FRONT MATTER

Title

- Full title: Transcriptional impacts of substance use disorder and HIV on human ventral midbrain neurons and microglia
- Short title: Transcriptional SUD and HIV impacts on human VM

Authors

Alyssa M. Wilson[1,2,*], Michelle M. Jacobs[1], Tova Y. Lambert[2, †], Aditi Valada[2, †], Gregory Meloni[1], Evan Gilmore[1], Jacinta Murray[1], Susan Morgello[1,3,4, ‡], Schahram Akbarian[2,3,5, ‡]

Affiliations

[1]Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.
[2]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.
[3]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.
[4]Department of Pathology, Molecular and Cell Based Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.
[5]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.
[†]Present address: Albert Einstein College of Medicine, New York, New York 10461, USA.
[‡]These authors contributed equally to this work.

[*]email: alyssa.wilson@mssm.edu

Abstract

For people with HIV (PWH), substance use disorders (SUDs) are a prominent neurological risk factor, and the impacts of both on dopaminergic pathways are a potential point of deleterious convergence. Here, we profile, at single nucleus resolution, the substantia nigra (SN) transcriptome of 90 postmortem donors in the context of chronic HIV and opioid/cocaine SUD, including 67 prospectively characterized PWH. We report altered microglial expression for hundreds of pro- and anti-inflammatory regulators attributable to HIV, and separately, to SUD. Stepwise progressive microglial dysregulation coupled to altered SN dopaminergic and GABAergic signaling was associated with SUD/HIV dual diagnosis and further with lack of viral suppression in blood. In virologically suppressed donors, SUD comorbidity was associated with microglial transcriptional changes permissive for HIV infection. We report HIV-related downregulation of monoamine reuptake transporters specifically in dopaminergic neurons regardless of SUD status or viral load, and additional transcriptional signatures consistent with selective vulnerability of SN dopamine neurons.

Teaser

Coordinated changes in neurons and microglia show disinhibition in SUD, and HIV/SUD synergy permissive for HIV infection.

MAIN TEXT

Introduction

The HIV pandemic continues to pose a global health burden, despite considerable advances in treatment and prevention (*1*, *2*). Although a large fraction of people with HIV (PWH) are receiving life-saving antiretroviral therapy (ART), treatment does not eradicate the virus, and HIV persists in multiple tissue reservoirs including the brain. PWH may develop neurologic disorders that are predicated in part on immune dysfunction, and that are presumed to result from the action of HIV on nervous system cells (*2–4*). An important co-morbidity relevant to HIV neurologic disease is substance use disorder (SUD), and in particular injection drug use (*5*), with its negative impact on health by a multitude of factors including risky behaviors, emergence of medical and psychiatric sequelae, and lack of compliance with therapy (*6–15*). Notably, HIV and SUD show a striking neuroanatomical convergence on dopamine-rich structures in brain. In ventral midbrain (*16*) this includes substantia nigra (SN) (*17*, *18*), a region important for mediating habitual behaviors and salience of cues associated with drug use (*18*), withdrawal-related anhedonia and dysphoria, and extra-pyramidal motor dysfunction (*19*, *20*). There is also evidence, mostly from preclinical work, that drug-of-abuse-induced dynamic adaptations in addiction-relevant dopaminergic circuitry could worsen HIV-associated cognitive and behavioral deficits (*21*), with multiple dopaminergic signaling pathways converging on susceptible myeloid cells to promote viral entry and replication (*22*, *23*).

Despite its importance, very little is known about cellular and molecular alterations in the ventral midbrain of PWH with or without SUD comorbidity, with the majority of studies limited to quantification of neurotransmitter (e.g., dopamine) in bulk tissue or neuroimaging and analysis of cerebrospinal fluid (CSF) in the context of advanced disease (AIDS) (*24–26*). Herein, we profile for the first time, the ventral midbrain transcriptomes from n=90 total PWH and people without HIV ("HIV-"), at single (cell) nucleus resolution, and deliver first insights into the cell-type-specific neurogenomic responses to chronic infection and SUD. We specifically examine responses to opioid and/or cocaine SUDs, given the high prevalence of these substances and their combination among PWH (*27–29*). Our analyses, focused on SN dopaminergic (DA) and GABAergic (GABA) neurons and their surrounding microglia, strongly suggest that SUD for opioids and cocaine creates, even in donors with antiretroviral therapy (ART)-mediated viral suppression, a molecular environment permissive for HIV viral replication and risk for cytotoxic damage to dopaminergic neurons. In addition, we report a stepwise progression of transcriptomic dysregulation for opioid/cocaine SUD and HIV comorbidity that culminates in widespread neuronal pathology and pronounced inflammatory signatures in microglia from individuals without viral suppression (viremia) in the context of SUD.

Results

SN cell type composition in the context of HIV and SUD

We dissected ventral midbrain for unilateral collection of the SN including *pars compacta* and *pars reticulata* from n=90 donors collected by the Manhattan HIV Brain Bank (MHBB) (39 female/ 51 male; mean+/-standard deviation age 54.0+/-11.4 years; 33% Hispanic, 38% non-Hispanic Black, 22% non-Hispanic White; Table 1), including 67 participants followed longitudinally in the MHBB with deep clinical phenotyping during years of study participation (median 6, maximum 21 years). We assigned each donor to one of three HIV statuses: negative ("HIV-", n=28 donors), HIV positive with "undetectable" final plasma viral load (VL) defined as <50 RNA copies/mL ("HIV+u", n=30 donors), or HIV positive with viremia, evidenced by a detectable final plasma VL ("HIV+d", n=32 donors [viral RNA copies/mL range 65 to >750,000, median 6,721]). We further grouped donors by substance use into "SUD-" or "SUD+" for any

combination of opioid/cocaine SUD (n=60 SUD+, including n=42 with a history of misusing opioids, of which n=28 had a history of misusing both opioids and cocaine). This grouping produced HIV-/SUD-defined subgroups that were overall closely matched with regard to demographic factors: SUD-HIV- (n= 13); SUD+HIV- (n=15); SUD-HIV+u (n=10); SUD+HIV+u (n=20); SUD-HIV+d (n=7); and SUD+HIV+d (n=25) (Fig. 1A, Tables S1-S3). Notably, the wider MHBB cohort from which these donors were selected had a high prevalence of polysubstance SUD, with approximately three quarters of SUD-positive individuals endorsing symptoms for two or more drug classes.

We processed nuclei in pools of up to 4 donors, with most pools containing mixed HIV/SUD donor phenotypes, using the 10x Chromium 3' system followed by Illumina sequencing. We then matched each single, sequenced nucleus to a specific donor using SNP-array genotyping and demuxlet (*30*), and applied a series of quality control (QC) filters to each nucleus (Figs. S1A-C), including an upper threshold on the proportion of mitochondrial reads (set at 0.0128; Fig. S1A). We note that 100% of glial and 98% of neuronal nuclei that passed QC had a minimum of 200 genes expressed (Figs. S1B, C), thereby exceeding standards from published SN reference sets (*31*, *32*) (compare Figs. S2A and S2B). After preprocessing, the final dataset amounted to a total of 200,732 QC-passing nuclei, with 81/90 donors contributing >1,000 nuclei (Fig. S1D; Table S4; Materials and Methods).

We first asked whether SN cell type composition is affected in the context of HIV and SUD. To maximize sensitivity for potential shifts in cellular subpopulations, we used as input 20,000/36,601 gene transcripts (Fig. S1E) as opposed to more typically suggested sizes of ~2,000 (*33*), and we performed a likelihood-based, agnostic approach for cell typing (Materials and Methods). Our approach produced 12 major cell types (Figs. 1B, C, Fig. S3) that by numbers and proportions closely matched those reported for human SN single-cell-based transcriptomic reference sets (*34*, *35*), with the largest share (65%) composed of oligodendrocytes and their precursors, and 14-15% made up of microglia and astrocytes. In contrast, SN neurons, endothelium, lymphocytes and macrophages each contributed a much smaller proportion of nuclei. There were no significant differences for 176/180 statistical comparisons of cell type proportions across HIV/SUD donor groups, indicating that the overall representation of major cell types in SN is preserved independent of HIV or SUD status (Figs. 1D, E; Table S5).

Cell-specific transcriptomic responses to HIV infection show progression in the context of SUD, as well as SUD-independent downregulation of monoamine transporters in DA neurons

To examine cell-level impacts on SN function due to SUD, HIV, or both, we performed differential expression analysis (DEA; Materials and Methods) by stratifying the entire cohort of n=90 SN samples first by HIV groupings to interrogate effects of SUD, thereby comparing SUD+ vs. SUD- for HIV- donors, SUD+ vs. SUD- for HIV+u donors, and SUD+ vs. SUD- for HIV+d donors (Fig. 2). Likewise, to examine the impacts of HIV status, we stratified by SUD grouping, comparing HIV+u vs. HIV- for SUD+ donors, then for SUD- donors, and repeating similar analyses for HIV+d vs. HIV- and HIV+d vs. HIV+u, first for SUD+ and then for SUD- donors (Figs. 3-5).

We focused our DEAs on SN DA neurons and GABA neurons, given their key role in addiction circuitry and potential sensitivity to HIV infection (*25*, *36–39*), and also on microglia, as mediators of neuroinflammation and host cells for HIV genomic integration (*39–41*). We reduced the rate of false-positive differentially expressed genes (DEGs) using multiple steps. First, we applied cross-fold validation (k=3 folds), creating for each comparison and cell type three separate subsets of nuclei, performing DEA on each, and retaining only DEGs appearing in all 3

subsets. Second, we applied DESeq2 DEA model fitting, to exclude genes with low expression dispersion estimates and low counts (*42*, *43*) (Materials and Methods).

Using the above DEA pipeline on DA neurons, and stratifying for HIV status (to assess the impact of SUD), we called 51 genes downregulated in SUD+HIV- vs. SUD-HIV- donors (false discovery rate [FDR] $p<0.05$; log2 fold change [l2fc] > variable values, dependent upon FDR, sample size, and dataset information content (*43*); Data S1). These DEGs enriched gene ontology (GO) terms (*44*) linked to neuronal function, including axon myelination, presynaptic organization, cell adhesion, and Na^+ transport (Fig. 2A; Data S2). Notably, SUD+HIV+u donors, while exhibiting significant sharing of these downregulated DEGs with SUD+HIV- donors (by permutation test, $p<10^{-5}$) also showed 82 uniquely upregulated DEGs, which enriched most prominently GO terms for axon growth and presynaptic transmission (Fig. 2B, J, K; Data S2). This observation suggests enhanced sensitivity to SUD in the context of HIV comorbidity, even in the face of viral suppression (for HIV+u donors). In contrast, SUD+HIV+d had 0 DEGs vs. SUD-HIV+d (Figs. 2C, P; Data S1), suggesting that for these two viremic donor groups, transcriptomic alterations of SN DA neurons were primarily driven by factors related to active HIV replication (viremia).

Indeed, this is what we observed when the DEA of DA neurons was stratified for SUD status (to assess the impacts of HIV), with SUD+HIV+d donors having 40 up-/ 38 down- regulated DEGs vs. SUD+HIV- (Fig. 3B). Furthermore, while downregulated DEGs were highly shared between SUD+HIV+d and SUD+HIV+u donors (35 shared, $p<10^{-5}$; Figs. 3A, B), many viremia-specific (HIV+d) effects were identified (56 downregulated and 5 upregulated DEGs that enriched different GO terms; Fig. 3C, Data S2). Importantly, the presence of SUD increased virus-associated transcriptomic dysregulation: DA neurons from SUD-HIV+d and SUD-HIV+u donors consistently showed a much smaller number of DEGs than their SUD+ counterparts (by approximately 1/3; Figs. 3D, E) and in the absence of SUD, HIV+u and HIV+d transcriptomes were indistinguishable (n DEGs = 0, Fig. 3F).

Despite the above SUD-attributable differences, a striking commonality in all HIV+ groupings also emerged from our analyses: DA neurons from all HIV+ groups, regardless of SUD status or VL, showed significant downregulation in *SLC6A3* dopamine and *SLC18A2* monoamine transporter genes (Figs. 3A, B, D, E; Data S1, S2), together with a more variable dysregulation of genes related to neuronal immune responses, presynaptic vesicle release, synaptic signaling, and cell-cell adhesion. From all the above observations, we draw two conclusions. First, the neurogenomic response to HIV infection, both for donors with viral suppression and without, alters fundamental signaling mechanisms in SN DA neurons including dopamine reuptake. Second, SUD comorbidity enhances abnormality in HIV+ donors, driving a stepwise progression in DA neuron dysregulation, in which abnormalities are more severe in SUD+ compared to SUD- donors and enhance the effects of systemic viral replication (Figs. 3C, F).

