CONTEXTUALIZING SUBSTANCE USE GENETIC RISK USING SINGLE-NUCLEUS RNA-SEQUENCING OF THE REWARD CIRCUITRY OF THE HUMAN BRAIN

by
Matthew Nguyen Tran

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
March 2022

Abstract

Novel single-cell technologies have afforded the generation of a vast amount of genomics data, including full-genome and -transcriptome sequencing, features of epigenetic and chromatin organization, and proteomics, all at the single-cell level. These new insights into human biology at such deep resolution have only begun to shed new light on tissue development and organization, as well as disease etiology, including in complex disorders or phenotypes of the human brain. Psychiatric genomics is one field exhibiting such progress in the unraveling of the complex genetics and molecular consequences underlying various neuropsychiatric disorders, whereas phenotypes characterized by other cognitive behaviors such as substance use and addiction, are yet to be understood at the cell-type-specific genetic level. In this work, I describe efforts to understand the complex transcriptomic architecture of cell populations within the reward circuitry in the human brain and demonstrate the utility of leveraging genomic association data to understand which of these cell populations identified genetic risk loci may be uniquely affecting.

Using single-nucleus RNA-sequencing (snRNA-seq), we profiled the single-nucleus-level transcriptomes of five regions embedded in the reward circuitry, including the nucleus accumbens, amygdala, hippocampus, subgenual anterior cingulate cortex, and dorsolateral prefrontal cortex. From this work, we characterized the transcriptomes of 107 robust cell classes at the region-specific level and performed cross-species analyses with existing data in rodent models, to assess the degree of convergence or divergence these established cell types

share with their corresponding rodent brain regions. Finally, we quantified genetic risk associations for five substance use phenotypes in each of the 107 cell profiles, identifying unique dopaminoceptive subpopulations in the nucleus accumbens that may be susceptible to genetic risk for increased levels of smoking. This work contributes substantially to the field of reward cognition and demonstrates the utility of integrating single-cell transcriptomics and population genomics.[1]

Primary Reader and Advisor: Keri Martinowich, PhD

Secondary Reader: Andrew E. Jaffe, PhD

[1] This work has been published at Tran, M.N., Maynard, K.R., et al. *Neuron* 2021 (https://doi.org/10.1016/j.neuron.2021.09.001). This dissertation features this text, reformatted and with supplemental words, as relevant. Abridged tables are as noted.

Acknowledgments

Kim Fitzgerald, my former high school World History AP teacher, and to whom I very much attribute my ending up in the sciences, as you ingrained into us the power of the question, “Why?”. As especially in my years of PhD training, I have developed the strength to not remaining silent about progressive or social issues, even if they didn’t directly impact me, I know this is something you would be proud to see. After all, is this *not* what PhD training is meant to teach us—to be the most critical and progressive thinkers in our respective fields?

Chefs Shawn Cirkiel and Justin Rupp, who trained me to work meticulously and methodically with my hands; to waste absolutely nothing; and started a pattern of giving someone who had zero experience a chance. I truly think my career in sciences started in the professional kitchen, so thank you.

Dr. Cassian Yee, who took me into his research lab at MD Anderson Cancer Center, regardless of having had no research experience and only seasoned hands from the professional kitchen—yet a thirst for understanding the potential of leveraging an individual’s immune system to help them battle their cancers.

Drs. Kristen Maynard and Leonardo Collado-Torres, who have provided not only a safe and patient space in navigating such brand new fields of neuroscience and biostatistics, but also comradery and encouragement anytime I ever expressed doubt.

Dr. Andrew Jaffe, my first PhD advisor, who truly believed I could learn `R`, probably without fully recognizing how little I actually knew. And providing an unending supply of code to kick off learning from.

And to Dr. Keri Martinowich, my formal PhD thesis advisor, who has been more than patient and trusting in my capacity to: perform single-cell transcriptomic data generation, including bringing in and applying skills from my previous research experience; learn, understand, and implement the complex methods to perform such big data analyses; and to teach about these and their nuances or caveats at every step of the way, to both colleagues and new trainees.

My dad, Nguyen Thanh Son, who makes half of who I am proud to be, and represents the epitome of unconditional love and devotion to family and your children/grandchildren. Your decades-long perseverance in creating a new and abundant life, after forced emigration as a Vietnam War refuge, never ceases to inspire me. I love you, Bố.

My mom, Tran Thu-Van, to whom I dedicate this work and this PhD, and for whom I adopted the professional moniker ‘Matthew Nguyen Tran’ after your services of 25+ years in pharmacy – I owe you everything. Your own story of not only survival, but

commitment and achievement, success, to finally and truly thriving in your last years – I am full of awe and gratitude. I miss you every day and will see you when the time comes for us to be reunited. Love you, Mẹ.

My siblings, Michelle, Quanny, Amanda – y'all are the best siblings this kid could ask for. Every day I legitimately miss being in Texas, closer to you guys, and hope that in the next chapter, there is more of seeing you and the kids, wherever we end up. Y'all stupid lol, but I love you for it, and I just love us. Jon, Genney, Chris, the kids – I freaking love you guys. The whole fam – thanks for all the support and love and truly just believing in me. I really mean it that I would not have gotten this far without all of you.

Dylan, my soon-to-be-fiancé; my partner in life; my rock; my stubbornness antidote; my number 1 fan. Thank you for embarking on this journey with, and just choosing, me. I have grown and learned so much more from being with you than I give you credit for. I am also so proud of you and have enjoyed being able to watch us grow professionally and personally, together. I cannot *wait* for what the next chapter brings us and Harley. I love you, Dylan.

Contents

List of Tables

List of Figures

Chapter 1. Introduction

Recent advances in single-cell and single-nucleus RNA-sequencing (scRNA-seq/snRNA-seq) technologies have facilitated the molecular characterization of diverse cell types in the postmortem human brain during development (Darmanis et al., 2015; Li et al., 2018a; Zhong et al., 2018, 2020), and have been used to assess cell type-specific gene expression differences in the context of several brain disorders, including Alzheimer's disease, autism spectrum disorder, multiple sclerosis, and major depressive disorder (Mathys et al., 2019; Nagy et al., 2020; Schirmer et al., 2019; Velmeshev et al., 2019). Identification of cell type-specific gene expression signatures has contributed to understanding the relationship between molecular identity and cell function as it relates to brain health, neurological disease, and genetic risk for neuropsychiatric disorders, such as schizophrenia (Skene et al., 2018).

While substantial advancements have been made in understanding cell type heterogeneity both within and across individual regions of the human brain, the majority of snRNA-seq reports are limited to a small number of brain areas. These primarily include the hippocampus (HPC) (Franjic et al., 2020; Habib et al., 2017) and several heavily studied sub-regions of the cortex (Lake et al., 2016), including the dorsolateral prefrontal cortex (DLPFC) (Li et al., 2018a; Nagy et al., 2020), medial temporal cortex (Darmanis et al., 2015; Hodge et al., 2019), entorhinal cortex (Grubman et al., 2019), and anterior cingulate cortex (Velmeshev et al., 2019). Molecular profiling of less studied cortical subregions including the subgenual anterior cingulate cortex (sACC), as well as striatal and limbic brain regions, including the nucleus accumbens (NAc) and the amygdala (AMY), is lacking in the human brain. The sACC, NAc, and AMY are interconnected within well-established circuit loops that mediate important behavioral

and neurobiological functions, including signaling for reward and motivation as well as processing emotional valence, particularly for fearful and threatening stimuli (Haber and Knutson, 2010; Janak and Tye, 2015; Russo and Nestler, 2013).

Importantly, the cellular composition of individual neuronal subtypes in these regions substantially differs from previously well-profiled cortical and hippocampal regions (Saunders et al., 2018; Zeisel et al., 2018). For example, the NAc contains dopaminoceptive populations of GABAergic medium spiny neurons (MSNs) - the principal projecting cell type comprising up to 95% of neurons in rodent - that harbor unique physiological and cellular properties (Gerfen et al., 1990; Kawaguchi, 1997; Kronman et al., 2019; Russo and Nestler, 2013). Early functional characterization of MSNs revealed two distinct classes of MSNs based on expression of D1 versus the D2 dopamine receptors (D1-MSNs and D2-MSNs, respectively) (Lobo, 2009; Lobo et al., 2006). However, recent sc/sn-RNAseq studies in the rodent striatum, and in the NAc specifically, revealed more complex transcriptional diversity within broader D1 and D2-MSN subclasses than was previously appreciated (Gokce et al., 2016; Saunders et al., 2018; Stanley et al., 2020; Zeisel et al., 2018). Moreover, subpopulations of MSNs are differentially recruited in response to cocaine exposure, and mediate divergent functional effects on behavioral responses to drugs of abuse (Savell et al., 2020). Similarly, single-cell profiling studies in the rodent AMY identified specialized populations of *Cck*-expressing neurons that are preferentially activated by behavioral experience, including exposure to acute stress (Wu et al., 2017). However, whether and to what extent this transcriptional diversity is conserved in these areas of the human NAc and amygdala has not yet been fully explored. Given evidence for the functional importance of specific cell types in these areas of the rodent brain, profiling these regions in human by snRNA-seq may identify analogous cell populations, which can then be analyzed in the context

of neurobiological dysfunction in human brain disorders.

Here we defined the molecular taxonomy of distinct cell types in subcortical regions (NAc and AMY), which act as key nodes within circuits that mediate critical brain and behavioral functions including reward signaling and emotional processing. We also validated molecular profiles for previously identified cell types in the HPC and DLPFC, and identified similar cell types in the sACC, an additional cortical region central to limbic system function that has been implicated in affective disorders. Furthermore, we evaluate cross-species conservation of NAc and AMY cell types between human and rodent, specifically focusing on comparisons of MSN sub-populations identified as playing key roles in reward-processing and addiction. We survey the transcriptomic architecture across 107 robust cell classes, identifying molecular relationships between cell populations and patterns of divergence within specialized MSNs. Finally, by integrating genetic studies for substance use and neuropsychiatric disorders, we show differential cell type association, or differential expression of risk loci-associated genes, with a number of neuropsychiatric or substance use phenotypes, highlighting the clinical relevance of understanding cell type- and region-specific expression in the human brain.

Chapter 2. Establishing a cell type transcriptomic atlas of the reward circuitry of human brain

We profiled 5 brain regions (DLPFC, HPC, sACC, NAc, and AMY) across up to 8 neurotypical, adult subjects using 10x Genomics Chromium technology. To minimize potential batch effects, regions/donors were split across Chromium runs, for a total of 24 samples (sample/demographic information found in **Table 1**). Nuclear preparations were generated and purified by flow cytometry using chromatin (DAPI or PI) staining (and NeuN enrichment for a subset of samples) to obtain nuclei from all cell types in a brain region. After sequencing, data processing and QC (Methods, [1]), we report a total of 70,615 high-quality nuclei, which were first analyzed in respective region-specific analyses, followed by an across-regions integration.

2.1 Nucleus accumbens

To evaluate the transcriptional landscape of MSNs and other cell populations in the human NAc, we analyzed 19,892 total nuclei from 8 donors. We performed data-driven clustering to generate 24 cell clusters across six broad cell types, including GABAergic inhibitory neurons, MSNs, oligodendrocytes, oligodendrocyte precursor cells, microglia, and astrocytes (**Figure 2.1.1A**). Of the 10 distinct neuronal clusters expressing established D1- and D2-MSNs markers (**Figure 2.1.1B**), including *PPP1R1B* (encoding DARPP-32), six of these MSN subclusters were enriched for *DRD1* (D1_A, D1_B, … through D1_F) and two were enriched for *DRD2* (D2_A, … D2_D). These MSN

[1] Please see **Table S2** from *Tran, M.N., Maynard, K.R., et al.* Neuron *2021* for Cell Ranger v3.0 QC metrics on these 24 samples.

subclusters collectively made up between 85-95% of neuronal nuclei from the neuron-enriched samples (**Table 2.1**), lending human evidence that, similar to the rodent, the vast majority of nuclei in this region of the striatum are composed of MSNs (Kawaguchi, 1997). Clusters D1_A and D2_A represented the largest D1-MSN (67%) and D2-MSN (87%) subclasses, respectively. As expected, MSN subclusters showed differential enrichment of several neuropeptides, including proenkephalin (*PENK)*, tachykinin 1 (*TAC1*), and prodynorphin (*PDYN)* (**Figure 2.1.2**) (Lobo, 2009; Lobo et al., 2006; Savell et al., 2020). Surprisingly, the classical D1-MSN marker, *TAC1,* was enriched in D2_C MSNs, while largely absent in the smaller D1_B , D1_C, and D1_F MSNs (**Figure 2.1.1B**). Similarly, the classical D2-MSN marker *PENK* was enriched in the large population of D2_A MSNs, in addition to D2_B and D2_D, but depleted in the smaller population of D2_C MSNs (Figure 2.1.2). Differential expression of these neuropeptides in D1 and D2 MSN subclasses was confirmed using single molecule fluorescent in situ hybridization (smFISH) with 4-plex RNAscope technology (Maynard et al., 2020); **Figure 2.1.2**).

Using differential expression analyses, we identified the most preferentially expressed genes in each MSN class and found tens to hundreds of unique markers for D1 and D2-MSN classes (at false discovery rate, or FDR, < 1e-6[1]). Among D1-MSNs, three classes (D1_B, D1_E, and D1_F) were enriched for the relaxin family peptide receptor 1, *RXFP1*, but only the *TAC1*-negative D1_F MSNs were enriched for the $GABA_A$ receptor subunit, *GABRQ,* and the relaxin family peptide receptor 1, *RXFP1* (**Figure 2.1.3**; **Figure 2.1.1C**). Similarly, only D1_E MSNs expressed substantial levels of *CRHR2*, encoding corticotropin releasing hormone receptor 2, a protein implicated in

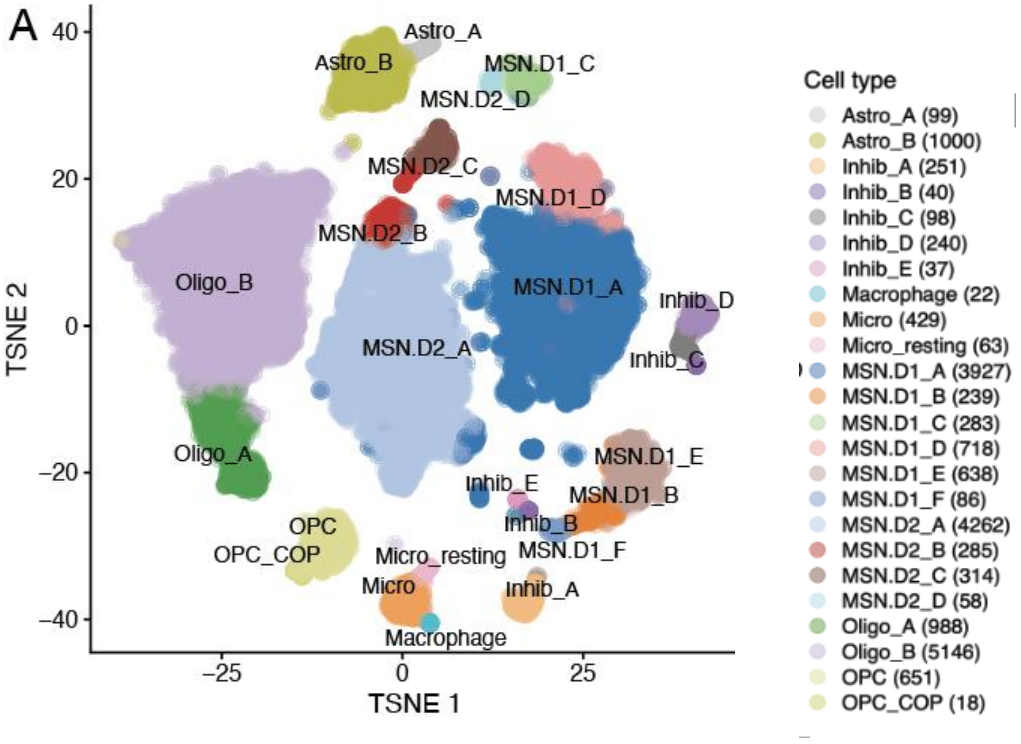
[1] Top 40 marker lists per cell class were published with this work (*Tran, M.N., Maynard, K.R., et al.* Neuron *2021*), and can be accessed at the associated public GitHub repository (https://github.com/LieberInstitute/10xPilot_snRNAseq-human)

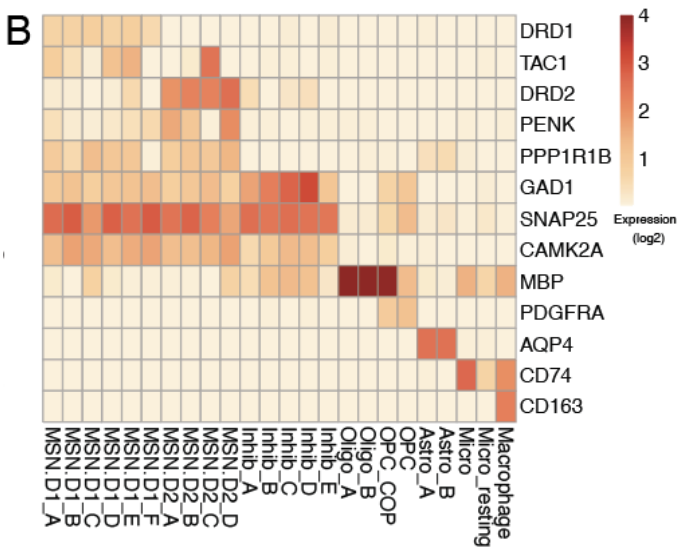
mediating the response to stress in the brain (**Figure 2.1.1D**). The *TAC1*-negative D1_C MSNs could be distinguished from all other MSN classes by elevated expression of relaxin family peptide receptor 2, *RXFP2,* and depletion of both *RXFP1* and *GABRQ* (**Figure 2.1.3**), though this small class of D1 MSNs was difficult to identify via smFISH. Consistent with the identification of a discrete D2-MSN subpopulation expressing *Htr7* in the mouse striatum (Gokce et al., 2016; Stanley et al., 2020), we identified enrichment of *HTR7* in D2_C (*TAC1*-positive; *PENK*-negative) MSNs, but not other classes of D2 MSNs (**Figure 2.1.4**). Similar to D1_E MSNs, the *HTR7*-positive D2_C cluster was the only D2-MSN class expressing *CRHR2*, though to a lesser degree. The existence of these novel D1 and D2 MSN classes was validated by smFISH on NAc brain sections derived from independent postmortem human brain donors (**Figure 2.1.1D-E; Figures 2.1.2-2.1.6**). Several other genes including *CASZ1*, *GPR6*, and *EBF1* were differentially expressed in unique D1 and/or D2-MSN subpopulations (**Figure 2.1.5**). *CASZ1* was highly enriched in the D1_B, D1_E, and D2_C subpopulations, *GPR6* in all D2 classes, and *EBF1* in the D1_C subpopulation.

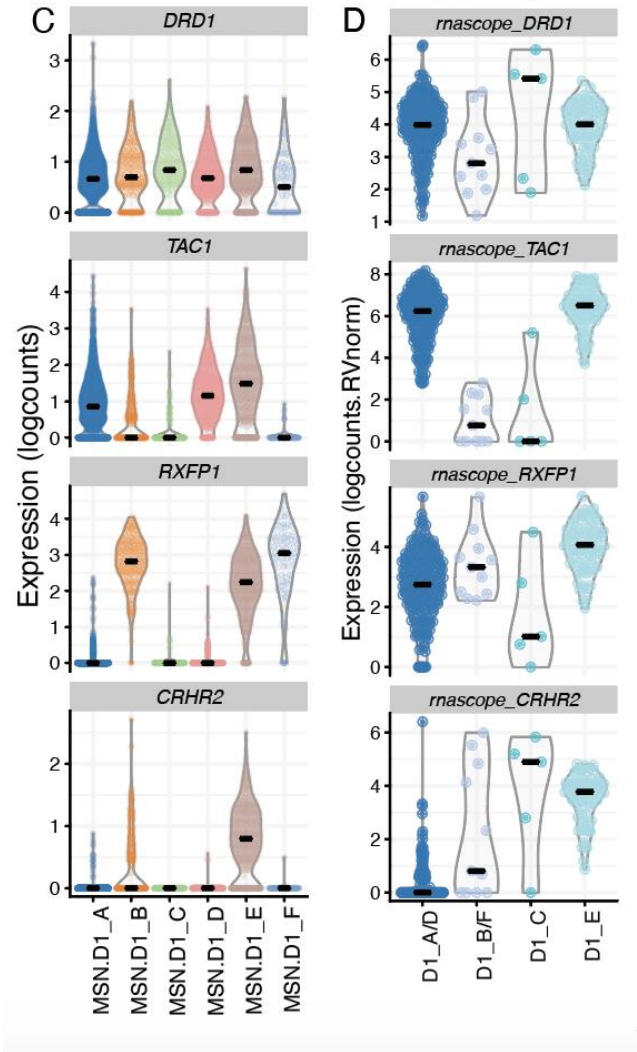
In addition to describing transcriptional diversity in D1 and D2 MSNs, we also identified 5 clusters of GABAergic inhibitory neurons expressing the marker genes *GAD1* and *GAD2*, but depleted for MSN marker genes (**Figure 2.1.1B; Figure 2.1.6**). These clusters contained different transcriptionally-defined classes, including GABAergic neurons expressing somatostatin (*SST*; Inhib_E), neuropeptide Y (*NPY*; Inhib_E), prepronociceptin (*PNOC*; Inhib_E), vasoactive intestinal peptide (*VIP*; Inhib_B), and tachykinin 3 (*TAC3*; Inhib_A; **Figure 2.1.6**; see **Chapter 7** (Methods)). While we did not observe robust expression of parvalbumin (*PVALB)* in any cluster, classes Inhib_C and Inhib_D showed high expression of *KIT*, encoding the protein c-Kit, which is frequently co-expressed in mouse *Pvalb/*PV-positive GABAergic neurons (Enterría-Morales et al.,

2020). smFISH for *PVALB* and other top marker genes for Inhib_C/_D (*PTHLH*, *KIT*, *GAD1*) confirmed that these GABAergic clusters likely represent unique PV-expressing interneuron classes (**Figure 2.1.6**).

Figure 2.1.1: Distinct subpopulations of D1- and D2-expressing MSNs in human NAc.




B
DRD1
TAC1
DRD2
PENK
PPP1R1B
GAD1
SNAP25
CAMK2A
MBP
PDGFRA
AQP4
CD74
CD163
4
3
2
1
Expression (log2)
MSN.D1_A
MSN.D1_B
MSN.D1_C
MSN.D1_D
MSN.D1_E
MSN.D1_F
MSN.D2_A
MSN.D2_B
MSN.D2_C
MSN.D2_D
Inhib_A
Inhib_B
Inhib_C
Inhib_D
Inhib_E
Oligo_A
Oligo_B
OPC_COP
OPC
Astro_A
Astro_B
Micro
Micro_resting
Macrophage


C
DRD1
TAC1
RXFP1
CRHR2
Expression (logcounts)
MSN.D1_A
MSN.D1_B
MSN.D1_C
MSN.D1_D
MSN.D1_E
MSN.D1_F
D
rnascope_DRD1
rnascope_TAC1
rnascope_RXFP1
rnascope_CRHR2
Expression (logcounts.RVnorm)
D1_A/D
D1_B/F
D1_C
D1_E

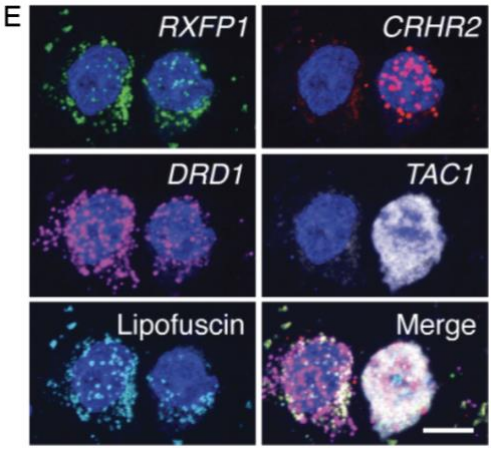

E
RXFP1
CRHR2
DRD1
TAC1
Lipofuscin
Merge

Figure 2.1.1 legend:

(A) tSNE plot of 19,789 nuclei (n=8 donors) across 21 clusters, including 6 clusters of D1 MSNs and 4 clusters of D2 MSNs.

(**B**) Heatmap depicting log_2 expression of canonical marker genes used to annotate each cluster.

(C) Violin plots for 4 genes differentially expressed (log2-normalized counts) in specific D1 classes (or class groups: *CRHR2, DRD1, RXFP1, and TAC1*) that were selected for validation using single molecule fluorescent in situ hybridization (smFISH).

(D) Log_2 expression of respective transcript counts per smFISH ROI (ROI size-normalized), post lipofuscin-masking (autofluorescence). Each *DRD1*+ ROI was classified into a Euclidean distance-predicted MSN class (or group of classes) and its(/their) respective expression.

(E) Multiplex smFISH in human NAc depicting a D1_C (left) and D1_E (right) MSN, side by side. Maximum intensity confocal projections showing expression of DAPI (nuclei), *CRHR2*, *DRD1*, *TAC1* and lipofuscin autofluorescence. Merged image without lipofuscin autofluorescence. Scale bar=10 μm.

Figure 2.1.2. Differential expression of neuropeptide genes *TAC1* and *PENK* in D1 and D2 MSN subpopulations.

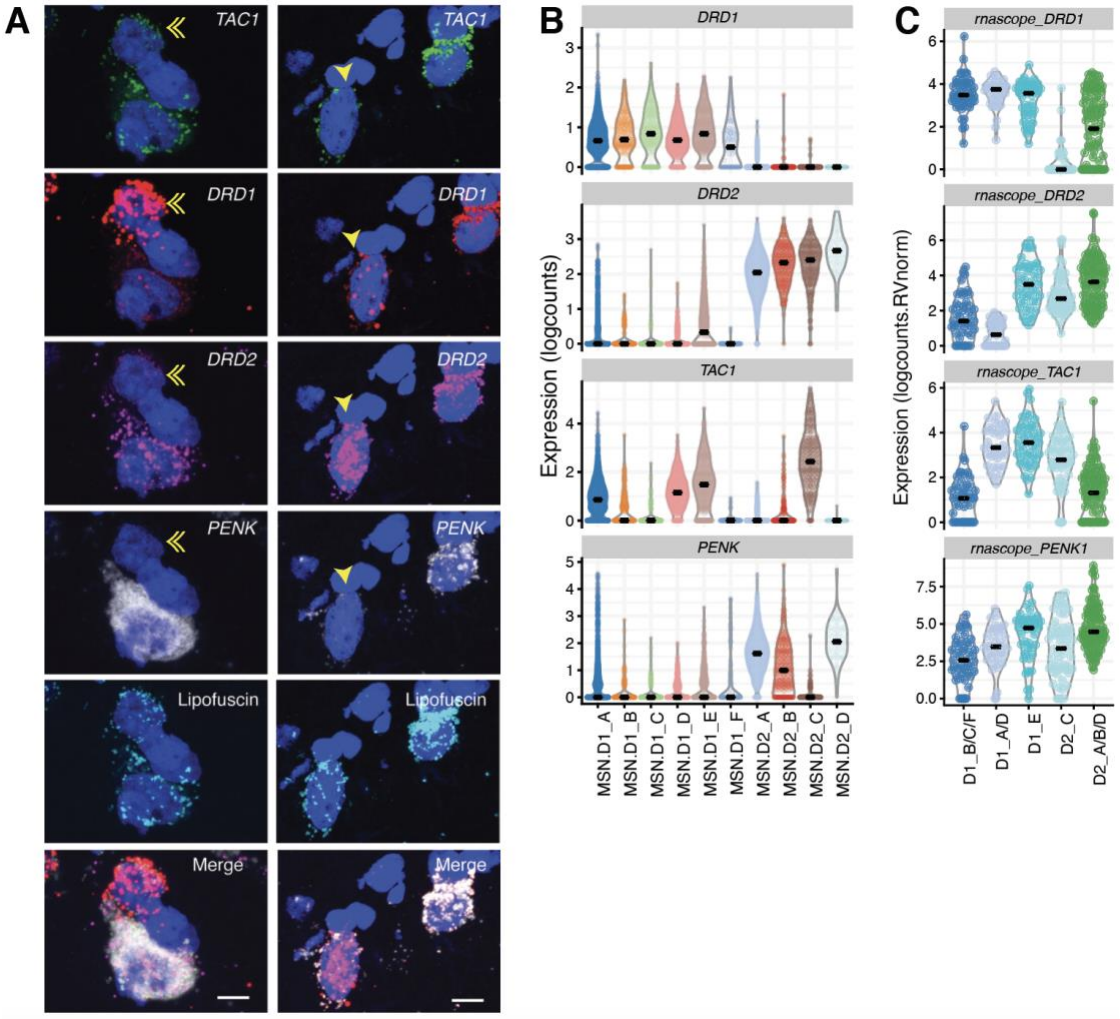


Figure 2.1.2 legend:

(**A**) Multiplex single molecule fluorescent in situ hybridization (smFISH) in human NAc, showing maximum intensity confocal projections showing expression of DAPI (nuclei), *DRD1*, *DRD2*, *TAC1, and PENK* and lipofuscin autofluorescence in two separate fields. Merged image without lipofuscin autofluorescence. Scale bar=10 µm. Double arrow indicates TAC1 negative D1 MSN. Single arrow indicates dual D1 and D2-expressing MSN.

(B) Corresponding violin plots showing differential (log_2) expression of *TAC1* and *PENK* in D1 and D2 MSN cell classes.

(C) Log_2 expression of respective transcript counts per smFISH ROI (ROI size-normalized), post lipofuscin-masking (autofluorescence). Each *DRD1*+ or *DRD2*+ ROI

was classified into a Euclidean distance-predicted MSN class (or group of classes) and its (/their) respective expression.

Figure 2.1.3. Further validation of D1 MSN subpopulations using smFISH.

