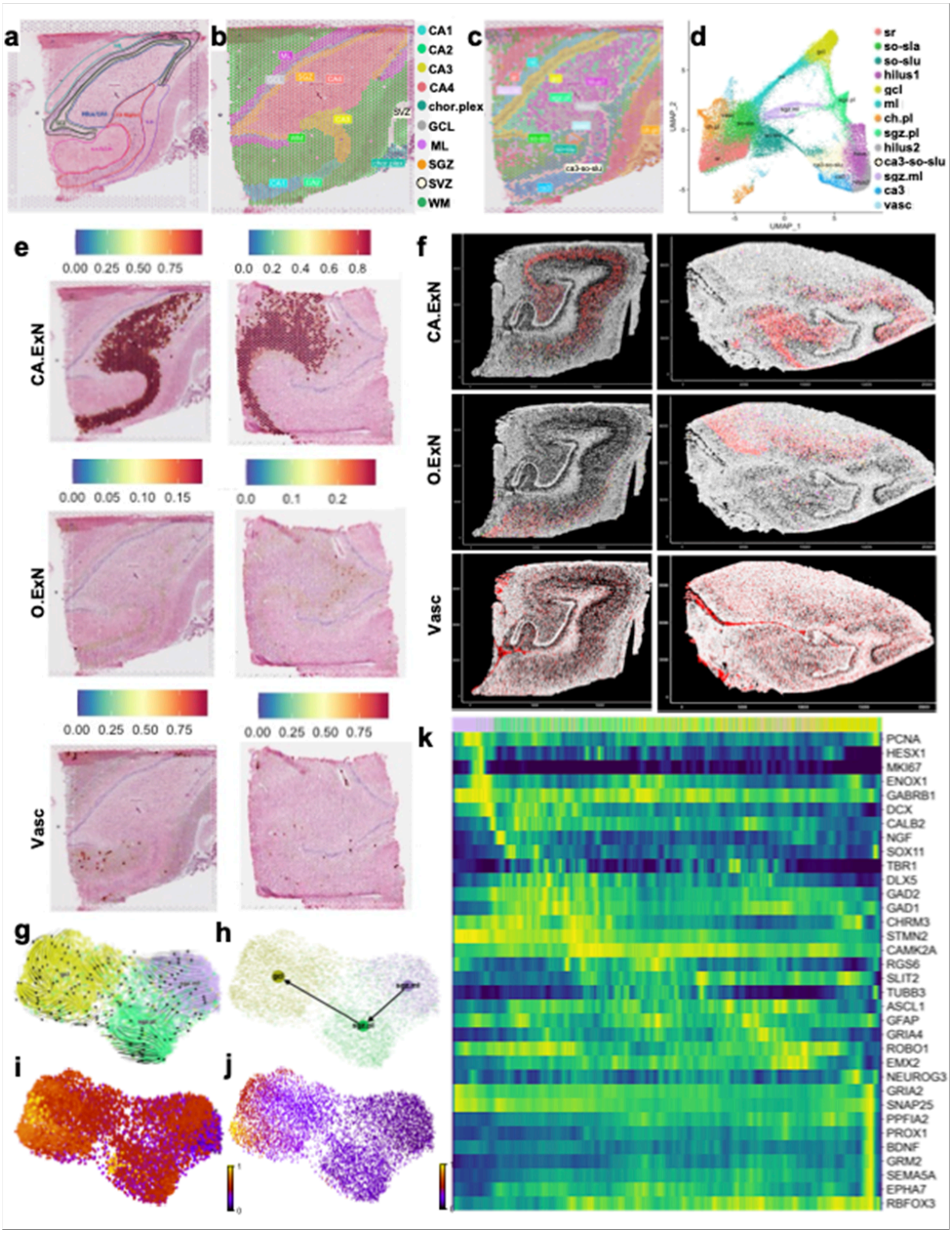

Figure 2

Visium spatial transcriptomics and Xenium in-situ hybridization, anchor-based integration of single nuclei data, and RNA Velocity. a. Example of Visium capture area with manual annotation of hippocampal subfields. **b.**Subfield annotations projected onto Visium section at a spot level. **c.** Visium unsupervised clustering annotation identifies 13 distinct clusters on the tissue section. **d.** UMAP visualization of 39,230 spots from 12 spatial gene expression samples. UMAP is derived from anchor-