Next, we ran our DEA pipeline on SN GABA neurons, which produced few DEGs across all comparisons (there were median [interquartile range or IQR] 0 [0, 0.75] GABA DEGs for each of our 6 DEAs, as opposed to 24 [0, 40.75] DA DEGs; compare Figs. 2A-C to D-F and Fig. 3 to Fig. 4). Stratification by HIV status produced only one DEG attributable to SUD: downregulation of *RBFOX1* GABAergic transcription factor in SUD+HIV+u vs. SUD-HIV+u donors (associated with decreased inhibition (*45*, *46*); Figs. 2D-F, L, M). However, stratification by SUD revealed severe HIV-associated GABA neuron transcriptomic dysregulation in viremic (HIV+d) donors with SUD, pointing to reduced inhibitory neurotransmission (Figs. 4B, C; Data S2). We counted 115 HIV-associated DEGs (96 down, 17 up) in SUD+HIV+d donors vs. SUD+HIV-, and 322

DEGs (266 down, 57 up) in SUD+HIV+d vs. SUD+HIV+u donors. There was significant overlap ($p<10^{-5}$) between these two comparisons, including downregulated DEGs involved in GABA production, glutamatergic G-protein-coupled receptor (GPCR) signaling, homophilic cell-cell adhesion, and action-potential-associated membrane depolarization (Data S1, S2). In addition, there was downregulation of several opioid-related transcripts, including of postsynaptic opioid-binding cell adhesion molecule (OBCAM) *OPCML*, which is downregulated during chronic opioid agonist exposure (*47–49*); the closely related neurotrimin (*NTM*); ephrin-B1 receptor (*EPHB1*), which colocalizes and interacts closely with ephrin-B2 receptor (*50*, *51*) as a mediator of *OPCML* signaling (*49*); and non-coding RNA *AC073225.1*, previously reported to show decreased expression in opioid use disorder (OUD) (*52*). These alterations were specific to SUD+HIV+d, given that SUD-HIV+d vs. SUD-HIV-, SUD-HIV+u vs. SUD-HIV-, and SUD+HIV+u vs. SUD+HIV- comparisons completely lacked significant differences in gene expression (n DEGs = 0).

Finally, we applied our stratified DEA pipeline to SN microglia, resulting in hundreds of DEGs attributable to SUD (Figs. 2G-I, N, O, Q). These transcriptional alterations were mostly enriched for various combinations of pro- and anti-inflammatory molecules (Data S1, S2), in broad agreement with recent studies that have reported upregulation of immune signaling genes and transcriptomic signatures indicative of microglial activation and altered glial motility in OUD, chronic opioid exposures, and overdose (*34*, *53–55*). However, in our analysis a substantial fraction of SUD-attributable microglial DEGs were unique to each HIV grouping (Figs. 2N, O), pointing to distinct microglial responses to SUD in the context of varying levels of HIV (HIV-, HIV+u, HIV+d (*56*)). For example, SUD-associated gene sets supporting reduced protein stabilization and interferon-1 production were seen only in HIV+d, and not in the other HIV groupings (Data S2).

Stratifying by SUD groupings (Fig. 5), we also counted hundreds of microglial DEGs attributable to HIV, including 446 down- and 89 up-regulated DEGs for SUD- subjects with viral suppression (SUD-HIV+u) vs. HIV- controls (SUD-HIV-), >50% of which were replicated in viremic SUD-HIV+d donors ($p<10^{-5}$). These shared changes primarily involved gene sets for increased inflammatory signaling, decreased phagocytosis and iron ion responses, and increased stress response protein folding/refolding (Data S2). For the SUD-HIV+d group, there were additional changes vs. SUD-HIV+u, including enhancements of immune signaling and telomere maintenance/remyelination processes, as well as antiviral-response-related reductions in glucose transport/GPCR signaling (*57*, *58*). In the context of SUD (Figs. 5A-C), when comparing SUD+HIV+u and SUD+HIV+d donors to SUD+HIV- controls, microglial transcriptional alterations included multiple gene sets related to immune activation, many of which were shared; further, some of these shared effects were opposite to shared HIV effects in the absence of SUD. For example, in the absence of SUD, HIV resulted in downregulated iron ion responses; in the context of SUD, iron ion responses were upregulated (Data S2). Additionally, in the context of SUD, microglia-mediated regulation of synaptic signaling was dysregulated in both HIV+u and HIV+d donors; this effect was absent in the SUD- comparisons.

Coordinated DEG clusters link microglial and neuronal function, and highlight microglial changes in SUD+HIV+u donors supportive for HIV replication

Next, we wanted to complement our pathway-by-cell type analyses with an alternative approach, by capturing DEGs with statistically linked, or "coordinated", expression levels across the majority of donors within an SUD/HIV group. To that end, for each DEA we performed, we searched for groups of DEGs that had a high degree of inter-patient covariance (Materials and Methods). We selected similarity-based DEG subclusters with average pairwise similarity scores

249 >0.75, our cutoff for identifying gene groups with the most ubiquitous coordinated transcriptional
250 alteration (Fig. 6A), for functional annotation. In the following, we discuss these DEG
251 subclusters, with a focus on microglia-neuron biological functions and interactions.
252
253 We first examined highly coordinated DEGs for DA neurons and microglia, using the 4 DEAs
254 that showed changes for either SUD, suppressed HIV (HIV+u), or both (Fig. 6B). We identified
255 117 highly coordinated DEG subclusters (with median [IQR] 10 [6,21] DEGs per subcluster;
256 Table S6, Data S4). Interestingly, 17 of these subclusters contained a mix of DA and microglial
257 DEGs, suggesting DA-microglial interactions. Two of the 17 were specific for the comparison of
258 SUD+HIV- vs. SUD-HIV- (related to impacts of SUD in the absence of HIV; Figs. 6D-G). Both
259 of these subclusters involved glycoproteins: one (C22S1) linked decreased N-glycan synthetic
260 enzyme *FUT8* fucosyltransferase (*59*) in both microglia and DA neurons to altered DA
261 neurotransmission (reduced membrane maintenance, reduced pre-/ post-synaptic machinery,
262 slowed presynaptic release, and increased excitability; Figs. 6D, E). Notably, *FUT8* has
263 previously been associated with increased basal microglial reactivity (*60*) and impaired neuronal
264 membrane maintenance (*61–63*). The other subcluster (C22S2) linked decreased DA neuron
265 expression of *GPM6B* glycoprotein to a microglial switch from oxidative phosphorylation to
266 aerobic glycolysis (*64*, *65*), as well as to deficits in the microglial antiviral response and in
267 transcripts associated with increased microglial-neuroligin-1-mediated excitatory synapse
268 formation (*MDGA2*↓ (*66*, *67*)) and increased catecholaminergic receptor responses (*SPATS2L*↓
269 (*68*, *69*)), among others (Figs. 6F, G).
270
271 For dual-diagnosis (SUD+HIV+u) donors, there were 9 mixed DA neuron/microglial subclusters.
272 Three of these closely resembled the subclusters described above for SUD without HIV,
273 specifically for DA DEGs (see "SUD conserved impacts," denoted by deep pink color, in Figs.
274 6B, D, F; Data S4). Additionally, and remarkably as these were virally suppressed donors, 5 out
275 of 9 had prominent representation of DEGs linked to HIV infection and replication (Data S4). For
276 example (subcluster C16S2, Figs. 7C, D), DA upregulation of the Na+-Ca2+ exchanger *SLC8A1*
277 was linked to 31 microglial DEGs important for HIV transcription and replication (*HIF1A*↑ (*70*,
278 *71*), *HIF1A-AS3*↑ (*72–74*), *HIVEP2*↑ (*75–78*), *SYTL3*↑ (*79*, *80*), *REL*↑ (*81–84*), *ELL2*↑ (*85*, *86*)),
279 HIV host entry and genome stabilization (*TFRC*↑ (*87*), *SEC14L1*↑ (*88*), *CD55*↑ (*89*, *90*)) or viral
280 stabilization (*SIPA1L1*↑ or *E6TP1*↑ (*91*)), and HIV-1 intracellular compartment release (*ATP1B3*↑
281 (*92*)). We note that Na+Ca2+-exchanger-mediated prolonged neuronal excitation was previously
282 linked to HIV infection in an *in vitro* model (*93*). Thus, these subclusters indicated that in virally
283 suppressed PWH, SUD may result in microglial changes permissive for HIV. Other DEG
284 subclusters for SUD+HIV+u donors also linked microglial dysregulation to increased excitability
285 in DA neurons (Figs. 7G, H; Data S4).
286
287 Another specific SUD+HIV+u subcluster (subcluster C17S0, Figs. 7E, F) linked microglial
288 upregulation of receptors for neurodegeneration-associated inflammatory phagocytosis (*OLR1*
289 (*94*, *95*) and *ITGAX*, which targets complement-coated particles (*96*)) to DA neuronal
290 downregulation of neuronal maintenance genes *CLMN* (*97*, *98*) and *PEX5L* (*99*)). Taken together,
291 this subcluster suggests that for individuals with undetectable HIV, SUD+/HIV+u synergy still
292 may be associated with damage to vulnerable DA neuron populations.
293
294 ***In SUD+HIV+d, impaired GABAergic function is coordinated with increased microglial pro-***
295 ***inflammatory signaling***
296 We also identified DEG subclusters for GABA neurons and microglial cells, to explore their
297 relationships in SUD+HIV+d, because in our DEAs this subgroup specifically showed signs of
298 disrupted GABA neuron function (Figs. 4B, C; Data S4). Compared to SUD+HIV+u,

SUD+HIV+d individuals had microglial upregulation of DEGs linked to pro-inflammatory signaling (*EPSTI1* (*100*), *IFI44L* (*101*, *102*)) together with alterations in GABAergic regulators of aging (*RANBP17*↓ (*103*)), cellular maintenance, GABAergic transmission (*SYT9*↓ (*104*); *DDAH1*↓ (*105*)), and ER-Ca2+-mediated apoptosis (*DPP10*↑, *RYR3*↑ (*106*, *107*)). Therefore, the impairment of GABAergic regulation in SUD+HIV+d is linked to pro-inflammatory responses in microglia.

Discussion

We have explored cell-type-specific transcriptional changes in the SN of donors with HIV, opioid/cocaine SUD, or both. Our rationale for HIV groupings (HIV-, HIV+u, HIV+d) was based on an extant literature that has linked plasma HIV to brain microglial activation and other aspects of brain pathology (*108*), and on the relevance of plasma HIV load as a standard for guiding patient care. Our rationale for using a polysubstance (opioid/cocaine) approach to SUD groupings was based on the pronounced polysubstance prevalence both in the MHBB cohort and overall in the United States (*109*), the importance of these particular drug classes to the HIV epidemic in New York City (*27–29*), and a call in existing literature for more research on polysubstance users, which reflects the lifetime exposures of the majority of individuals with chronic SUDs (*109–111*).

Our findings, summarized in Fig. 8, demonstrated altered expression of hundreds of pro- and anti-inflammatory transcripts in microglia attributable both to SUD and HIV, with the combined effects of SUD and HIV resulting in a more pronounced dysregulation of the microglial inflammatory transcriptome, particularly in those donors lacking viral suppression (HIV+d). These changes were significantly coupled with abnormalities in neuronal integrity, excitability, and neurotransmission. Perhaps most strikingly, we observed that dysregulated DA neuron transcription in the SN of virally suppressed, opioid/cocaine-SUD+ HIV+u donors was linked to upregulated microglial transcription of genes permissive for HIV infection, replication, and reactivation, along with an altered microglial antiviral response and increased neurodegeneration-associated inflammatory phagocytosis (Data S4). These findings, in a carefully phenotyped, prospectively characterized human population, are striking given the previously reported evidence of a selectively expanded brain viral reservoir in ART-treated, Simian Immunodeficiency Virus (SIV)-infected non-human primates with chronic morphine exposure (*112*, *113*). In this SIV model, reservoir expansion occurred in conjunction with microglial transcriptional changes indicative of immunosuppression and degeneration, relative to SIV+ opioid-negative comparators (*112*, *113*). With the techniques utilized in the present study, we are unable to comment on the brain viral reservoir in our donors, and we identified only 2 HIV transcripts in our entire collection of nuclei (Materials and Methods). This is not surprising, given that in a previous study with similar snRNA-seq methodology, HIV transcripts in a subset of microglial nuclei were limited to donor brains displaying the histopathology of active brain viral replication or HIV encephalitis (HIVE) (*114*). None of the donors in our present study had histopathologic evidence of HIVE, assessed as part of the standard MHBB neuropathology protocol (*115*, *116*). The lack of viral transcript detection constitutes a limitation of our present work. Furthermore, it contrasts with recent study in which HIV RNA was detected in whole-cell, sorted brain microglia from 2 of 3 virally suppressed HIV+ donors (*117*, *118*). The donors in this prior study were exclusively male and described as having been on ART "until at least 2 wk before death". The discrepancy between our study and this previous study warrants further investigation, and it could reflect differences in nuclear vs. cytoplasmic substrates for analysis; the time at which ART therapy was discontinued in donors prior to death; or possibly either geographical variabilities in circulating and evolving HIV strains or differing genetic determinants of immune function and viral control in donor populations.