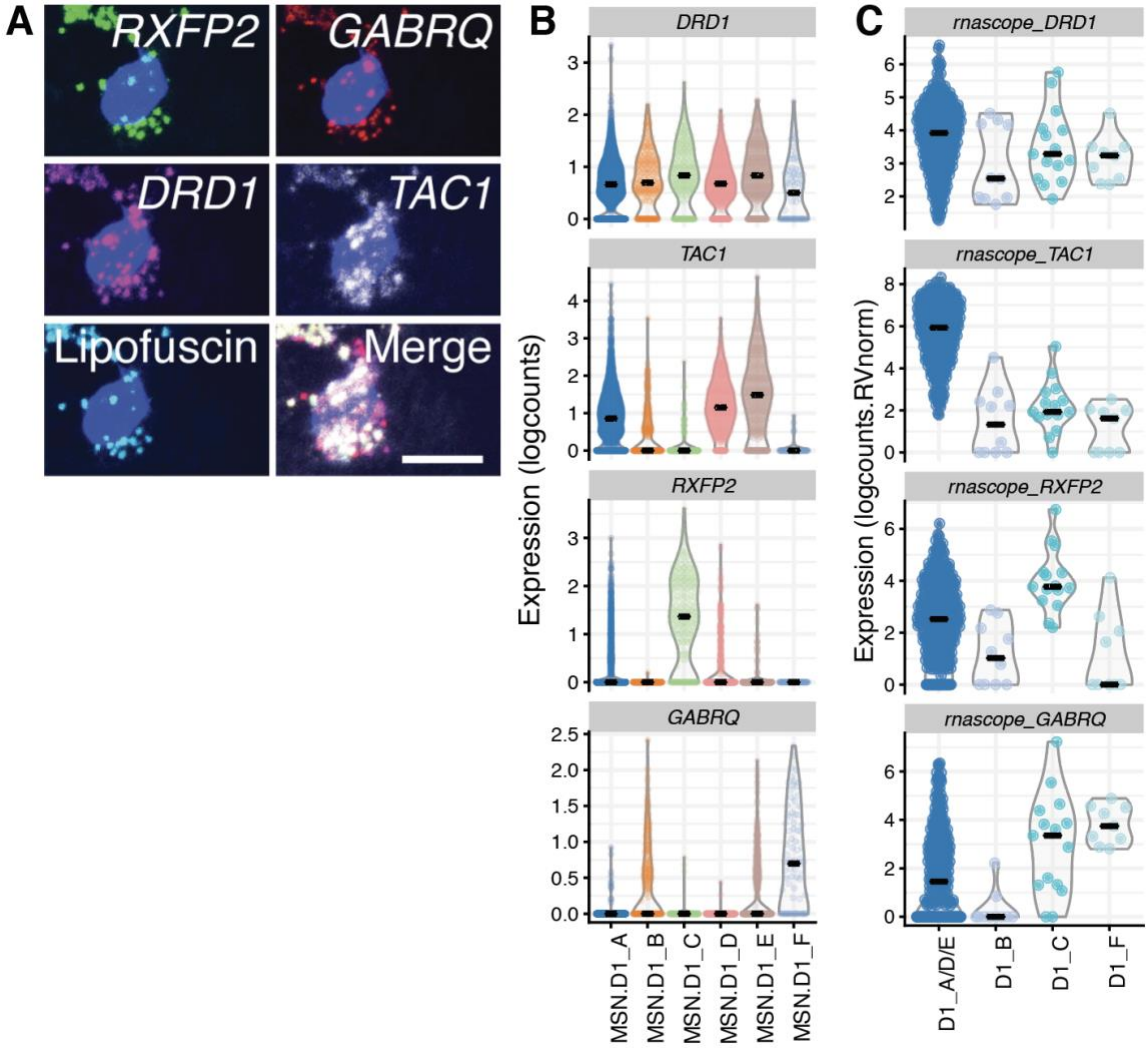


Figure 2.1.3 legend:

(A) Multiplex single molecule fluorescent in situ hybridization (smFISH) in human NAc depicting an D1_A, _E, or _E MSN. Maximum intensity confocal projections showing expression of DAPI (nuclei), *RXFP2*, *GABRQ*, *DRD1*, *TAC1* and lipofuscin autofluorescence. Merged image without lipofuscin autofluorescence. Scale bar=10 μm.

(B) Corresponding violin plots showing differential (log_2) expression of these three genes in specific D1 subpopulations by snRNAseq.

(C) Log_2 expression of respective transcript counts per smFISH ROI (ROI size-normalized), post lipofuscin-masking (autofluorescence). Each *DRD1*+ ROI was assigned to a Euclidean distance-predicted D1 MSN class (or group of classes) and its (/their) respective expression, showing possible identification of the less abundant D1_C class.

Figure 2.1.4. Confirmation of *HTR7*-expressing D2 MSNs in human NAc by smFISH.

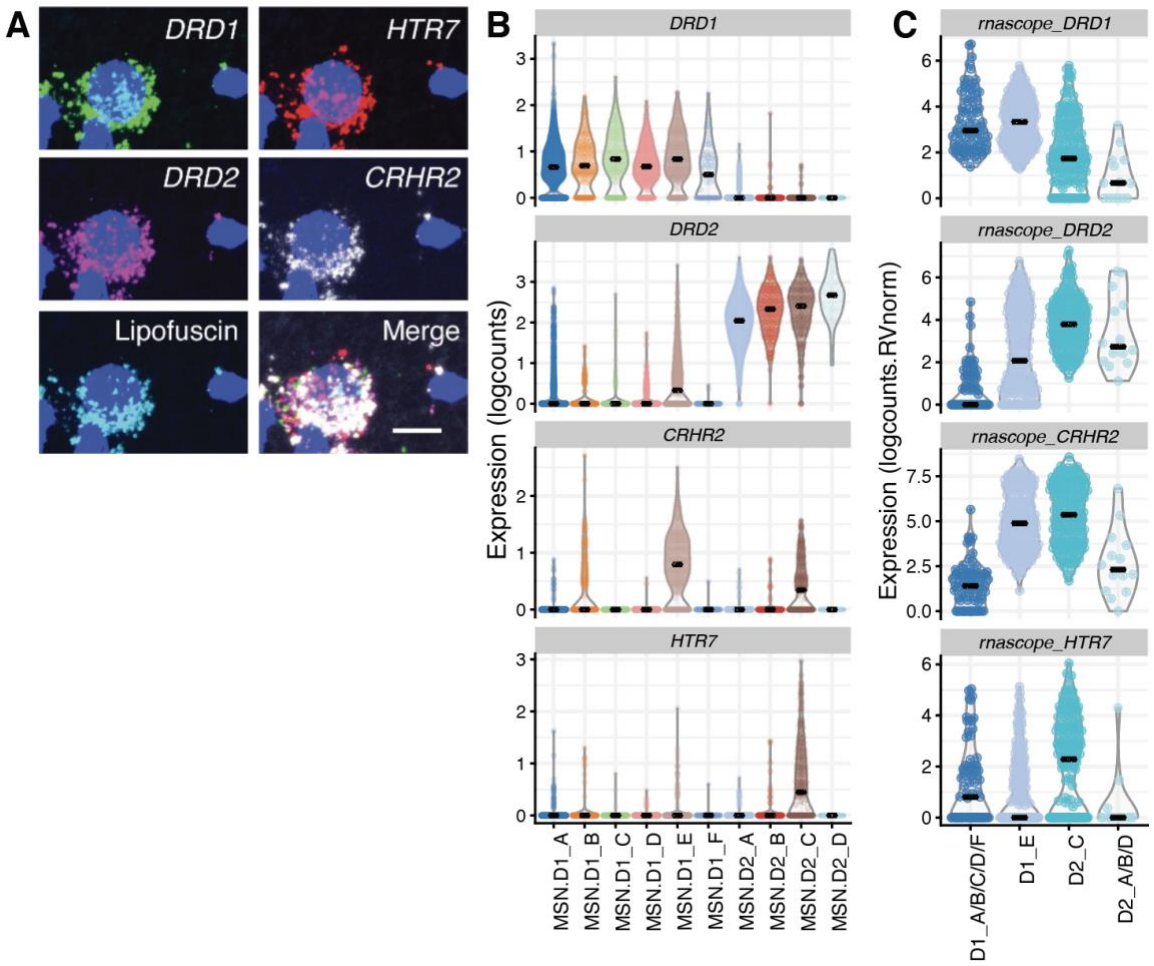


Figure 2.1.4 legend:

(A) Multiplex single molecule fluorescent in situ hybridization (smFISH) in human NAc depicting expression of *HTR7* in a D2_C MSN. Maximum intensity confocal projections

showing expression of DAPI (nuclei), *DRD1*, *HTR7*, *DRD2*, *CRHR2* and lipofuscin autofluorescence. Merged image without lipofuscin autofluorescence. Scale bar=10 µm.

(B) Corresponding violin plots showing differential expression of *HTR7* and *CRHR2* in D1 and D2 MSNs subpopulations by snRNA-seq.

(C) Log_2 expression of respective transcript counts per smFISH ROI (ROI size-normalized), post lipofuscin-masking (autofluorescence). Each *DRD1*+ or *DRD2*+ ROI was assigned to a Euclidean distance-predicted MSN class (or group of classes) and its (/their) respective expression.

Figure 2.1.5. Other differentially expressed MSN markers and similarity between largest D1/D2 subpopulations.

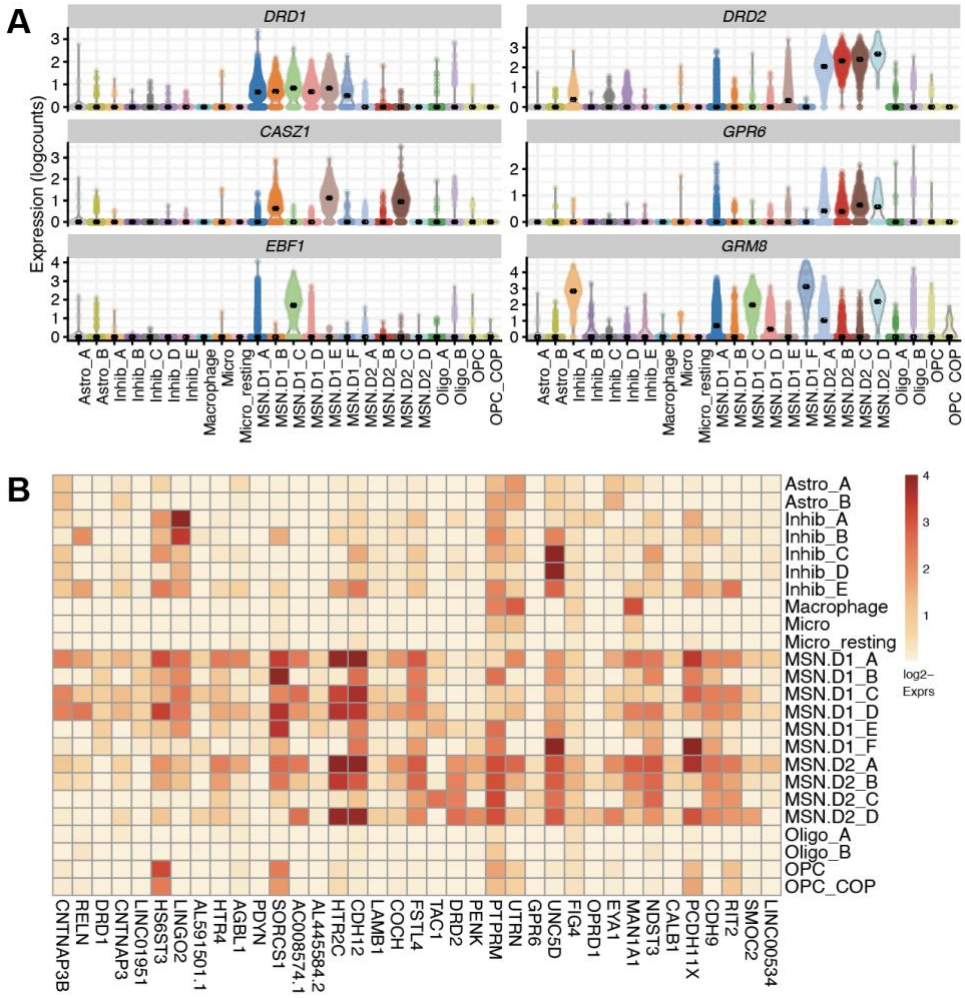


Figure 2.1.5 legend:

(A) Log_2-normalized counts of other markers for MSN subpopulations not prioritized for smFISH validation, as above. *GRM8* is included to show specific enrichment in a variety of D1 or D2 classes.

(B) Heatmap of mean snRNA-seq expression, showing broad coexpression of the combined top 20 markers for classes D1_A and D2_A (scale thresholded to mean log_2-normalized counts = 4.0).

Figure 2.1.6. Characterization of interneuron subpopulations in human NAc.

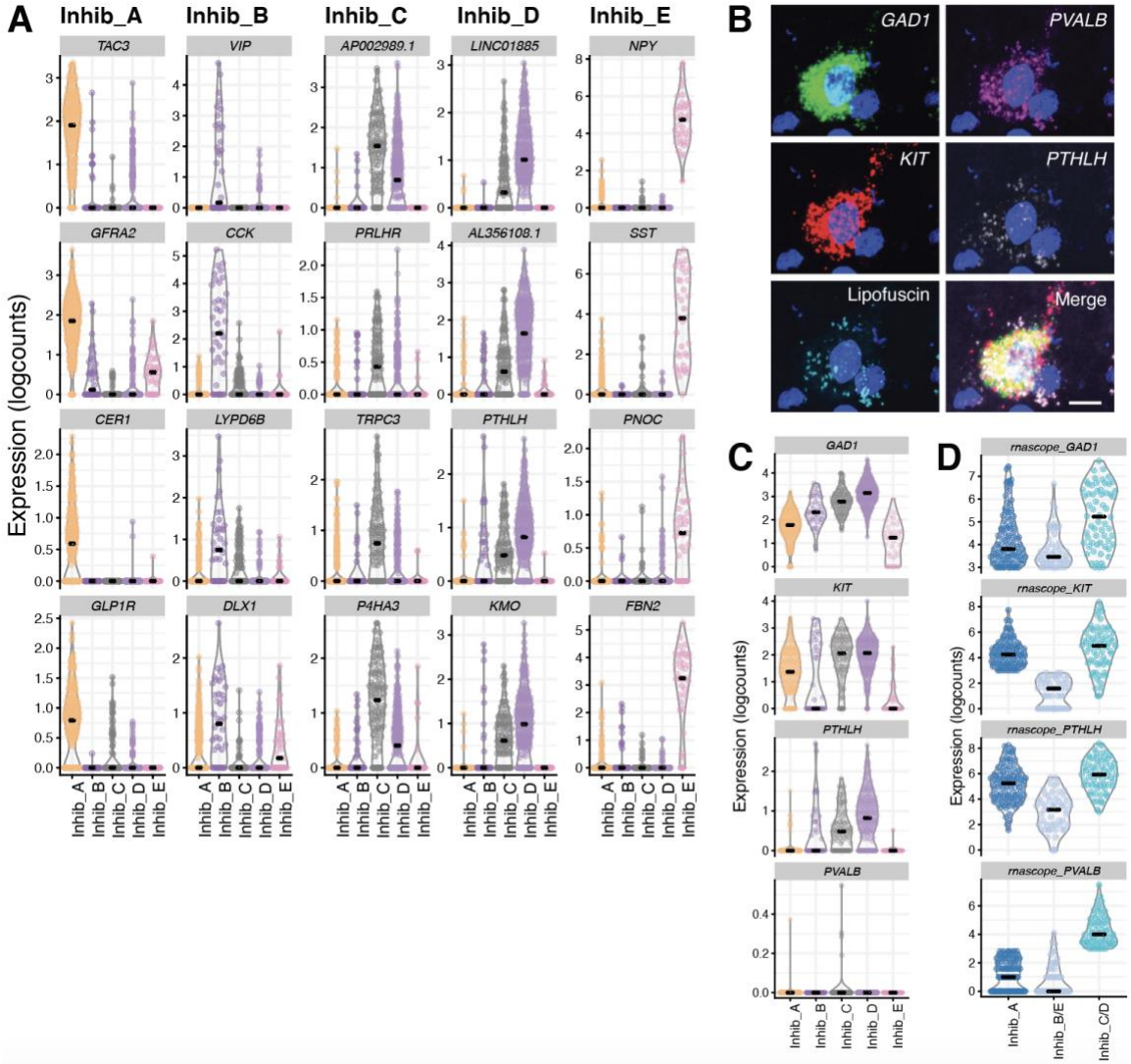


Figure 2.1.6 legend:

(**A)** Violin plots depicting top 4 genes in each GABAergic inhibitory neuron class (columns) in NAc snRNA-seq.

(B) Multiplex single molecule fluorescent in situ hybridization (smFISH) in human NAc depicting co-expression of *PVALB, KIT,* and *PTHLH* in *GAD1*+ inhibitory neurons. Maximum intensity confocal projections showing expression of DAPI (nuclei), *GAD1*, *PVALB*, *KIT*, *PTHLH* and lipofuscin autofluorescence. Merged image without lipofuscin autofluorescence. Scale bar=10 μm.

(**C)** Corresponding violin plots showing (log_2) expression of these genes in different interneuron classes by snRNA-seq.

(D) Log_2 expression of respective transcript counts per smFISH ROI (ROI size-normalized), post lipofuscin-masking (autofluorescence). Each *GAD1*+ ROI was assigned to a Euclidean distance-predicted interneuron class (or group of classes) and its(/their) respective expression.

2.2 Amygdala

The amygdala (AMY), a medial structure of the temporal lobe, is noted for its role in processing emotional valence, particularly for both fear and reward (Janak and Tye, 2015; Wassum and Izquierdo, 2015). Dysfunction in amygdalar signaling is implicated in major depressive disorder, bipolar disorder and posttraumatic stress disorder (PTSD) (Fenster et al., 2018; Garrett and Chang, 2008; Murray et al., 2011). The human amygdala can be subdivided into a number of distinct regions based on histology, immunohistochemical classifications, connectivity, and neural activation patterns as revealed by functional magnetic resonance imaging (fMRI) of the brain (Barger et al., 2012; Schumann and Amaral, 2005; Sorvari et al., 1995; Tyszka and Pauli, 2016; Zhang et al., 2018). Studies in the rodent and non-human primate amygdala have identified different cell compositions across the amygdala, which likely correspond to differential patterns of synaptic connections between cell types across amygdalar subregions, and

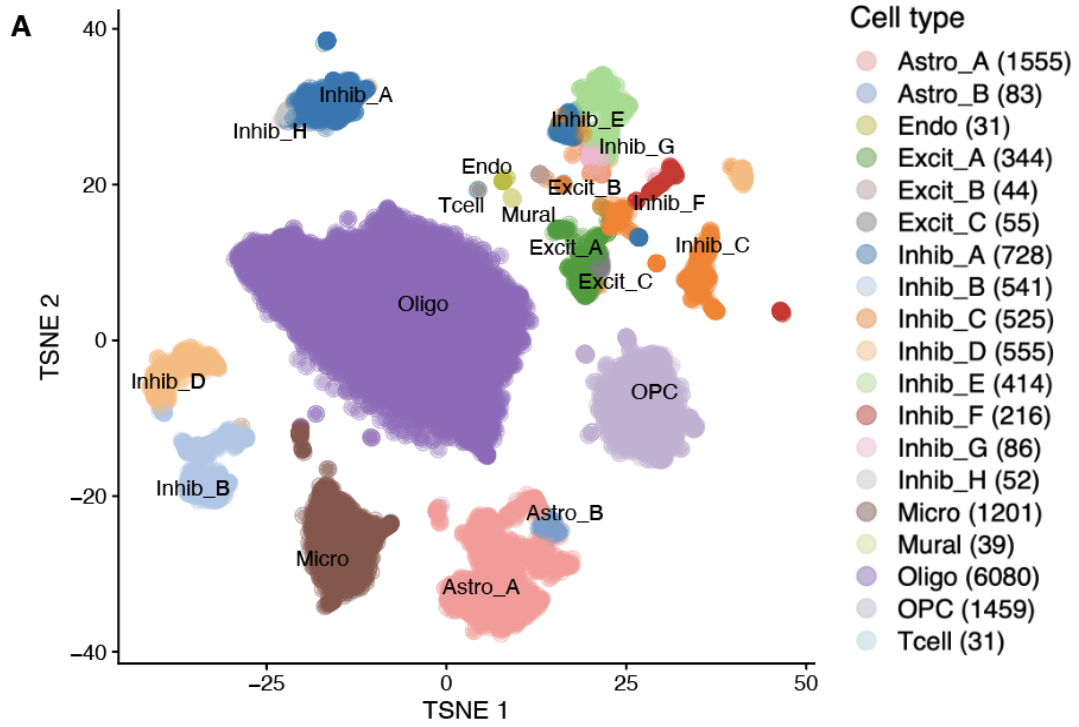
with extra-amygdalar brain regions (Chareyron et al., 2011). Hence, it is likely that various cell types with unique molecular signatures also exist within the human amygdala, which can be surveyed by snRNA-seq. We analyzed 14,039 nuclei from the amygdala of five adult neurotypical donors to create a molecular taxonomy of cell types in this brain region. We identified 19 clusters that corresponded to four broad glial cell types (Oligo, Astro, Micro, and OPC), stromal (endothelial, 'Endo'; and mural cells) or immune cell populations (Macrophages and T cells), and 11 neuronal classes (Figure 2.2.1A). Glial cell populations were present at similar proportions between the non-NeuN-enriched donors (mean 54.4% Oligo; 12.3% Astro; 10.7% Micro; 11.5% OPC), but we observed a varied distribution of neuronal classes between donors, and the stromal/immune cell classes, as these were more rare (see Discussion; Table 2.2). Despite this, after correcting for donor batch effects, we identified hundreds of genes enriched in each broad glial, stromal, and neuronal classes at FDR < 1e-6 (for information on top markers, see **Chapter 7** (Methods)).

Within the 11 neuronal classes expressing the pan neuronal marker gene *SNAP25*, three clusters were enriched for excitatory neuronal markers (*SLC17A7*, *SLC17A6*) and eight clusters were enriched for inhibitory GABAergic markers (*GAD1, GAD2;* **Figure 2.2.2**). The three excitatory classes comprised different functional classes of neurons (referred to as 'Excit_A' to '_C'), with top markers including *NRN1*, *NPTX1* and *SLC30A3* (encoding neuritin, neuronal pentraxin 1, and zinc transporter 3, respectively) for Excit_A, and *SLC17A6* and *VCAN* (Versican, typically associated with OPCs) for Excit_B (Figure 2.2.1B). *NRN1*, *NPTX1*, *SLC30A3*, and *VCAN* have all been implicated in modulation of synaptic plasticity and memory (Figueiro-Silva et al., 2015; Horii-Hayashi et al., 2008; Sindreu and Storm, 2011; Yao et al., 2018). Top markers for class Excit_C included *MCHR2* (melanin-concentrating hormone receptor 2) and *CDH22*

(pituitary/brain-cadherin). Additionally, Excit_A is a large class made up of four subpopulations (see Discussion).

Compared to the excitatory neuron classes, we identified a greater diversity of inhibitory GABAergic classes and subclasses (see Methods). Those classes of note include cholecystokinin (*CCK*)-containing regular-spiking interneurons (Inhib_B, Inhib_D) evident by high expression of *CCK* (Figure 2.2.1B). Of these *CCK*-expressing GABAergic classes, Inhib_B was also enriched in *VIP* and *CALB2* (calretinin), whereas Inhib_D showed enrichment for *KIT*. *NPY* was specific for the smaller of two *PENK*+ classes, Inhib_A and Inhib_H (Figure 2.2.1B), whereas *SST* and *TAC3* were enriched in specific subclusters of some of these interneuron classes (data not shown). More functional characterization includes Inhib_B enrichment for *CRH* (corticotropin release hormone/factor)-expressing subpopulations. CRH is a key regulator of the hypothalamic-pituitary-adrenal (HPA) axis, which is critical for both the acute stress response and adaptation to chronic stress. Finally, *NPFFR2* and *TLL1*, additional genes associated with HPA axis regulation, were selectively expressed in Inhib_C (Lin et al., 2016; Tamura et al., 2005). These classes of neurons reflect most known anxiety-related neuronal subclasses as reviewed in (Babaev et al., 2018), and also better clarifies some expected patterns of molecular identity, such as that *SST* and *PRKCD* (which characterize striatal-like GABAergic neurons of the centrolateral amygdala) are not necessarily mutually exclusive at the subclass level (data not shown).

Figure 2.2.1: Atlas of molecularly-defined cell types in human AMY.



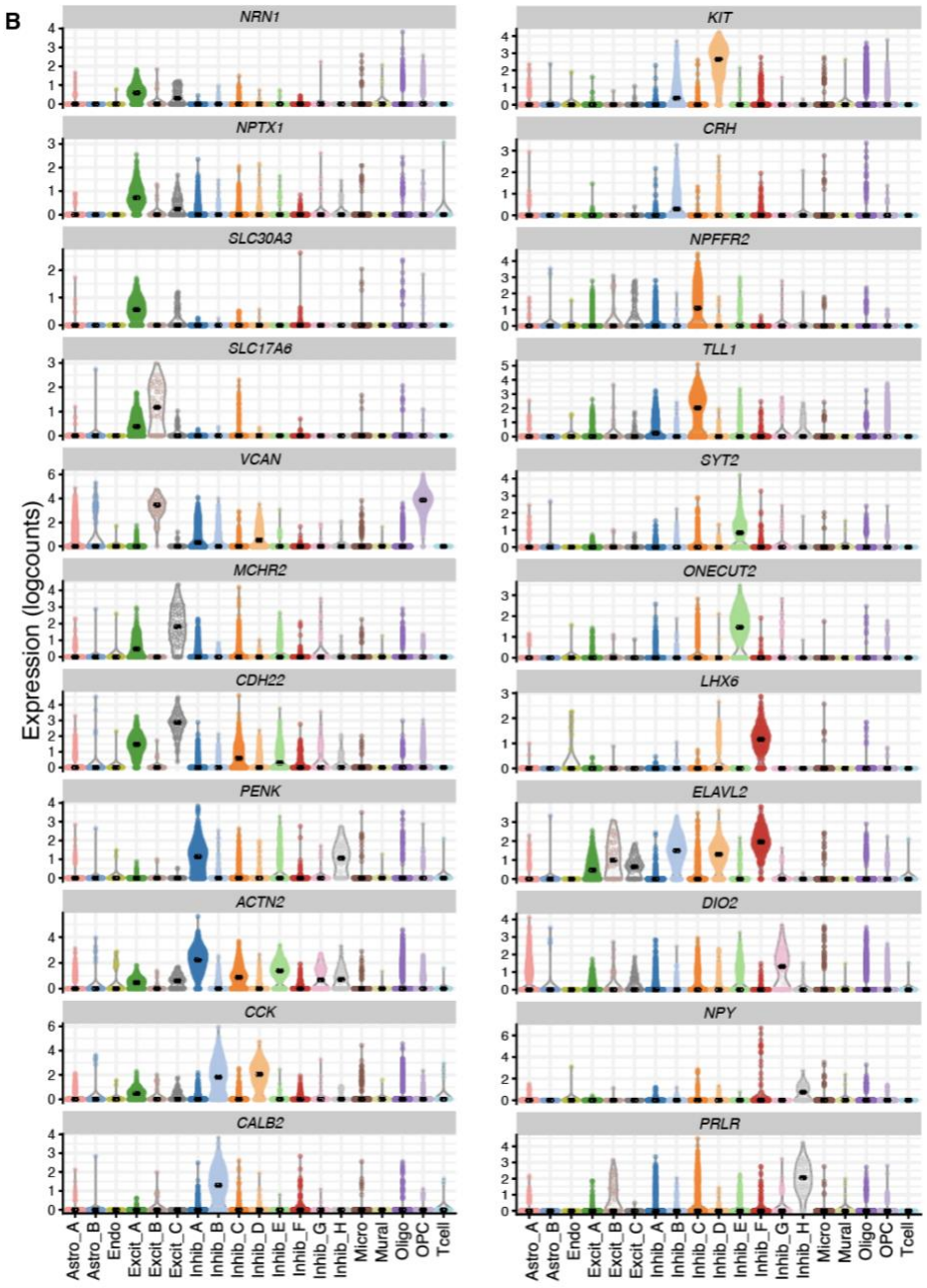


Figure 2 legend:

(A) tSNE plot of 14,039 nuclei (*n*=5 donors) across 19 clusters, including 3 clusters of excitatory neurons and 8 clusters of GABAergic inhibitory neurons.

(B) Expression violin plots for the top 2-3 genes for each of the neuronal subpopulations (log2-normalized counts).

Figure 2.2.2. Broad cell type marker expression for AMY cell classes.

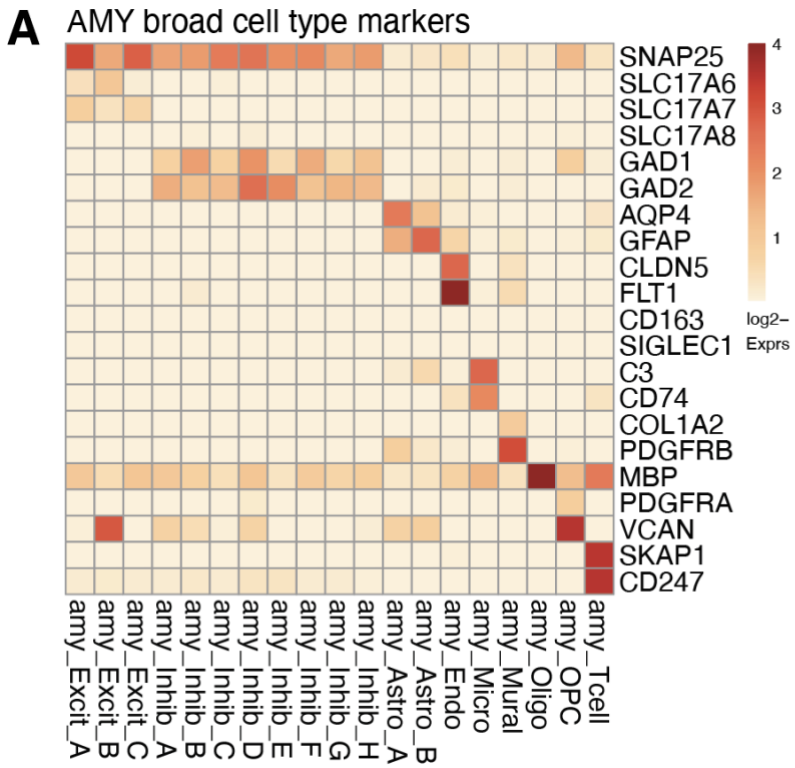


Figure 2.2.2 legend:

(A) Mean $\log_2$-normalized expression for broad cell type markers, used for annotation of AMY cell classes.