based integration. **c-d.** Cluster color-coding is maintained. **e.** Anchor-based integration of whole transcriptome spatial gene expression and single-nuclei (sn) gene expression provides a probabilistic transfer of annotations from reference (sn) to query (spatial) to annotate 55um spatial spots. Color shows probability score. From top to bottom: *Cornu Ammonis* (CA) excitatory neurons (CA.ExN) have high probability score on CA regions, other excitatory neuron (O.ExN) have low probability score on CA regions, vasculature (Vasc) is diffuse across section. **f**. Anchor-based integration of snRNA-seq data (reference) with single-cell in-situ hybridization using 266 probes over 12x24 mm capture area (query) The larger capture area allows the spatial differentiation of CA.ExN in CA regions from O.ExN in subiculum. X and Y-axes represent microns. **g-j.** UMAP projection of subset of Visium unsupervised clusters gcl, sgz.pl, and sgz.ml **g.** with stream of velocity vectors from sgz.ml to sgz.pl to gcl estimated using RNA Velocity dynamical modeling. **h.** Partition-based graph abstraction (PAGA) with velocity-directed edges, a much simpler abstracted graph, in which edge weight represents confidence in the presence of connections, providing an easily interpretable overview of lineage trajectory. **i.** UMAP projection colored by latent time which represents the internal time of the spot and is based only on its transcriptional dynamics. **j.** UMAP colored by pseudotime which is calculated from the velocity graph. The root spot, or ‘zero’ is implicitly inferred (i.e. not used as an input parameter). **k.**Heatmap of gene expression (yellow = high, dark blue = low) by latent time, showing spots ordered by latent time on x-axis and genes on the y-axis.