Another important aspect of our work was the demonstration that SUD comorbidity in the context of HIV status exerted impacts on SN neuronal transcriptomes in a stepwise progression. In HIV-donors, we observed only downregulation of DA neuronal transcripts that potentially induce impaired function, senescence-like changes, and increased excitability (Fig. 6, Data S4). In virally suppressed individuals (HIV+u), we observed both a shared (with HIV-) downregulation of DA neuronal genes potentially increasing DA neuron excitability, as well as uniquely upregulated DEGs that suggested enhanced sensitivity to SUD (Fig. 2). Similarly, we observed evidence of enhanced sensitivity to HIV in SUD+HIV+u donors (Fig. 3). With viremia/absence of viral suppression (HIV+d), we further observed changes affecting hundreds of GABAergic transcripts and potentially compromising inhibitory neurotransmission. As microglial transcriptional dysregulation had unique components and also progressed with each HIV grouping, this raises the question of what drives progression, and whether it is unidirectional (driven by one cell type) or bidirectional (driven reciprocally). Recent work in mixed cell cultures (containing microglia with either DA or GABA neurons) has shown that healthy neurons may drive HIV into latency in microglia; and conversely, neuronal damage, including substance-related damage, fosters microglial HIV reactivation and viral expression, thereby further compromising neuronal health due to enhanced microglia-mediated neurotoxicity (*119*). Our findings in postmortem human brain raise the possibility that similar bidirectional phenomena may be relevant in human, but future studies will be needed to further examine the mechanisms driving neuronal and glial dysfunction, expanding on our observations of transcriptomic signatures of coordinated detrimental interactions between HIV and SUD (for a more complete listing of significant coordinated abnormalities, see Data S3, S4).

Finally, while we have focused on effects of HIV and SUD comorbidity in altering SN neuronal and microglial transcriptomes, it is noteworthy that in all our HIV+ donor groups, regardless of viral load or SUD status, DA neurons showed a consistent deficit in the expression of the dopamine and monoamine reuptake transporters *SLC6A3* (encoding the protein "DAT") and *SLC18A2*, respectively, two key regulators of DA turnover at synapses. Combination cocaine-and-raclopride positron emission tomography (PET) studies of viremic PWH, both with and without SUD, have demonstrated similar abnormalities at the protein level *in vivo*, showing decreased DAT signal in the dopamine-terminal-rich basal ganglia (*120*, *121*). The prevailing hypothesis for this phenomenon was that HIV viral proteins, including its transactivator of transcription or Tat and others, directly bind to monoamine transporters, thereby inhibiting monoamine (including dopamine) reuptake and damaging DA neurons (*122*, *123*). However, downregulation of DAT and monoamine reuptake transporter transcripts occurred in our study even in the virally suppressed HIV+ SUD- donor group; and furthermore, evidence of viral Tat transcripts in our brains was lacking. The DA-neuron-specific RNA results presented here point to a transcriptional mechanism resulting in decreased monoamine transporter expression that, while linked to some unified aspect of HIV infection, does not require evidence of ongoing viral replication or transcription.

In summary, the strengths of our study include: a large number of carefully phenotyped brain donors with large amounts of prospectively collected data on SUD and immunovirologic status; novel analytic pipelines that are tolerant to abnormality yet rigidly subjected to quality assurance/QC procedures that exceed other published norms; analytic groups parsed by clinically relevant factors in HIV and SUD; and in-depth annotation of DEG groups that show coordinated expression in most members of their associated donor group (Data S3, S4). Limitations include a lack of viral transcript detection and the lack of mechanistic exploration of the multiple transcriptomic abnormalities described in these donor groupings. Future studies are indicated to

399 further understand the significance of these genomic adaptations for future development of
400 therapies for PWH and SUD.
401
402 **Materials and Methods**
403 ***Brain donor tissue and clinical data collection***
404 Brain donors were autopsied and their donations curated by the MHBB between the years 1999
405 and 2021, using study protocols and consent documents approved by the Icahn School of
406 Medicine at Mount Sinai (ISMMS) Institutional Review Board (IRB; IRB Approval Number
407 STUDY-11-00388). Briefly, at the time of autopsy, the entire brain was removed, photographed,
408 and bisected, and one hemisphere was placed in phosphate-buffered formalin for subsequent
409 processing, gross neuropathologic review, sectioning, and histological analysis (*115*). The other
410 hemisphere was immediately sectioned, each section photographed, and then was snap-frozen
411 between aluminum plates chilled to a temperature of -85° Celsius (C). Frozen brain sections were
412 stored at -85°C in a monitored facility. For this study, samples of ventral midbrain containing
413 recognizable pigmentation indicative of SN were dissected, using a rotary saw on a cold table
414 (chilled to -20°C). As there was some rostro-caudal variation in the anatomical level at which
415 each donor's SN was dissected, we accounted for the relative location as rostral midbrain (SN at
416 the transverse level of the red nucleus) or caudal midbrain (at the level of the cerebellar
417 peduncular decussation) and included this as a covariate in our differential expression analysis
418 (see below).
419
420 Prospectively collected clinical information was available for 67 donors (62 HIV+, 5 HIV-) as a
421 result of their participation in the MHBB longitudinal, observational study. For 23 other HIV-
422 donors, information was extracted by medical record review at the time of demise. Substance use
423 characterization (DSM IV dependency diagnoses) for those followed in the MHBB study was
424 obtained with either the Psychiatric Research Interview for Substance and Mental Disorders
425 (PRISM) versions 1.9b and SL (*124*) or the World Health Organization Composite International
426 Diagnostic Interview (CIDI) version 2.1 (*125*), and drug utilization was monitored by urine
427 toxicologic analysis for opiates, cocaine, and other psychoactive substances at each study visit, as
428 described previously (*126*). At the time of urine toxicology, the donor's reported medications
429 were examined to determine whether psychoactive substances were prescribed secondary to
430 medical illness. For donors not enrolled in the prospective MHBB study, DSM IV diagnoses and
431 problematic utilization were obtained from medical records, inclusive of urine toxicologic
432 analysis when available. Immunovirologic data (CD4 T-cell counts and plasma HIV VL) were
433 similarly obtained either as a prospective research study laboratory or retrospectively extracted
434 from medical records; all were performed in the CLIA-certified laboratories of the Mount Sinai
435 Hospital. In this study, for HIV+ individuals, we used a VL of 50 copies/mL as the threshold for
436 HIV detectability, to be consistent with the quantitative lower limit of the earliest assays utilized
437 in the MHBB cohort.
438
439 ***snRNA-seq***
440 We performed sample processing consisting of nuclei purification, RNA extraction, and
441 subsequent generation of snRNA-seq libraries, following almost exactly a previously described
442 method (*34*) for sequencing of pooled tissue samples. Briefly, we homogenized tissue aliquots
443 containing pooled ventral midbrain from SN using a douncer in a mixture of lysis buffer and
444 RNase inhibitor. In each library, we pooled tissue from 1 to 4 unique donors to increase
445 sequencing throughput; we also added a small volume of mouse tissue to most libraries (28/33) as
446 a negative control to create a threshold for spurious HIV reads (for future work). On pooled
447 tissue, we used a lysis and sucrose buffer gradient to isolate nuclei during ultracentrifugation. We
448 resuspended nuclei in a BSA solution containing RNase inhibitor, stained them with DAPI

nucleophilic dye, and used fluorescence-activated nuclei sorting (FANS) to collect 300,000 DAPI+ nuclei for input into the 10x Chromium NEXT GEM Single 3' v3.1 (Dual Index) Protocol according to manufacturer instructions. We quantified libraries using qPCR prior to sequencing on the Illumina NovaSeq platform. Sequencing was performed by the New York Genome Center. As described below, each library was separated by donor after sequencing and prior to analysis, using genotype-based demultiplexing, such that we performed our analyses on a per-donor basis (see section below).

We generated the first round of snRNA-seq libraries by pooling tissue from 3 to 4 patients per library, then performed QC and generated additional snRNA-seq libraries for donors with <1,000 QC-passing nuclei, in each pooling tissue from only 1 to 2 patients to increase the per-patient nuclear yield. We added libraries for low-yield donors until the tissue sample was exhausted or the total number of quality-passing nuclei exceeded 1,000 for 90% of donors, resulting in a total of 51 libraries across the 90 study donors. We determined the threshold of 1,000 QC-passing nuclei per donor using a power analysis (see below).

Imputed donor SNP arrays for demultiplexing

We used donor SNP arrays as a reference to demultiplex pooled snRNA-seq data, generating them from donor cerebellum tissue using the Illumina Infinium Global Screening Array platform (GSA-24v3.0). We prepared DNA for each donor at a final concentration of 20 ng/µL. Genotyping was performed at the Center for Applied Genomics Core at Children's Hospital of Pennsylvania (CHOP). We performed QC on SNP arrays using GenomeStudio v2.0 (*127*, *128*), retaining patient genotypes with call rate of >95% of SNPs, and SNPs with minor allele frequency (MAF) >0.2, call rate >99%, and clustering ("GenTrain") score >0.7. We performed a second round of SNP genotyping for three donors with SNP call rates <95%, resulting in two panels with 517,091 SNPs (614,679 variants total) and 416,166 SNPs (603,685 variants total) across chromosomal designations 1-22, X, Y, XY, PAR, and MT. To create a demultiplexing reference, we retained SNPs from chromosomes 1-22 for each panel, and we imputed larger arrays for these to improve demultiplexing accuracy (*30*) using the Trans-Omics for Precision Medicine (TOPMed) Imputation Server (*129–131*) with TOPMed r2 reference, EAGLE v2.4 phasing (*132*, *133*), no r^2 filter, data-TOPMed reference panel allele frequency comparisons, and pre-run QC. After pre-run QC, 482,995 and 212,783 SNPs from the two panels were used for phasing and imputation; after imputation the panels had 27,749,087 and 11,426,093 SNPs. We combined these imputed genotypes for donors into a single reference and ran demuxlet on each group of sequenced libraries we produced as described above (4 total). For each group, we filtered the demultiplexing reference using bcftools (*134*) to retain only biallelic SNPs that had a depth of coverage across donors >9, as well as estimated imputation accuracy r^2 >0.5 and MAF >0.2. Our resulting reference panels included 568,817; 480,835; 676,322; and 291,170 SNPs.

Alignment and cleaning of snRNA-seq data

We pre-processed sequenced data using multiple steps, first performing alignment, demultiplexing, and expression noise/doublet removal on each pooled library separately, then performing heuristic-based filtering on the data from all libraries combined, to generate per-donor data cleaned of ambient RNA, doublets, and other low-quality information.