2.3 sACC, DLPFC, and HPC

To complement the cell class populations described in the previous sections for the NAc and AMY, we additionally defined the catalog of cell type clusters and cluster-specific genes within the other brain regions in our dataset (sACC, DLPFC, and HPC), separately (**Figure 2.3.1**; **Figure 2.3.2**). We further benchmarked our transcriptomic

profiles against other published datasets that profiled similar regions in the postmortem human brain. Overall, our HPC subpopulations correlated well with the broad cell classes as reported in (Habib et al., 2017); **Figure 2.3.3**). We additionally observed strong overlap between our DLPFC to the reported PFC profiles from (Velmeshev et al., 2019); **Figure 2.3.4**), or similarly, sACC to the ACC set (**Figure 2.3.5**). Interestingly, our sACC subpopulations did not correlate more strongly with the ACC subpopulation profiles than their co-reported PFC profiles from (Velmeshev et al., 2019), whereas our DLPFC subclusters generally correlated only slightly more strongly to the reported PFC than ACC subpopulations. This suggests that these cortical regions share a high degree of overlap in their nuclear transcriptomic profiles. The strength of correlation to these benchmark datasets demonstrates the robustness and utility of our pipeline, and the presented data significantly expand the existing repository of postmortem human brain snRNA-seq datasets.

From this work, we established a cell type atlas of five key brain regions embedded within the reward circuitry: the NAc, AMY, sACC, DLPFC, and HPC, annotating cell classes identified within each region, separately. This was a practical and intentional approach, to allow for flexibility in downstream usage of specific datasets of interest with respect to the individual investigator's research questions. Additionally, we generated public, interactive web apps for each of these regions (e.g. https://libd.shinyapps.io/tran2021_NAc/; see **Chapter 7** (Methods) for more information), to explore the data without needing to be adept at the command line. This includes various dimensionality reduction principal components or visualizations (t-SNE, UMAP), gene expression heatmaps and violin plots, and gene information, including the proportion of nuclei in a given cell class that exhibit non-0 expression for gene *X*. These resources reflect this work's publication as a *Neuron* NeuroResource, at (Tran et al.,

2021).

Figure 2.3.1. Region-specific analyses reveals a total of 107 cell classes in the reward circuitry.

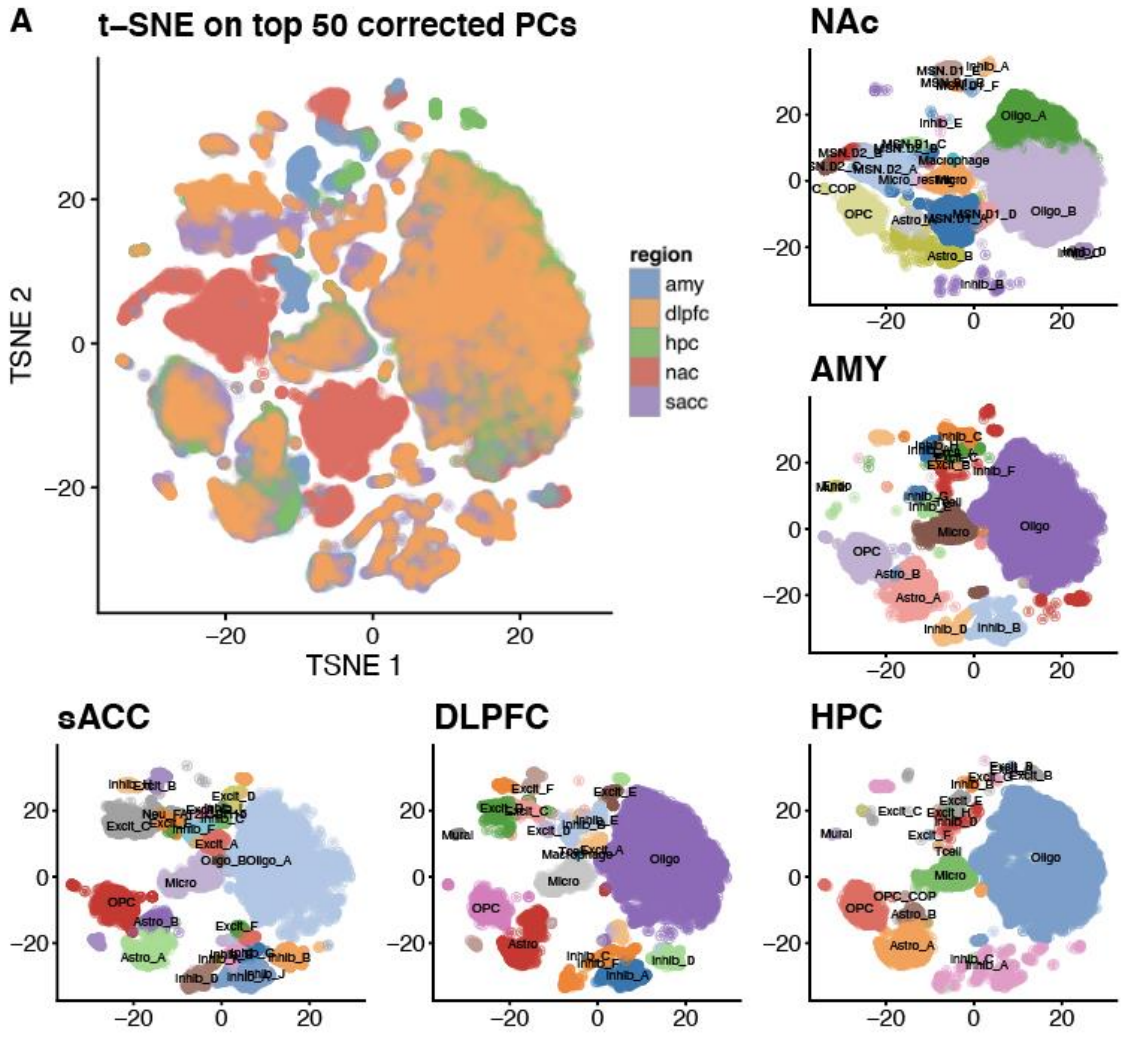


Figure 2.3.1 legend:

(A) tSNE array of a total of 70,615 nuclei, paneled by each brain region and their regionally-defined cell classes (a total of 107 cell classes)

Figure 2.3.2. Broad cell type marker expression for pan-brain-defined clusters or regionally-defined populations.

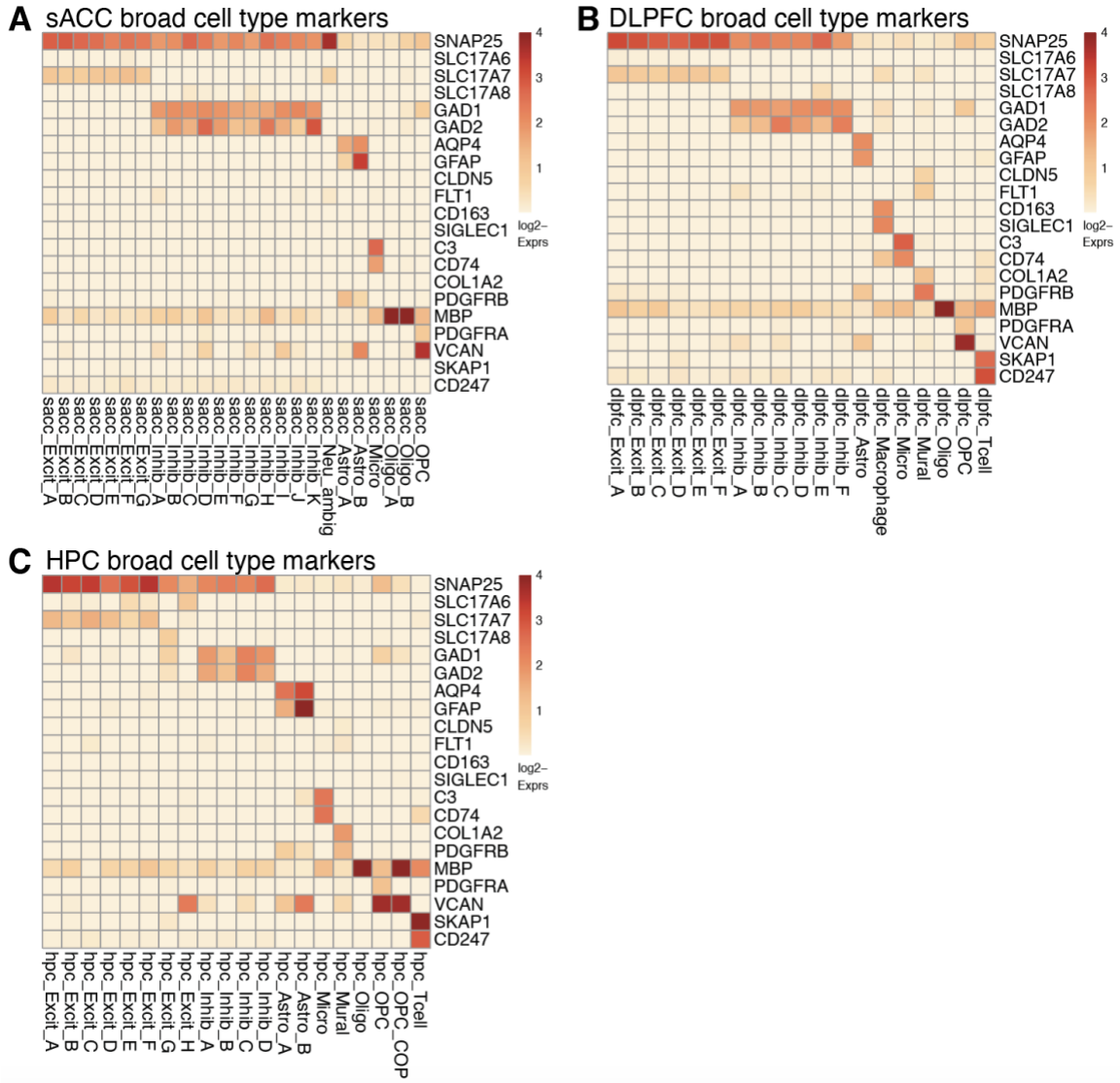


Figure 2.3.2 legend:

(A) Mean $\log_2$-normalized expression for broad cell type markers, used for annotation (or identified, post hoc), in clusters defined within sACC nuclei.

(B) Same as (A), but for DLPFC, and

(C) HPC.

Figure 2.3.3. Benchmarking of HPC cell classes to published data.

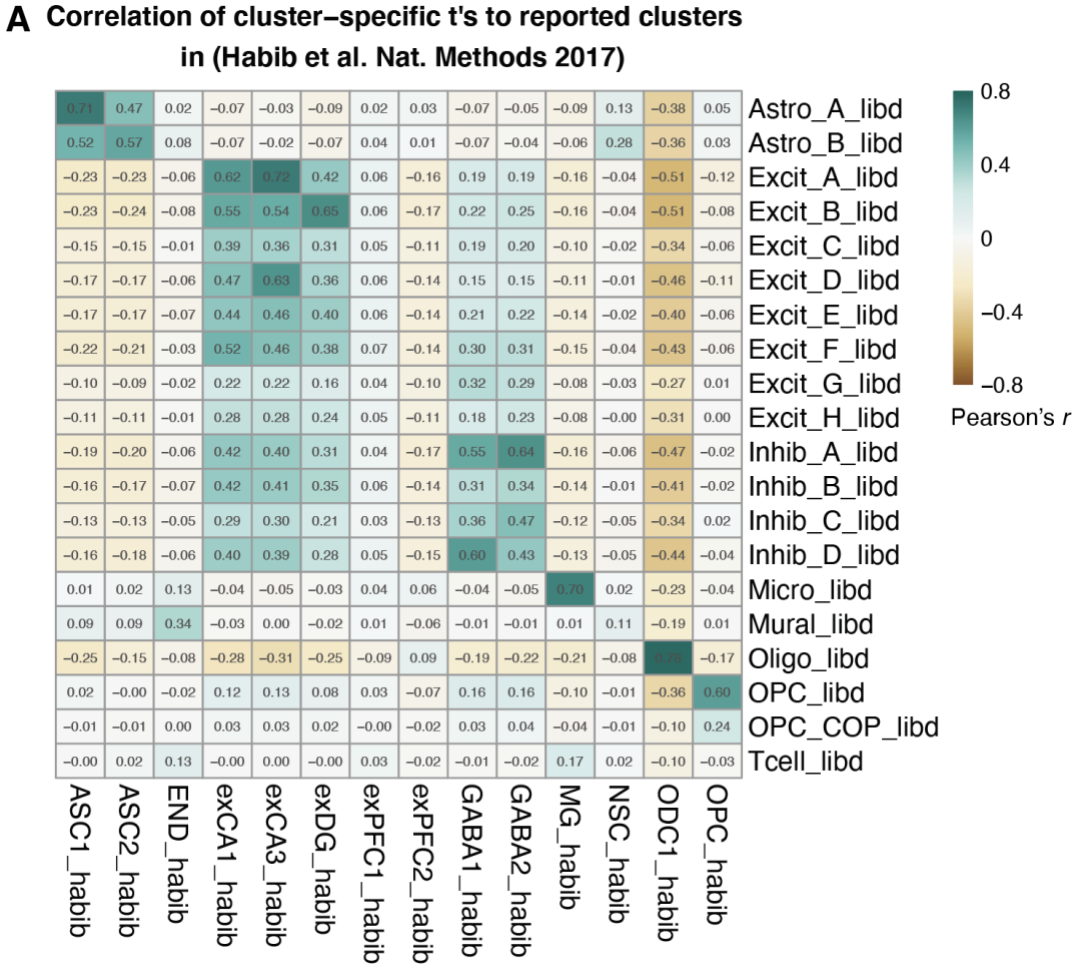


Figure 2.3.3 legend:

(A) Correlation heatmap between HPC subclusters (rows) and the reported HPC populations in (Habib et al., 2017); columns). Printed values and scales show the Pearson correlation coefficient (*r*), correlating across all shared expressed genes and the *t*-statistics of their specificity test.

Figure 2.3.4. Benchmarking of DLPFC subpopulations to published data.

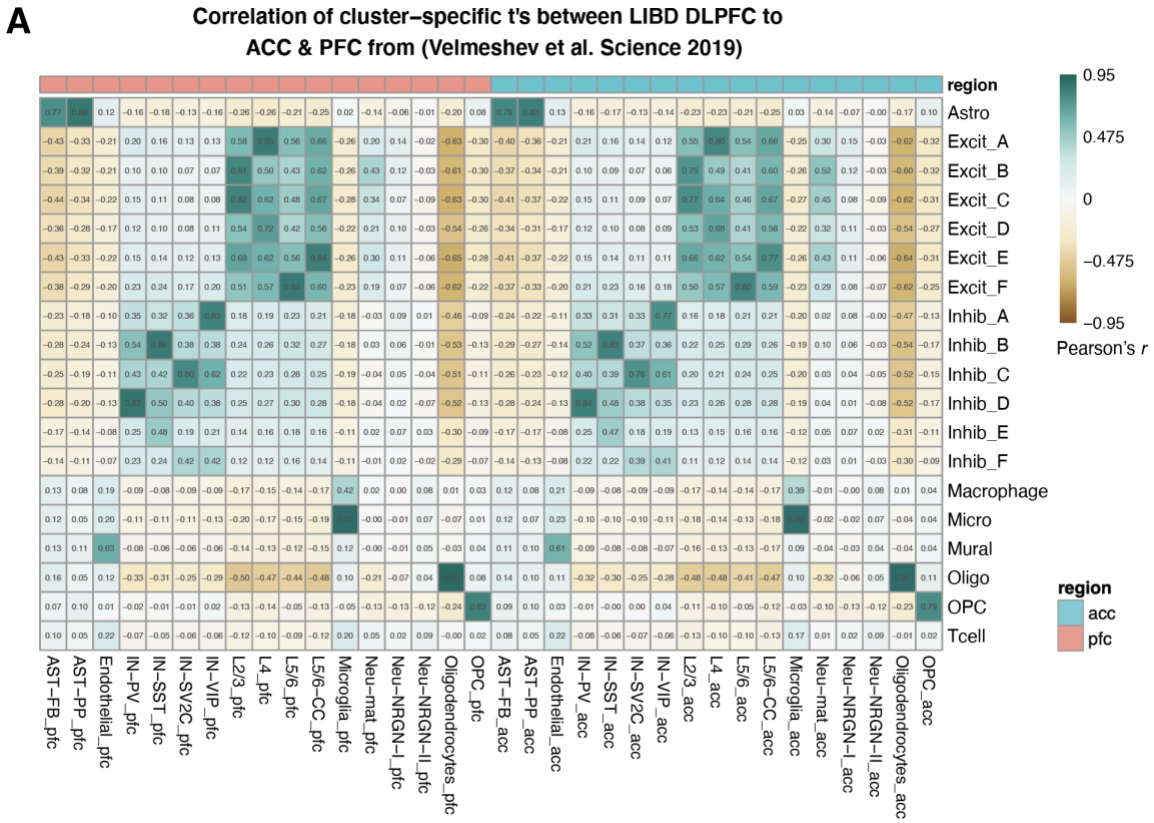


Figure 2.3.4 legend:

(A) Correlation heatmap between DLPFC spatially-registered subpopulations (rows) and split-PFC and ACC 10x snRNA-seq clusters (columns) from (Velmeshev et al., 2019). Printed values and scales show the Pearson correlation coefficient, correlating across all shared expressed genes (26,970) and the *t*-statistics of their specificity test.

Figure 2.3.5. Benchmarking of sACC subpopulations to published data.

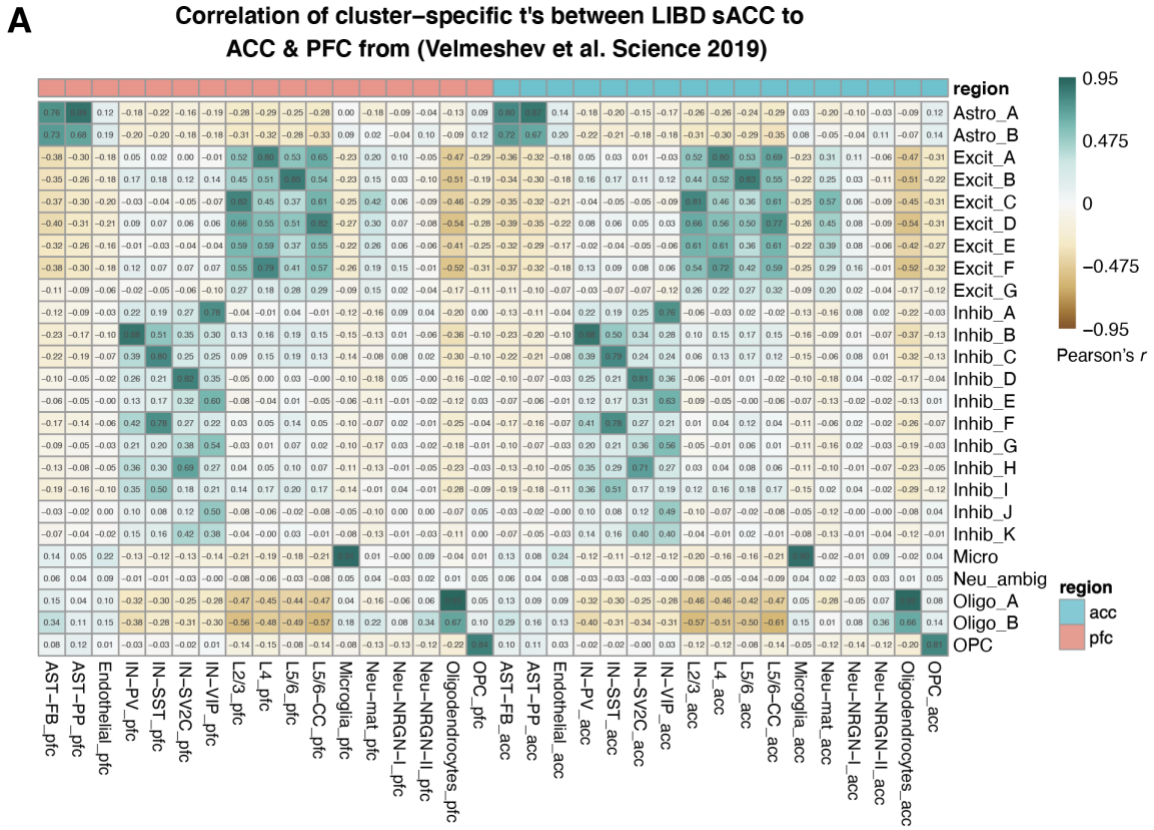


Figure 2.3.5 legend:

(A) Correlation heatmap between sACC subpopulations (rows) and split-PFC and ACC 10x snRNA-seq clusters (columns) from (Velmeshev et al., 2019). Printed values and scales show the Pearson correlation coefficient, correlating across all shared expressed genes (27,890) and the *t*-statistics of their specificity test.

Chapter 3. Assessment of cell type conservation with published rodent model single-cell/-nucleus RNA-seq data

We next evaluated the conservation of NAc cell types across species by comparing our cluster-level transcriptional profiles with those generated in a previous snRNA-seq study of the NAc following cocaine experience in a rat model system, which analyzed a total of 16 subclusters across 15,631 rat NAc nuclei (Savell et al., 2020). Correlation analyses between our NAc cell classes with those derived from rat NAc revealed that glial populations, including astrocytes, microglia, oligodendrocytes, and oligodendrocyte progenitor cells, were highly conserved (**Figure 3.1A**). GABAergic inhibitory neuron populations were also well-correlated across species as rat *Sst*-expressing and likely-*Pvalb*-expressing clusters overlapped with human Inhib_E and Inhib_C/_D classes, respectively (Pearson's r = 0.63, 0.63, and 0.62, respectively). We also observed substantial correlation between rat and human D1 and D2-MSNs, especially between rat *Drd1*-expressing MSNs and human D1_A/_D1_D MSNs (r = 0.74, 0.74). Beyond the overlap of rat *Drd2*-expressing MSNs in the human D2_A and D2_B MSN classes (r = 0.77, 0.70, respectively), we additionally saw positive correlations across D1 and D2 MSN subtypes, such that rat *Drd2*-expressing MSNs also showed enrichment in our human D1_A/_D MSNs. This result is not likely fully explained by co-expression of *DRD1* and *DRD2* in the same nucleus because, while we we did find that ~11.2% of all MSNs expressed both *DRD1* and *DRD2* to some degree, these dual-expressing nuclei were by far the most enriched in the D1_E class (43.4% of D1_E nuclei expressing both *DRD1*/*DRD2*) (**Figure 2.1.1B;** Figure 2.1.2). Additionally, many of the top markers for either the D1_A or D2_A classes were highly expressed in both MSN

clusters (**Figure 2.1.5**), suggesting that the majority of canonically dichotomous D1 or D2 MSNs may be more molecularly similar than previously appreciated. We did not observe strong enrichment for rat *Drd3*- and *Grm8*-expressing MSNs in any human MSN subclusters. Likewise, a few human D1 (D1_B, _C, _E_ and _F) and D2 (D2_C and _D) MSN classes did not appear to be convincingly represented in rat MSN subtypes (**Figure 3.1A**; see Discussion).

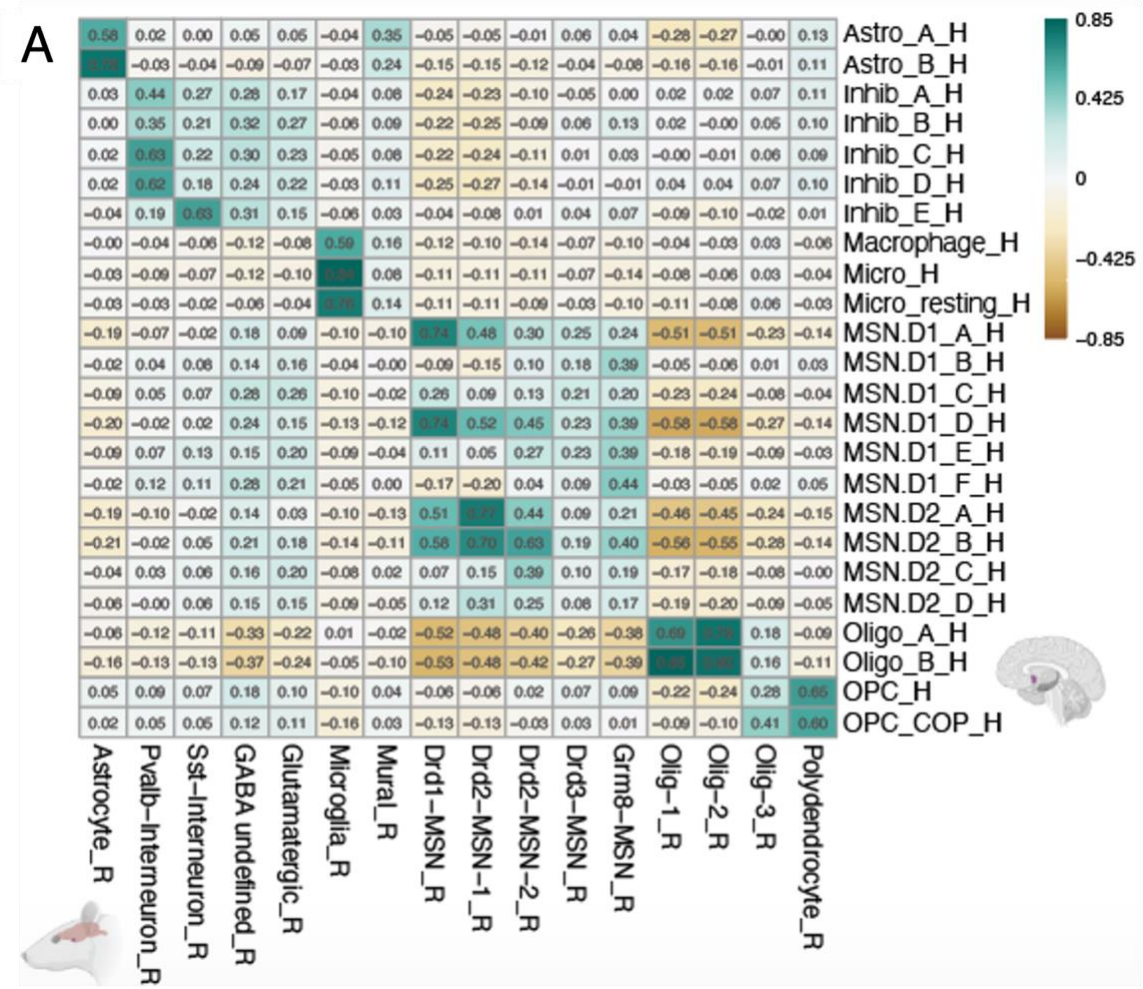
Taken together, while these data suggest strong overall conservation between rat and human NAc cell types, there appear to be transcriptional features that are unique among specialized subpopulations of rodent and human MSNs. This work also highlights a key caveat of single-nucleus molecular data, which, namely, may not capture transcripts associated with well-characterized and expected cell populations, despite multiple lines of evidence for their biological identity. This is potentially due to differential nuclear export dynamics of RNA transcripts, as discussed in **Chapter 6**, *Takeaways and expectation for future work*.

We next compared our AMY cluster-level transcriptomic profiles to those of a previously published single-cell dataset derived from the mouse medial amygdala (MeA) (Chen et al., 2019) to evaluate conservation of amygdalar cell types between humans and rodents (**Figure 3.1B**). Across the top shared homologous genes (see Methods), we observed substantial correlation between several mouse and human amygdala cell types. For example, our human glutamatergic class Excit_A (*SLC17A6*+, *SLC17A7*+) most closely correlated with the mouse MeA glutamatergic subcluster ‘N.11’ (Pearson correlation: r = 0.60). Indeed the marker genes that were most highly conserved between these subclusters included *SLC30A3*, *NPTX1*, and *NRN1*. Another notable pair of cell classes conserved between species was human inhibitory neuronal class, Inhib_C, and mouse inhibitory subcluster MeA ‘N.8’ (r = 0.61). The top shared genes

between these clusters included *NPFFR2*, *GRM8*, and *FOXP2*. Though we observed selective co-expression of *NPFFR2* and *TLL1* in human Inhib_C, we note absence of orthologous *Tll1* expression in all mouse MeA neuronal subclusters (**Figure 3.2**), including the corresponding cluster 'N.8', suggesting species differences in the molecular characteristics of neuronal subpopulations. Evidence supporting this is that *MCHR2* (a top marker for Excit_C) is restricted to higher-order mammals, whereas rodent genomes only encode the related *Mchr1(Tan et al., 2002)*. Indeed, in assessing the extent to which marker genes (per human AMY cell class) overlapped with this shared homology gene space, an average of 10.4% cell type markers appeared to be human-specific (data not shown). Importantly, several neuronal subpopulations in the mouse and human datasets lacked strong correlation with each other (e.g. human Excit_B, mouse 'N.3', N.7', and 'N.12'), either suggesting possible molecular divergence between species, or unique differences between the cell-type makeup of amygdalar subregions, such that all subpopulations may not be fully represented in our human amygdala samples compared to mouse MeA samples. Our cross-species analysis demonstrates the potential conservation of neuronal subtypes between human amygdala and mouse MeA, but highlights potential differences in the cellular distribution and transcriptomic profiles across neuronal subtypes.