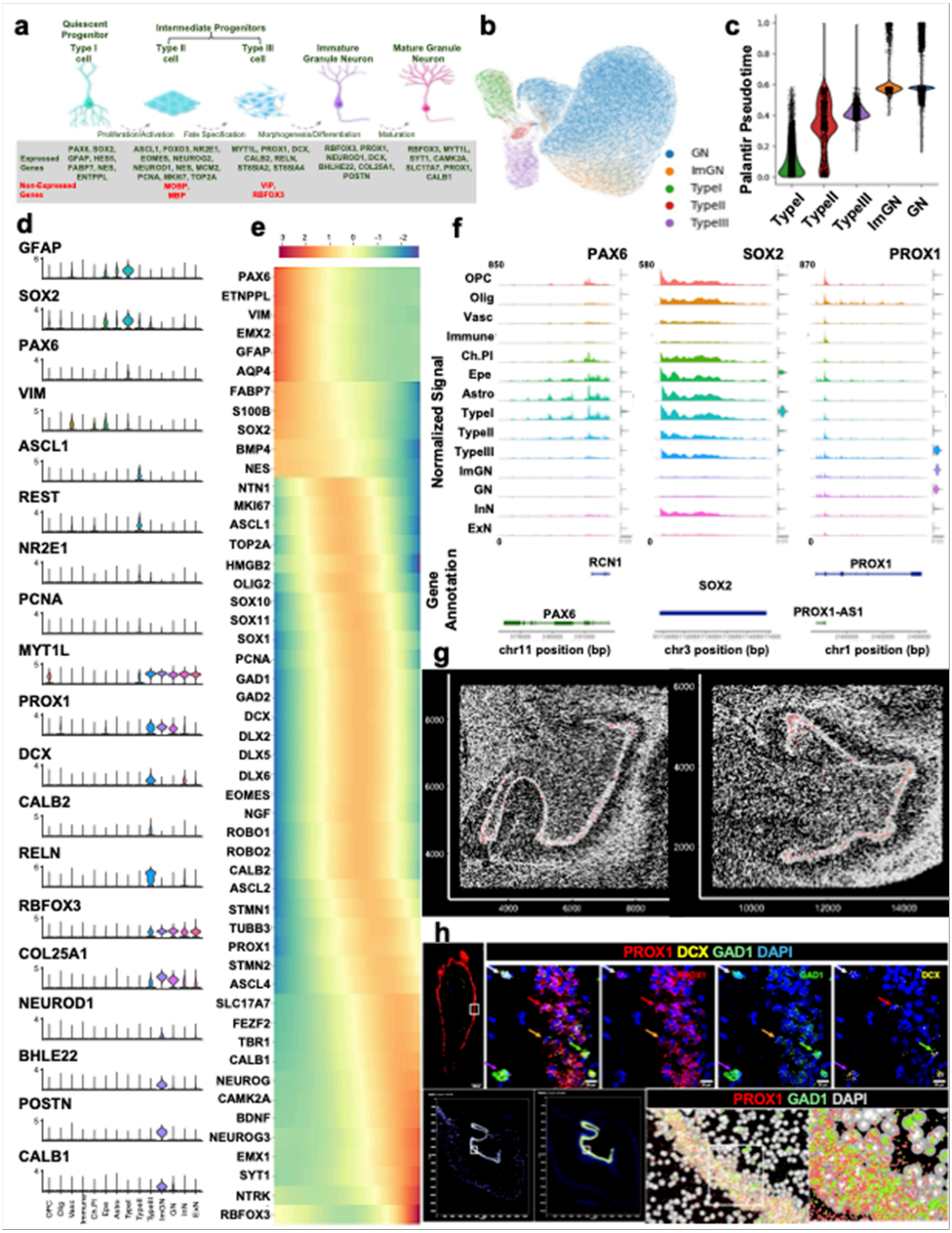

Figure 3

achine learning approach and Palantir trajectory inference to investigate neurogenic lineage | . a. Schematic representation of gene expression in prototypical cells used to train and test the LASSO regression. **b.** UMAP plot of TypeI, TypeII, TypeIII, ImGN, and GN cells using batch-, sample-, and donor-corrected Harmony principal components, colored by cell type. **c.** Palantir-inferred pseudotime showing TypeI as youngest, TypeII and TypeIII following, and ImGN and GN as oldest. **d.** Violin plots of gene

expression in cell types identified as TypeI, TypeII, TypeIII, and ImGN by LASSO regression. CA.ExN and O.ExN are merged to ExN. **e.** Gene expression trends for the neurogenic lineage using Palantir Generalized Additive Models. Color represents the z-scored scale gene trends of imputed expression. **f.** Coverage plots showing accessible chromatin peaks in each cell type with violin plot to showing RNA expression. From left to right: *PAX6*, *SOX2*, and *PROX1*. CA.ExN and O.ExN are merged. **g.** Anchor based integration of snRNA-seq data (reference) with Xenium single-cell in-situ hybridization (query). ImGN in red are in the subgranular zone (SGZ) and granule cell layer (GCL). X and Y-axes represent microns. **h.** Two probe-based protocols (RNAScope and Xenium) used to validate co-expression of *PROX1* (red), *GAD1* (green), and *DCX* (yellow) in cells located in the SGZ and GCL of the dentate gyrus.