1. Read alignment

We first aligned reads per library to a composite genome using Cellranger version 6.1.2 (using *count* with option *--include-introns* (*135*, *136*)). We built the composite genome reference using CellRanger's *mkref*, to be a combination of the GRCh38 human genome, the mm10 mouse genome, and a split version of the HIV-1 reference genome (previously shown to reduce

499 ambiguities in alignment of HIV-associated reads, increasing detection (*137*)). Notably, we only
500 detected two HIV reads in our data, likely related to both the presence of highly mutated HIV
501 dissimilar to the reference sequence and lower viral levels in our chronically infected, non-
502 encephalitic HIV+ donors (Discussion).
503
504 *2. Demultiplexing/multidonor doublet removal (demuxlet) + expression noise removal*
505 *(CellBender)*
506 Second, we performed two processes in parallel, demultiplexing and ambient RNA/noise
507 removal, using as input for each the human-genome-aligned reads for each library. We
508 demultiplexed data with demuxlet (*30*) and popscle (https://github.com/statgen/popscle), using
509 our donor SNP reference panel described above, doublet mixing fraction α ranging from 0 to 0.5,
510 and 0.1 doublet prior, which produced for each nuclear barcode both a label of either "singlet",
511 "doublet", or "ambiguous" and its maximum-likelihood donor ID(s). We confirmed for each
512 library that the donor IDs returned matched those expected.
513
514 We performed ambient RNA/noise removal using CellBender remove-background (*138*, *139*)
515 with the *remove-background-v3-alpha* workflow on the Terra.bio platform (*140*, *141*).
516 CellBender uses unsupervised learning to model and remove expression count noise, due to
517 ambient RNA molecules and random barcode swapping. For most libraries we ran CellBender
518 using the following parameters: false positive rate 0.01; learning rate 0.00005; posterior batch size
519 5; 150 epochs; and 2 retries. For a few libraries we made adjustments: CellBender failed to
520 converge during learning for library SNr11 with these parameters but achieved a reasonable
521 learning loss (indicative of successful learning) with learning rate 0.000025 over 200 epochs.
522 Another library (SNr27) showed signs of overfitting that was corrected in a re-run with learning
523 rate 0.00005 and 60 epochs. For each library, the output from CellBender was a cleaned
524 expression matrix in anndata (*142*) format.
525
526 We combined demuxlet and CellBender results by filtering each library's CellBender-cleaned
527 expression data to retain only demuxlet-assigned, nuclear singlet barcodes, then adding the
528 corresponding donor IDs to the expression data as barcode annotations.
529
530 *3. Multi-cell-type doublet removal (Scrublet)*
531 Third, we used Scrublet (*143*) to remove remaining doublets (barcodes with reads from multiple
532 cell types from one patient; running demuxlet removed multi-patient doublets), processing each
533 patient's nuclei within each library separately (134 patient groups across 51 libraries). Scrublet
534 removed 4,372 doublets in total, corresponding to median (IQR) 33 (8, 37) doublets per group, or
535 1.45% (0.82%, 2.22%) of barcodes per group. After this step, we presumed that most barcodes
536 corresponded to single nuclei and referred to them as nuclei.
537
538 *4. Heuristic filtering*
539 Fourth, we combined expression data across all libraries and performed heuristic filtering to
540 remove potentially low-quality or dying nuclei. We filtered based on fraction of mitochondrial
541 genes/nucleus, number of genes/nucleus, and number of reads (denoted by unique molecular
542 identifiers or "UMIs")/nucleus. For each heuristic, we produced the distribution of values across
543 all dataset nuclei, then performed filtering at thresholds that excluded nuclei in distribution tails
544 and were consistent with previously reported findings in human SN (*31*, *32*). As a result, we
545 removed nuclei with mitochondrial fractions >1.28% (this value marked an upper tail formed by
546 10% of our nuclei; 90% of all nuclei were at or below this value; Fig. S1A). Notably, our
547 threshold is stricter than conventional thresholds of 5% or 10% that are used to identify data from
548 potentially dying cells (*31*). We also removed nuclei with >4,396 genes (the value marking an

549 upper tail formed by 5% of all nuclei; 95% of nuclei passed this threshold; Fig. S1B) and any
550 nuclei with >17,153 UMIs (the value marking an upper tail formed by 5% of all nuclei; Fig. S1C).
551 Notably, we identified these thresholds each time a new group of sequenced data was added to
552 our dataset (during data generation; see above) and identified roughly the same thresholds as
553 those above during each of these intermediate stages.
554
555 ***snRNA-seq data dimensionality reduction, batch correction, and nucleus clustering***
556 To produce a finalized dataset for analysis, we performed three additional steps. First, we further
557 reduced expression count noise by performing two forms of linear dimensionality reduction,
558 retaining only genes with the highest expression variance (HVGs), and using principal component
559 analysis (PCA) on expression counts, keeping only principal components with non-negligible
560 explained variance. Second, we performed batch correction, to facilitate cross-library dataset
561 integration, and third, we performed clustering to generate initial nuclei groups for cell typing.
562 We explain several choices made in these steps to accommodate expected transcriptional
563 abnormalities for donors. Many analyses were performed using scanpy (*144*) and other packages
564 as noted.
565
566 We selected HVGs using scanpy.pp.highly_variable_genes() with the "seurat_v3" option, which
567 ranks genes by cross-nucleus variance (*33*). We found that using the suggested 2,000 HVGs
568 excluded many marker genes we used for cell typing and instead included genes, like
569 inflammatory markers, that might relate more to cell state. To retain more cell type marker genes
570 but still reduce noise, we ordered all 36,601 genes in our data by increasing HVG rank (i.e. by
571 decreasing expression variance) and inspected the simultaneous decrease in minimum expression
572 variance and increase in number of cell type marker genes that occurred as we included more
573 genes as HVGs (Fig. S1E). We selected the number of HVGs to be a value past the elbow of the
574 "number of cell type marker genes" curve, such that beyond this value including additional HVGs
575 would introduce cell type marker genes at slower rates, at the cost of adding progressively lower-
576 variance data (a diminishing return). We kept 20,000 HVGs, which included 67% of the cell type
577 marker genes in our database (Data S6). Of note, working with this number of HVGs posed a
578 substantial memory burden, since several scanpy functions by default convert expression counts
579 to high-precision data types (64-bit). To alleviate this burden, we created modified versions of
580 those scanpy functions for downstream analysis, which required expression data to be maintained
581 at a user-specified precision, then built a function to determine the smallest data precision that
582 would accommodate our expression data (32-bit) and used this value as input to those functions.
583
584 We performed PCA on HVG-selected expression data (scanpy.tl.pca()) and retained the first 25
585 components (Fig. S1F) for batch correction, subsequent UMAP (Uniform Manifold
586 Approximation and Projection) (*145*, *146*) generation, clustering, as well as, notably, for
587 downstream analysis. We preserved PCA-reduced, batch-corrected expression data for
588 downstream analysis by projecting it back to the barcodes-by-genes basis from the truncated PCA
589 basis (details below). To facilitate this last step, we stored the means and standard deviations used
590 to scale data prior to PCA, and PCA loadings. We performed scaling and stored scaling
591 parameters using a data-type-modified version of scanpy.pp.scale(), with zero centering and no
592 expression count capping (to recover the original counts data in the limit of no PCA truncation or
593 batch correction).
594
595 We ran batch correction using Harmony (*147*) with snRNA-seq library ID as the batch variable
596 and 20 maximum iterations; Harmony converged in 3 iterations. We inspected batch correction
597 quality by comparing 2D UMAP localizations for nuclei from the same patient but different

libraries before vs. after Harmonization. We found that post-correction colocalization was improved (Data S5).

Batch-corrected data in the PCA basis was used to make a UMAP-based neighborhood connectivity graph using scanpy.pp.neighbors() with 20 neighbors for local manifold approximation; we then generated 2D UMAPs using scanpy.tl.umap() and initial Leiden clusters for cell typing using scanpy.tl.leiden() with resolution parameter 0.5.

We projected batch-corrected, scaled expression data back to the gene basis from the PCA basis using the 25 stored PCA loadings, then made the resulting scaled expression data count-like, ensuring non-negative counts, then rescaled it using stored means and standard deviations (Fig. S4). Batch-corrected distributions sometimes had non-negligibly negative values (indicative of barcodes with minimal, effectively zero, HVG counts relative to others). Most of these values were between -1 and 0 (Fig. S4B), and any adjusted count values of <-1 occupied <1% of any given HVG's corrected distribution (Fig. S4C). We thus set these negative values to zero; our approach for doing so was to zero the lowest 1% of each HVG's count distribution, and notably, this approach had no discernible effect on existing non-negative count values because a substantial percentage of each HVG's distribution already consisted of zero counts. We then shifted each distribution so its minimum matched its uncorrected counterpart and would produce non-negative values upon rescaling, and finally rescaled the values, set any remaining negative values to zero, and converted to integer-type.

Likelihood-based cell typing

Because we observed indications of abnormal transcription, we wanted to avoid assumptions about cell types that were present in our data or the most appropriate marker genes to use for cell typing. We thus developed an approach for cell typing based on the principle employed in ScType, a computational cell typing tool (*148*). ScType estimates cell types using a large, multi-tissue, multi-organism database of cell-type-specific "positive" and "negative" marker genes that are expressed in or suppressed in a cell type, respectively. For a group of unidentified nuclei or cells and specified tissue(s), ScType computes a "cell type score" for each nucleus and possible cell type, which is a weighted sum of the nucleus's standardized expression levels across the possible cell type's positive and negative marker genes. Weights each reflect a "marker gene specificity score", reducing the impact of less-specific marker genes, and have negative sign for negative marker genes to penalize the score when these are expressed; ScType assigns the cell type with the maximum score to each nucleus.

Our approach is similar to ScType's in that we compare gene expression levels against a cell type marker gene database, but in other respects it differs considerably. We perform cell typing on nuclear clusters rather than individual nuclei, with the reasoning that clusters, by definition, group nuclei with similar transcriptional profiles, and considering them may highlight genes with characteristic expression levels that may be more helpful for cell typing. We identify such "distinguishing HVGs" for each cluster by finding HVGs that are differentially overexpressed in the cluster vs. all other clusters combined, using scanpy.tl.rank_genes_groups() with a k-fold, leave-one-out cross-fold validation (k=10) scheme on scaled (using scanpy.pp.normalize_total()), log1p-transformed (using scanpy.pp.log1p()) data. We ran this function with the Wilcoxon-rank-sum comparison and Benjamini-Hochberg p-value correction. We considered genes that were overexpressed across all k=10 calculations with log2 fold change (l2fc)>1.0 and adjusted p<0.01 to be distinguishing HVGs, and we used these (as opposed to all HVGs) for comparison of the nuclear cluster against our cell type marker gene database.

The cell type database we use in this study is a modified subset of the ScType database (*148*) that contains marker genes for relevant human tissues (brain, immune tissue, and smooth muscle tissue, which has cell types resembling those in the blood-brain-barrier). Some ScType cell types appeared in multiple tissues, and we removed any redundancies in these, either merging types with identical marker gene lists or assigning tissue-specific names to cell types with differing lists (e.g. renaming "Endothelial Cell" to "Brain Endothelial Cells"). We then added marker genes from recent postmortem human SN snRNA-seq studies (*34*, *149*, *150*) to the database. We also retained cell types not expected for SN (e.g. glutamatergic neurons, fibroblasts, cancer cells) to detect potential sample issues and to better characterize unexpected phenotypes. Finally, although we pulled positive and negative marker genes from ScType and have included an option to use them in our code, we only used positive marker genes in this study, to avoid assuming that negative markers would be absent for cells in non-normative states. The final database that we used in our analysis here has 61 cell types, with median (IQR) 14 (6, 27) positive marker genes per cell type, and 515 positive marker genes total (Data S6).

In our approach, we use likelihoods to assess whether a cluster is of a particular cell type. Explicitly, for a cluster *n* of nuclei, and cell type *c*, we define the likelihood *L* as the joint probability that each of *c*'s markers, mg_c^i (where *i* is an identifying index running from 1 to the total number of marker genes *M*) is a true, defining marker gene for *n* (meaning that it is in the set of *n*'s distinguishing HVGs, which we denote as mg_n):

$$mg_c^i \in \{mg_n^j\} \tag{1}$$

(where *j* is another identifying index). We assume that the probability of mg_c^i being a marker for *n* is a monotonically increasing function of its overexpression level in *n*, which we denote as $o_{c,i}^n$. By this, we mean that if mg_c^i was not overexpressed in *n*, its probability of being a marker for *n* is ~0, and if it is overexpressed, its probability of being a marker increases with increasing $o_{c,i}^n$, saturating at ~1 above some overexpression level:

$$Prob(mg_c^i \in \{mg_n^j\}) \sim f(o_{c,i}^n), f \text{ increasing with } o_{c,i}^n \tag{2}$$

We make an approximation of linear dependence on $o_{c,i}^n$,

$$Prob(mg_c^i \in \{mg_n^j\}) \sim f(o_{c,i}^n) \sim o_{c,i}^n \tag{3}$$

and we make the simplifying assumption that the probability functions for $\{mg_c^i\}$ can reasonably be treated as independent. The likelihood then takes the form

$$L(n,c) = \prod_{i=1}^{M} o_{c,i}^n \tag{4}$$

In our calculations we use the log-likelihood *LL*:

$$LL(n,c) = \sum_{i=1}^{M} log(o_{c,i}^n) \tag{5}$$

If negative marker genes were to be used (they are not in this study), the log likelihood becomes

$$LL(n,c) = \sum_{i=1}^{Mpos} log(o_{c,i}^n) - \sum_{j=1}^{Mneg} log(o_{c,j}^n) \tag{6}$$

We do not apply cell-type-specificity or other weighting to marker gene likelihood terms, to avoid inadvertently making biased assumptions. For example, by not applying weights, we avoid inappropriately upweighting marker genes that appear for fewer database cell types as a result of missing data rather than biological specificity. We also avoid discounting less-specific marker genes that may nevertheless be informative for cell typing.