This work demonstrates a statistically simple and efficient method to interrogate the overlap of different datasets - in this case, comparisons of cell populations between species. However, many caveats need to be noted, the most important being that this is not a formal assessment of species conservation; that the approach relies on up-to-date bioinformatics information for orthologous gene matching, and for the investigator to make a choice on which paralog to represent said gene, where there exist many; and different technologies employed for single-cell/-nucleus transcriptome capture.

Figure 3.1. Cross-species assessment of conservation between human and rodent brain regions.



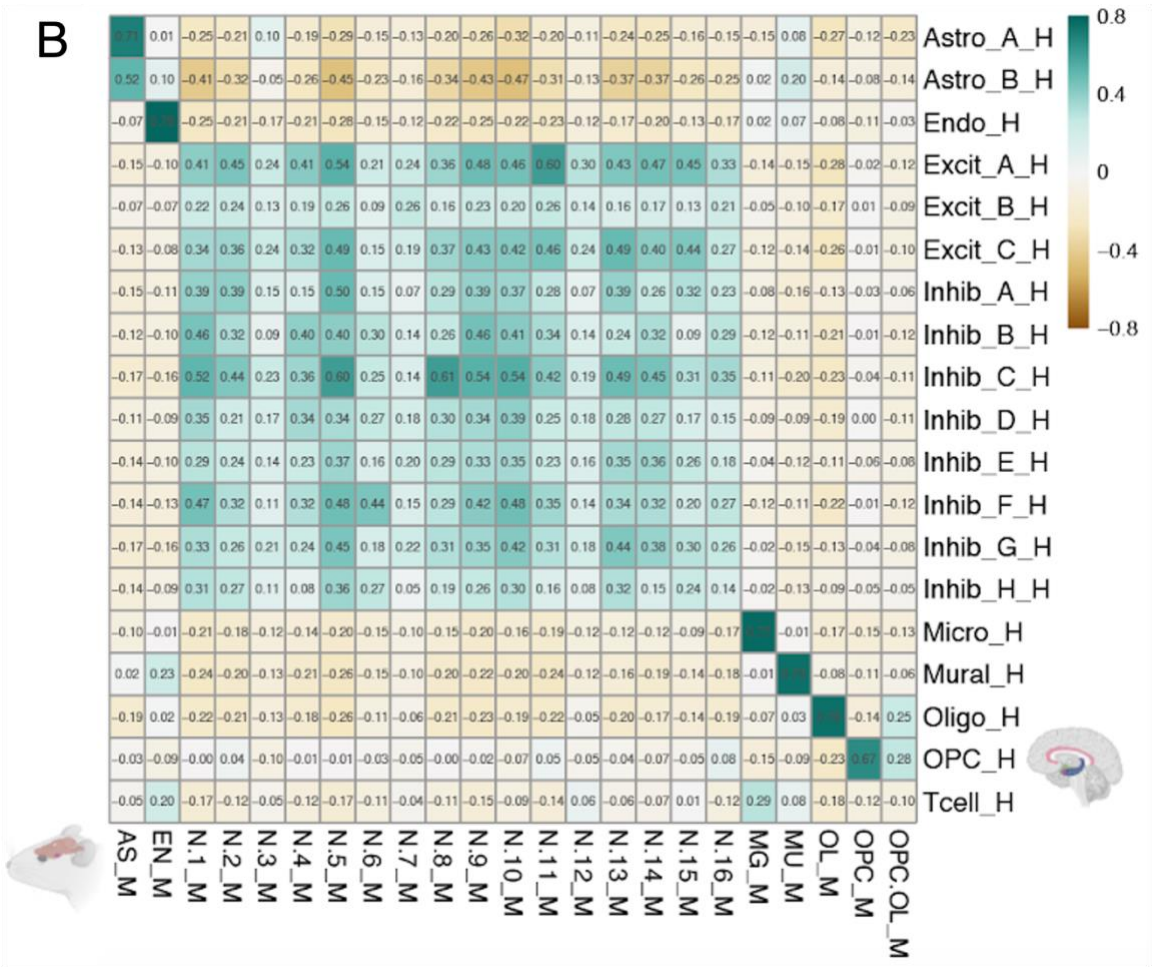


Figure 3.1 legend:

(A) Heatmap of Pearson correlation values (*r*, scale) evaluating the relationship between our human-derived NAc cell classes (rows) and reported rat NAc populations from (Savell et al., 2020). Correlation was performed on the combined top-100 markers/cell population, where annotated homology exists (here, 582 genes; see Methods).

(B) Heatmap of Pearson correlation values (*r*, scale) evaluating the relationship between our human-derived amygdala cell classes (rows) vs. the cell populations reported in (Chen et al., 2019), derived from mouse medial amygdala (MeA). Correlation was performed on the combined top-100 markers/cell population, where annotated homology exists (here, 480 genes; see Methods).

Figure 3.2. AMY ‘Inhib.5’ vs. corresponding MeA ‘N.8’ shared markers.

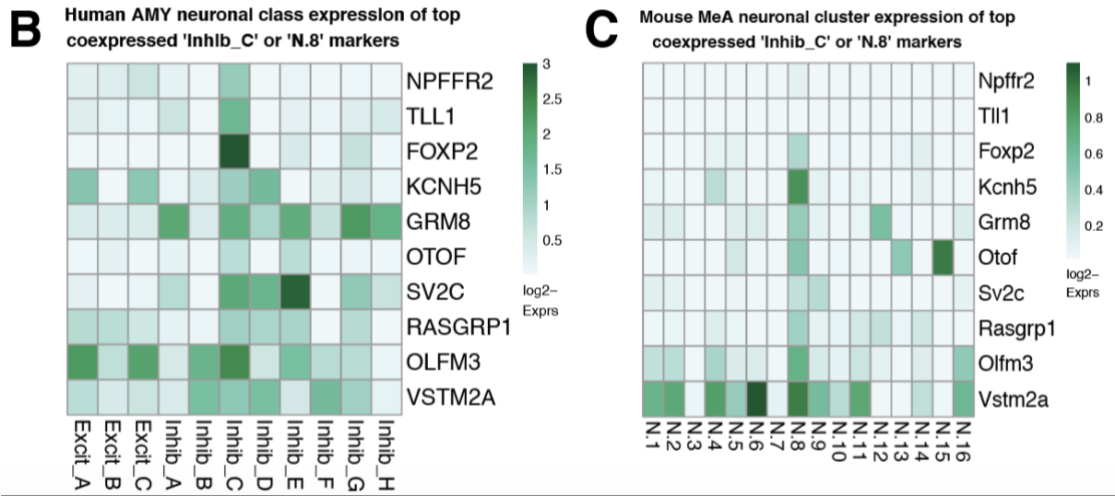


Figure 3.2 legend:

(B) Mean expression of top enriched markers for human AMY subpopulation Inhib_C shared with

(C) mouse MeA neuronal subclusters, as reported in (Chen et al., 2019). *Tll1*, however, was not defined as a marker of MeA ‘N.8’.

Chapter 4. Integrating region-specific cell class profiles to characterize the transcriptomic architecture of the reward circuitry

To generate a global view of the transcriptomic architecture across the five brain regions, we compared gene expression patterns across all 107 regionally-defined cell classes. Overall, each glial or stromal cell subpopulation (Oligos, Astros, OPCs, and Micros, Endothelial cells, Mural cells, Macrophages and T cells) showed broadly consistent gene expression patterns across all five brain regions (**Figure 4.1**), in line with previous analyses of broad non-neuronal cell populations using DNA methylation data (Rizzardi et al., 2019). However, in cases which yielded multiple classes of glial cells, such as astrocytes, there were unique blocks of shared transcriptional programs between subclasses. For example, the sACC 'Astro_B', HPC 'Astro_C', and NAc 'Astro_A' are tightly correlated and more distinct from the other astrocyte subclasses, whereas the 'Astro_B' from the AMY seem to be most distinct, in that it doesn't cluster with any of the rest. Amongst the annotated 'Astro' classes, these are characterized by the highest expression of *DST*, *COL19A1*, and *MACF1* (**Figure 4.2**). Though most of the other astrocyte classes (aside from the sACC 'Astro_B'), themselves, are characterized by specific or higher expression of unique markers, the AMY 'Astro_B' showed a distinctly lower transcriptional activity state (**Figure 4.2**), likely related to this somewhat divergent astrocyte class from the amygdala, as seen with 'Micro_resting', or putative resting or dormant microglia, found in the NAc (**Figure 2.1.1**A; **Figure 4.2**). Finally, we highlight that most of these reported classes of glial cell populations comprise a greater diversity of subclusters (see Discussion), the characterization of which is beyond the scope of this paper.

Within the neuronal set of region-specific annotations, totaling 69 neuronal classes across 28,150 nuclei, most inhibitory or excitatory populations showed transcriptional patterns that clustered these broad classes together across brain regions, as expected (**Figure 4.3**), although this wasn't exclusively the case in all instances. For example the AMY Inhib_A, _C, _E and _G, in addition to HPC Inhib_B, more strongly correlated with the excitatory branch, though they distinctly express more canonically inhibitory-classifying markers (**Figure 2.2.2**; **Figure 2.3.2**). We also observed strong similarities between unique pairs of neuronal subpopulations across regions, such as between AMY and DLPFC ('In_B_amy' and 'In_A_dlpfc'; r = 0.86). Indeed, this DLPFC inhibitory subpopulation shares many top markers with its *CALB2*+, *VIP*+ AMY counterpart (see **Chapter 7** (Methods)), including *CALB2* and *VIP*, in addition to selective expression of *CRH*. In addition, we saw a variety of strongly correlating DLPFC : sACC neuronal class pairs, such as 'Ex_F_dlpfc' and 'Ex_B_sacc' (r = 0.94), and 'Ex_E_dlpfc' and 'Ex_D_sacc' (r = 0.94). This suggests potentially overlapping layer-specific identities in these regions, as evident, for example, by 'Ex_F_dlpfc' and 'Ex_B_sacc' most highly correlating with the corresponding and reported 'L5/6' cluster from (Velmeshev et al., 2019) (r= 0.84, 0.83, respectively; Figure 2.3.4; Figure 2.3.5).

Integrating these neuronal classes across regions also suggested an excitatory transcriptomic signature in the NAc-specific MSN.D1_A/_D classes, in addition to D2_A/_B MSNs, as their profiles clustered with the broad excitatory branch of neuronal classes (**Figure 4.3**), whereas the remaining, less abundant MSN classes (MSN.D1_B/_C/_E/_F, and D2_C/_D) clustered in the general inhibitory branch. Strikingly, the former MSN classes negatively correlate with most other MSN and GABAergic inhibitory populations within the NAc, suggesting potentially divergent transcriptional programs across NAc neuronal classes. This observation was supported

by hierarchical NAc cluster relationships, where the four D1/D2 MSN classes carrying more excitatory signature (collectively termed ‘MSN.excit’) are seen to be more related than the remaining six (‘MSN.inhib’; **Figure 4.4**), and further by separation between these groupings by the top principal components (PCs) describing variance across all NAc nuclei (**Figure 4.4**). Further investigation will be needed to assess this divergent feature of MSN classes and identify what unique roles these dopaminoceptive neurons, with respect to their general D1/D2 classification, play in reward and emotional valence processing. These analyses illustrate the utility of integrating cell-type profiles from across datasets or related tissue types/regions, to better understand cell type identity at a full transcriptome level.

Figure 4.1. Comparison across all non-neuronal, regionally-defined subpopulations.

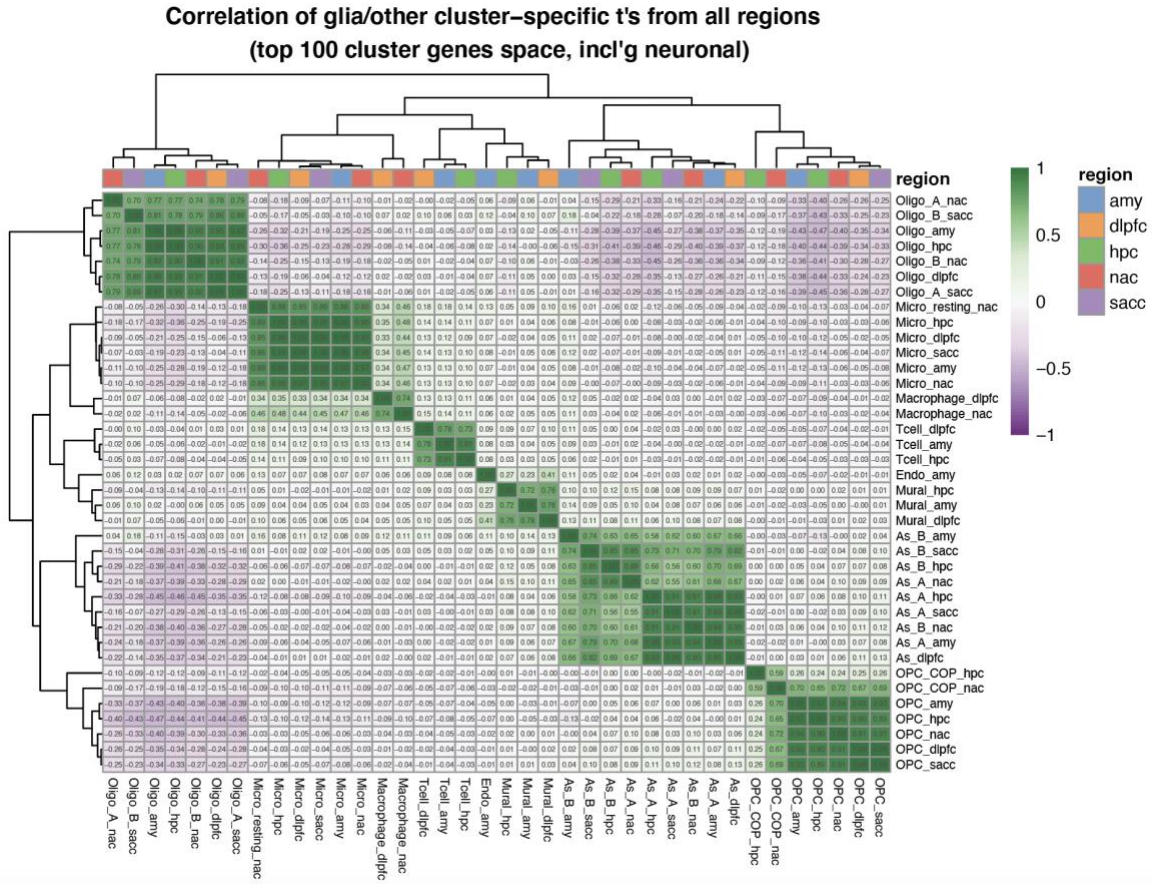


Figure 4.1 legend:

Pairwise correlation of *t*-statistics, comparing the top cell class marker genes of the total of 107 classes reported across the five brain regions (total of 3,715 genes). Here, only the 38 non-neuronal (glial, stromal, or immune) classes are shown. Regions are colored and labeled in the suffix. Scale values are of Pearson correlation coefficient (*r*).

Figure 4.2. Across-regions astrocyte differential expression analysis and astrocyte/microglia QC.

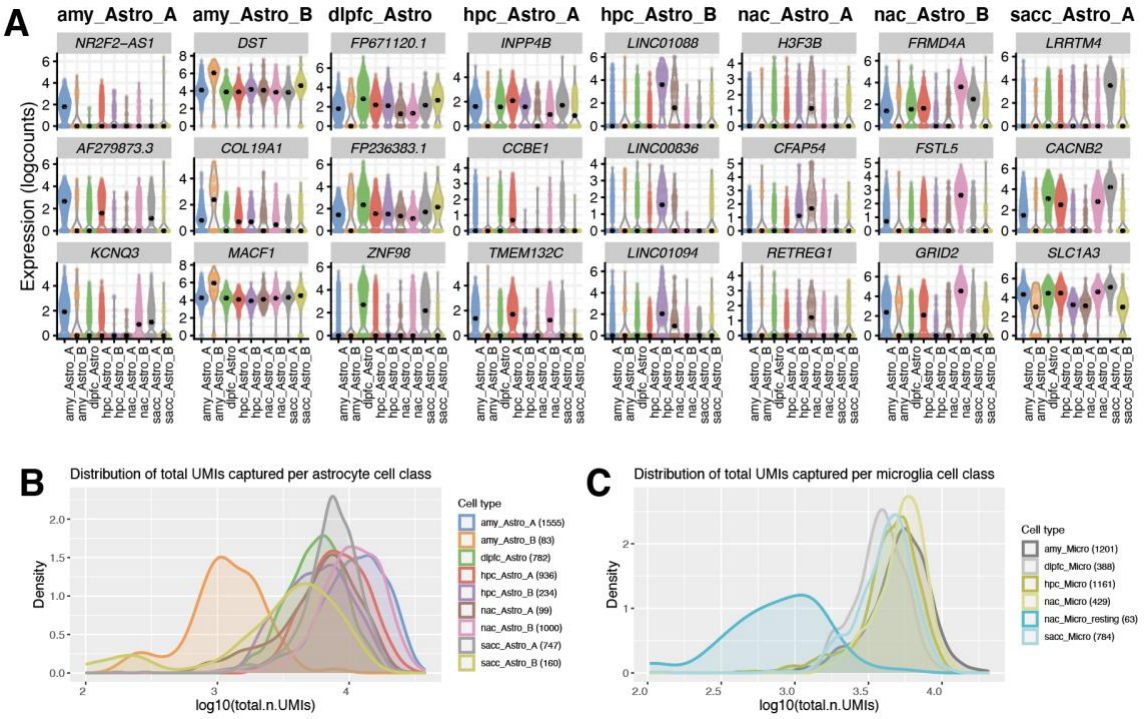


Figure 4.2 legend:

(A) Violin plots (scale normalized, log2-transformed, 'logcounts') of the top 3 pairwise-defined markers per regionally-defined astrocyte class. The 'Astro_B' from the sACC had no statistically significant pairwise test-defined markers, so doesn't have its own column (but still showing its corresponding expression in the presented 24 marker genes).

(B) Density plots for total number of UMIs captured per astrocyte cell class (including sACC 'Astro_B'). Here, we see the AMY 'Astro_B' shows a 10-fold magnitude lower in transcriptional activity than the rest of the astrocytes.

(C) Similarly, for resting or dormant microglia in the NAc, or 'Micro_resting', its distribution of total number of UMIs captured is an order of magnitude lower than the rest of the defined microglia classes.

Figure 4.3: Across-regions analyses reveal whole brain transcriptomic architecture and neuronal subtype similarities across regions.

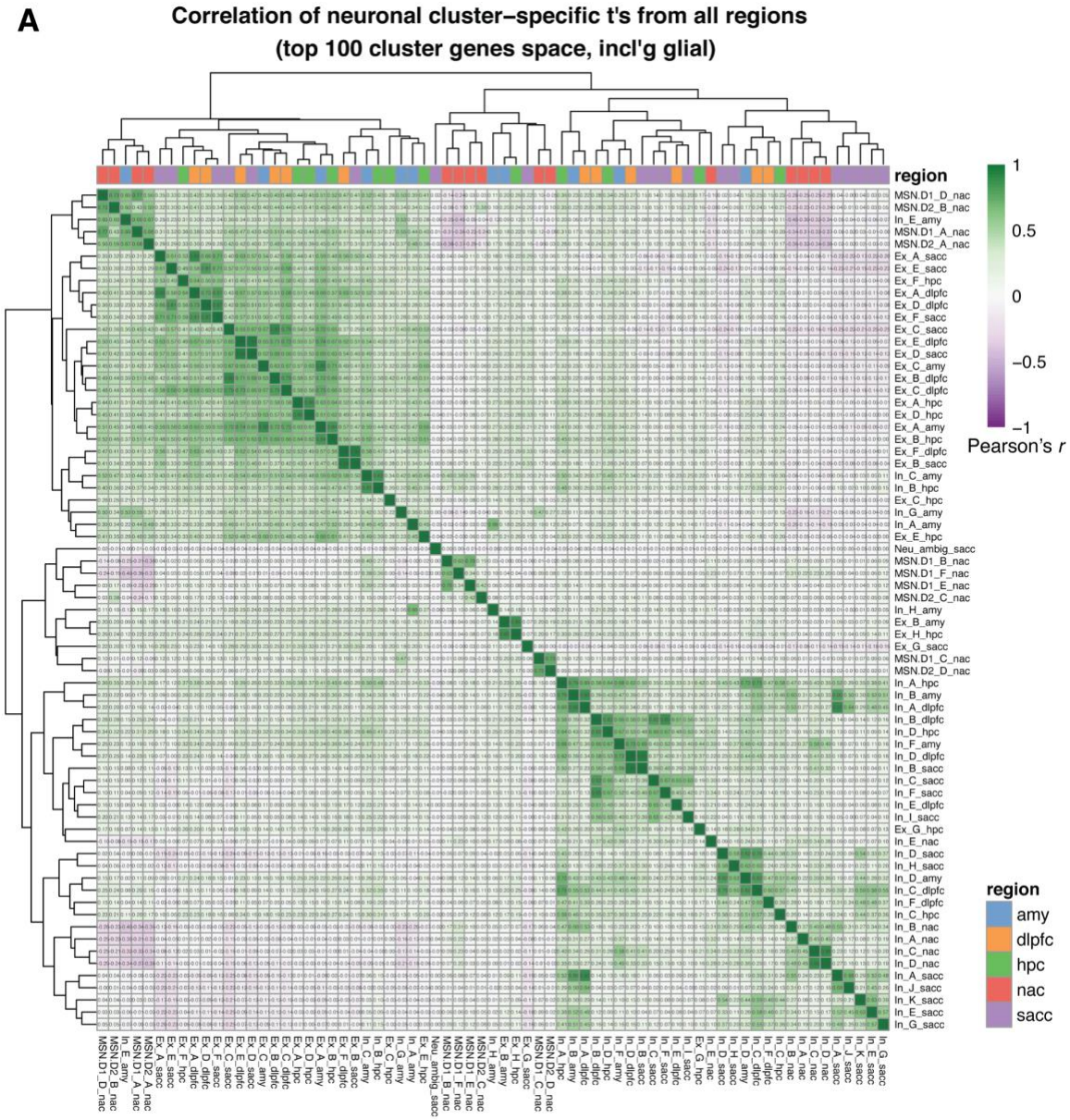


Figure 4.3 legend:

(A) Pairwise correlation of *t*-statistics, comparing the top cell class marker genes of the 107 classes (total of 3,715 genes). Here, only the 66 neuronal classes are shown. Regions are colored and labeled in lowercase as the suffix (e.g. as '_hpc' for HPC, etc.); 'Excit_' is abbreviated as 'Ex_', and 'Inhib_' as 'In_'. Scale values are of Pearson correlation coefficient (*r*).

Figure 4.4. Divergence of MSN class groups by excitatory/inhibitory signature.

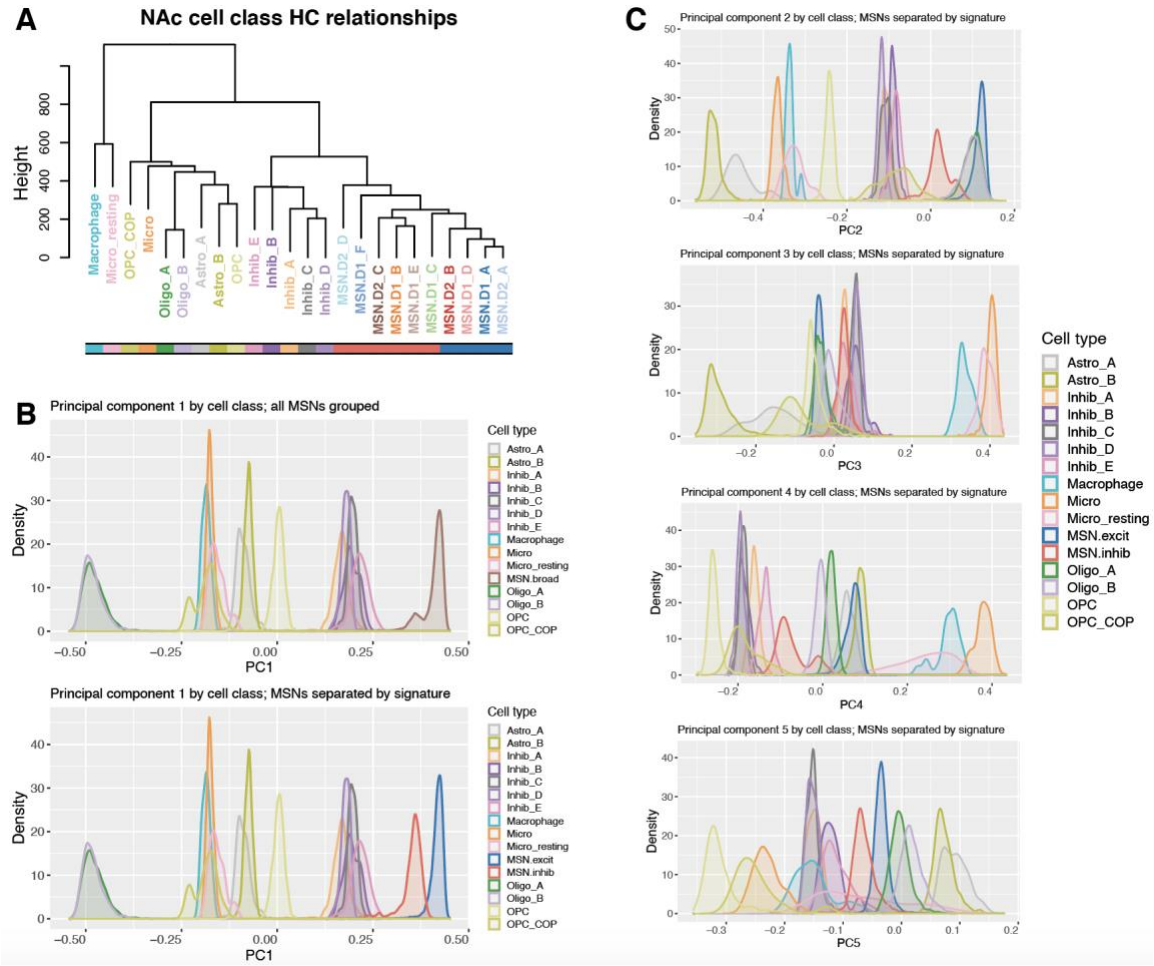


Figure 4.4 legend:

(A) Hierarchical clustering dendrogram characterizing the relationship between pseudo-bulked profiles of the 24 reported NAc cell classes in **Figure 2.1.1** (see Methods). The color bar below the labels shows separation into those D1/D2 MSNs with excitatory transcriptomic signatures (blue, 'MSN.excit': MSN.D1_A/_D and MSN.D2_A/_B) or more inhibitory (red, 'MSN.inhib': MSN.D1_B/_C/_E/_F and MSN.D2_C/_D) when compared to all other neuronal cell classe from the five regions, as per **Figure 2.3.1B**.

(B) Principal component (PC) 1, describing the largest component of variance across all NAc nuclei (see Methods), which separates NAc neuronal classes from glial cell types. MSNs further separate as per these groupings. Above: all MSNs combined (brown); below: binned via their corresponding color bars from the (A).

(C) PCs 2 through 5, conveying further separation of MSNs by their 'MSN.excit' and 'MSN.inhib' signatures.

Chapter 5. Quantifying genetic association from GWAS of neuropsychiatric and substance use phenotypes with the newly established cell type atlas

Genome-wide association studies (GWAS) have identified a plethora of genetic risk variants or loci (segregating variants in linkage disequilibrium, or LD) for common psychiatric disorders, including schizophrenia (SCZ:(Pardiñas et al., 2018; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014)), autism spectrum disorder (ASD:(Grove et al., 2019)), bipolar disorder (BIP:(Stahl et al., 2019)), major depressive disorder (MDD:(Howard et al., 2019; Wray et al., 2018)), posttraumatic stress disorder (PTSD:(Nievergelt et al., 2019)), Alzheimer's disease (AD:(Jansen et al., 2019)), and attention deficit/hyperactivity disorder (ADHD:(Demontis et al., 2019)). Additionally, a large GWAS was recently performed with 1.2 million individuals to identify the genetic risk and correlations for alcohol and tobacco use (Liu et al., 2019). Approaches have been developed to identify the biological context or relevance of the hundreds of risk loci that are often identified for a given disorder or phenotype, such as LD score regression (Finucane et al., 2015), which assesses the heritability of complex phenotypes across input categories/genomic regions and their measured LD with single nucleotide polymorphism (SNP)-level variants. Multi-marker Analysis of GenoMic Annotation (MAGMA)(de Leeuw et al., 2015) is an alternative approach that defines gene-level localization of GWAS risk, then integrates this with gene set observations, affording flexibility to assess a variety of marker lists, such as for brain region-specific snRNA-seq subcluster profiles, in two separable analyses.