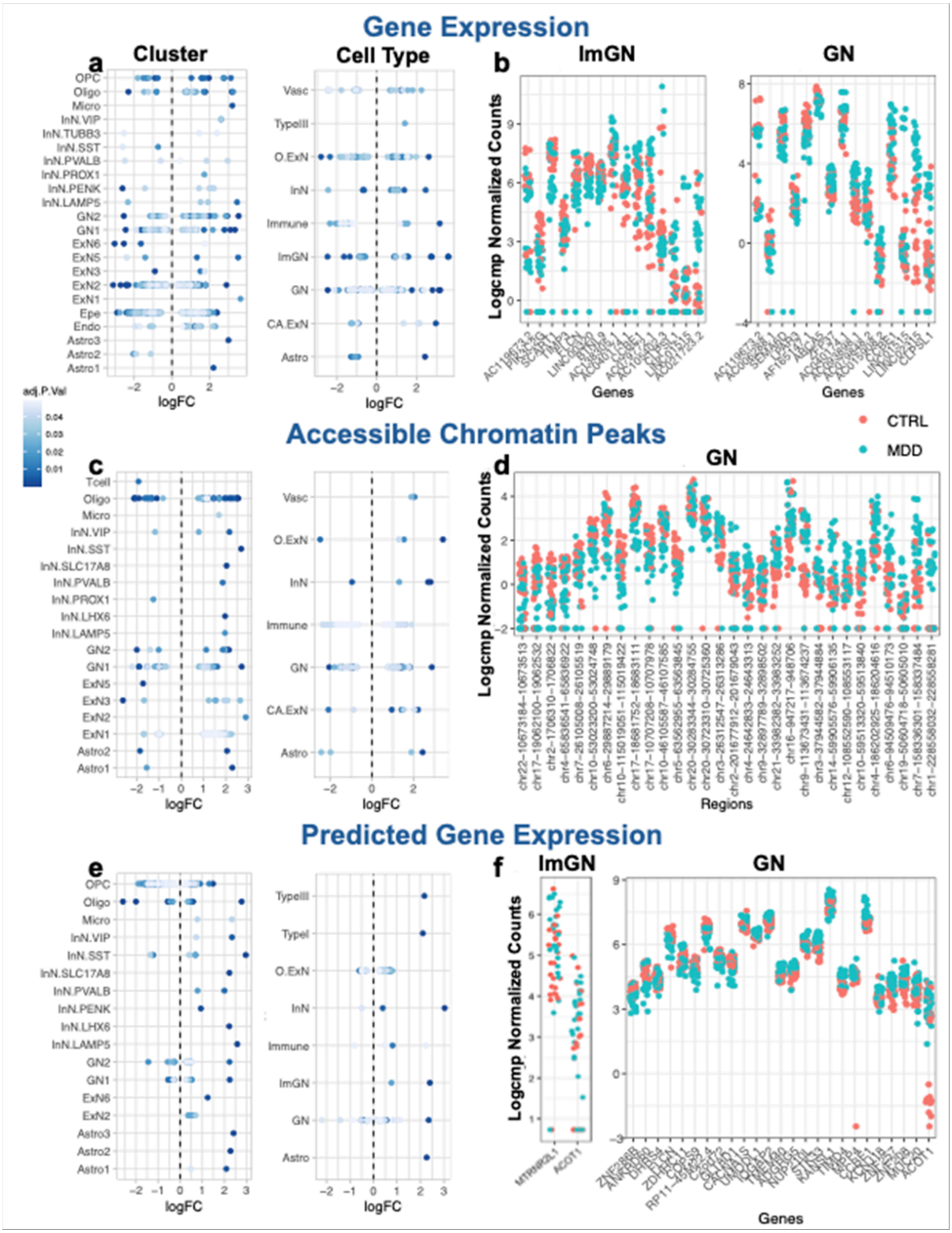

Figure 4

Cell type-specific differential gene expression, chromatin accessibility, and predicted gene activity in MDD versus CTRL. **a.** Distribution of differentially expressed genes in clusters (left) and cell types (right). Upregulated genes on the right of the dashed line and downregulated gene on the left. Clusters Olig, OPC, Epe, and Ch.PI were not aggregated, so they are not shown again in the cell type graph. Clusters or cell types with no differential gene expression in MDD are not shown. **b.** Dotplot showing

log(cpm)-adjusted counts for top 15 differentially expressed genes in MDD ImGN (left) and GN (right). **c.** Distribution of differentially accessible chromatin peaks in MDD in clusters (left) and cell types (right). Regions more accessible are on the right of the dashed line and less accessible on the left. **d.** Dotplot showing log(cpm)-adjusted counts for top 30 differentially expressed accessible chromatin peaks in MDD GN. **e.**Distribution of differentially predicted gene activity in MDD calculated from accessible peaks counts in clusters (left) and cell types (right) with increased predicted gene activity in MDD on the right of the dashed line and decreased on the right. **f.** Dotplot showing log(cpm)-adjusted counts for ImGN and GN differentially predicted gene activity in MDD calculated from accessible peaks counts.