We construct the log-likelihood in Eq. (5) by setting $log(o_{c,i}^{n})$, the log overexpression for marker gene *i* of cell type *c*, to either the mean (across *k* calculations) l2fc expression of that gene in cluster *n* if the gene was a distinguishing HVG (and thus significantly overexpressed), or 0 if not (overexpression level $o_{c,i}^{n}$=1). Using this measure, we observe substantial dropoff in log-likelihood as the overlap between a cluster's distinguishing HVG's and a cell type's marker genes decreases (Fig. S5), leaving only a few putative cell types. We used the highest-likelihood cell types and the contributing database marker genes for each to assign a cell type to each nuclear cluster, and we used this approach in multiple rounds to arrive at final cell types. First, we ran cell typing on our original Leiden clusters, then combined contiguous clusters with the same cell type, then performed subclustering on any of the resulting clusters that had visible substructure (8 total), again using a Leiden algorithm but with resolution 0.1 on each cluster in isolation. Subclustering had the effect of splitting nuclei groups with different cell types that were originally merged, including microglia and macrophages, and GABAergic and dopaminergic neurons. Finally, we performed cell typing again on these re-clustered nuclei, generating the cell types in Fig. 1. We found our cell types to be in good agreement with previously published human SN snRNA-seq data (Fig. S3 (*34*)).

Differential expression analysis (DEA)

We performed Wald-formulation DEAs on each cell type separately with *k*-fold cross-validation (*k*=3 set by power analysis; below), using DESeq2 (*43*) ported into Python with the package diffexpr (specifically the module pyDESeq2). For SUD+ vs. SUD- comparisons, the generalized linear model (GLM) we used as DESeq2 input was

$$\begin{gathered} gene\ expression\ level \sim HIV\ level, SUD_{opc}\ donor\ group\ + SUD_{other} + \\ age\ +\ sex\ +\ race\ +\ ethnicity\ + \\ overdose\ death\ status\ +\ PMI\ +\ tissue\ location\ + \\ CD4\ nadir \end{gathered} \tag{7}$$

For HIV comparisons (HIV+u vs. HIV-; HIV+d vs. HIV-; HIV+d vs. HIV+u), our GLM was

$$\begin{gathered} gene\ expression\ level \sim HIV\ level, SUD_{opc}\ donor\ group +\ SUD_{other} + \\ age\ +\ sex\ +\ race\ +\ ethnicity\ + \\ overdose\ death\ status\ +\ PMI\ +\ tissue\ location \end{gathered} \tag{8}$$

Covariates are defined as follows:

HIV level: Categorical. Final HIV viral load in copies RNA/mL. (Options are "HIV-", "HIV+u" for <50 copies/mL, "HIV+d" for ≥50 copies/mL)

SUD_{opc}: Categorical. Presence of opioid and/or cocaine SUD diagnosis. ("yes", "no")

SUD_{other}: Categorical. Presence of SUD diagnosis for substance other than opioids/cocaine. ("yes", "no")

Age: Numeric. Donor age in years.

Sex: Categorical. Donor sex assigned at birth. ("female", "male")

Race: Categorical. Donor race described in clinical records. ("Black or African American", "White", "Asian", "Black and White", "Black and Indian", "Unknown or Not Reported")
Ethnicity: Categorical. Donor ethnicity described in clinical records. ("Hispanic or Latino", "Not Hispanic or Latino")
Overdose Death Status: Categorical. Presence of overdose as cause of death. ("yes", "no")
Postmortem Interval (PMI): Numeric. Donor post-mortem interval in hours.
Tissue Location: Categorical. Approximate anatomical location of SN sample. ("coronal" for rostral midbrain, "midbrain" for caudal midbrain)
CD4 Nadir: Categorical. Lowest donor CD4 T-cell count. ("<200 cells/μL", "[200,500] cells/μL", ">500 cells/μL"). HIV- donors are assigned ">500 cells/μL".

Prior to running DESeq2, numerical covariates were standardized, and categorical covariates that covaried perfectly were identified, and all but one were removed, to avoid ambiguity in fitting. None of the DEAs in this study required us to drop covariates.

DESeq2 implements multiple controls to remove false positives (*43*), and we used k-fold cross-validation as an additional layer of false positive control for each DEA. In k-fold cross-validation, random subsetting of data is expected to randomly alter low gene dispersions. Low dispersions can exist for genes with prominent within-donor correlations (e.g. if there are many similar nuclei per donor relative to the rest of the DEA data) and/or low counts, and they can inflate the apparent effect size of expression changes, leading to false positives. The randomly altered dispersions across k-fold cross-validation computations alters the degree to which such genes are false positives. We filtered to include only DEGs that show up in all *k* random subsets (below), which helps to remove these impacts. DESeq2 can also be run on single-nucleus data rather than pseudobulked data (i.e. with nuclei aggregated per cell type and donor); the single-nucleus approach can help to highlight DEGs that may be pertinent to subsets of a cell type if the nuclei exhibit biological variability, and for this reason we ran DESeq2 on our single-nucleus data, without pseudobulking. DESeq2 authors have made recommendations for run parameters that increase the number of "true positive" DEGs returned when running on single-nucleus data by handling zero inflation characteristic of single-cell approaches (*42*); we have confirmed that those recommendations reproduce our results and add additional DEGs. Importantly, our procedure is more restrictive in terms of DEGs returned but still benefits from false positive controls and does not produce DEGs not seen when run with DESeq2-author-recommended parameters. Below, we show one example of this comparison.

Parameters used in our study: Wald test, useT=False, default sizeFactors, and pseudocount addition for dispersion modeling. Parameters recommended in (*42*): likelihood ratio test, useT=True, minmu=1e-6, sizeFactors computed using scran::computeSumFactors.

Cell type	DEA	N up DEGs reported in our paper	N up DEGs common to both approaches	N down DEGs reported in our paper	N down DEGs common to both approaches
DA	SUD+HIV- vs. SUD-HIV-	0	0	51	49
DA	SUD+HIV+u vs. SUD-HIV+u	82	76	41	41
DA	SUD+HIV+d vs. SUD-HIV+d	0	0	0	0

We subsetted all our data for k-fold cross-validation at once. Then, for each of our *k* computations per DEA, we fit the appropriate GLM to the data subset for all donor groups, and we then pulled the results of each DEA from the fitted model. We included an extra term for CD4 nadir in our SUD comparisons (Eq. 7). This is because, although we split HIV groups by the clinically relevant measure of final VL, we realized that other longer-term aspects of a donor's HIV infection history may impact transcription, most notably, damage incurred during periods of low T-cell counts. We thus added CD4 nadir to account for this type of long-term impact and aid detection of SUD impacts. CD4 nadir tracked closely with final HIV VL for HIV+ donors, accounting for a small degree of HIV-related variation. For HIV comparisons, we excluded CD4 nadir because it closely recapitulated HIV VL variation.

In each DEA, we identified differentially expressed genes (DEGs) as those with adjusted $p<0.05$ across all *k* calculations. For significant DEGs, we use the cross-fold means of l2fc and normalized counts in Figs. 2-5 and for gene set enrichment analysis.

Gene set enrichment analysis (GSEA)

We performed GSEA with Enrichr (*44*) using GSEApy, a Python wrapper for R-based GSEA tools (*151*). For each DEA, we ran Enrichr separately for up- and down- regulated DEGs (gseapy.enrichr()), using our 20,000 HVGs as background and human annotated gene set reference "GO_Biological_Process_2023" (*152*, *153*). We saved all gene sets returned regardless of significance and considered gene sets with adjusted $p<0.05$ to be significantly endorsed (Data S2). We also tested the tool GSEA (*154*, *155*) in GSEApy, and found that it identified similar but generally fewer significant gene sets for DEGs (Table S7, Data S7).

Identification of highly coordinated DEG subclusters

To find DEG subclusters for a particular DEA and cell types of interest, we first pulled the normalized (using scanpy.pp.normalize_total()) count data (not any result from the DEA itself) for all nuclei of each cell type, across all that DEA's case group donors. Then, for each cell type, we computed, for each of its DEGs, the mean normalized expression count across that cell type's nuclei for each case donor. This step produced a matrix of cell-type-specific DEGs per row, and the mean normalized expression of that DEG for each case donor per column. To distinguish DEGs that arose for multiple cell types, and to keep track of the direction of dysregulation, we named DEGs using the convention c*ell-type_DEG-direction_gene-name*. For all DEGs across all cell types, we then computed cross-donor Pearson correlations in mean expression for DEG pairs using scipy.stats.pearsonr(), using the built-in beta-assumption hypothesis test to determine significance (two-tailed alternative of nonzero correlation) with significance level $p=0.05$. Across all our computations (Table S8) there were median (IQR) 30% (28%, 34%) DEG pairs with significant correlations.

We computed the similarity for each DEG pair to be the correlation magnitude if it was significant and 0 if not; or, for DEGs *i,j*

$$similarity_{i,j} = sim_{i,j} = abs(r_{i,j}), p_{i,j} < 0.05;\ 0\ otherwise \tag{9}$$

In order to use hierarchical agglomerative clustering to group DEGs, we transformed similarities to distances using

$$dist_{i,j} = \frac{1}{(1 + sim_{i,j})} \tag{10}$$

adding 1 to the similarity to avoid division by zero. We used Eq. 10 instead of $dist_{i,j} = 1 - sim_{i,j}$ because its nonlinearity required DEG clusters to have more pronounced similarity to be detected, by inflating distances between higher-similarity DEG pairs, and thus by suppressing the dynamic range of distances used to parse clusters. We performed clustering using sklearn.cluster.AgglomerativeClustering() with average linkage distance, using the parameters "distance_threshold"=0, "n_clusters"=None to compute the entire linkage tree and return algorithm-generated DEG clusters. The resulting DEG clusters typically consisted of noticeable subclusters with inter-subcluster distances substantially larger than intra-group distances, so we split them at points where the distance between adjacent DEGs was ≥3 standard deviations above the cluster's mean adjacent DEG distance. To keep track of original cluster identity, we named DEG subclusters using the convention C*[cluster-index]*S*[subcluster-index]* (e.g. "C17S2" for cluster 17, subcluster 2). We characterized subcluster coordination strength by computing the average similarity across all DEG pairs in a subcluster. Some subclusters had exceptionally low average similarities (~0.2 vs. ~0.8; Fig. 6A, right), indicating weaker overall relationships with other DEGs. To focus on annotating the most highly coordinated DEG groups, we removed these low-similarity clusters by excluding anything in the lower tail of the distribution of average similarities across all DEG subclusters (<0.75; Fig. 6A right).

DEG subcluster annotations

We produced a skeleton annotation .csv file for each DEG subcluster using "GO_Biological_Process_2023" annotations and the acyclic graph of their relationships ("go-basic.obo"). For each DEG, we found all related GO terms and filtered them to include only the most subordinate annotation represented on each graph branch (most specific process). We split remaining terms into those explicitly mentioning positive regulation of a process, explicit negative regulation, explicit response, and others, as an initial basis for annotation (Data S3). We used DEG naming convention *cell-type_DEG-direction_gene-name* to facilitate interpretation.

For focused DEG subcluster annotations, two experts (MJ, AW) independently searched different gene databases for each DEG (MJ, GeneCards; AW, National Center for Biotechnology Information gene database) and pulled the functional summary. They then performed a literature search for the DEG using terms related to HIV, SUD, brain location ("substantia nigra", "ventral midbrain"), and cell type (e.g. "gene *x* in neurons", "gene *x* in dopaminergic neurons", "gene *x* in microglia"). Relevant references were collected and used to annotate potentially relevant functions for each gene; these results were then used to generate a consensus across annotators that was used in each DEG subcluster (Data S4).

Statistical Analysis

snRNA-seq sample size selection

We used a power analysis simulation to estimate the number of nuclei per donor that would be needed to power detection of cell-type-specific expression differences. We simulated the case where expression data was compared across two groups for one gene, in one cell type, and there was a small but true difference in the two distributions. We defined each group's underlying expression distribution for the gene to have a lognormal distribution, i.e.

$$Prob(x\ counts) \sim Lognormal(\mu, \sigma^2) \ = \ exp(N(\mu, \sigma^2)) \tag{11}$$

We let both distributions have the same variance of 1 and means differing by some l2fc value. The difference between means μ_1 and μ_2 was related to the l2fc by

$$abs(\mu_1 - \mu_2) = log_e(2^{log2fc}) \quad (12)$$

For each l2fc we tested (0.5, 1.0, 1.5), we defined the two underlying distribution means using Eq. (12), then over 10,000 iterations sampled distributions of size *s* (nuclei) from each and performed a Wilcoxon rank sum test. We computed the power as the fraction of iterations that the true difference was detected with p<0.05. We repeated this simulation beginning with *s*=10 and incrementing *s* by 2 until reaching a power of at least 80%. We achieved sufficient detection power with *s*=158 for l2fc 0.5; *s*=42 for l2fc 1.0; and *s*=20 for l2fc 1.5. Based on this simplified situation, we set a threshold of ~100 nuclei per cell type, aiming for detection of expression differences at l2fc 1.0. Prior to analysis, we estimated there would be ~10 expected cell types in SN (DA and GABA neurons, microglia, astrocytes, ODCs, OPCs, multiple blood-brain-barrier cells, multiple immune cells (*149*, *156*)) and thus aimed to sequence 1,000 nuclei per donor.