We used MAGMA to identify which cell subtypes in this study harbored aggregated genetic risk for psychiatric disorders, and found robust signals across many nuclear transcriptomic profiles in each of the five profiled brain regions. As expected, many DLPFC and HPC neuronal subtypes exhibited significant effect sizes for both SCZ and BIP GWAS, spanning both excitatory and inhibitory subpopulations (**Figure 5.2**), extending and strengthening previous findings in (Bryois et al., 2020; Skene et al., 2018). After controlling with the strict Bonferroni multiple test correction across all MAGMA gene set tests (threshold *p*-value < 3.89e-5), nine of ten DLPFC neuronal cell classes, which were FDR-significant (controlling for false discovery rate < 0.05), retained significant association for both SCZ and BIP. This near-uniform pattern of SCZ and BIP risk association to neuronal DLPFC cell classes was similar in the sACC (**Figure 5.2**), though only one inhibitory cell class in the sACC retained Bonferroni significance. This suggests potential regional differences in inhibitory cell classes between the two cortical brain regions, in their manifestation of genetic risk for bipolar disorder. Contrary to these patterns of risk association in the cortex, the HPC showed disorder-specific patterns of Bonferroni-significant risk (**Figure 5.2**). For example, HPC Excit_D significantly associated with only BIP, at this threshold (*p*-value = 2.96e-5, $\boldsymbol{\beta}$ = 0.17), whereas Excit_H associated with only SCZ (*p*-value = 4.48e-13). This rare excitatory class (**Table 2.5**) additionally showed the strongest effect size ($\boldsymbol{\beta}$ = 0.34) for SCZ amongst all GWAS tested for these three regions' cell classes. Interestingly, this small, hippocampal population was most enriched for *VCAN*, *SLC17A6* (VGLUT2), and both SoxC transcription factors, *SOX4* and *SOX11*. None of the cell class profiles in these cortical or hippocampal regions retained significant signal for aggregated GWAS risk in MDD, PTSD, ADHD, AD, or ASD (other than in DLPFC OPCs), after Bonferroni correction, though there were some differential FDR-significant signals across certain cell classes

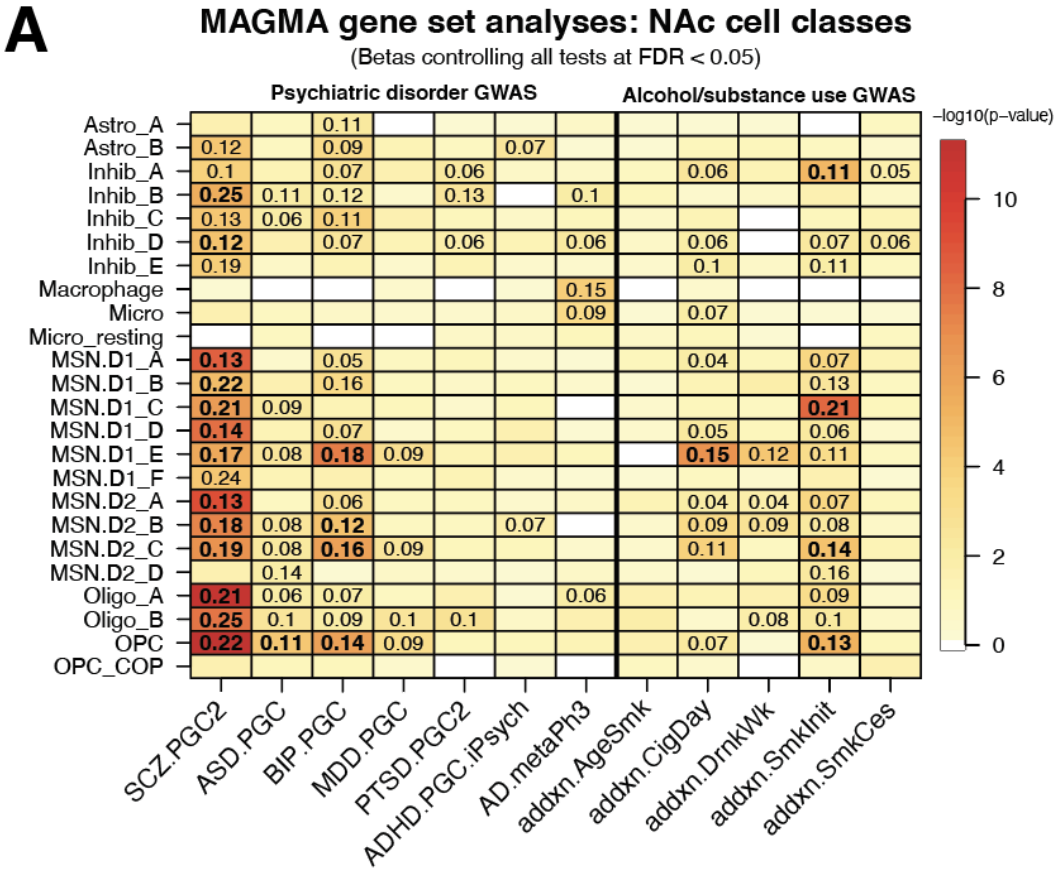
for these disorders.

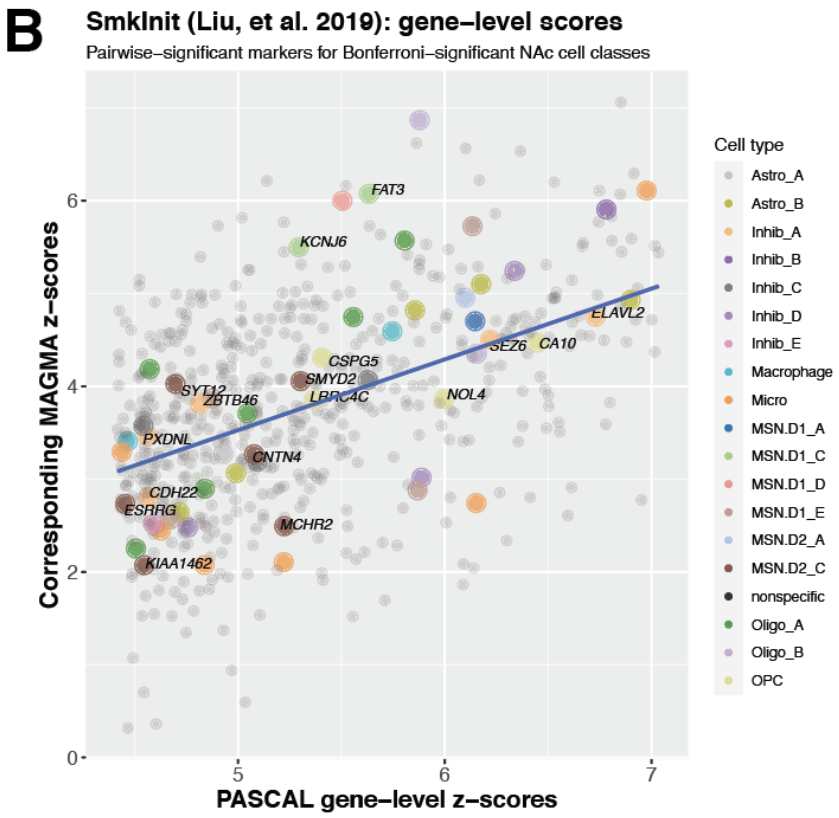
Previously, it has been shown that broad mouse striatal neuronal populations (GABAergic inhibitory neurons, Drd1, and Drd2-expressing medium spiny neurons, or MSNs) additionally associated with SCZ (Skene et al., 2018) and BIP (Bryois et al., 2020) genetic risk. We demonstrate that most of our refined MSN classes in the human NAc exhibited strong associations to SCZ risk with variable effect sizes, even at the Bonferroni correction threshold (**Figure 5.1A**), aside from D1_F and D2_D. Classes MSN.D1_E, MSN.D2_B and D2_C additionally associated with BIP at this threshold, and interestingly, the smallest D2 class, MSN.D2_D, which showed associations with neither SCZ nor BIP, was FDR-significant for association to ASD (p-value = 0.0076, $\boldsymbol{\beta}$ = 0.14). NAc GABAergic inhibitory neuronclasses Inhib_B and _D showed strong SCZ associations, but none to BIP, at the Bonferroni threshold. Within AMY, we observed significant associations in most of our neuronal classes to SCZ and BIP (**Figure 5.1C**). As with all other regions (other than DLPFC and NAc OPCs to ASD), our AMY cell classes exhibited differential, albeit weaker, FDR-significant associations to the other psychiatric disorders assessed. In summary, the NAc analyses showed complementary findings of Drd1 and Drd2-expressing striatal MSN associations with SCZ and BIP. We also dissected these mouse association signals with more relevant human GABAergic inhibitory neuron and MSN subpopulations in the NAc, and further extended this analysis to human AMY snRNA-seq-defined cell classes.

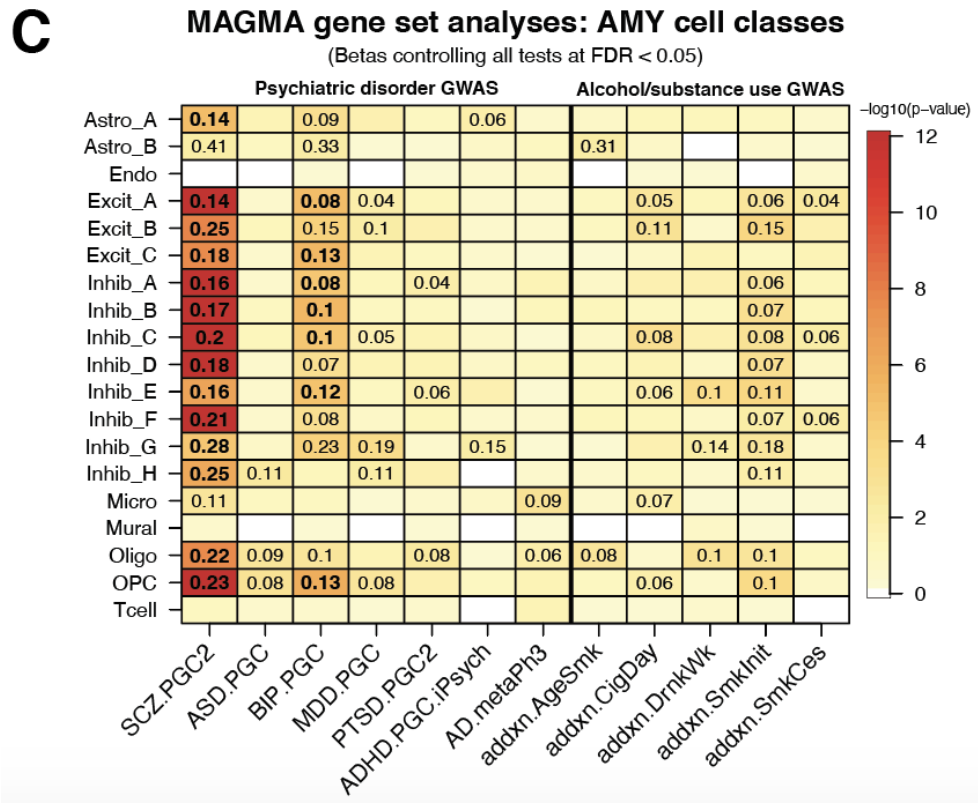
We further tested for alcohol and tobacco use GWAS (Liu et al., 2019) genetic risk associations across subcluster profiles from each of our brain regions, focusing on the subcortical regions centered in reward circuitry, the NAc and AMY, and their subcluster profiles described above. This highlighted various MSN and inhibitory subpopulations in the NAc differentially associated with genetic risk for regular smoking

behavior ('SmkInit') and heaviness of smoking 'CigDay' at the Bonferroni correction threshold (**Figure 5.1A**). In addition to their gene set markers collectively contributing to risk associations to these substance use phenotypes, we saw that many genes with strong gene-level risk scores for a given phenotype were, themselves, also markers for those and additional cell classes (e.g. 'SmkInit' in **Figure 5.1B**; see **Chapter 7** (Methods)). Although no AMY cell classes associated with any of the five substance use phenotypes at the strict Bonferroni correction threshold, they were still differentially associated with the non-'SmkInit' phenotypes. One of these classes was *DRD2*-expressing Inhib_E (**Figure 5.1C**), which was the strongest AMY cell class associated with heaviness of drinking ('DrnkWk', *p*-value = 0.00031, $\boldsymbol{\beta}$ = 0.10). As with the NAc, many other cell classes were characterized by selectively enriched genes harboring local, gene-level risk (e.g. 'DrnkWk' in **Figure 5.1D**), even though the gene set analyses did not assign strict Bonferroni-significant cell class association. Collectively, these results provide complementary human findings for genetic risk associations to those previously described for psychiatric disease, further identifying subpopulations in the NAc and AMY harboring aggregated risk for substance use behaviors.

Figure 5.1. Genetic associations of NAc and AMY cell populations with psychiatric disease and addiction phenotypes.




B
SmkInit (Liu, et al. 2019): gene–level scores
Pairwise–significant markers for Bonferroni–significant NAc cell classes
Corresponding MAGMA z–scores
PASCAL gene–level z–scores
FAT3
KCNJ6
ELAVL2
SEZ6
CA10
CSPG5
SMYD2
LRRC4C
NOL4
SYT12
ZBTB46
PXDNL
CNTN4
CDH22
ESRRG
MCHR2
KIAA1462
Cell type
Astro_A
Astro_B
Inhib_A
Inhib_B
Inhib_C
Inhib_D
Inhib_E
Macrophage
Micro
MSN.D1_A
MSN.D1_C
MSN.D1_D
MSN.D1_E
MSN.D2_A
MSN.D2_C
nonspecific
Oligo_A
Oligo_B
OPC


C
MAGMA gene set analyses: AMY cell classes
(Betas controlling all tests at FDR < 0.05)
Psychiatric disorder GWAS
Alcohol/substance use GWAS
–log10(p–value)
Astro_A
Astro_B
Endo
Excit_A
Excit_B
Excit_C
Inhib_A
Inhib_B
Inhib_C
Inhib_D
Inhib_E
Inhib_F
Inhib_G
Inhib_H
Micro
Mural
Oligo
OPC
Tcell
SCZ.PGC2
ASD.PGC
BIP.PGC
MDD.PGC
PTSD.PGC2
ADHD.PGC.iPsych
AD.metaPh3
addxn.AgeSmk
addxn.CigDay
addxn.DrnkWk
addxn.SmkInit
addxn.SmkCes

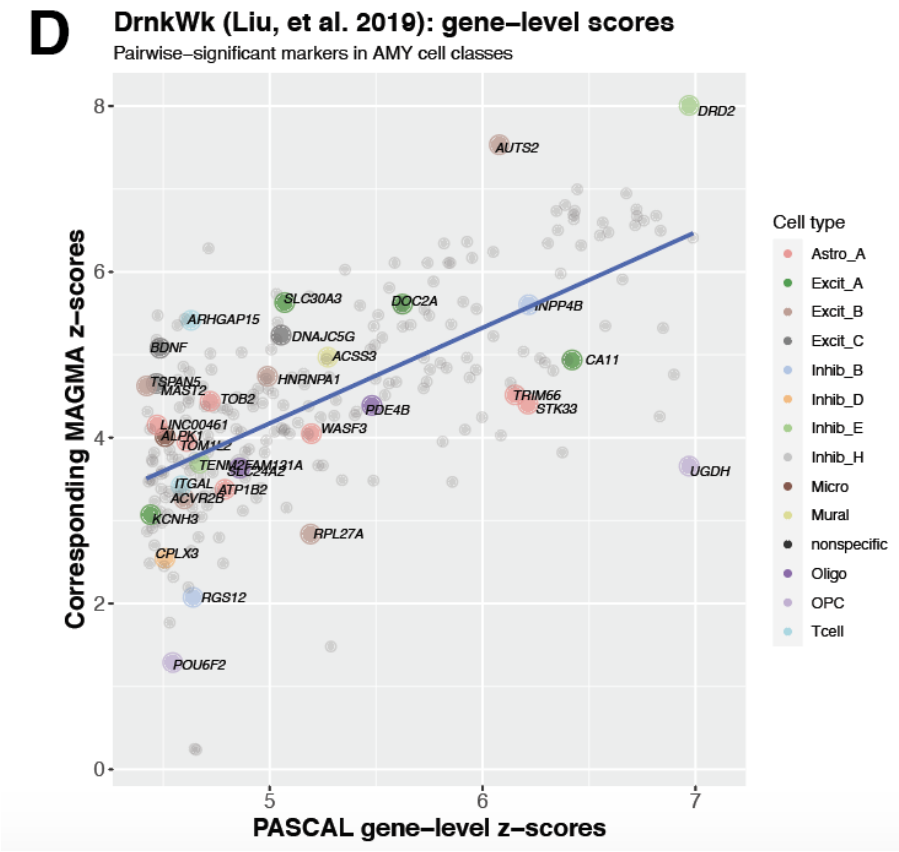


Figure 5.1 legend:

(A) MAGMA associations of 12 GWAS for each of 24 cell classes profiled in human NAc. See below for abbreviations used.

(B) MAGMA-computed, gene-level *z*-scores, compared to their reported significant PASCAL scores, for 'SmkInit' from (Liu et al., 2019). Genes are colored if they were statistically significant for pairwise marker tests, for the corresponding NAc cell class, and additionally labeled if that cell class was Bonferroni-significant in MAGMA association with the phenotype.

(C) MAGMA associations for each of 16 cell classes profiled in human AMY.

(D) Same as (B) but for 'DrnkWk' and colored/labeled by AMY pairwise cell class markers (no MAGMA-gene set analysis result restriction).

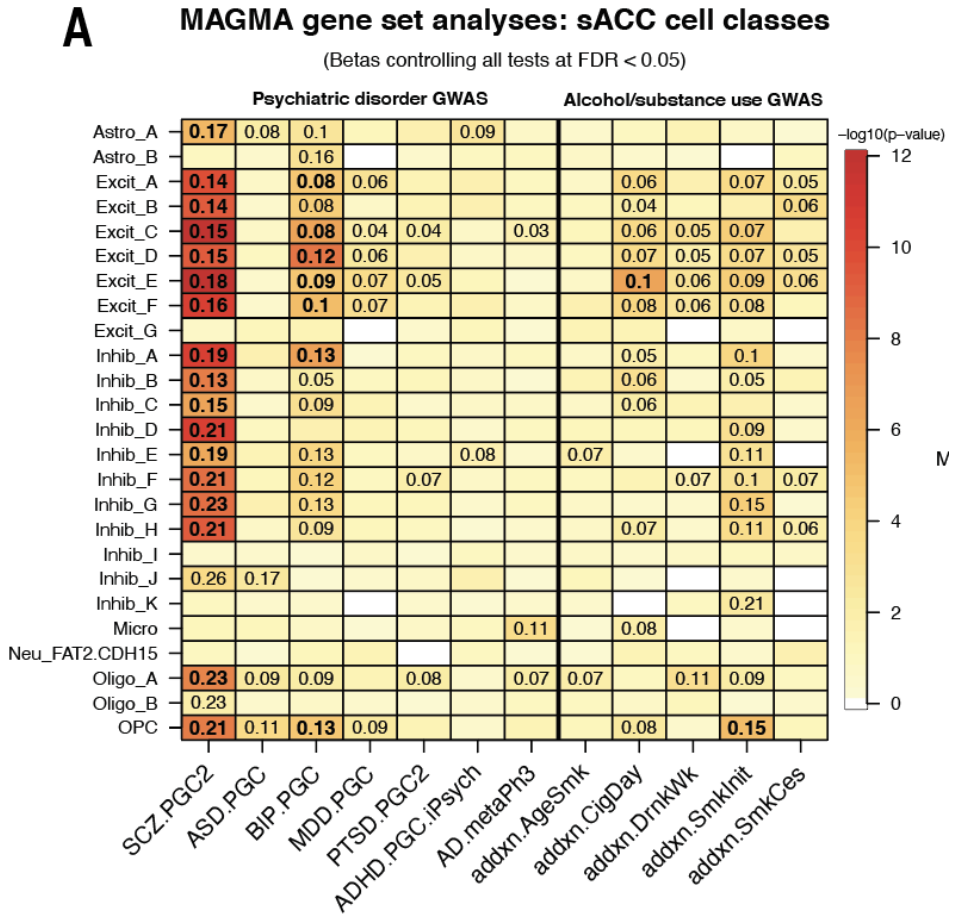
For the MAGMA heatmaps: Displayed numbers are the effect size (***β***) for significant associations (controlled for false discovery rate, FDR < 0.05), on a *Z* (standard normal) distribution. Bolded numbers are those that additionally satisfy a strict Bonferroni correction threshold of $p < 3.89e\text{-}5$. Heatmap is colored by empirical -log10(*p-value*) for each association test.

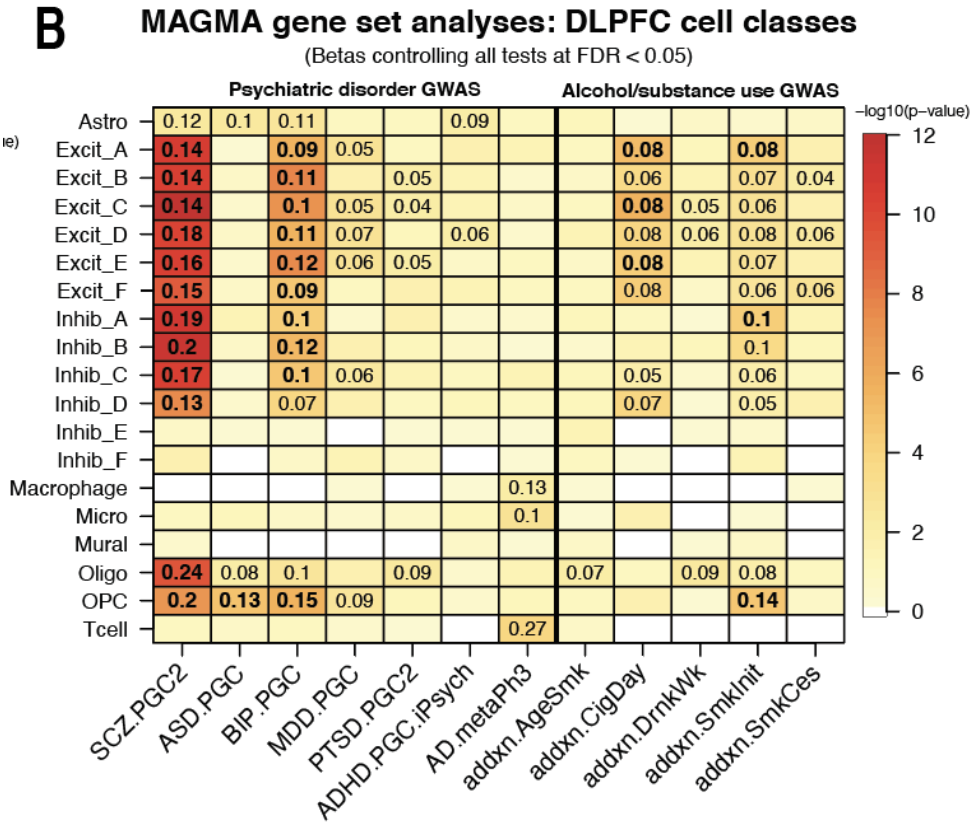
Abbreviations:

SCZ: schizophrenia, ASD: autism spectrum disorder, BIP: bipolar disorder, MDD: major depressive disorder, PTSD: posttraumatic stress disorder, ADHD: attention deficit/hyperactivity disorder, AD: Alzheimer's disease. The suffix for these (e.g. '.PGC2') reference the specific study (see Methods).

For the (Liu et al., 2019) phenotypes, 'addxn.': "addiction"; 'AgeSmk': age of initiation of regular smoking, 'CigDay': number of cigarettes per day, 'DrnkWk': number of drinks per week, 'SmkInit': whether regular smoking was ever reported (binary variable), 'SmkCes': if so, had an individual stopped smoking (binary variable).

Figure 5.2. Genetic associations for HPC and cortical regions with psychiatric disease and addiction phenotypes.




B
MAGMA gene set analyses: DLPFC cell classes
(Betas controlling all tests at FDR < 0.05)
Psychiatric disorder GWAS
Alcohol/substance use GWAS
−log10(p−value)
12
10
8
6
4
2
0
Astro 0.12 0.1 0.11 0.09
Excit_A 0.14 0.09 0.05 0.08 0.08
Excit_B 0.14 0.11 0.05 0.06 0.07 0.04
Excit_C 0.14 0.1 0.05 0.04 0.08 0.05 0.06
Excit_D 0.18 0.11 0.07 0.06 0.08 0.06 0.08 0.06
Excit_E 0.16 0.12 0.06 0.05 0.08 0.07
Excit_F 0.15 0.09 0.08 0.06 0.06
Inhib_A 0.19 0.1 0.1
Inhib_B 0.2 0.12 0.1
Inhib_C 0.17 0.1 0.06 0.05 0.06
Inhib_D 0.13 0.07 0.07 0.05
Inhib_E
Inhib_F
Macrophage 0.13
Micro 0.1
Mural
Oligo 0.24 0.08 0.1 0.09 0.07 0.09 0.08
OPC 0.2 0.13 0.15 0.09 0.14
Tcell 0.27
SCZ.PGC2
ASD.PGC
BIP.PGC
MDD.PGC
PTSD.PGC2
ADHD.PGC.iPsych
AD.metaPh3
addxn.AgeSmk
addxn.CigDay
addxn.DrnkWk
addxn.SmkInit
addxn.SmkCes

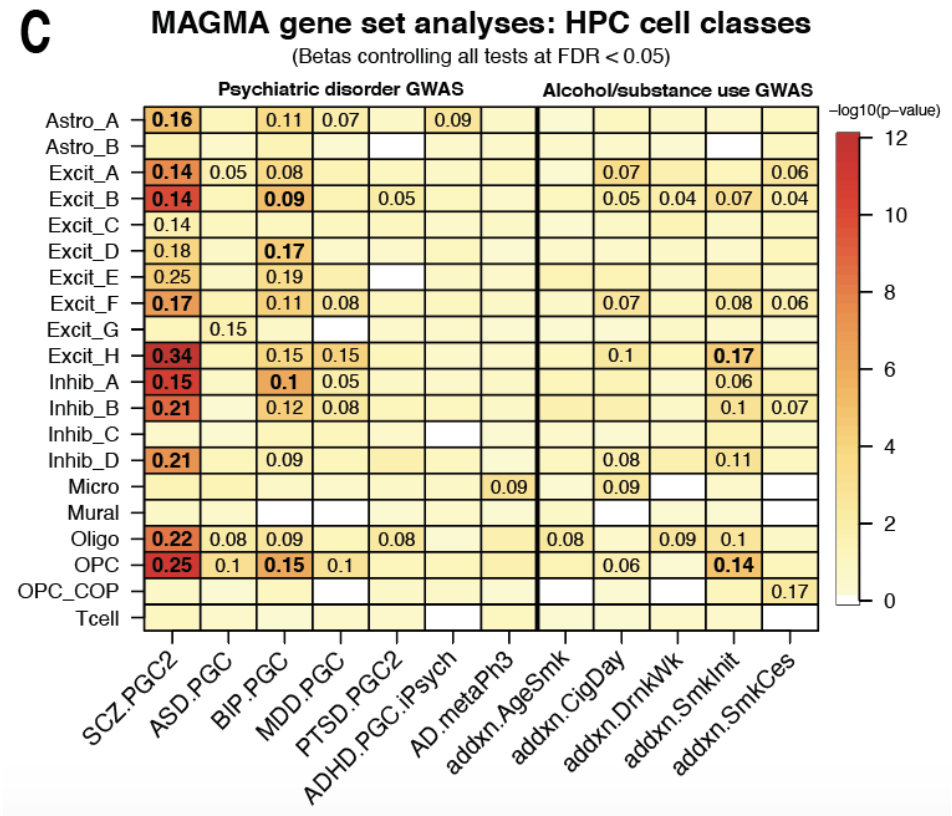

C
MAGMA gene set analyses: HPC cell classes
(Betas controlling all tests at FDR < 0.05)
Psychiatric disorder GWAS
Alcohol/substance use GWAS
−log10(p−value)
12
10
8
6
4
2
0
Astro_A 0.16 0.11 0.07 0.09
Astro_B
Excit_A 0.14 0.05 0.08 0.07 0.06
Excit_B 0.14 0.09 0.05 0.05 0.04 0.07 0.04
Excit_C 0.14
Excit_D 0.18 0.17
Excit_E 0.25 0.19
Excit_F 0.17 0.11 0.08 0.07 0.08 0.06
Excit_G 0.15
Excit_H 0.34 0.15 0.15 0.1 0.17
Inhib_A 0.15 0.1 0.05 0.06
Inhib_B 0.21 0.12 0.08 0.1 0.07
Inhib_C
Inhib_D 0.21 0.09 0.08 0.11
Micro 0.09 0.09
Mural
Oligo 0.22 0.08 0.09 0.08 0.08 0.09 0.1
OPC 0.25 0.1 0.15 0.1 0.06 0.14
OPC_COP 0.17
Tcell
SCZ.PGC2
ASD.PGC
BIP.PGC
MDD.PGC
PTSD.PGC2
ADHD.PGC.iPsych
AD.metaPh3
addxn.AgeSmk
addxn.CigDay
addxn.DrnkWk
addxn.SmkInit
addxn.SmkCes

Figure 5.2 legend:

(A) MAGMA associations for each of 25 cell classes profiled in sACC,

(B) 19 DLPFC cell classes, and

(C) 20 HPC cell classes.

See **Figure 5.1** legend for abbreviations.

Heatmap is colored by empirical -log10(*p-value*) for each association test. Displayed numbers are the effect size (***β***) for significant associations (controlled for false discovery rate, FDR < 0.05), on a *Z* (standard normal) distribution. Bolded numbers are those that additionally satisfy a strict Bonferroni correction threshold of $p < 3.89e\text{-}5$.

Chapter 6. Takeaways and expectation for future work

In this study, we used snRNA-seq to profile five human brain regions within the ventral striatum (NAc), limbic system (AMY and HPC), and two cortical subregions (sACC, DLPFC) that are interconnected within the larger reward circuitry. While single-nucleus transcriptomic profiling in the postmortem human brain has rapidly accelerated, most efforts to date (Mathys et al., 2019; Nagy et al., 2020; Velmeshev et al., 2019) have focused on cortical regions and the HPC. While efforts to generate a diverse human cell type atlas at the single-cell level are underway (Han et al., 2020), the landscape of specialized molecular cell types across the complex human brain remains largely unexplored at this level of number of samples and regional diversity. This study is the first, to our knowledge, to systematically profile and compare across multiple interconnected cortical and sub-cortical human brain areas, selected for their function and association with risk for neuropsychiatric disorders and addiction. We placed special emphasis on analyses in the NAc and the AMY given their roles in emotional processing and reward signaling, and the lack of any human snRNA-seq reference data in these regions. While this study was performed in neurotypical donors, the strong cell type-specific associations to genetic risk for these disorders provide important information about disease etiology. This link to genetic risk is important, given that differential gene expression identified in case-control studies of postmortem tissue are difficult to interpret as signals may more likely represent consequences, rather than causes, of these disorders (Collado-Torres et al., 2019; Jaffe et al., 2020). More generally, understanding the transcriptomic architecture and cell type composition across the normal human brain is crucial to understanding the etiology of disease and the molecular pathology observed

in postmortem tissues, in order to identify and prioritize potential novel disease targets. Our study is a significant contribution as it demonstrates differential enrichment of disease risk in snRNA-seq-defined cell populations across multiple brain regions, including the NAc and AMY, which have not yet been profiled at the single-cell/-nucleus level in the human brain.