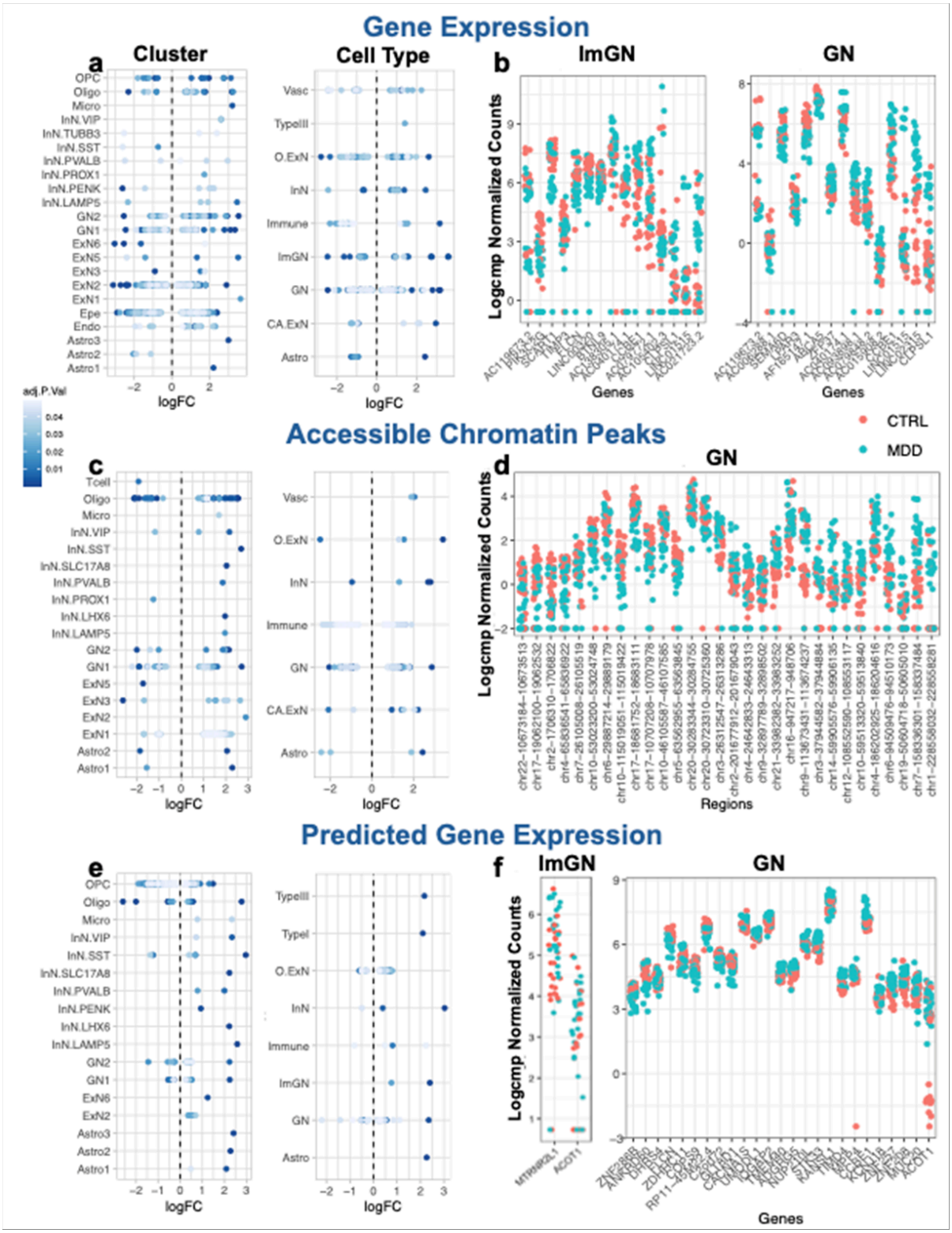

Figure 5

Pathway dysregulation in MDD. a. PsyGeNET database reported gene-disease association network plot for all DEGs from snRNA-seq data in our sample and genes previously linked to psychiatric disorders. **b.** Bar graph of gene-disease association showing number of genes associated with disease (blue) and number of genes uniquely associated with each disease (yellow). **c.** Gene-disease association heatmap for GN DEGs (note that THEMIS is also a DEG in ImGN). Evidence index of 1 (blue) indicates that all

literature support association. Evidence index between 0 and 1 (grey) indicate partial support for association.

d-g. Gene set enrichment analysis for ImGN and GN. Genes are ranked using log(fold change)-log10 (Benjamini-Hochberg adjusted p-value). Terms in blue are enriched for genes upregulated in MDD. Terms in red are enriched for genes downregulated in MDD. **d.** KEGG and REACTOME pathway dysregulation in ImGN**. e.** Hallmark pathway terms dysregulated in ImGN. **f.** KEGG and Reactome pathway dysregulation in GN. **g.** Hallmark pathway terms dysregulated in GN **h.** Weighted gene co-expression network analysis (WGCNA) heatmap showing module eigengenes (ME) correlation and associated *p*-value (in parentheses) for ImGN, with MDD-CTRL status and covariates (age, RIN, PMI). **i.** The turquoise module was positively associated with MDD in ImGN and was enriched for lipid modification and metabolism and the blue module was enriched for RNA splicing, RNA metabolism, translation and response to stress, and protein folding. **j.** The blue module negatively associated with MDD in ImGN was enriched metabolism, and lipid modifications. **k.** Heatmap showing the correlation and associated *p*-value (in parentheses) of GN WGCNA module eigengenes with MDD-CTRL status and covariates (age, pH, PMI). **l.** The turquoise module positively associated with MDD in GN was enriched for cell morphogenesis and projection development, and VEGF pathways. **m.** The blue module negatively associated with MDD in GN was enriched for lipid and metabolic processes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- BoldriniHippocampusMultiomeExtendedDataFigures.docx