Overexpression level and k selection for cluster "distinguishing HVG" detection

We used a similar power analysis to estimate the l2fc at which distinguishing HVG detection was powered for every cluster of nuclei. We set a significance threshold of p=0.01 to restrict detected distinguishing HVGs to a higher confidence subset, since these could impact cell typing and thus all downstream analysis. We let distributions represent expression of a gene for one cluster vs. the rest, choosing normal distributions with equal variance and differing means to reflect log1p pre-processing applied for scanpy.tl.rank_genes_groups(). Means and l2fc were related by

$$\mu_1 = 1, \mu_2 = 2^{log2fc} \quad (13)$$

At l2fc=1.0, every cluster exceeded the minimum sample size for 80% detection power. We determined the number of folds *k* for cross-validation, by increasing *k* from 2 in increments of 1 until every cluster with one of its *k* folds removed had enough nuclei to also power detection. Our code also had an option for the resulting *k* to be increased by 1, to put each one-fold-removed group of nuclei further over the power detection threshold during cross-fold validation calculations. We enabled this option, finding *k*=9 and using *k*=10.

Overexpression threshold and k selection for DEA

We used the same power simulation to estimate the values of l2fc and *k* that would power DEAs for each cell type at p=0.05, using *k*=3 and l2fc=1.0.

Significance testing of cell type proportions across donor groups

We tested whether nuclei for each cell type were differently represented across donor groups by first using the binomial distribution to model cell type proportions for each donor group. For a donor with *k* nuclei of a specified cell type and *n* nuclei total, the maximum likelihood estimation for the probability of that cell type's occurrence, $\hat{p}$, is

$$\hat{p} = \frac{k}{n} \quad (14)$$

Within each donor group, we computed for each donor the $\hat{p}$ for each cell type, producing a distribution of each cell type's occurrence probability across donors in the group. We performed Kolmogorov-Smirnov tests to compare these distributions per cell type for each pair of donor groups, using the Benjamini-Hochberg procedure to correct for multiple comparisons (12 comparisons for each donor group pair, one per cell type), and considering donor group pairs with adjusted p<0.05 to have significant differences (Table S5).

Significance test for DEG overlaps

935 We performed random sampling tests to determine whether fractions of overlapping DEGs
936 observed for related DEAs (Figs. 2-5) were significantly different from expected outcomes of
937 random sampling. For each DEA pair, we sampled the observed numbers of DEGs per DEA from
938 the 20,000-HVG background, and we recorded the fraction of overlapping DEGs between the two
939 simulated populations. We repeated sampling for 100,000 iterations and computed the upper limit
940 on the p-value to be the fraction of iterations with as least as many overlapping DEGs as
941 observed. In all tests we found 0 such iterations, such that $p<10^{-5}$.
942
943 **References**

944 1. The path that ends AIDS: UNAIDS Global AIDS Update 2023.
945 https://www.unaids.org/en/resources/documents/2023/global-aids-update-2023.

946 2. Full report — In Danger: UNAIDS Global AIDS Update 2022.
947 https://www.unaids.org/en/resources/documents/2022/in-danger-global-aids-update.

948 3. L. T. Le, S. S. Spudich, HIV-Associated Neurologic Disorders and Central Nervous System
949 Opportunistic Infections in HIV. *Semin. Neurol.* **36**, 373–381 (2016).

950 4. D. B. Clifford, B. M. Ances, HIV-associated neurocognitive disorder. *Lancet Infect. Dis.* **13**,
951 976–986 (2013).

952 5. S.-C. Wang, B. Maher, Substance use disorder, intravenous injection, and HIV infection: A
953 review. *Cell Transplant.* **28**, 1465–1471 (2019).

954 6. R. M. Klevens, D. J. Hu, R. Jiles, S. D. Holmberg, Evolving epidemiology of hepatitis C
955 virus in the United States. *Clin. Infect. Dis.* **55 Suppl 1**, S3-9 (2012).

956 7. R. Moss, B. Munt, Injection drug use and right sided endocarditis. *Heart* **89**, 577–581 (2003).

957 8. T. M. Kelly, D. C. Daley, Integrated treatment of substance use and psychiatric disorders.
958 *Soc. Work Public Health* **28**, 388–406 (2013).

959 9. A. Gupta, D. Ashour, F. M. Shebl, L. Platt, J. J. Chiosi, S. B. Nelson, K. L. Ard, A. Y. Kim,
960 I. V. Bassett, Evaluation of Hepatitis C Treatment Outcomes Among Patients Enrolled in
961 Outpatient Parenteral Antibiotic Therapy-Boston, Massachusetts, 2016-2021. *Open Forum*
962 *Infect Dis* **10**, ofad342 (2023).

963 10. N. D. Volkow, C. Blanco, Substance use disorders: a comprehensive update of classification,
964 epidemiology, neurobiology, clinical aspects, treatment and prevention. *World Psychiatry* **22**,
965 203–229 (2023).

966 11. Y. Zhang, T. E. Wilson, A. Adedimeji, D. Merenstein, J. Milam, J. Cohen, M. Cohen, E. T.
967 Golub, The Impact of Substance Use on Adherence to Antiretroviral Therapy Among HIV-
968 Infected Women in the United States. *AIDS Behav.* **22**, 896–908 (2018).

969 12. C. H. Hinkin, T. R. Barclay, S. A. Castellon, A. J. Levine, R. S. Durvasula, S. D. Marion, H.
970 F. Myers, D. Longshore, Drug use and medication adherence among HIV-1 infected
971 individuals. *AIDS Behav.* **11**, 185–194 (2007).

972 13. S. E. Cohn, H. Jiang, J. A. McCutchan, S. L. Koletar, R. L. Murphy, K. R. Robertson, A. M.
973 de St Maurice, J. S. Currier, P. L. Williams, Association of ongoing drug and alcohol use
974 with non-adherence to antiretroviral therapy and higher risk of AIDS and death: results from
975 ACTG 362. *AIDS Care* **23**, 775–785 (2011).

976 14. D. R. Bangsberg, K. Ragland, A. Monk, S. G. Deeks, A single tablet regimen is associated
977 with higher adherence and viral suppression than multiple tablet regimens in HIV+ homeless
978 and marginally housed people. *AIDS* **24**, 2835–2840 (2010).

979 15. H. de Wit, Impulsivity as a determinant and consequence of drug use: a review of underlying
980 processes. *Addict. Biol.* **14**, 22–31 (2009).

981 16. V. Purohit, R. Rapaka, D. Shurtleff, Drugs of abuse, dopamine, and HIV-associated
982 neurocognitive disorders/HIV-associated dementia. *Mol. Neurobiol.* **44**, 102–110 (2011).

983 17. A. Ilango, A. J. Kesner, K. L. Keller, G. D. Stuber, A. Bonci, S. Ikemoto, Similar roles of
984 substantia nigra and ventral tegmental dopamine neurons in reward and aversion. *J. Neurosci.*
985 **34**, 817–822 (2014).

986 18. R. A. Wise, Roles for nigrostriatal--not just mesocorticolimbic--dopamine in reward and
987 addiction. *Trends Neurosci.* **32**, 517–524 (2009).

988 19. G. F. Koob, N. D. Volkow, Neurocircuitry of addiction. *Neuropsychopharmacology* **35**, 217–
989 238 (2010).

990 20. K. C. Berridge, The debate over dopamine's role in reward: the case for incentive salience.
991 *Psychopharmacology (Berl.)* **191**, 391–431 (2007).

992 21. M. D. Namba, Q. Xie, J. M. Barker, Advancing the preclinical study of comorbid neuroHIV
993 and substance use disorders: Current perspectives and future directions. *Brain Behav. Immun.*
994 **113**, 453–475 (2023).

995 22. E. A. Nickoloff-Bybel, P. Mackie, K. Runner, S. M. Matt, H. Khoshbouei, P. J. Gaskill,
996 Dopamine increases HIV entry into macrophages by increasing calcium release via an
997 alternative signaling pathway. *Brain Behav. Immun.* **82**, 239–252 (2019).

998 23. P. J. Gaskill, T. M. Calderon, A. J. Luers, E. A. Eugenin, J. A. Javitch, J. W. Berman, Human
999 immunodeficiency virus (HIV) infection of human macrophages is increased by dopamine: a
1000 bridge between HIV-associated neurologic disorders and drug abuse. *Am. J. Pathol.* **175**,
1001 1148–1159 (2009).

1002 24. M. Obermann, M. Küper, O. Kastrup, O. Yaldizli, S. Esser, J. Thiermann, E. Koutsilieri, G.
1003 Arendt, H.-C. Diener, M. Maschke, German Competence Network HIV/AIDS, Substantia
1004 nigra hyperechogenicity and CSF dopamine depletion in HIV. *J. Neurol.* **256**, 948–953
1005 (2009).

1006 25. V. Purohit, R. Rapaka, J. Frankenheim, A. Avila, R. Sorensen, J. Rutter, National Institute on
1007 Drug Abuse symposium report: drugs of abuse, dopamine, and HIV-associated
1008 neurocognitive disorders/HIV-associated dementia. *J. Neurovirol.* **19**, 119–122 (2013).

1009 26. A. M. Kumar, J. B. Fernandez, E. J. Singer, D. Commins, D. Waldrop-Valverde, R. L.
1010 Ownby, M. Kumar, Human immunodeficiency virus type 1 in the central nervous system
1011 leads to decreased dopamine in different regions of postmortem human brains. *J. Neurovirol.*
1012 **15**, 257–274 (2009).

1013 27. R. J. Battjes, R. W. Pickens, H. W. Haverkos, Z. Sloboda, HIV risk factors among injecting
1014 drug users in five US cities. *AIDS* **8**, 681–687 (1994).

1015 28. N. Bertholet, M. R. Winter, T. Heeren, A. Y. Walley, R. Saitz, Polysubstance Use Patterns
1016 Associated With HIV Disease Severity Among Those With Substance Use Disorders: A
1017 Latent Class Analysis. *J. Stud. Alcohol Drugs* **84**, 79–88 (2023).

1018 29. M. Karamouzian, A. Pilarinos, K. Hayashi, J. A. Buxton, T. Kerr, Latent patterns of
1019 polysubstance use among people who use opioids: A systematic review. *Int. J. Drug Policy*
1020 **102**, 103584 (2022).

1021 30. H. M. Kang, M. Subramaniam, S. Targ, M. Nguyen, L. Maliskova, E. McCarthy, E. Wan, S.
1022 Wong, L. Byrnes, C. M. Lanata, R. E. Gate, S. Mostafavi, A. Marson, N. Zaitlen, L. A.
1023 Criswell, C. J. Ye, Multiplexed droplet single-cell RNA-sequencing using natural genetic
1024 variation. *Nat. Biotechnol.* **36**, 89–94 (2018).

1025 31. A. Subramanian, M. Alperovich, Y. Yang, B. Li, Biology-inspired data-driven quality control
1026 for scientific discovery in single-cell transcriptomics. *Genome Biol.* **23**, 267 (2022).

1027 32. J. D. Welch, V. Kozareva, A. Ferreira, C. Vanderburg, C. Martin, E. Z. Macosko, Single-cell
1028 multi-omic integration compares and contrasts features of brain cell identity. *Cell* **177**, 1873-
1029 1887.e17 (2019).

1030 33. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck 3rd, Y. Hao, M.
1031 Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell* **177**,
1032 1888-1902.e21 (2019).

1033 34. J. Wei, T. Y. Lambert, A. Valada, N. Patel, K. Walker, J. Lenders, C. J. Schmidt, M.
1034 Iskhakova, A. Alazizi, H. Mair-Meijers, D. C. Mash, F. Luca, R. Pique-Regi, M. J. Bannon,
1035 S. Akbarian, Single nucleus transcriptomics of ventral midbrain identifies glial activation
1036 associated with chronic opioid use disorder. *Nat. Commun.* **14**, 5610 (2023).

1037 35. T. Kamath, A. Abdulraouf, S. J. Burris, J. Langlieb, V. Gazestani, N. M. Nadaf, K.
1038 Balderrama, C. Vanderburg, E. Z. Macosko, Single-cell genomic profiling of human
1039 dopamine neurons identifies a population that selectively degenerates in Parkinson's disease.
1040 *Nat. Neurosci.* **25**, 588–595 (2022).

1041 36. N. L. Ostrowski, A. Pert, Substantia nigra opiate receptors on basal ganglia efferents. *Brain*
1042 *Res. Bull.* **38**, 513–523 (1995).

1043 37. R. Ting-A-Kee, D. van der Kooy, The neurobiology of opiate motivation. *Cold Spring Harb.*
1044 *Perspect. Med.* **2** (2012).