The NAc is a central hub for reward signaling, and altered function in circuits encompassing the NAc is implicated in a number of psychiatric disorders as well as drug addiction. Hence, we sought to define molecular profiles for NAc cell types, with a specific focus on functionally dichotomous subtypes of DRD1- and DRD2-expressing MSNs. Consistent with studies that used single-cell sequencing to profile the mouse striatum, including ~1000 striatal cells in each study (Gokce et al., 2016; Stanley et al., 2020), we identified several discrete subpopulations of D1 and D2-expressing MSNs in human NAc. In contrast to these studies, we identified six discrete classes of D1-MSNs, and four distinct D2-MSN classes, which we validated by smFISH. Several reasons may explain why we identified different discrete D1/D2 subpopulations, including differences in species (human vs. mouse), region (NAc-specific vs. striatum-wide), sample preparation (whole cells vs. nuclei), number of MSNs profiled (about 10x greater in our dataset) and single cell technology employed (10x Genomics Chromium vs. SMART-Seq v2). However, in agreement with these studies, we also observed co-expression of *DRD1* and *DRD2* in a small subset of MSNs. While these dual-expressing neurons did not emerge as their own cluster, they were largely found in the D1_E subpopulation (**Figure 2.1.1B; Figure 2.1.2**). Interestingly, this cluster showed the strongest enrichment of genes associated with psychiatric and substance use phenotypes, indicating that this particular subpopulation might be especially vulnerable to dysfunction in these disorders. Indeed, among D1 subtypes, D1_E MSNs (along with *SST*+, *NPY*+

interneuron class Inhib_E, and to some degree MSN.D2_D) show selective expression of *CRHR2*, a gene encoding corticotropin releasing hormone receptor 2, suggesting that they may be particularly susceptible to the effects of corticotropin-releasing hormone (CRH), which is released and mediates the physiological and behavioral response to stress, modulating several neurotransmitter systems, including dopamine release (Bonfiglio et al., 2011; Payer et al., 2017). Given that dysfunction of the CRH system has been associated with many psychiatric disorders, including depression, anxiety, and PTSD (Claes, 2004), understanding which cell types express CRH receptors may aid in more specific targeting of the stress axis for therapeutic developments.

Similar to Gokce et al., we also observed promiscuous expression of "typical" D1 and D2 neuropeptide marker genes (*TAC1* and *PENK*, respectively) in both D1 and D2 MSN subpopulations, providing further evidence that these classic markers may not be as selectively expressed as previously understood. Future studies using spatial transcriptomic approaches will be important to clarify whether *TAC1*-expressing D1 and D2 MSN subpopulations show topographical organization in the NAc core vs. shell. Anatomical location may explain differences in *TAC1* and *PENK* expression in specific MSN subpopulations, as it is well established that specific neuropeptides are expressed in a spatial gradient across the core and shell (Prensa et al., 2003; Salgado and Kaplitt, 2015; Stanley et al., 2020; Voorn et al., 1989). To better interpret clinical implications of studies that focus on circuitry encompassing the NAc in animal models, further understanding of similarities and differences across species for cell types that contribute to NAc function are important. While many cell populations were conserved between rat and human NAc (Savell et al., 2020), we did observe differences in specific MSN subpopulations, which may indicate unique molecular features between analogous MSN subpopulations and/or the existence of divergent MSN subclasses, as exemplified by the lack of a specific human MSN subpopulations with strong correlation to rat '*Grm8-MSN*'

subpopulation (**Figure 3.1F**). However, given the small positive correlations measured with human D2_B (r = 0.40) and various D1 classes, it is possible that this *Grm8*-expressing population encompasses a broader species-equivalent of these less abundant D1/D2 classes, which express variable *GRM8* (**Figure 2.1.5**). We also were unable to identify a population of cholinergic interneurons. While cholinergic interneurons are thought to be more abundant in the human neostriatum compared to the rodent, where they only account for ~0.3% of neurons (Graveland et al., 1985; Rymar et al., 2004; Tepper and Bolam, 2004), it is likely that the low rate of sampling and this population's relative rarity accounts for this lack of observation.

In addition to profiling NAc cell types, we also generated a molecular taxonomy of human amygdala cell types. We identified 11 distinct neuronal subpopulations as well as accompanying gene marker annotations, including *NRN1* (neuritin) and *NPTX1* (neuronal pentraxin 1) for the AMY Excit_A subcluster. Neuritin is a neurotrophic factor which modulates neurite outgrowth and plasticity (Yao et al., 2018), whereas neuronal pentraxin 1 regulates neuron excitability via synapse density (Figueiro-Silva et al., 2015). Additionally, the highest *SLC17A6* (VGLUT2)-expressing subcluster, Excit_B, specifically expressed high levels of *VCAN* (Versican) amongst other neuronal subpopulations, which has multiple isoforms exhibiting different mechanisms for synaptic regulation (Horii-Hayashi et al., 2008). Among the diverse set of inhibitory subpopulations in the AMY, the stress modulator *CRH* was specifically enriched in Inhib_B. Top markers in the AMY Inhib_C subcluster included *NPFFR2* (Neuropeptide FF Receptor 2) and *TLL1* (Tolloid-Like 1), which are both associated with glucocorticoid signaling and the response to stress (Lin et al., 2016; Tamura et al., 2005). Comparing human AMY subcluster profiles to data from the mouse medial amygdala (MeA;(Chen et al., 2019)), we found that Inhib_C and its corresponding population in mouse (MeA 'N.8' subcluster, **Figure 3.1C**) were the most strongly correlated neuronal subpopulations.

While *Tll1* expression was notably absent in mouse MeA, *Npffr2* and other top MeA 'N.8' marker genes were shared with Inhib_C (**Figure 3.2**). These insights highlight the importance of deriving reference snRNA-seq datasets across the human brain, as molecular gene markers may not be shared across species between analogous neuronal subpopulations.

Integrating the transcriptomic profiles across our total of 107 reported cell classes across the NAc, AMY, sACC, DLPFC, and HPC showed patterns of expected similarity, particularly amongst glial cell classes. However, this analysis also revealed some within-cell-type substructure that highlight unique relationships between these regionally-defined subpopulations. As an example, we noted a cluster of strongly correlated astrocyte classes from each of the five brain regions (**Figure 4.1**), and a small cluster made up of the sACC 'Astro_B', HPC 'Astro_B', and NAc 'Astro_A'. Interestingly these latter three represent the smaller of two astrocyte classes from their corresponding regionally-defined cell type catalog (160, 234, and 99 nuclei, respectively). The most unique astrocyte population, 'Astro_B' from AMY (83 nuclei), appeared to be a metabolically low class of astrocytes, suggested by its low distribution of total UMI capture (**Figure 4.2**), even though it is represented across all donors (**Figure 6.1**). These observations complement recent work focused on surveying astrocyte diversity in the mammalian brain (Batiuk et al., 2020), or across other glial cell types and their developmental origin (Chamling et al., 2021; Masuda et al., 2020). We additionally observed a variety of shared neuronal cell classes across regions. Most strikingly, this revealed that despite their broad D1/D2 classification, MSNs separate into divergent groups that exhibit a more excitatory versus a more inhibitory signature, and these respective groups of MSN classes are generally negatively correlated (**Figure 4.3; Figure 4.4**). A comprehensive characterization of the molecular pathways driving this

divergence at the transcriptome level, in addition to how this influences their potential roles within the existing micro- and macro-circuitry, is beyond the scope of this work. Taken together, however, this full integration of single-nucleus profiles across regions demonstrates a practical method of assessing cell type relationships and elucidating patterns across the cell type manifold, while maintaining the molecular resolution of transcriptomic signatures related to their tissues of origin.

We finally used the snRNA-seq data from the five profiled regions to ask whether specific cell classes harbored aggregate genetic risk for various neuropsychiatric disorders and/or features of substance use. We confirmed previous findings by identifying strong associations for neuronal subpopulations in the DLPFC and HPC with both schizophrenia (SCZ) and bipolar disorder (BIP) (Bryois et al., 2020; Skene et al., 2018), and significantly extended these findings by providing associations with specific sACC excitatory and inhibitory populations (**Figure 5.2**). Additionally, we not only confirmed previously observed associations to broad striatal populations defined in the mouse, but showed, for the first time, that individual populations of dopaminoceptive (DRD1/2) neurons in the human NAc may be differentially associated with SCZ and BIP (**Figure 5.1**). We also found that specific subpopulations of GABAergic inhibitory neurons in the human AMY were preferentially associated with SCZ that weren't significantly associated with BIP. These observations highlight a potential role for these subcortical brain regions in mediating genetic risk for SCZ and BIP.

As both the NAc and AMY play critical roles in reward signaling, we also evaluated enrichment of genetic risk for addiction or substance use behaviors (Liu et al., 2019). Intriguingly, the genetic risk for adopting regular smoking associated more broadly across most neuronal populations, whereas other phenotypes assessed in this addiction GWAS showed more preferential associations to certain subpopulations. This

suggests that the risk for adopting addictive-like behaviors might affect these brain regions more broadly than specific features of addiction (**Figure 5.1A/C**). With regard to the other features, the MSN.D1_E subpopulation significantly associated with genetic risk for heaviness of smoking ('CigDay') and drinking ('DrnkWk'), after Bonferroni and FDR correction, respectively. As a top marker for this subpopulation was *CRHR2*, this might be a key population in understanding these features of addiction. Indeed, many rodent studies have implicated CRH receptors in alcohol consumption and alcohol dependence (Heilig and Koob, 2007; Yong et al., 2014). Finally, though no neuronal AMY subpopulations met our strict Bonferroni threshold for association, two neuronal classes drew our attention, due to association with multiple phenotypes, including AMY 'Inhib_C' and 'Inhib_E'. Due to their marker expression of *NPFFR2*/*TLL1* and serotonin receptors *HTR4*/*HTR2C*, respectively (data not shown; **see Chapter 7 (Methods)**), these GABAergic classes might be of interest in understanding amygdalar circuits underlying genetic risk for substance use. From these analyses, we surveyed our diversity of neuronal subpopulations profiled in the NAc and AMY for their clinical relevance in psychiatric disease and addiction behaviors. Additionally, we have extended such analyses for these regions, which have formerly only been performed on cell-type profiles defined in murine models (Bryois et al., 2020; Skene et al., 2018) to their relevant human context, and with increased resolution of molecularly-defined subpopulations. Finally, we narrowed down on those subpopulations manifesting the greatest genetic risk, potentially highlighting some neuronal subclasses mediating certain substance use behaviors.

While we identified and characterized a diversity of robust neuronal subpopulations with our analytical pipelines for this study, we recognize that our sample sizes may not fully capture all cell types or subpopulations, such as striatal cholinergic

interneurons, mentioned above. The most direct evidence for this is that there remains some bias in donor makeup of certain subpopulations (**Tables 2.1-2.5; Figure 6.1**), keeping NeuN enrichment for a subset of samples in mind. However, despite steps to mitigate the impact of the small input for our sample processing protocol (see Methods), we expect some degree of sampling bias since cell type makeup is not expected to be homogeneous within a single region. For example, the NAc core or shell have different functional properties, and differ in regards to their afferent and efferent connections, and thus differences in cell composition across these two subregions is expected (Heimer et al., 1991; Li et al., 2018; Zahm and Heimer, 1993). Integration of spatial transcriptomic technologies with snRNA-seq data in these regions (Maynard et al., 2021) will help resolve expected heterogeneity across these adjacent subregions. Further, while many groups have recently begun to identify sex-specific differences in specific roles or hormonal responses of neuronal subpopulations (Cao et al., 2018; Chen et al., 2019), we believe our study remains underpowered to potentially recapitulate these observations. It is noted that capture of certain non-neuronal cell classes was observed in only select brain regions, namely endothelial cells, as these were only identified in the AMY. On the other hand, we identified mural cells (comprising of pericytes and vascular smooth muscle cells) in the AMY, DLPFC, and HPC. With these sample sizes being still limited, and an emphasis on NeuN enrichment in a subset of our data, we believe that these smaller, stromal populations were inherently not captured by our protocols in some regions. On the other hand, we highlight that most of the cell classes we report are made up of their own set of subclusters (see Methods and *Data and code availability*).

Another caveat to these snRNA-seq data is the lack of gene expression information from the cytosolic compartment, such as the neuropil. This is an important caveat given that synaptic signaling is implicated in neuropsychiatric disorders, and gene

products localized to the synapse are enriched for SCZ genetic risk (Skene et al., 2018). In addition, mRNA from some expected marker genes, e.g. *PVALB*, may be preferentially localized to the cytosol, as demonstrated with smFISH for the *GAD1*+ interneuron 'Inhib_C/_D' classes in the NAc (**Figure 2.1.6**). However, this seems to be cell population and/or area-specific, with regards to the transcriptional and nuclear-export dynamics of the respective cell population, as *PVALB* was highly expressed in some DLPFC subpopulations (data not shown; see *Data and code availability*). These and observations by others thus emphasize that snRNA-seq will not capture the full transcriptomic profile of cell populations, including activation-induced or disease-associated molecular changes restricted to the cytosol (Thrupp et al., 2020). However, as we previously demonstrated (Maynard et al., 2021), snRNA-seq-defined cell populations can be registered to spatial transcriptomic data, which does retain such information, for further characterization of transcriptomic profiles.

In summary, we used snRNA-seq to profile five human brain regions with roles in the reward circuitry. We defined transcriptomic profiles for 107 regionally-defined cell type classes and characterized the architecture of molecular relationships across these brain regions. We finally identified associations with genetic risk for neuropsychiatric disorders and substance use phenotypes in unique neuronal subpopulations in the NAc and AMY. This study provides a significant step towards constructing a single-nucleus transcriptomic atlas of the human brain and illustrates the utility of this type of data in understanding the diversity of cell populations, as well as their roles in biology and disease.

Figure 6.1. Distribution of regionally-defined cell classes by donor.

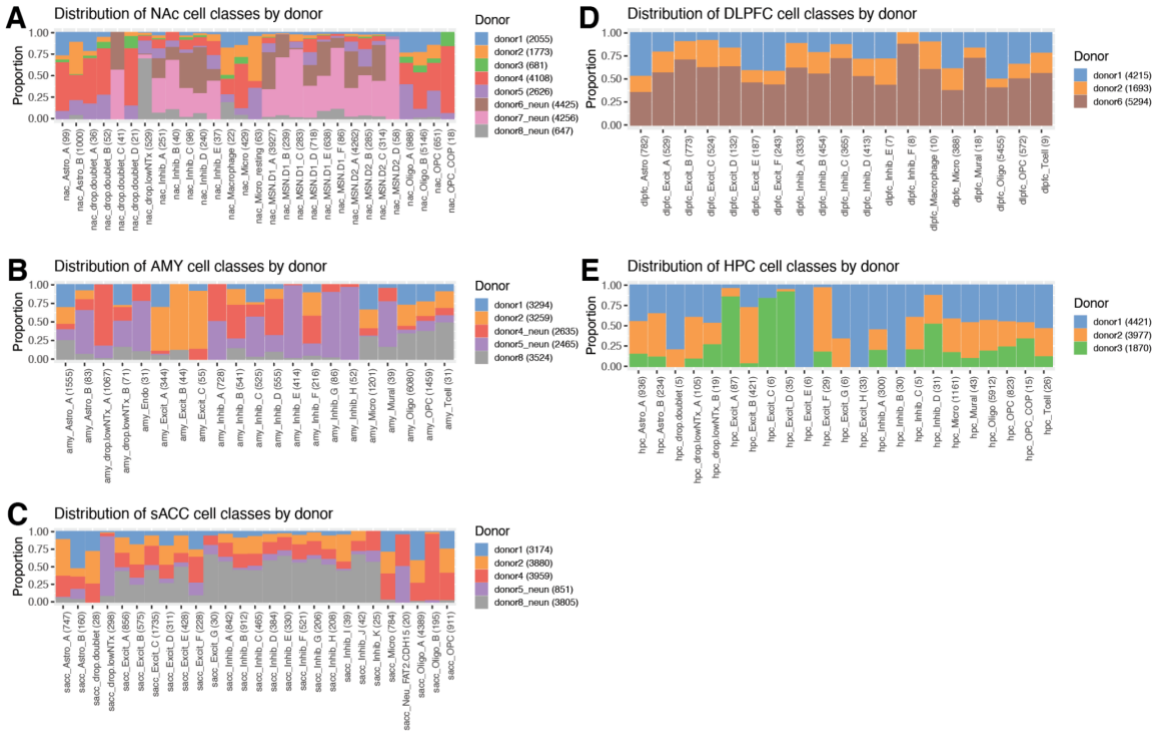


Figure 6.1 legend: Distribution of all 107 regionally-defined cell classes and their proportions by donor for

(A) NAc, **(B)** AMY, **(C)** sACC, **(D)** DLPFC, and **(E)** HPC.

Included are each region's technical artifact-driven clusters (total of 12), which are annotated with the 'drop.' prefix and: 'doublet', if they were flagged for high median 'doubletScore' (see Methods), in addition to expressing multiple broad cell class markers (not shown); or 'lowNTx' (for "low number of transcripts"): clusters driven by low quality nuclei or those that captured ambient transcripts/UMIs, yet passed nuclei calling (Methods)

Chapter 7. Methods

7.1 Experimental methods

Post-mortem human tissue

Post-mortem human brain tissue from eight neurotypical donors of European ancestry from age 40 to 69 (**Table 1**) was obtained by autopsy from the Office of the Chief Medical Examiner for the State of Maryland under State of Maryland Department of Health and Mental Hygiene Protocol 12-24. Clinical characterization, diagnoses, and macro- and micro-scopic neuropathological examinations were performed on all samples using a standardized paradigm, and subjects with evidence of macro- or micro-scopic neuropathology were excluded. Details of tissue acquisition, handling, processing, dissection, clinical characterization, diagnoses, neuropathological examinations, RNA extraction and quality control measures have been described previously (Lipska et al., 2006). Dorsolateral prefrontal cortex (DLPFC, n=3) and hippocampus (HPC, n=3) tissue was microdissected using a hand-held dental drill as previously described (Collado-Torres et al., 2019). The subgenual anterior cingulate cortex (sACC, n=5) was dissected under visual guidance from the medial aspect of the forebrain at the level of the rostrum of the corpus callosum. Dissections were performed ventral to the corpus callosum, and dorsal to the orbital frontal cortex (BA11). Medially it was bounded by the interhemispheric fissure, while laterally it was bounded by the corona radiata/centrum semiovale. For the amygdala (AMY, n=5), a block containing the structure was dissected under visual guidance at the level of its maximal size, taken from a 1 cm thick slab of one hemisphere, and sectioned in the coronal plane. The amygdala block was chosen by visual inspection at a level that contained the maximal number of subnuclei.

Landmarks for selection of the amygdala block included presence of the internal and external segments of the globus pallidus, the anterior commissure, and optic tract. The block containing the nucleus accumbens was taken from a 1 cm thick slab of one hemisphere, and sectioned in the coronal plane. The nucleus accumbens (NAc, n=8) block was chosen at a level where the putamen and caudate are joined by the accumbens at the ventral aspect of the striatum, with clear striations separating the putamen from the caudate. Additional landmarks include the presence of the anterior aspect of the temporal lobe and the claustrum.

snRNAseq data generation

We performed single-nucleus RNA-seq (snRNA-seq) on 24 samples from 3-8 individual donors, per region (*n*=3 DLPFC, *n*=3 HPC, *n*=5 AMY, *n*=5 sACC, *n*=8 NAc), using 10x Genomics Chromium Single Cell Gene Expression V3 technology (Zheng et al., 2017). Nuclei were isolated using a “Frankenstein” nuclei isolation protocol developed by Martelotto *et al.* for frozen tissues (Habib et al., 2016, 2017; Hu et al., 2017; Lacar et al., 2016; Lake et al., 2016). Briefly, ~40mg of frozen, ground tissue was homogenized in chilled Nuclei EZ Lysis Buffer (MilliporeSigma #NUC101) using a glass dounce with ~15 strokes per pestle. Homogenate was filtered using a 70µm-strainer mesh and centrifuged at 500 x g for 5 minutes at 4°C in a benchtop centrifuge. Nuclei were resuspended in the EZ lysis buffer, centrifuged again, and equilibrated to nuclei wash/resuspension buffer (1x PBS, 1% BSA, 0.2U/µL RNase Inhibitor). Nuclei were washed and centrifuged in this nuclei wash/resuspension buffer three times, before labeling with DAPI (10µg/mL) or propidium iodide (PI) (depending on processing batch). For 3 NAc, 2 sACC, and 2 AMY samples from individual donors, nuclei were additionally labeled with Alexa Fluor 488-conjugated anti-NeuN (MilliporeSigma cat. #MAB377X), at

1:1000 in the same wash/resuspension buffer, for 30 minutes on ice, to facilitate enrichment of neurons during fluorescent activated cell sorting (FACS). Samples were then filtered through a 35µm-cell strainer and sorted on a BD FACS Aria II Flow Cytometer (Becton Dickinson) at the Johns Hopkins University Sidney Kimmel Comprehensive Cancer Center (SKCCC) Flow Cytometry Core, or Bio-Rad S3e Cell Sorter (depending on processing batch) into 10X Genomics reverse transcription reagents. Gating criteria hierarchically selected for whole, singlet nuclei (by forward/side scatter), G_0/G_1 nuclei (by DAPI or PI fluorescence), and NeuN-positive cells for the respective NeuN-enriched samples. A “null” sort of nuclei into the wash buffer was additionally performed from the same preparation, for quantification of nuclei concentration and to ensure that sorted nuclei were intact and free of debris. For each sample, approximately 8,500 single nuclei were sorted directly into 25.1µL of reverse transcription reagents from the 10x Genomics Single Cell 3’ Reagents kit (without enzyme). The 10x Chromium process was performed and libraries prepared, according to manufacturer’s instructions (10x Genomics), and finally sequenced on the Next-seq (Illumina) at the Johns Hopkins University Transcriptomics and Deep Sequencing Core.

RNAscope single molecule fluorescent in situ hybridization (smFISH)

Fresh frozen NAc from two independent donors was sectioned at 10µm and stored at -80°C. *In situ* hybridization assays were performed with RNAscope technology utilizing the RNAscope Fluorescent Multiplex Kit V2 and 4-plex Ancillary Kit (Cat # 323100, 323120 ACD, Hayward, California) according to the manufacturer's instructions. Briefly, tissue sections were fixed with a 10% neutral buffered formalin solution (Cat # HT501128 Sigma-Aldrich, St. Louis, Missouri) for 30 minutes at room temperature (RT), series dehydrated in ethanol, pretreated with hydrogen peroxide for 10 minutes at RT,

and treated with protease IV for 30 minutes. Sections were incubated with 5 different probe combinations to assess MSN and inhibitory neuron subtypes: 1) "Square": *DRD1, TAC1, RXFP2, GABRQ* (Cat 524991-C4, 310711-C3, 452201, 483171-C2, ACD, Hayward, California); 2) "Circle": *DRD1, TAC1, CRHR2, RXFP1* (Cat 524991-C4, 310711-C3, 469621, 422821-C2); 3) "Triangle": *DRD1, DRD2, TAC1, PENK* (Cat 524991-C4, 553991, 310711-C2, 548301-C3); 4) "Star": *DRD1, DRD2, CRHR2, HTR7* (Cat 524991-C4, 553991-C3, 469621, 413041-C2). 5) "Swirl": *PVALB, GAD1, PTHLH, KIT* (Cat 422181-C4, 404031-C3, PTHLH, 606401-C2). Following probe labeling, sections were stored overnight in 4x SSC (saline-sodium citrate) buffer. After amplification steps (AMP1-3), probes were fluorescently labeled with Opal Dyes (Perkin Elmer, Waltham, MA; 1:500) and stained with DAPI (4′,6-diamidino-2-phenylindole) to label the nucleus. Lambda stacks were acquired in *z*-series using a Zeiss LSM780 confocal microscope equipped with a 63x x 1.4NA objective, a GaAsP spectral detector, and 405, 488, 555, and 647 lasers as previously described (Maynard et al., 2020). All lambda stacks were acquired with the same imaging settings and laser power intensities. For each subject, high magnification 63x images were randomly acquired in the NAc (n= 2 subjects, n=2 sections per subject, n=12 images per section).

7.2 Quantification and statistical analyses

snRNA-seq raw data processing

We processed the sequencing data with the 10x Genomics' Cell Ranger v3.0 pipeline, aligning to the human reference genome GRCh38, with a reconfigured GTF such that intronic alignments were additionally counted given the nuclear context, to generate UMI/feature-barcode matrices (https://support.10xgenomics.com/single-cell-gene-

expression/software/pipelines/latest/advanced/references#premrna). Per the output metrics of Cell Ranger, each sample was sequenced to a median depth of 284.3M reads (IQR: 253.7M-419.0M). We started with raw feature-barcode matrices from this output for analysis with the Bioconductor suite of R packages for single-cell RNA-seq analysis (Amezquita et al., 2020) using Bioconductor (Huber et al., 2015) version 3.12. For quality control (QC) and nuclei calling, we first used a Monte Carlo simulation-based approach to assess and exclude empty droplets or those with random ambient transcriptional noise, such as from debris (Griffiths et al., 2018; Lun et al., 2019). This was then followed by mitochondrial rate adaptive thresholding, which, though expected to be near-zero in this nuclear context, we applied a 3x median absolute deviation (MAD) threshold, to allow for flexibility in output/purity of nuclear enrichment by FACS using *scater*'s 'isOutlier'(Lun et al., 2016). Finally, within each sample, we computed doublet scores implemented with R package *scDblFinder*'s 'computeDoubletDensity' function (Dahlin et al., 2018), to assess putative doublet-driven clustering (see below). This QC pipeline yielded 11,202 high-quality nuclei from the DLPFC, 10,268 nuclei from HPC, 15,669 nuclei from AMY, 15,669 nuclei from sACC, and 20,571 nuclei from NAc. Collectively, these exhibited a median unique molecular identifier (UMI) count of 9,450(interquartile range, IQR: 5,513-23,078 UMIs) per nucleus, and a median detected gene count of 3,225 (IQR: 2,292-5,739) genes captured per nucleus. These feature-barcode gene counts were then rescaled across all nuclear libraries, using *batchelor*'s 'multiBatchNorm'(Haghverdi et al., 2018). Finally, these rescaled counts were $\log_2$-transformed for identification of highly-variable genes (HVGs) with *scran*'s 'modelGeneVar'(Lun et al., 2016), taking all genes with a greater variance than the fitted trend.

Dimensionality reduction and clustering

Principal components analysis (PCA) was then performed on the HVGs to reduce the high dimensionality of nuclear transcriptomic data for each region, implementing *batchelor*'s 'fastMNN' PC coordinate correction to remove batch effects at the donor (highest variance-contributing)-level (Amezquita et al., 2020; Haghverdi et al., 2018). The optimal principal component (PC) space was defined with iterative graph-based clustering to determine the *d* PCs where resulting *n* clusters stabilize, with the constraint that *n* clusters </= (*d* + 1) PCs (Lun et al., 2016), resulting in a chosen *d* between 59-99 PCs. In this PC-reduced space, graph-based clustering was performed to identify what we classified as preliminary clusters; specifically, *k*-nearest neighbors with *k*=20 neighbors and the Walktrap method from R package *igraph* (Csardi and Nepusz, 2006) for community detection. We then took all feature counts for these assignments and pseudo-bulked counts (Crowell et al., 2019; Kang et al., 2018; Lun and Marioni, 2017) across these preliminary nuclear clusters, rescaling for combined library size and log-transformed normalized counts, using *scater's* 'librarySizeFactors'(Lun et al., 2016). With the pseudo-bulked count profiles, we then performed hierarchical clustering to identify preliminary cluster relationships, and finally merged with the 'cutreeDynamic' function of R package *dynamicTreeCut* (Langfelder, et al., 2016), or keeping split clusters at the preliminary resolution, if generally well-represented across donors, as this suggested biologically valid subpopulations (for example, neuronal subtypes) as opposed to more likely batch-driven preliminary clusters. However, in some cases, cluster marker identification (see below) suggested sample bias in true, biological subpopulations (see Discussion). The final clusters merged at the appropriate tree height were then annotated for broad cell type identity with well-established cell type markers (Mathys et al., 2019), and with a letter suffix where multiple broad cell class populations were

defined ('*Excit_A*', '*Excit_B*', etc.). We also used Bioconductor package *scater*'s (Lun et al., 2016) implementation of non-linear dimensionality reduction techniques, *t*-SNE (van der Maaten and Hinton, 2008) and UMAP (McInnes et al., 2018), with default parameters and within the aforementioned optimal PC space, simply for visualization of the high-dimensional structure in the data, which generally complemented the clustering results. Additionally, in the HPC, AMY, sACC, and NAc analyses, we flagged clusters that were driven by low transcript capture or doublets (suggested by 'dual' cell-type marker expression, but confirmed with high doublet scores), and these were removed prior to downstream analyses and from the *t*-SNE display, resulting in a final *n* nuclei analyzed per region of: 11,202 from the DLPFC, 10,139 nuclei from HPC, 14,039 nuclei from AMY, 15,343 nuclei from sACC, and 19,892 nuclei from NAc (an average of 96.8% nuclei kept post-QC, above). These final numbers of nuclei analyzed per regionally-defined cell class by donor (and subcluster[1]) can be found in **Tables 2.1-2.5**.

Cluster marker identification

For marker identification with our final clusters defined in each brain region, we utilized *scran*'s 'findMarkers'(Lun et al., 2016) function for two sets of statistics:

1) Pairwise *t*-tests, to identify differences between each cluster, or
2) Implementing the function 'findMarkers' to perform a cluster-vs-all-other-nuclei *t*-test iteration

In both cases, we re-computed non-scale-matched $\log_2$-transformed counts (from 'logNormCounts'), including a donor covariate to properly model linearly (in the 'design=' parameter) on this unwanted batch effect (as 'multiBatchNorm', which is preferred as the

[1] Subcluster-level distribution per reported cell classes in this study can be accessed from the full publication of this work (*Tran, M.N., Maynard, K.R., et al.* Neuron *2021*) and were omitted from this dissertation for brevity.

input to 'fastMNN', above, removes much of the sequencing depth differences being modeled). The latter approach, 2), we consider a less-stringent marker test for enriched genes in a given cluster, but which would not necessarily differentiate between said cluster and all others. In addition to these statistics, for each cell class, we computed a Boolean parameter for non-0 median expression of each gene, to differentiate between noise-driven statistics. We used the results from both tests to interpret cell type identity beyond the broad classes (excitatory vs. inhibitory neuron), and to identify markers to probe via smFISH (below). The top 40 markers from each test result (including the respective non-0-median expression filter applied) are provided for each regionally-defined cell class in the Table S5 of this published work at *Tran, M.N., Maynard, K.R., et al.* Neuron *2021* (regions separated by worksheet), where the *'_pw'* suffix corresponds to the pairwise tests (set *1*), and *'_1vAll'* to the enriched expression test (set *2*). Importantly, 2) can be used to return a statistic, Cohen's D, or the standardized log-fold change, which we used to back-compute a single *t*-statistic for each cluster per gene, using:

`t = std.logFC * sqrt(n)`, where `n =` the total *n* nuclei (per region/dataset)

* Back-computing a single *t*-statistic cannot be generated with the result of 1) due to pairwise testing.

Comparing cell class conservation between datasets or across species

The *t*-statistics, described above, can then be used to compare such 'transcriptomic profiles' to those we computed for publicly-available postmortem human datasets, using the provided cell type annotations (or across our 5 regions), and compute the Pearson correlation coefficient (*r*), as was done in the spatial registration

approaches in *spatialLIBD (Maynard et al., 2021)*, across all shared expressed genes. To perform cross-species conservation analyses, we generated these *t*-statistics (from marker test *2.*, above) per gene, per reported cell annotation, subsetting on shared homologous genes between our human data and rat or mouse, using the 'DB Class Key' identifier provided by (http://www.informatics.jax.org/downloads/reports/HOM_AllOrganism.rpt), before computing the pairwise correlations. In the case of "many-to-many" orthalog scenarios, we took the highest-expressing paralog as the surrogate for each homologous pair, though these were small sets of genes in both rat and mouse cases. Correlation Pearson's *r* for both the human-vs-rat NAc and human AMY vs mouse MeA sections were performed in the gene space defined by the combined top-100 markers per cell class/subcluster (whether markers for the human cell classes or the reported subpopulations in each respective rodent model), where species homology information was known. In the NAc comparison, this was across 582 homologous marker genes, and for the AMY, 480 homologous genes.

GWAS association analyses with MAGMA

The latest version (*v1.08*) of Multi-marker Analysis of GenoMic Annotation (MAGMA;(de Leeuw et al., 2015) was used to test for genetic risk association of our 107 regionally-defined cell classes with schizophrenia (SCZ: (Pardiñas et al., 2018; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014)), autism spectrum disorder (ASD: (Grove et al., 2019)), bipolar disorder (BIP: (Stahl et al., 2019)), major depressive disorder (MDD:(Wray et al., 2018)), posttraumatic stress disorder (PTSD: (Nievergelt et al., 2019)); Alzheimer's disease (AD: (Jansen et al., 2019)); attention deficit/hyperactivity disorder (ADHD: (Demontis et al., 2019)); and for alcohol

and tobacco use phenotypes (Liu et al., 2019). For the marker gene sets, we used any genes defined as enriched per subpopulation (using marker test set *2*, from above), at the Benjamini & Hochberg false discovery rate (FDR) < 1e-6(Benjamini and Hochberg, 1995) and a restriction that the median expression of putative marker gene per cell class be > 0. SNPs were first annotated to genes, using window sizes from -10kb to +35kb of each gene, with the 1000 Genomes EUR reference panel, and gene-level analyses were performed, using provided summary statistics from each of the above listed GWAS (via https://www.med.unc.edu/pgc/download-results/ or https://genome.psych.umn.edu/index.php/GSCAN for results from (Liu et al., 2019)) and the 'snp-wise=mean' model, to test whether there was enrichment of genetic risk for disease/phenotype in each gene. Following this, we performed the default competitive gene set analysis with the 107 regionally-defined marker sets, testing for association of gene-level risk and whether genes were enriched/specific to each subpopulation. From the empirical *p*-value of the gene set analysis, we performed multiple test correction with both false-discovery rate (FDR) and the stricter Bonferroni procedure (threshold $p < 3.89e-5$) across all 1284 (107 regionally-defined subpopulations and 12 GWAS phenotypes tested) tests. All genetic association test results were published as Table S6 of this work (*Tran, M.N., Maynard, K.R., et al.* Neuron *2021*); Bonferroni-significant phenotype-cell class pairs, however, were included here, in **Table 3**.

RNAscope data analysis

Following image acquisition, lambda stacks in *z*-series were linearly unmixed in Zen software (weighted; no autoscale) using reference emission spectral profiles previously created in Zen (Maynard et al., 2020) and saved as Carl Zeiss Image "*.czi*" files. Images were segmented and quantitatively analyzed in MATLAB using *dotdotdot* software

(Maynard et al., 2020) and statistical analyses were performed in R v4.0.4: For each of the five experiments (see below for elaboration), we combined *DAPI*-defined region of interest (ROI)-level data from all respective images, and used data-driven cutoffs based on distributional overlap to determine binary expression levels (i.e. expressed or unexpressed) for each gene/channel, for cell class [group] prediction. In each experiment, each *GAD1*+ (or *DRD1*+ or *DRD1*+/*DRD2*+, pending on the experiment) ROI was classified into a Euclidean distance-predicted neuronal cell class (or group of classes, as accordingly, due to 4-plex limitations), based on the lowest distance. Probe counts were then quantified as the number of dots per 10,000 ROI pixels, post-lipofuscin masking, then log2-transformed (annotated as '*rnascope_[GENE]*' in **Figures 2.1.1D** and **2.1.2, 2.1.3, 2.1.4,** & **2.1.6**).

Experiment-specific information:

- *Circle*: 1033 ROIs were quantified across 48 images taken from 4 tissue sections across from 2 donors (two sections/donor). 251 ROIs were classified as *DRD1*+ with >3 dots post lipofuscin masking, and among these ROIs, *RXFP1* and *CRHR2* binarized expression, for prediction only, was classified as >3 dots and *TAC1* expression was classified as >6 dots. Corresponds to experiment shown in **Figure 2.1.1D**.
- *Square*: 1126 ROIs were quantified across 48 images taken from 4 tissue sections across from 2 donors (two sections/donor). 341 ROIs were classified as *DRD1*+ with >3 dots post lipofuscin masking, and among these ROIs, *RXFP2, GABRQ,* and *TAC1* binarized expression, for prediction 2nly, were each classified as >6 dots. Corresponds to experiment shown in **Figure 2.1.32**.
- *Triangle*: 1039 ROIs were quantified across 47 images taken from 4 tissue sections across from 2 donors (two sections/donor). 271 ROIs were classified as either *DRD1*+ or *DRD2*+ with >3 dots post lipofuscin masking in either gene, and among these ROIs,

TAC1 and *PENK* binarized expression, for prediction only, were classified as >6 dots. Corresponds to experiment shown in **Figure 3.2**.

- *Star*: 1003 ROIs were quantified across 44 images taken from 4 tissue sections across from 2 donors (two sections/donor). 482 ROIs were classified as either *DRD1*+ or *DRD2*+ with >3 dots (post lipofuscin masking) in either gene, and among these ROIs, *HTR7* and *CRHR21* binarized expression, for prediction only, were classified as >6 dots. Corresponds to experiment shown in **Figure 2.1.43**.
- *Swirl*: 989 ROIs were quantified across 44 images taken from 4 tissue sections across from 2 donors (two sections/donor). 212 ROIs were classified as *GAD1*+ inhibitory neurons with >6 dots post lipofuscin masking, and among these ROIs, *PVALB, KIT* and *PTHLH* binarized expression, for prediction only, were classified as >6 dots. Corresponds to experiment shown in **Figure 2.1.6**.

Public resources generated from this work

1. Raw single-nucleus RNA-seq read data is publicly available from the Globus endpoint 'jhpce#tran2021', linked from https://research.libd.org/globus. Data files containing de-identified and processed 'SingleCellExperiment' objects are hosted on Amazon S3, and the links are available on the README.md of the GitHub repository for this project (https://github.com/LieberInstitute/10xPilot_snRNAseq-human). RNAscope data generated in this study will be shared by the lead contact upon request.
2. All code for processing and analyzing the data has been archived at the time of submission at Zenodo: https://doi.org/10.5281/zenodo.5149046
3. For each of the five brain regions in this study, we created an interactive website with the data using *iSEE (Rue-Albrecht et al., 2018)* and deployed at the LIBD

shinyapps.io account at URLs such as https://libd.shinyapps.io/tran2021_NAc/ (and accordingly, /tran2021_sACC, /tran2021_DLPFC, /tran2021_AMY, and /tran2021_HPC).

Tables

Table 1. Donor demographic information.[1]

DonorID.pub	Sex	AgeDeath_yrs	Race	PrimaryDx	PMI_hrs	Smoking	Codeine	Morphine	BMI
donor1	M	54.43	CAUC	Control	21.5	FALSE	FALSE	FALSE	36.6
donor2	M	51.63	CAUC	Control	38.5	TRUE	FALSE	FALSE	24.4
donor3	M	40.08	CAUC	Control	28	FALSE	FALSE	FALSE	53.2
donor4	F	52.69	CAUC	Control	23.5	TRUE	FALSE	FALSE	26.0
donor5	M	59.86	CAUC	Control	26	FALSE	FALSE	FALSE	38.8
donor6	M	61.95	CAUC	Control	29	FALSE	FALSE	FALSE	30.4
donor7	M	42.05	CAUC	Control	28.5	FALSE	FALSE	FALSE	37.4
donor8	F	68.69	CAUC	Control	20.5	FALSE	FALSE	FALSE	30.6

Table 2.1. NAc cell classes by donor (*n*=8).

CellType	*donor1*	*donor2*	*donor4*	*donor8*	*donor5*	*donor6*	*donor3*	*donor7*
Astro_A	27	5	56	0	8	0	3	0
Astro_B	115	377	294	33	173	0	8	0
drop.doublet_A	8	2	19	0	6	0	1	0
drop.doublet_B	3	3	28	0	14	0	4	0
drop.doublet_C	0	0	0	0	0	18	0	23
drop.doublet_D	0	1	14	0	3	0	3	0
drop.lowNTx	5	14	30	366	81	5	4	24
Inhib_A	10	8	26	15	29	58	4	101
Inhib_B	0	0	4	4	3	6	0	23
Inhib_C	4	2	11	1	4	40	3	33
Inhib_D	4	5	24	13	4	128	6	56
Inhib_E	1	3	6	0	6	5	0	16
Macrophage	4	7	4	4	2	0	1	0
Micro	66	59	222	15	33	0	34	0
Micro_resting	3	33	5	0	22	0	0	0
MSN.D1_A	96	129	258	41	419	1680	35	1269
MSN.D1_B	4	2	0	7	1	64	0	161

[1] This is an abridged table, from Table 1 from *Tran, M.N., Maynard, K.R., et al.* Neuron *2021*

MSN.D1_C	8	22	2	2	21	4	0	224
MSN.D1_D	27	26	111	1	108	173	5	267
MSN.D1_E	15	6	30	51	15	255	6	260
MSN.D1_F	2	0	2	9	2	10	0	61
MSN.D2_A	95	101	252	57	488	1819	29	1421
MSN.D2_B	9	7	36	7	58	29	5	134
MSN.D2_C	9	3	18	5	15	131	3	130
MSN.D2_D	3	0	0	0	2	0	0	53
Oligo_A	237	50	247	0	385	0	69	0
Oligo_B	1202	804	2186	10	523	0	421	0
OPC	98	104	209	6	200	0	34	0
OPC_COP	0	0	14	0	1	0	3	0

Table 2.2. AMY cell classes by donor (*n*=5).

CellType	***donor1***	***donor2***	***donor4***	***donor8***	***donor5***	***donor6***	***donor3***	***donor7***
Astro_A	484	350	111	380	230	0	0	0
Astro_B	7	10	12	5	49	0	0	0
drop.lowNTx_A	4	0	880	4	179	0	0	0
drop.lowNTx_B	20	2	13	11	25	0	0	0
Endo	0	0	7	3	21	0	0	0
Excit_A	106	203	14	16	5	0	0	0
Excit_B	0	39	0	5	0	0	0	0
Excit_C	5	43	7	0	0	0	0	0
Inhib_A	0	0	362	0	366	0	0	0
Inhib_B	36	115	245	74	71	0	0	0
Inhib_C	128	17	85	11	284	0	0	0
Inhib_D	36	75	271	49	124	0	0	0
Inhib_E	0	0	7	2	405	0	0	0
Inhib_F	24	68	81	7	36	0	0	0
Inhib_G	0	0	9	1	76	0	0	0
Inhib_H	0	0	2	0	50	0	0	0
Micro	411	304	117	355	14	0	0	0
Mural	2	0	7	6	24	0	0	0
Oligo	1688	1736	309	2043	304	0	0	0
OPC	340	290	93	537	199	0	0	0
Tcell	3	7	3	15	3	0	0	0

Table 2.3 sACC cell classes by donor (*n*=5).

CellType	donor1	donor2	donor4	donor8	donor5	donor6	donor3	donor7
Astro_A	87	390	224	38	8	0	0	0
Astro_B	85	19	28	5	23	0	0	0
drop.doublet	8	13	7	0	0	0	0	0
drop.lowNTx	6	4	13	22	253	0	0	0
Excit_A	79	189	175	367	46	0	0	0
Excit_B	110	165	108	133	59	0	0	0
Excit_C	113	257	474	771	120	0	0	0
Excit_D	65	85	61	80	20	0	0	0
Excit_E	28	107	57	210	26	0	0	0
Excit_F	60	24	84	19	41	0	0	0
Excit_G	2	0	4	20	4	0	0	0
Inhib_A	34	108	168	476	56	0	0	0
Inhib_B	87	212	180	407	26	0	0	0
Inhib_C	34	117	80	205	29	0	0	0
Inhib_D	17	44	65	223	35	0	0	0
Inhib_E	11	19	61	213	26	0	0	0
Inhib_F	30	70	111	286	24	0	0	0
Inhib_G	4	26	41	124	11	0	0	0
Inhib_H	14	40	29	108	17	0	0	0
Inhib_I	2	15	4	17	1	0	0	0
Inhib_J	0	6	6	28	2	0	0	0
Inhib_K	0	0	7	14	4	0	0	0
Micro	232	243	292	14	3	0	0	0
Neu_FAT2.CDH15	1	0	9	0	10	0	0	0
Oligo_A	1833	1408	1132	12	4	0	0	0
Oligo_B	3	5	184	0	3	0	0	0
OPC	229	314	355	13	0	0	0	0

Table 2.4 DLPFC cell classes by donor (*n*=3).

CellType	donor1	donor2	donor4	donor8	donor5	donor6	donor3	donor7
Astro	371	137	0	0	0	274	0	0
Excit_A	111	120	0	0	0	298	0	0
Excit_B	75	154	0	0	0	544	0	0
Excit_C	44	155	0	0	0	325	0	0
Excit_D	22	27	0	0	0	83	0	0
Excit_E	77	25	0	0	0	85	0	0
Excit_F	102	36	0	0	0	105	0	0
Inhib_A	39	89	0	0	0	205	0	0

Inhib_B	98	106	0	0	0	250	0	0
Inhib_C	47	56	0	0	0	262	0	0
Inhib_D	119	78	0	0	0	216	0	0
Inhib_E	2	2	0	0	0	3	0	0
Inhib_F	0	1	0	0	0	7	0	0
Macrophage	1	3	0	0	0	6	0	0
Micro	152	92	0	0	0	144	0	0
Mural	3	2	0	0	0	13	0	0
Oligo	2754	517	0	0	0	2184	0	0
OPC	196	91	0	0	0	285	0	0
Tcell	2	2	0	0	0	5	0	0

Table 2.5 HPC cell classes by donor (*n*=3).

CellType	***donor1***	***donor2***	***donor4***	***donor8***	***donor5***	***donor6***	***donor3***	***donor7***
Astro_A	424	375	0	0	0	0	137	0
Astro_B	83	125	0	0	0	0	26	0
drop.doublet	4	1	0	0	0	0	0	0
drop.lowNTx_A	42	54	0	0	0	0	9	0
drop.lowNTx_B	9	5	0	0	0	0	5	0
Excit_A	4	9	0	0	0	0	74	0
Excit_B	118	291	0	0	0	0	12	0
Excit_C	1	0	0	0	0	0	5	0
Excit_D	2	1	0	0	0	0	32	0
Excit_E	6	0	0	0	0	0	0	0
Excit_F	1	23	0	0	0	0	5	0
Excit_G	4	2	0	0	0	0	0	0
Excit_H	33	0	0	0	0	0	0	0
Inhib_A	166	76	0	0	0	0	58	0
Inhib_B	30	0	0	0	0	0	0	0
Inhib_C	2	2	0	0	0	0	1	0
Inhib_D	4	11	0	0	0	0	16	0
Micro	487	481	0	0	0	0	193	0
Mural	20	19	0	0	0	0	4	0
Oligo	2586	2235	0	0	0	0	1091	0
OPC	374	255	0	0	0	0	194	0
OPC_COP	7	3	0	0	0	0	5	0
Tcell	14	9	0	0	0	0	3	0

Table 3. MAGMA phenotype-cell class association statistics (Bonferroni-significant).

Region	CellType	GWAS[1]	Beta	P	P.adj.fdr
dlpfc	Excit_A	addxn.CigDay	0.075536	6.32E-06	9.66E-05
dlpfc	Excit_C	addxn.CigDay	0.077921	2.09E-06	3.49E-05
dlpfc	Excit_E	addxn.CigDay	0.079377	3.72E-05	0.00041761
sacc	Excit_E.1	addxn.CigDay	0.098092	3.36E-07	6.86E-06
nac	MSN.D1_E	addxn.CigDay	0.14977	1.57E-07	3.35E-06
dlpfc	Excit_A	addxn.SmkInit	0.077226	3.51E-05	0.00040014
dlpfc	Inhib_A	addxn.SmkInit	0.10242	3.36E-05	0.00039251
dlpfc	OPC	addxn.SmkInit	0.14199	3.16E-05	0.00037252
sacc	OPC.1	addxn.SmkInit	0.14615	7.27E-06	0.00010733
hpc	Excit_H	addxn.SmkInit	0.16928	3.52E-05	0.00040014
hpc	OPC.2	addxn.SmkInit	0.14486	1.12E-05	0.00015474
nac	Inhib_A.3	addxn.SmkInit	0.10779	1.16E-05	0.0001587
nac	MSN.D1_C	addxn.SmkInit	0.21169	5.70E-09	1.83E-07
nac	MSN.D2_C	addxn.SmkInit	0.13958	2.93E-05	0.00035146
nac	OPC.3	addxn.SmkInit	0.13329	6.82E-06	0.00010302
dlpfc	Excit_A	SCZ.PGC2	0.13825	2.21E-11	1.58E-09
dlpfc	Excit_B	SCZ.PGC2	0.14336	3.44E-11	1.92E-09
dlpfc	Excit_C	SCZ.PGC2	0.1445	1.27E-12	1.48E-10
dlpfc	Excit_D	SCZ.PGC2	0.1793	7.05E-11	3.62E-09
dlpfc	Excit_E	SCZ.PGC2	0.15862	2.80E-11	1.71E-09
dlpfc	Excit_F	SCZ.PGC2	0.14873	5.13E-10	2.13E-08
dlpfc	Inhib_A	SCZ.PGC2	0.18842	1.00E-11	8.04E-10
dlpfc	Inhib_B	SCZ.PGC2	0.20057	4.02E-12	4.31E-10
dlpfc	Inhib_C	SCZ.PGC2	0.16933	4.25E-11	2.27E-09
dlpfc	Inhib_D	SCZ.PGC2	0.12639	2.86E-08	7.07E-07
dlpfc	OPC	SCZ.PGC2	0.20255	6.91E-08	1.56E-06
dlpfc	Oligo	SCZ.PGC2	0.24113	2.82E-10	1.30E-08
sacc	Astro_A	SCZ.PGC2	0.16774	3.95E-06	6.34E-05
sacc	Excit_A.1	SCZ.PGC2	0.13611	1.28E-10	6.08E-09
sacc	Excit_B.1	SCZ.PGC2	0.13544	3.79E-10	1.68E-08
sacc	Excit_C.1	SCZ.PGC2	0.15436	2.15E-15	2.77E-12
sacc	Excit_D.1	SCZ.PGC2	0.14667	4.05E-10	1.73E-08
sacc	Excit_E.1	SCZ.PGC2	0.1803	1.00E-13	3.77E-11
sacc	Excit_F.1	SCZ.PGC2	0.1639	2.99E-11	1.75E-09
sacc	Inhib_A.1	SCZ.PGC2	0.19199	2.51E-11	1.64E-09
sacc	Inhib_B.1	SCZ.PGC2	0.13024	7.08E-09	2.22E-07
sacc	Inhib_C.1	SCZ.PGC2	0.15393	4.22E-07	8.33E-06
sacc	Inhib_D.1	SCZ.PGC2	0.20559	2.55E-11	1.64E-09
sacc	Inhib_E.1	SCZ.PGC2	0.19153	9.27E-07	1.68E-05
sacc	Inhib_F.1	SCZ.PGC2	0.20658	3.28E-09	1.14E-07
sacc	Inhib_G	SCZ.PGC2	0.23418	3.25E-09	1.14E-07
sacc	Inhib_H	SCZ.PGC2	0.2103	5.85E-10	2.35E-08
sacc	OPC.1	SCZ.PGC2	0.20613	8.91E-09	2.72E-07
sacc	Oligo_A	SCZ.PGC2	0.23094	1.33E-08	3.80E-07

[1] This is an abridged table, from Table S6 from *Tran, M.N., Maynard, K.R., et al.* Neuron *2021*. Please see legend for associated Figure 5.1 for description of abbreviations.

hpc	Astro_A.1	SCZ.PGC2	0.15694	7.63E-06	0.00011011
hpc	Excit_A.2	SCZ.PGC2	0.14364	2.26E-08	5.81E-07
hpc	Excit_B.2	SCZ.PGC2	0.14109	8.62E-11	4.26E-09
hpc	Excit_E.2	SCZ.PGC2	0.24608	3.78E-05	0.00041862
hpc	Excit_F.2	SCZ.PGC2	0.16529	1.18E-07	2.61E-06
hpc	Excit_H	SCZ.PGC2	0.34191	4.48E-13	8.21E-11
hpc	Inhib_A.2	SCZ.PGC2	0.15178	1.70E-11	1.28E-09
hpc	Inhib_B.2	SCZ.PGC2	0.20854	1.28E-09	4.84E-08
hpc	Inhib_D.2	SCZ.PGC2	0.20881	4.54E-08	1.06E-06
hpc	OPC.2	SCZ.PGC2	0.24907	5.52E-12	5.45E-10
hpc	Oligo.1	SCZ.PGC2	0.21552	5.42E-09	1.78E-07
nac	Inhib_B.3	SCZ.PGC2	0.25446	3.14E-06	5.18E-05
nac	Inhib_D.3	SCZ.PGC2	0.12151	2.45E-05	0.00030822
nac	MSN.D1_A	SCZ.PGC2	0.12561	2.79E-09	1.02E-07
nac	MSN.D1_B	SCZ.PGC2	0.21885	1.55E-05	0.0002069
nac	MSN.D1_C	SCZ.PGC2	0.2051	3.37E-07	6.86E-06
nac	MSN.D1_D	SCZ.PGC2	0.13892	9.76E-09	2.91E-07
nac	MSN.D1_E	SCZ.PGC2	0.16778	1.51E-06	2.62E-05
nac	MSN.D2_A	SCZ.PGC2	0.13307	8.52E-10	3.31E-08
nac	MSN.D2_B	SCZ.PGC2	0.18381	3.89E-08	9.24E-07
nac	MSN.D2_C	SCZ.PGC2	0.19412	1.36E-07	2.95E-06
nac	OPC.3	SCZ.PGC2	0.22329	6.41E-12	5.88E-10
nac	Oligo_A.1	SCZ.PGC2	0.21488	9.54E-12	8.04E-10
nac	Oligo_B.1	SCZ.PGC2	0.24562	1.51E-08	4.23E-07
amy	Astro_A.3	SCZ.PGC2	0.13993	8.33E-06	0.00011755
amy	Excit_A.3	SCZ.PGC2	0.1446	1.04E-12	1.34E-10
amy	Excit_B.3	SCZ.PGC2	0.24528	1.83E-08	5.00E-07
amy	Excit_C.3	SCZ.PGC2	0.18371	2.09E-08	5.60E-07
amy	Inhib_A.4	SCZ.PGC2	0.15633	1.17E-13	3.77E-11
amy	Inhib_B.4	SCZ.PGC2	0.16963	8.78E-13	1.28E-10
amy	Inhib_C.4	SCZ.PGC2	0.20357	1.88E-13	4.83E-11
amy	Inhib_D.4	SCZ.PGC2	0.17613	2.32E-13	4.97E-11
amy	Inhib_E.3	SCZ.PGC2	0.16303	4.05E-07	8.12E-06
amy	Inhib_F.2	SCZ.PGC2	0.20582	8.97E-13	1.28E-10
amy	Inhib_G.1	SCZ.PGC2	0.27639	1.99E-05	0.00025794
amy	Inhib_H.1	SCZ.PGC2	0.25004	7.61E-07	1.42E-05
amy	OPC.4	SCZ.PGC2	0.23242	3.62E-14	2.32E-11
amy	Oligo.2	SCZ.PGC2	0.21521	2.35E-08	5.91E-07
dlpfc	OPC	ASD.PGC	0.13119	1.45E-05	0.00019561
nac	OPC.3	ASD.PGC	0.10672	3.74E-05	0.00041761
dlpfc	Excit_A	BIP.PGC	0.085783	2.08E-06	3.49E-05
dlpfc	Excit_B	BIP.PGC	0.10859	1.32E-08	3.80E-07
dlpfc	Excit_C	BIP.PGC	0.097418	4.80E-08	1.10E-06
dlpfc	Excit_D	BIP.PGC	0.11389	2.02E-06	3.46E-05
dlpfc	Excit_E	BIP.PGC	0.1162	3.19E-08	7.74E-07
dlpfc	Excit_F	BIP.PGC	0.088368	2.57E-05	0.00031768
dlpfc	Inhib_A	BIP.PGC	0.10028	3.43E-05	0.00039716
dlpfc	Inhib_B	BIP.PGC	0.11715	4.96E-06	7.68E-05
dlpfc	Inhib_C	BIP.PGC	0.096701	2.39E-05	0.00030342
dlpfc	OPC	BIP.PGC	0.15209	3.26E-06	5.30E-05
sacc	Excit_A.1	BIP.PGC	0.076044	2.75E-05	0.0003366

sacc	Excit_C.1	BIP.PGC	0.084701	4.95E-07	9.64E-06
sacc	Excit_D.1	BIP.PGC	0.12209	3.94E-09	1.33E-07
sacc	Excit_E.1	BIP.PGC	0.085839	2.91E-05	0.00035146
sacc	Excit_F.1	BIP.PGC	0.097209	7.25E-06	0.00010733
sacc	Inhib_A.1	BIP.PGC	0.1299	2.59E-07	5.46E-06
sacc	OPC.1	BIP.PGC	0.12934	2.56E-05	0.00031768
hpc	Excit_B.2	BIP.PGC	0.085862	7.81E-06	0.00011139
hpc	Excit_D.2	BIP.PGC	0.17231	2.96E-05	0.00035146
hpc	Inhib_A.2	BIP.PGC	0.099111	8.59E-07	1.58E-05
hpc	OPC.2	BIP.PGC	0.15325	9.55E-07	1.70E-05
nac	MSN.D1_E	BIP.PGC	0.17659	2.25E-08	5.81E-07
nac	MSN.D2_B	BIP.PGC	0.12283	2.06E-05	0.00026502
nac	MSN.D2_C	BIP.PGC	0.16273	6.68E-07	1.26E-05
nac	OPC.3	BIP.PGC	0.14102	5.52E-07	1.06E-05
amy	Excit_A.3	BIP.PGC	0.079011	7.51E-06	0.00010955
amy	Excit_C.3	BIP.PGC	0.13354	4.05E-06	6.42E-05
amy	Inhib_A.4	BIP.PGC	0.078296	1.65E-05	0.00021874
amy	Inhib_B.4	BIP.PGC	0.095296	4.43E-06	6.94E-05
amy	Inhib_C.4	BIP.PGC	0.10418	1.81E-05	0.00023654
amy	Inhib_E.3	BIP.PGC	0.12419	1.09E-05	0.00015148
amy	OPC.4	BIP.PGC	0.12914	1.28E-06	2.25E-05

Bibliography

Amezquita, R.A., Lun, A.T.L., Becht, E., Carey, V.J., Carpp, L.N., Geistlinger, L., Marini, F., Rue-Albrecht, K., Risso, D., Soneson, C., et al. (2020). Orchestrating single-cell analysis with Bioconductor. Nat. Methods *17*, 137–145.

Babaev, O., Piletti Chatain, C., and Krueger-Burg, D. (2018). Inhibition in the amygdala anxiety circuitry. Exp. Mol. Med. *50*, 1–16.