1045 38. E. Galaj, X. Han, H. Shen, C. J. Jordan, Y. He, B. Humburg, G.-H. Bi, Z.-X. Xi, Dissecting
1046 the Role of GABA Neurons in the VTA versus SNr in Opioid Reward. *J. Neurosci.* **40**,
1047 8853–8869 (2020).

1048 39. E. A. Nickoloff-Bybel, T. M. Calderon, P. J. Gaskill, J. W. Berman, HIV Neuropathogenesis
1049 in the Presence of a Disrupted Dopamine System. *J. Neuroimmune Pharmacol.* **15**, 729–742
1050 (2020).

1051 40. A. Plaza-Jennings, S. Akbarian, Genomic Exploration of the Brain in People Infected with
1052 HIV-Recent Progress and the Road Ahead. *Curr. HIV/AIDS Rep.* **20**, 357–367 (2023).

1053 41. D. M. Thomas, D. M. Francescutti-Verbeem, D. M. Kuhn, Gene expression profile of
1054 activated microglia under conditions associated with dopamine neuronal damage. *FASEB J.*
1055 **20**, 515–517 (2006).

1056 42. K. Van den Berge, F. Perraudeau, C. Soneson, M. I. Love, D. Risso, J.-P. Vert, M. D.
1057 Robinson, S. Dudoit, L. Clement, Observation weights unlock bulk RNA-seq tools for zero
1058 inflation and single-cell applications. *Genome Biol.* **19**, 24 (2018).

1059 43. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for
1060 RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

1061 44. E. Y. Chen, C. M. Tan, Y. Kou, Q. Duan, Z. Wang, G. V. Meirelles, N. R. Clark, A.
1062 Ma'ayan, Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool.
1063 *BMC Bioinformatics* **14**, 128 (2013).

1064 45. L. Gu, J. Caprioli, N. Piri, Rbfox1 expression in amacrine cells is restricted to GABAergic
1065 and VGlut3 glycinergic cells. *Biosci. Rep.* **42** (2022).

1066 46. C. K. Vuong, W. Wei, J.-A. Lee, C.-H. Lin, A. Damianov, L. de la Torre-Ubieta, R. Halabi,
1067 K. O. Otis, K. C. Martin, T. J. O'Dell, D. L. Black, Rbfox1 Regulates Synaptic Transmission
1068 through the Inhibitory Neuron-Specific vSNARE Vamp1. *Neuron* **98**, 127-141.e7 (2018).

1069 47. P. Y. Law, D. S. Hom, H. H. Loh, Opiate receptor down-regulation and desensitization in
1070 neuroblastoma X glioma NG108-15 hybrid cells are two separate cellular adaptation
1071 processes. *Mol. Pharmacol.* **24**, 413–424 (1983).

1072 48. C. M. Lane, R. Elde, H. H. Loh, N. M. Lee, Regulation of an opioid-binding protein in
1073 NG108-15 cells parallels regulation of delta-opioid receptors. *Proc. Natl. Acad. Sci. U. S. A.*
1074 **89**, 11234–11238 (1992).

1075 49. Z. Zhang, M. Ye, Q. Li, Y. You, H. Yu, Y. Ma, L. Mei, X. Sun, L. Wang, W. Yue, R. Li, J.
1076 Li, D. Zhang, The Schizophrenia Susceptibility Gene OPCML Regulates Spine Maturation
1077 and Cognitive Behaviors through Eph-Cofilin Signaling. *Cell Rep.* **29**, 49-61.e7 (2019).

1078 50. T. K. Darling, T. J. Lamb, Emerging Roles for Eph Receptors and Ephrin Ligands in
1079 Immunity. *Front. Immunol.* **10**, 1473 (2019).

1080 51. A. C. McClelland, S. I. Sheffler-Collins, M. S. Kayser, M. B. Dalva, Ephrin-B1 and ephrin-
1081 B2 mediate EphB-dependent presynaptic development via syntenin-1. *Proc. Natl. Acad. Sci.*
1082 *U. S. A.* **106**, 20487–20492 (2009).

1083 52. U. Radhakrishna, S. K. Nath, S. Vishweswaraiah, L. V. Uppala, A. Forray, S. B. Muvvala, N.
1084 K. Mishra, S. Southekal, C. Guda, H. Govindamangalam, D. Vargas, W. G. Gardella, R. C.
1085 Crist, W. H. Berrettini, R. P. Metpally, R. O. Bahado-Singh, Maternal opioid use disorder:
1086 Placental transcriptome analysis for neonatal opioid withdrawal syndrome. *Genomics* **113**,
1087 3610–3617 (2021).

1088 53. M. L. Seney, S.-M. Kim, J. R. Glausier, M. A. Hildebrand, X. Xue, W. Zong, J. Wang, M. A.
1089 Shelton, B. N. Phan, C. Srinivasan, A. R. Pfenning, G. C. Tseng, D. A. Lewis, Z. Freyberg,
1090 R. W. Logan, Transcriptional alterations in dorsolateral prefrontal cortex and nucleus
1091 accumbens implicate neuroinflammation and synaptic remodeling in opioid use disorder.
1092 *Biol. Psychiatry* **90**, 550–562 (2021).

1093 54. B. N. Phan, M. H. Ray, X. Xue, C. Fu, R. J. Fenster, S. J. Kohut, J. Bergman, S. N. Haber, K.
1094 M. McCullough, M. K. Fish, J. R. Glausier, Q. Su, A. E. Tipton, D. A. Lewis, Z. Freyberg,
1095 G. C. Tseng, S. J. Russek, Y. Alekseyev, K. J. Ressler, M. L. Seney, A. R. Pfenning, R. W.
1096 Logan, Single nuclei transcriptomics in human and non-human primate striatum in opioid use
1097 disorder. *Nat. Commun.* **15**, 878 (2024).

1098 55. C. Walss-Bass, "Features of opioid use disorder: Interlinking transcriptomics, epigenetics,
1099 and postmortem studies highlight opioid-induced neurovascular alterations" in *Handbook of*
1100 *the Biology and Pathology of Mental Disorders* (Springer International Publishing, Cham,
1101 2024), pp. 1–15.

1102 56. D. Byrd, J. Murray, G. Safdieh, S. Morgello, Impact of opiate addiction on
1103 neuroinflammation in HIV. *J. Neurovirol.* **18**, 364–373 (2012).

1104 57. D. Sumbria, E. Berber, M. Mathayan, B. T. Rouse, Virus Infections and Host Metabolism-
1105 Can We Manage the Interactions? *Front. Immunol.* **11**, 594963 (2020).

1106 58. M. S. Maginnis, β-arrestins and G protein-coupled receptor kinases in viral entry: A
1107 graphical review. *Cell. Signal.* **102**, 110558 (2023).

1108 59. E. Bieberich, Synthesis, Processing, and Function of N-glycans in N-glycoproteins. *Adv*
1109 *Neurobiol* **9**, 47–70 (2014).

1110 60. X. Lu, D. Zhang, H. Shoji, C. Duan, G. Zhang, T. Isaji, Y. Wang, T. Fukuda, J. Gu,
1111 Deficiency of α1,6-fucosyltransferase promotes neuroinflammation by increasing the
1112 sensitivity of glial cells to inflammatory mediators. *Biochim. Biophys. Acta Gen. Subj.* **1863**,
1113 598–608 (2019).

1114 61. M. M. Bradberry, T. M. Peters-Clarke, E. Shishkova, E. R. Chapman, J. J. Coon, N-
1115 glycoproteomics of brain synapses and synaptic vesicles. *Cell Rep.* **42**, 112368 (2023).

1116 62. W. Gu, T. Fukuda, T. Isaji, H. Hashimoto, Y. Wang, J. Gu, α1,6-Fucosylation regulates
1117 neurite formation via the activin/phospho-Smad2 pathway in PC12 cells: the implicated dual
1118 effects of Fut8 for TGF-β/activin-mediated signaling. *FASEB J.* **27**, 3947–3958 (2013).

1119 63. T. Fukuda, H. Hashimoto, N. Okayasu, A. Kameyama, H. Onogi, O. Nakagawasai, T.
1120 Nakazawa, T. Kurosawa, Y. Hao, T. Isaji, T. Tadano, H. Narimatsu, N. Taniguchi, J. Gu,
1121 Alpha1,6-fucosyltransferase-deficient mice exhibit multiple behavioral abnormalities
1122 associated with a schizophrenia-like phenotype: importance of the balance between the
1123 dopamine and serotonin systems. *J. Biol. Chem.* **286**, 18434–18443 (2011).

1124 64. S. Guan, L. Zhao, R. Peng, Mitochondrial Respiratory Chain Supercomplexes: From
1125 Structure to Function. *Int. J. Mol. Sci.* **23** (2022).

1126 65. D. Xia, L. Esser, W.-K. Tang, F. Zhou, Y. Zhou, L. Yu, C.-A. Yu, Structural analysis of
1127 cytochrome bc1 complexes: implications to the mechanism of function. *Biochim. Biophys.*
1128 *Acta* **1827**, 1278–1294 (2013).

1129 66. S. A. Connor, I. Ammendrup-Johnsen, A. W. Chan, Y. Kishimoto, C. Murayama, N.
1130 Kurihara, A. Tada, Y. Ge, H. Lu, R. Yan, J. M. LeDue, H. Matsumoto, H. Kiyonari, Y.
1131 Kirino, F. Matsuzaki, T. Suzuki, T. H. Murphy, Y. T. Wang, T. Yamamoto, A. M. Craig,
1132 Altered Cortical Dynamics and Cognitive Function upon Haploinsufficiency of the Autism-
1133 Linked Excitatory Synaptic Suppressor MDGA2. *Neuron* **91**, 1052–1068 (2016).

1134 67. A. Toledo, M. Letellier, G. Bimbi, B. Tessier, S. Daburon, A. Favereaux, I. Chamma, K.
1135 Vennekens, J. Vanderlinden, M. Sainlos, J. de Wit, D. Choquet, O. Thoumine, MDGAs are
1136 fast-diffusing molecules that delay excitatory synapse development by altering neuroligin
1137 behavior. *Elife* **11** (2022).

1138 68. B. E. Himes, X. Jiang, R. Hu, A. C. Wu, J. A. Lasky-Su, B. J. Klanderman, J. Ziniti, J.
1139 Senter-Sylvia, J. J. Lima, C. G. Irvin, S. P. Peters, D. A. Meyers, E. R. Bleecker, M. Kubo,
1140 M. Tamari, Y. Nakamura, S. J. Szefler, R. F. Lemanske Jr, R. S. Zeiger, R. C. Strunk, F. D.
1141 Martinez, J. P. Hanrahan, G. H. Koppelman, D. S. Postma, M. A. E. Nieuwenhuis, J. M.
1142 Vonk, R. A. Panettieri Jr, A. Markezich, E. Israel, V. J. Carey, K. G. Tantisira, A. A.
1143 Litonjua, Q. Lu, S. T. Weiss, Genome-wide association analysis in asthma subjects identifies
1144 SPATS2L as a novel bronchodilator response gene. *PLoS Genet.* **8**, e1002824 (2012).

1145 69. K. D. Broad, J. Hassell, B. Fleiss, G. Kawano, M. Ezzati, E. Rocha-Ferreira, M. Hristova, K.
1146 Bennett, I. Fierens, R. Burnett, B. Chaban, D. Alonso-Alconada, A. Oliver-Taylor, I.
1147 Tachsidis, J. Rostami, P. Gressens, R. D. Sanders, N. J. Robertson, Isoflurane Exposure
1148 Induces Cell Death, Microglial Activation and Modifies the Expression of Genes Supporting
1149 Neurodevelopment and Cognitive Function in the Male Newborn Piglet Brain. *PLoS One* **11**,
1150 e0166784 (2016).

1151 70. G. Duette, P. Pereyra Gerber, J. Rubione, P. S. Perez, A. L. Landay, S. M. Crowe, Z. Liao, K.
1152 W. Witwer, M. P. Holgado, J. Salido, J. Geffner, O. Sued, C. S. Palmer, M. Ostrowski,
1153 Induction of HIF-1α by HIV-1 Infection in CD4+ T Cells Promotes Viral Replication and
1154 Drives Extracellular Vesicle-Mediated Inflammation. *MBio* **9** (2018).

1155 71. S. L. Deshmane, S. Amini, S. Sen, K. Khalili, B. E. Sawaya, Regulation of the HIV-1
1156 promoter by HIF-1α and Vpr proteins. *Virol. J.* **8**, 477 (2011).

1157 72. W. Xie, W. Wang, S. Meng, X. Wu, X. Liu, Y. Liu, X. Kang, Y. Su, X. Lv, L. Guo, C.
1158 Wang, A novel hypoxia-stimulated lncRNA HIF1A-AS3 binds with YBX1 to promote
1159 ovarian cancer tumorigenesis by suppressing p21 and AJAP1 transcription. *Mol. Carcinog.*
1160 **62**, 1860–1876 (2023).