Barger, N., Stefanacci, L., Schumann, C.M., Sherwood, C.C., Annese, J., Allman, J.M., Buckwalter, J.A., Hof, P.R., and Semendeferi, K. (2012). Neuronal populations in the basolateral nuclei of the amygdala are differentially increased in humans compared with apes: a stereological study. J. Comp. Neurol. *520*, 3035–3054.

Batiuk, M.Y., Martirosyan, A., Wahis, J., de Vin, F., Marneffe, C., Kusserow, C., Koeppen, J., Viana, J.F., Oliveira, J.F., Voet, T., et al. (2020). Identification of region-specific astrocyte subtypes at single cell resolution. Nat. Commun. *11*, 1220.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Statist. Soc. B *57*, 289–300.

Bonfiglio, J.J., Inda, C., Refojo, D., Holsboer, F., Arzt, E., and Silberstein, S. (2011). The corticotropin-releasing hormone network and the hypothalamic-pituitary-adrenal axis: molecular and cellular mechanisms involved. Neuroendocrinology *94*, 12–20.

Bryois, J., Skene, N.G., Hansen, T.F., Kogelman, L.J.A., Watson, H.J., Liu, Z., Eating Disorders Working Group of the Psychiatric Genomics Consortium, International Headache Genetics Consortium, 23andMe Research Team, Brueggeman, L., et al. (2020). Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. Nat. Genet. *52*, 482–493.

Cao, J., Willett, J.A., Dorris, D.M., and Meitzen, J. (2018). Sex Differences in Medium Spiny Neuron Excitability and Glutamatergic Synaptic Input: Heterogeneity Across Striatal Regions and Evidence for Estradiol-Dependent Sexual Differentiation. Front Endocrinol (Lausanne) *9*, 173.

Chamling, X., Kallman, A., Fang, W., Berlinicke, C.A., Mertz, J.L., Devkota, P., Pantoja, I.E.M., Smith, M.D., Ji, Z., Chang, C., et al. (2021). Single-cell transcriptomic reveals molecular diversity and developmental heterogeneity of human stem cell-derived oligodendrocyte lineage cells. Nat. Commun. *12*, 652.

Chareyron, L.J., Banta Lavenex, P., Amaral, D.G., and Lavenex, P. (2011). Stereological analysis of the rat and monkey amygdala. J. Comp. Neurol. *519*, 3218–3239.

Chen, P.B., Hu, R.K., Wu, Y.E., Pan, L., Huang, S., Micevych, P.E., and Hong, W. (2019). Sexually dimorphic control of parenting behavior by the medial amygdala. Cell *176*, 1206-1221.e18.

Claes, S.J. (2004). Corticotropin-releasing hormone (CRH) in psychiatry: from stress to psychopathology. Ann. Med. *36*, 50–61.

Collado-Torres, L., Burke, E.E., Peterson, A., Shin, J., Straub, R.E., Rajpurohit, A., Semick, S.A., Ulrich, W.S., BrainSeq Consortium, Price, A.J., et al. (2019). Regional Heterogeneity in Gene Expression, Regulation, and Coherence in the Frontal Cortex and Hippocampus across Development and Schizophrenia. Neuron *103*, 203-216.e8.

Crowell, H.L., Soneson, C., Germain, P.-L., Calini, D., Collin, L., Raposo, C., Malhotra, D., and Robinson, M.D. (2019). On the discovery of population-specific state transitions from multi-sample multi-condition single-cell RNA sequencing data. BioRxiv.

Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. http://www.interjournal.org/manuscript_abstract.php? 361100992.

Dahlin, J.S., Hamey, F.K., Pijuan-Sala, B., Shepherd, M., Lau, W.W.Y., Nestorowa, S., Weinreb, C., Wolock, S., Hannah, R., Diamanti, E., et al. (2018). A single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in Kit mutant mice. Blood *131*, e1–e11.

Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., et al. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat. Genet. *51*, 63–75.

Enterría-Morales, D., Del Rey, N.L.-G., Blesa, J., López-López, I., Gallet, S., Prévot, V., López-Barneo, J., and d'Anglemont de Tassigny, X. (2020). Molecular targets for endogenous glial cell line-derived neurotrophic factor modulation in striatal parvalbumin interneurons. Brain Commun. *2*, fcaa105.

Fenster, R.J., Lebois, L.A.M., Ressler, K.J., and Suh, J. (2018). Brain circuit dysfunction in post-traumatic stress disorder: from mouse to man. Nat. Rev. Neurosci. *19*, 535–551.

Figueiro-Silva, J., Gruart, A., Clayton, K.B., Podlesniy, P., Abad, M.A., Gasull, X., Delgado-García, J.M., and Trullas, R. (2015). Neuronal pentraxin 1 negatively regulates excitatory synapse density and synaptic plasticity. J. Neurosci. *35*, 5504–5521.

Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. *47*, 1228–1235.

Garrett, A., and Chang, K. (2008). The role of the amygdala in bipolar disorder development. Dev. Psychopathol. *20*, 1285–1296.

Gerfen, C.R., Engber, T.M., Mahan, L.C., Susel, Z., Chase, T.N., Monsma, F.J., and Sibley, D.R. (1990). D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. Science *250*, 1429–1432.

Gokce, O., Stanley, G.M., Treutlein, B., Neff, N.F., Camp, J.G., Malenka, R.C., Rothwell, P.E., Fuccillo, M.V., Südhof, T.C., and Quake, S.R. (2016). Cellular Taxonomy of the Mouse Striatum as Revealed by Single-Cell RNA-Seq. Cell Rep. *16*, 1126–1137.

Graveland, G.A., Williams, R.S., and DiFiglia, M. (1985). A Golgi study of the human neostriatum: neurons and afferent fibers. J. Comp. Neurol. *234*, 317–333.

Griffiths, J.A., Richard, A.C., Bach, K., Lun, A.T.L., and Marioni, J.C. (2018). Detection and removal of barcode swapping in single-cell RNA-seq data. Nat. Commun. *9*, 2667.

Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R., et al. (2019). Identification of common genetic risk variants for autism spectrum disorder. Nat. Genet. *51*, 431–444.

Habib, N., Li, Y., Heidenreich, M., Swiech, L., Avraham-Davidi, I., Trombetta, J.J., Hession, C., Zhang, F., and Regev, A. (2016). Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. Science *353*, 925–928.

Habib, N., Avraham-Davidi, I., Basu, A., Burks, T., Shekhar, K., Hofree, M., Choudhury, S.R., Aguet, F., Gelfand, E., Ardlie, K., et al. (2017). Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat. Methods *14*, 955–958.

Haghverdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat. Biotechnol. *36*, 421–427.

Han, X., Zhou, Z., Fei, L., Sun, H., Wang, R., Chen, Y., Chen, H., Wang, J., Tang, H., Ge, W., et al. (2020). Construction of a human cell landscape at single-cell level. Nature *581*, 303–309.

Heilig, M., and Koob, G.F. (2007). A key role for corticotropin-releasing factor in alcohol dependence. Trends Neurosci. *30*, 399–406.

Heimer, L., Zahm, D.S., Churchill, L., Kalivas, P.W., and Wohltmann, C. (1991). Specificity in the projection patterns of accumbal core and shell in the rat. Neuroscience *41*, 89–125.

Horii-Hayashi, N., Okuda, H., Tatsumi, K., Ishizaka, S., Yoshikawa, M., and Wanaka, A. (2008). Localization of chondroitin sulfate proteoglycan versican in adult brain with special reference to large projection neurons. Cell Tissue Res. *334*, 163–177.

Howard, D.M., Adams, M.J., Clarke, T.-K., Hafferty, J.D., Gibson, J., Shirali, M., Coleman, J.R.I., Hagenaars, S.P., Ward, J., Wigmore, E.M., et al. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat. Neurosci. *22*, 343–352.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods *12*, 115–121.

Hu, P., Fabyanic, E., Kwon, D.Y., Tang, S., Zhou, Z., and Wu, H. (2017). Dissecting Cell-Type Composition and Activity-Dependent Transcriptional State in Mammalian Brains by Massively Parallel Single-Nucleus RNA-Seq. Mol. Cell *68*, 1006-1015.e7.

Jaffe, A.E., Hoeppner, D.J., Saito, T., Blanpain, L., Ukaigwe, J., Burke, E.E., Collado-Torres, L., Tao, R., Tajinda, K., Maynard, K.R., et al. (2020). Profiling gene expression in

the human dentate gyrus granule cell layer reveals insights into schizophrenia and its genetic risk. Nat. Neurosci. *23*, 510–519.

Janak, P.H., and Tye, K.M. (2015). From circuits to behaviour in the amygdala. Nature *517*, 284–292.

Jansen, I.E., Savage, J.E., Watanabe, K., Bryois, J., Williams, D.M., Steinberg, S., Sealock, J., Karlsson, I.K., Hägg, S., Athanasiu, L., et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat. Genet. *51*, 404–413.

Kang, H.M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E., Wan, E., Wong, S., Byrnes, L., Lanata, C.M., et al. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nat. Biotechnol. *36*, 89–94.

Kawaguchi, Y. (1997). Neostriatal cell subtypes and their functional roles. Neurosci. Res. *27*, 1–8.

Kronman, H., Richter, F., Labonté, B., Chandra, R., Zhao, S., Hoffman, G., Lobo, M.K., Schadt, E.E., and Nestler, E.J. (2019). Biology and Bias in Cell Type-Specific RNAseq of Nucleus Accumbens Medium Spiny Neurons. Sci. Rep. *9*, 8350.

Lacar, B., Linker, S.B., Jaeger, B.N., Krishnaswami, S.R., Barron, J.J., Kelder, M.J.E., Parylak, S.L., Paquola, A.C.M., Venepally, P., Novotny, M., et al. (2016). Nuclear RNA-seq of single neurons reveals molecular signatures of activation. Nat. Commun. *7*, 11022.

Lake, B.B., Ai, R., Kaeser, G.E., Salathia, N.S., Yung, Y.C., Liu, R., Wildberg, A., Gao, D., Fung, H.-L., Chen, S., et al. (2016). Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. Science *352*, 1586–1590.

Langfelder, P., Zhang, B., and Horvath, S. (2016). dynamicTreeCut: Methods for Detection of Clusters in Hierarchical Clustering Dendrograms. https://rdrr.io/cran/dynamicTreeCut/.

de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. *11*, e1004219.

Lin, Y.-T., Liu, T.-Y., Yang, C.-Y., Yu, Y.-L., Chen, T.-C., Day, Y.-J., Chang, C.-C., Huang, G.-J., and Chen, J.-C. (2016). Chronic activation of NPFFR2 stimulates the stress-related depressive behaviors through HPA axis modulation. Psychoneuroendocrinology *71*, 73–85.

Lipska, B.K., Deep-Soboslay, A., Weickert, C.S., Hyde, T.M., Martin, C.E., Herman, M.M., and Kleinman, J.E. (2006). Critical factors in gene expression in postmortem human brain: Focus on studies in schizophrenia. Biol. Psychiatry *60*, 650–658.

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D.M., Chen, F., Datta, G., Davila-Velderrain, J., McGuire, D., Tian, C., et al. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat. Genet. *51*,

237–244.

Li, Z., Chen, Z., Fan, G., Li, A., Yuan, J., and Xu, T. (2018). Cell-Type-Specific Afferent Innervation of the Nucleus Accumbens Core and Shell. Front. Neuroanat. *12*, 84.

Lobo, M.K. (2009). Molecular profiling of striatonigral and striatopallidal medium spiny neurons past, present, and future. Int. Rev. Neurobiol. *89*, 1–35.

Lobo, M.K., Karsten, S.L., Gray, M., Geschwind, D.H., and Yang, X.W. (2006). FACS-array profiling of striatal projection neuron subtypes in juvenile and adult mouse brains. Nat. Neurosci. *9*, 443–452.

Lun, A.T.L., and Marioni, J.C. (2017). Overcoming confounding plate effects in differential expression analyses of single-cell RNA-seq data. Biostatistics *18*, 451–464.

Lun, A.T.L., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. [version 2; peer review: 3 approved, 2 approved with reservations]. F1000Res. *5*, 2122.

Lun, A.T.L., Riesenfeld, S., Andrews, T., Dao, T.P., Gomes, T., participants in the 1st Human Cell Atlas Jamboree, and Marioni, J.C. (2019). EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. Genome Biol. *20*, 63.

van der Maaten, L., and Hinton, G. (2008). Visualizing Data using t-SNE. J Mach Learn Res *9*, 2579–2605.

Masuda, T., Sankowski, R., Staszewski, O., and Prinz, M. (2020). Microglia Heterogeneity in the Single-Cell Era. Cell Rep. *30*, 1271–1281.

Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J.Z., Menon, M., He, L., Abdurrob, F., Jiang, X., et al. (2019). Single-cell transcriptomic analysis of Alzheimer's disease. Nature *570*, 332–337.

Maynard, K.R., Tippani, M., Takahashi, Y., Phan, B.N., Hyde, T.M., Jaffe, A.E., and Martinowich, K. (2020). dotdotdot: an automated approach to quantify multiplex single molecule fluorescent in situ hybridization (smFISH) images in complex tissues. Nucleic Acids Res.

Maynard, K.R., Collado-Torres, L., Weber, L.M., Uytingco, C., Barry, B.K., Williams, S.R., Catallini, J.L., Tran, M.N., Besich, Z., Tippani, M., et al. (2021). Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nat. Neurosci. *24*, 425–436.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection. ArXiv.

Murray, E.A., Wise, S.P., and Drevets, W.C. (2011). Localization of dysfunction in major depressive disorder: prefrontal cortex and amygdala. Biol. Psychiatry *69*, e43-54.

Nagy, C., Maitra, M., Tanti, A., Suderman, M., Théroux, J.-F., Davoli, M.A., Perlman, K., Yerko, V., Wang, Y.C., Tripathy, S.J., et al. (2020). Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells

and excitatory neurons. Nat. Neurosci. *23*, 771–781.

Nievergelt, C.M., Maihofer, A.X., Klengel, T., Atkinson, E.G., Chen, C.-Y., Choi, K.W., Coleman, J.R.I., Dalvie, S., Duncan, L.E., Gelernter, J., et al. (2019). International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. Nat. Commun. *10*, 4558.

Pardiñas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S.E., Bishop, S., Cameron, D., Hamshere, M.L., et al. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat. Genet. *50*, 381–389.

Payer, D., Williams, B., Mansouri, E., Stevanovski, S., Nakajima, S., Le Foll, B., Kish, S., Houle, S., Mizrahi, R., George, S.R., et al. (2017). Corticotropin-releasing hormone and dopamine release in healthy individuals. Psychoneuroendocrinology *76*, 192–196.

Prensa, L., Richard, S., and Parent, A. (2003). Chemical anatomy of the human ventral striatum and adjacent basal forebrain structures. J. Comp. Neurol. *460*, 345–367.

Rizzardi, L.F., Hickey, P.F., Rodriguez DiBlasi, V., Tryggvadóttir, R., Callahan, C.M., Idrizi, A., Hansen, K.D., and Feinberg, A.P. (2019). Neuronal brain-region-specific DNA methylation and chromatin accessibility are associated with neuropsychiatric trait heritability. Nat. Neurosci. *22*, 307–316.

Rue-Albrecht, K., Marini, F., Soneson, C., and Lun, A.T.L. (2018). iSEE: Interactive SummarizedExperiment Explorer. [version 1; peer review: 3 approved]. F1000Res. *7*, 741.

Russo, S.J., and Nestler, E.J. (2013). The brain reward circuitry in mood disorders. Nat. Rev. Neurosci. *14*, 609–625.

Rymar, V.V., Sasseville, R., Luk, K.C., and Sadikot, A.F. (2004). Neurogenesis and stereological morphometry of calretinin-immunoreactive GABAergic interneurons of the neostriatum. J. Comp. Neurol. *469*, 325–339.

Salgado, S., and Kaplitt, M.G. (2015). The nucleus accumbens: A comprehensive review. Stereotact. Funct. Neurosurg. *93*, 75–93.

Saunders, A., Macosko, E.Z., Wysoker, A., Goldman, M., Krienen, F.M., de Rivera, H., Bien, E., Baum, M., Bortolin, L., Wang, S., et al. (2018). Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. Cell *174*, 1015-1030.e16.

Savell, K.E., Tuscher, J.J., Zipperly, M.E., Duke, C.G., Phillips, R.A., Bauman, A.J., Thukral, S., Sultan, F.A., Goska, N.A., Ianov, L., et al. (2020). A dopamine-induced gene expression signature regulates neuronal function and cocaine response. Sci. Adv. *6*, eaba4221.

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.

Schumann, C.M., and Amaral, D.G. (2005). Stereological estimation of the number of neurons in the human amygdaloid complex. J. Comp. Neurol. *491*, 320–329.

Sindreu, C., and Storm, D.R. (2011). Modulation of neuronal signal transduction and memory formation by synaptic zinc. Front. Behav. Neurosci. *5*, 68.

Skene, N.G., Bryois, J., Bakken, T.E., Breen, G., Crowley, J.J., Gaspar, H.A., Giusti-Rodriguez, P., Hodge, R.D., Miller, J.A., Muñoz-Manchado, A.B., et al. (2018). Genetic identification of brain cell types underlying schizophrenia. Nat. Genet. *50*, 825–833.

Sorvari, H., Soininen, H., Paljärvi, L., Karkola, K., and Pitkänen, A. (1995). Distribution of parvalbumin-immunoreactive cells and fibers in the human amygdaloid complex. J. Comp. Neurol. *360*, 185–212.

Stahl, E.A., Breen, G., Forstner, A.J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J.R.I., Gaspar, H.A., et al. (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat. Genet. *51*, 793–803.

Stanley, G., Gokce, O., Malenka, R.C., Südhof, T.C., and Quake, S.R. (2020). Continuous and discrete neuron types of the adult murine striatum. Neuron *105*, 688-699.e8.

Tamura, G., Olson, D., Miron, J., and Clark, T.G. (2005). Tolloid-like 1 is negatively regulated by stress and glucocorticoids. Brain Res. Mol. Brain Res. *142*, 81–90.

Tan, C.P., Sano, H., Iwaasa, H., Pan, J., Sailer, A.W., Hreniuk, D.L., Feighner, S.D., Palyha, O.C., Pong, S.-S., Figueroa, D.J., et al. (2002). Melanin-concentrating hormone receptor subtypes 1 and 2: species-specific gene expression. Genomics *79*, 785–792.

Tepper, J.M., and Bolam, J.P. (2004). Functional diversity and specificity of neostriatal interneurons. Curr. Opin. Neurobiol. *14*, 685–692.

Thrupp, N., Sala Frigerio, C., Wolfs, L., Skene, N.G., Fattorelli, N., Poovathingal, S., Fourne, Y., Matthews, P.M., Theys, T., Mancuso, R., et al. (2020). Single-Nucleus RNA-Seq Is Not Suitable for Detection of Microglial Activation Genes in Humans. Cell Rep. *32*, 108189.

Tran, M.N., Maynard, K.R., Spangler, A., Huuki, L.A., Montgomery, K.D., Sadashivaiah, V., Tippani, M., Barry, B.K., Hancock, D.B., Hicks, S.C., et al. (2021). Single-nucleus transcriptome analysis reveals cell-type-specific molecular signatures across reward circuitry in the human brain. Neuron *109*, 3088-3103.e5.

Tyszka, J.M., and Pauli, W.M. (2016). In vivo delineation of subdivisions of the human amygdaloid complex in a high-resolution group template. Hum. Brain Mapp. *37*, 3979–3998.

Velmeshev, D., Schirmer, L., Jung, D., Haeussler, M., Perez, Y., Mayer, S., Bhaduri, A., Goyal, N., Rowitch, D.H., and Kriegstein, A.R. (2019). Single-cell genomics identifies cell type-specific molecular changes in autism. Science *364*, 685–689.

Voorn, P., Gerfen, C.R., and Groenewegen, H.J. (1989). Compartmental organization of the ventral striatum of the rat: immunohistochemical distribution of enkephalin, substance P, dopamine, and calcium-binding protein. J. Comp. Neurol. *289*, 189–201.

Wassum, K.M., and Izquierdo, A. (2015). The basolateral amygdala in reward learning and addiction. Neurosci. Biobehav. Rev. *57*, 271–283.

Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., et al. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat. Genet. *50*, 668–681.

Yao, J.-J., Zhao, Q.-R., Lu, J.-M., and Mei, Y.-A. (2018). Functions and the related signaling pathways of the neurotrophic factor neuritin. Acta Pharmacol. Sin. *39*, 1414–1420.

Yong, W., Spence, J.P., Eskay, R., Fitz, S.D., Damadzic, R., Lai, D., Foroud, T., Carr, L.G., Shekhar, A., Chester, J.A., et al. (2014). Alcohol-preferring rats show decreased corticotropin-releasing hormone-2 receptor expression and differences in HPA activation compared to alcohol-nonpreferring rats. Alcohol. Clin. Exp. Res. *38*, 1275–1283.

Zahm, D.S., and Heimer, L. (1993). Specificity in the efferent projections of the nucleus accumbens in the rat: comparison of the rostral pole projection patterns with those of the core and shell. J. Comp. Neurol. *327*, 220–232.

Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L.E., La Manno, G., et al. (2018). Molecular architecture of the mouse nervous system. Cell *174*, 999-1014.e22.

Zhang, X., Cheng, H., Zuo, Z., Zhou, K., Cong, F., Wang, B., Zhuo, Y., Chen, L., Xue, R., and Fan, Y. (2018). Individualized Functional Parcellation of the Human Amygdala Using a Semi-supervised Clustering Method: A 7T Resting State fMRI Study. Front. Neurosci. *12*, 270.

Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049.

Curriculum Vitae

Matthew Nguyen Tran (Nam-Phong Nguyen)
ng.nam109@gmail.com | Twitter: @mattntran

Profile

- Human geneticist with a focus on functional genomics & expertise in single-cell-level transcriptomics
- Active 7+ years doing bench research, from molecular work to clinical-grade production
- Can lead major projects in conception, experimentation, analyses, and publication process
- Emphasis on collaboration, open science, learning and exchanging knowledge

Education

JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE **2017 – MAR 2022**

- Ph.D. in Human Genetics (Department of Genetic Medicine)
- Thesis: Contextualizing genetic risk for substance use in the reward circuitry of the human brain

THE UNIVERSITY OF TEXAS AT AUSTIN **2010 - 2014**

- B.S. with Honors in Biology (Microbiology)

Research Experience

JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE **2017 - CURRENT**

Ph.D. Candidate | Principal Investigator: Keri Martinowich

- Performed single-nucleus RNA-sequencing to create a novel resource for single-cell-level transcriptomic profiles in the reward circuitry of the postmortem human brain, across five brain regions
- Led the entire QC and analyses of a comprehensive snRNA-seq dataset to create publicly available interactive web apps
- Assessed genetic risk for various substance use and neuropsychiatric disorder phenotypes in the context of the reward circuitry and its molecularly defined cell classes

Graduate Student | Pre-thesis advisor: Andrew Jaffe

- Modeled stem cell differentiation and performed differential expression analyses in iPSC-derived neurons from tissue for both postmortem samples of Pitt-Hopkins syndrome and schizophrenia
- Performed weighted gene co-expression network analysis (WGCNA) and created a consensus scoring algorithm to assess performance of different WGCNA models

THE UNIVERSITY OF TEXAS MD ANDERSON CANCER CENTER **2014 - 2017**

Clinical Cell Therapy Specialist | Principal Investigator: Cassian Yee, M.D.

- Clinical manufacture of patient tumor antigen-specific T cells after assessing the immunogenicity of over 100 tumor peptide antigens for prioritization of antigen targeting, including functional analyses of T cell products
- Investigatory research in non-canonical CD4/CD8 T cell MHC class restriction

Publications & Presentations

ORCID iD: **0000-0001-9694-7378**

PUBLISHED

Page S.C., Rao, Sripathy, S.R., Farinelli, F., Ye, Z., Wang, Y., Hiler, D.J., Pattie, E.A., Nguyen, C.V., Tippani, M., Moses, R.L., Chen, H., **Tran, M.N.**, Eagles, N.J., Stolz, J.M., Catallini II, J.L., Soudry, O.R., Dickinson, D., Berman, K.F., Apud, J.A., Weinberger, D.R., Martinowich, K., Jaffe, A.E., Straub, R.E., Maher, B.J. (2022). Electrophysiological measures from human iPSC-derived neurons are associated with schizophrenia clinical status and predict individual cognitive performance. *Proceedings of the National Academy of Sciences* 119 (3). https://doi.org/10.1073/pnas.2109395119

Tran, M.N., Maynard, K.R., Spangler, A., Huuki, L.A., Montgomery, K.D., Sadashivaiah, V., Tippani, M., Barry, B.K., Hancock, D.B., Hicks, S.C., Kleinman, J.E., Hyde, T.M., Collado-Torres, L., Jaffe, A.E., Martinowich, K. (2021). Single-nucleus transcriptome analysis reveals cell-type-specific molecular signatures across reward circuitry in the human brain. *Neuron* (109), 3013-3178. https://doi.org/10.1016/j.neuron.2021.09.001

Maynard, K.R., Collado-Torres, L., Weber, L.M., Uytingco, C., Barry, B.K., Williams, S.R., Catallini II, J.L., **Tran, M.N.**, Besich, Z., Tippani, M., Chew, J.,Yin, Y., Kleinman, J.E., Hyde, T.M., Rao, N., Hicks, S.C., Martinowich, K., Jaffe, A.E. (2021). Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci* (24), 425-436. https://doi.org/10.1038/s41593-020-00787-0

Jaffe, A.E., Hoeppner, D.J., Saito, T., Blanpain, L., Ukaigwe, J., Burke, E.E., Collado-Torres, L., Tao, R., Tajinda, K., Maynard, K.R., **Tran, M.N.**, Martinowich, K., Deep-Soboslay, A., Shin, J., Kleinman, J.E., Weinberger, D.R., Matsumoto, M., Hyde, T.M. (2020). Profiling gene expression in the human dentate gyrus granule cell layer reveals insights into schizophrenia and its genetic risk. *Nat Neurosci* (23), 510-519. https://doi.org/10.1038/s41593-020-0604-z

IN REVIEW

Sosina, O., **Tran, M.N.**, Maynard, K.R., Tao, R., Taub, M.A., Martinowich, K., Semick, S.A., Quach, B.C., Weinberger, D.R., Hyde, T.M., Hancock, D.B., Kleinman, J.E., Leek, J.T., Jaffe, A.E. (2021). Strategies for cellular deconvolution in human brain RNA sequencing data. *F1000Research*. https://doi.org/10.12688/f1000research.50858.1

PRESENTATIONS

National Institute on Drug Abuse Genetics Consortium Meeting **MAR 2021**

Poster title: Contextualizing substance use genetic risk with snRNA-seq of the reward circuitry.

Cold Spring Harbor Laboratory: The Biology of Genomes Conference **MAY 2020**

Poster title: An expanded single-nucleus RNA-sequencing resource elucidates regional transcriptomic diversity in the brain

Skills & Abilities

COMPUTATIONAL

- Single-cell & bulk RNA-sequencing (Bioconductor and CRAN tools)
- Functional genomics tools: GWAS, MAGMA, LDSC, TWAS
- Imaging data analysis: RNAscope (single-molecule FISH)
- Dimensionality reduction & clustering techniques: PCA, NMF, HCA, k-means, nearest-neighbor
- Regression analyses, ANOVA
- Gene network analysis: WGCNA
- Bash scripting for batch processing
- R Markdown for generation of presentations and research reports
- Version control (GitHub)
- Creation of public datasets/interactive web apps

LABORATORY

- Techniques in surface and intracellular staining; flow cytometry & FACS
- New FACS machine QC & implementation into ongoing workflows
- Tissue homogenization and single cell/nuclei suspension
- Single-cell/-nucleus RNA-sequencing protocols: 10x Genomics & SPLiT-seq workflows
- Library preparation for high-throughput sequencing: Illumina and PacBio
- Single-cell cloning
- Molecular biology/cloning
- Aseptic handling of a variety of primary and transformed human cell lines
- Immune based assays: western blot and ELISA
- Radioactive assays for measuring metabolism and cytotoxicity
- Whole blood & leukapheresis processing for PBMCs; plasma; serum
- Sterile gowning and ISO class 7 GMP processing

GENERAL

- Manuscript writing and navigating submission & peer-review process
- Mentorship and training of students or lab personnel in both lab-based and computational protocols

LANGUAGES

CODING

- R (including Markdown & LaTeX)
- Unix
- Python

OTHER

- Spanish: Professional working proficiency
- German: Elementary proficiency

Teaching & Leadership

TEACHING EXPERIENCE

Johns Hopkins University School of Medicine

Teaching Assistant | Computational Bioinformatics and Bioinformatics **NOV-DEC 2019**

- Assisted students' learning of R syntax and navigating the RStudio Cloud
- Co-created learning modules for data visualization & differential expression analysis

Teaching Assistant | Pathology for Graduate Students: Basic Mechanisms **AUG-SEP 2019**

- Developed a presentation and lectured on basic anatomy & physiology of the cardiovascular system
- Designed & graded quiz questions; contributed examination questions

LEADERSHIP & MENTORSHIP EXPERIENCE

Community Outreach Chair | Human Genetics Program EIG Committee **AUG 2020 - CURRENT**

- Current chair for the community outreach role for the program's equity, diversity & inclusion (EDI) student committee, Equity in Genetics (EIG)
- Organized and ran student body-wide climate surveys for feedback to program leadership & administration

Head of Family | Human Genetics Program 'family' cohort **AUG 2019 - CURRENT**

- Organize and lead year-wide training opportunities for a cohort of program students
- Provide peer support and informal mentorship training

Graduate Program Student Representative | Human Genetics, JHU SOM **AUG 2018 - CURRENT**

- 2017 Class student representative with various year-long roles, including co-running recruitment and maintaining training opportunities for students at various stages in their pre-doctoral training
- Represent students in various instances of conflict with program administration or leadership
- Co-created and oversaw the organization of the program's first EDI student committee

Head of Family | Thread, Paul Laurence Dunbar High School **JAN 2018 - CURRENT**

- Provide academic and life mentorship support to inner-city youth
- Attend regular meetings and contribute to end goals of organization