1161 73. M. Gedam, M. M. Comerota, N. E. Propson, T. Chen, F. Jin, M. C. Wang, H. Zheng,
1162 Complement C3aR depletion reverses HIF-1α-induced metabolic impairment and enhances
1163 microglial response to Aβ pathology. *J. Clin. Invest.* **133** (2023).

1164 74. C. J. Day, R. L. Hardison, B. L. Spillings, J. Poole, J. A. Jurcisek, J. Mak, M. P. Jennings, J.
1165 L. Edwards, Complement Receptor 3 Mediates HIV-1 Transcytosis across an Intact Cervical
1166 Epithelial Cell Barrier: New Insight into HIV Transmission in Women. *MBio* **13**, e0217721
1167 (2022).

1168 75. N. Nomura, M. J. Zhao, T. Nagase, T. Maekawa, R. Ishizaki, S. Tabata, S. Ishii, HIV-EP2, a
1169 new member of the gene family encoding the human immunodeficiency virus type 1
1170 enhancer-binding protein. Comparison with HIV-EP1/PRDII-BF1/MBP-1. *J. Biol. Chem.*
1171 **266**, 8590–8594 (1991).

1172 76. A. J. Henderson, K. L. Calame, CCAAT/enhancer binding protein (C/EBP) sites are required
1173 for HIV-1 replication in primary macrophages but not CD4(+) T cells. *Proc. Natl. Acad. Sci.*
1174 *U. S. A.* **94**, 8714–8719 (1997).

1175 77. J. R. Engler, A. E. Robinson, I. Smirnov, J. G. Hodgson, M. S. Berger, N. Gupta, C. D.
1176 James, A. Molinaro, J. J. Phillips, Increased microglia/macrophage gene expression in a
1177 subset of adult and pediatric astrocytomas. *PLoS One* **7**, e43339 (2012).

1178 78. K. M. Roman, A. K. Jenkins, D. A. Lewis, D. W. Volk, Involvement of the nuclear factor-κB
1179 transcriptional complex in prefrontal cortex immune activation in bipolar disorder. *Transl.*
1180 *Psychiatry* **11**, 40 (2021).

1181 79. T. S. Kuroda, M. Fukuda, H. Ariga, K. Mikoshiba, The Slp homology domain of
1182 synaptotagmin-like proteins 1-4 and Slac2 functions as a novel Rab27A binding domain. *J.*
1183 *Biol. Chem.* **277**, 9212–9218 (2002).

1184 80. P. P. Gerber, M. Cabrini, C. Jancic, L. Paoletti, C. Banchio, C. von Bilderling, L. Sigaut, L. I.
1185 Pietrasanta, G. Duette, E. O. Freed, G. de S. Basile, C. F. Moita, L. F. Moita, S. Amigorena,
1186 P. Benaroch, J. Geffner, M. Ostrowski, Rab27a controls HIV-1 assembly by regulating
1187 plasma membrane levels of phosphatidylinositol 4,5-bisphosphate. *J. Cell Biol.* **209**, 435–452
1188 (2015).

1189 81. Z. Wang, H. Dong, J. Wang, Y. Huang, X. Zhang, Y. Tang, Q. Li, Z. Liu, Y. Ma, J. Tong, L.
1190 Huang, J. Fei, M. Yu, J. Wang, F. Huang, Pro-survival and anti-inflammatory roles of NF-κB
1191 c-Rel in the Parkinson's disease models. *Redox Biol* **30**, 101427 (2020).

1192 82. S. M. T. Rahman, M. Aqdas, E. W. Martin, F. Tomassoni Ardori, P. Songkiatisak, K.-S. Oh,
1193 S. Uderhardt, S. Yun, Q. C. Claybourne, R. A. McDevitt, V. Greco, R. N. Germain, L.
1194 Tessarollo, M.-H. Sung, Double knockin mice show NF-κB trajectories in immune signaling
1195 and aging. *Cell Rep.* **41**, 111682 (2022).

1196 83. B. Romani, R. H. Glashoff, S. Engelbrecht, Functional integrity of naturally occurring
1197 mutants of HIV-1 subtype C Vpr. *Virus Res.* **153**, 288–298 (2010).

1198 84. S. Z. Khan, S. Gasperino, S. L. Zeichner, Nuclear Transit and HIV LTR Binding of NF-κB
1199 Subunits Held by IκB Proteins: Implications for HIV-1 Activation. *Viruses* **11** (2019).

1200 85. S. Qi, Z. Li, U. Schulze-Gahmen, G. Stjepanovic, Q. Zhou, J. H. Hurley, Structural basis for
1201 ELL2 and AFF4 activation of HIV-1 proviral transcription. *Nat. Commun.* **8**, 14076 (2017).

1202 86. D. Yu, R. Liu, G. Yang, Q. Zhou, The PARP1-Siah1 Axis Controls HIV-1 Transcription and
1203 Expression of Siah1 Substrates. *Cell Rep.* **23**, 3741–3749 (2018).

1204 87. H. Ohno, R. C. Aguilar, M. C. Fournier, S. Hennecke, P. Cosson, J. S. Bonifacino,
1205 Interaction of endocytic signals from the HIV-1 envelope glycoprotein complex with
1206 members of the adaptor medium chain family. *Virology* **238**, 305–315 (1997).

1207 88. M.-T. Li, W. Di, H. Xu, Y.-K. Yang, H.-W. Chen, F.-X. Zhang, Z.-H. Zhai, D.-Y. Chen,
1208 Negative regulation of RIG-I-mediated innate antiviral signaling by SEC14L1. *J. Virol.* **87**,
1209 10037–10046 (2013).

1210 89. P. Marschang, J. Sodroski, R. Würzner, M. P. Dierich, Decay-accelerating factor (CD55)
1211 protects human immunodeficiency virus type 1 from inactivation by human complement.
1212 *Eur. J. Immunol.* **25**, 285–290 (1995).

1213 90. Z. Tang, L. Wang, Z. Huang, H. Guan, W. Leung, X. Chen, H. Xia, W. Zhang, CD55 is
1214 upregulated by cAMP/PKA/AKT and modulates human decidualization via Src and ERK
1215 pathway and decidualization-related genes. *Mol. Reprod. Dev.* **89**, 256–268 (2022).

1216 91. C. Lee, T. R. Wooldridge, L. A. Laimins, Analysis of the roles of E6 binding to E6TP1 and
1217 nuclear localization in the human papillomavirus type 31 life cycle. *Virology* **358**, 201–210
1218 (2007).

1219 92. H. Nishitsuji, R. Sugiyama, M. Abe, H. Takaku, ATP1B3 Protein Modulates the Restriction
1220 of HIV-1 Production and Nuclear Factor κ Light Chain Enhancer of Activated B Cells (NF-
1221 κB) Activation by BST-2. *J. Biol. Chem.* **291**, 4754–4762 (2016).

1222 93. R. B. Meeker, J. C. Boles, K. R. Robertson, C. D. Hall, Cerebrospinal fluid from human
1223 immunodeficiency virus--infected individuals facilitates neurotoxicity by suppressing
1224 intracellular calcium recovery. *J. Neurovirol.* **11**, 144–156 (2005).

1225 94. Q. S. Li, L. De Muynck, Differentially expressed genes in Alzheimer's disease highlighting
1226 the roles of microglia genes including OLR1 and astrocyte gene CDK2AP1. *Brain Behav*
1227 *Immun Health* **13**, 100227 (2021).

1228 95. Y. Aoki, H. Dai, F. Furuta, T. Akamatsu, T. Oshima, N. Takahashi, Y.-I. Goto, A. Oka, M.
1229 Itoh, LOX-1 mediates inflammatory activation of microglial cells through the p38-
1230 MAPK/NF-κB pathways under hypoxic-ischemic conditions. *Cell Commun. Signal.* **21**, 126
1231 (2023).

1232 96. D. M. Borucki, A. Toutonji, C. Couch, K. Mallah, B. Rohrer, S. Tomlinson, Complement-
1233 Mediated Microglial Phagocytosis and Pathological Changes in the Development and
1234 Degeneration of the Visual System. *Front. Immunol.* **11**, 566892 (2020).

1235 97. M. A. Marzinke, M. Clagett-Dame, The all-trans retinoic acid (atRA)-regulated gene Calmin
1236 (Clmn) regulates cell cycle exit and neurite outgrowth in murine neuroblastoma (Neuro2a)
1237 cells. *Exp. Cell Res.* **318**, 85–93 (2012).

1238 98. M. Takaishi, Z. Ishisaki, T. Yoshida, Y. Takata, N.-H. Huh, Expression of calmin, a novel
1239 developmentally regulated brain protein with calponin-homology domains. *Brain Res. Mol.*
1240 *Brain Res.* **112**, 146–152 (2003).

1241 99. B. Y. Salmani, L. Lahti, L. Gillberg, J. K. Jacobsen, I. Mantas, P. Svenningsson, T.
1242 Perlmann, Transcriptomic atlas of midbrain dopamine neurons uncovers differential
1243 vulnerability in a Parkinsonism lesion model (2023). https://doi.org/10.7554/elife.89482.1.

1244 100. A. Kari, Z. M, Z. Aili, A. Adili, N. Hairula, A. Abuduhaer, Knockdown of EPSTI1
1245 alleviates lipopolysaccharide-induced inflammatory injury through regulation of NF-κB
1246 signaling in a cellular pneumonia model. *Allergol. Immunopathol.* **50**, 106–112 (2022).

1247 101. Y. Wang, W. Jia, Q. Ma, F. Li, Z. Ma, M. Yang, J. Pu, Z. Huo, J. Dang, Identification of
1248 IFI44L as a new candidate molecular marker for systemic lupus erythematosus. *Clin. Exp.*
1249 *Rheumatol.* **41**, 48–59 (2023).

1250 102. Y. Li, J. Zhang, C. Wang, W. Qiao, Y. Li, J. Tan, IFI44L expression is regulated by IRF-1
1251 and HIV-1. *FEBS Open Bio* **11**, 105–113 (2021).

1252 103. J. Mertens, A. C. M. Paquola, M. Ku, E. Hatch, L. Böhnke, S. Ladjevardi, S. McGrath, B.
1253 Campbell, H. Lee, J. R. Herdy, J. T. Gonçalves, T. Toda, Y. Kim, J. Winkler, J. Yao, M. W.
1254 Hetzer, F. H. Gage, Directly Reprogrammed Human Neurons Retain Aging-Associated
1255 Transcriptomic Signatures and Reveal Age-Related Nucleocytoplasmic Defects. *Cell Stem*
1256 *Cell* **17**, 705–718 (2015).

1257 104. M. J. Seibert, C. S. Evans, K. S. Stanley, Z. Wu, E. R. Chapman, Synaptotagmin 9
1258 Modulates Spontaneous Neurotransmitter Release in Striatal Neurons by Regulating
1259 Substance P Secretion. *J. Neurosci.* **43**, 1475–1491 (2023).

1260 105. A. A. Kozlova, E. Rubets, M. R. Vareltzoglou, N. Jarzebska, V. N. Ragavan, Y. Chen, J.
1261 Martens-Lobenhoffer, S. M. Bode-Böger, R. R. Gainetdinov, R. N. Rodionov, N. Bernhardt,
1262 Knock-out of the critical nitric oxide synthase regulator DDAH1 in mice impacts
1263 amphetamine sensitivity and dopamine metabolism. *J. Neural Transm. (Vienna)* **130**, 1097–
1264 1112 (2023).

1265 106. K. M. Chung, E.-J. Jeong, H. Park, H.-K. An, S.-W. Yu, Mediation of Autophagic Cell
1266 Death by Type 3 Ryanodine Receptor (RyR3) in Adult Hippocampal Neural Stem Cells.
1267 *Front. Cell. Neurosci.* **10**, 116 (2016).

1268 107. D. Balschun, D. P. Wolfer, F. Bertocchini, V. Barone, A. Conti, W. Zuschratter, L.
1269 Missiaen, H. P. Lipp, J. U. Frey, V. Sorrentino, Deletion of the ryanodine receptor type 3
1270 (RyR3) impairs forms of synaptic plasticity and spatial learning. *EMBO J.* **18**, 5264–5273
1271 (1999).

1272 108. J. Murray, G. Meloni, E. P. Cortes, A. KimSilva, M. Jacobs, A. Ramkissoon, J. F. Crary,
1273 S. Morgello, Frontal lobe microglia, neurodegenerative protein accumulation, and cognitive
1274 function in people with HIV. *Acta Neuropathol Commun* **10**, 69 (2022).

1275 109. A. J. Bailey, R. K. McHugh, Why do we focus on the exception and not the rule?
1276 Examining the prevalence of mono- versus polysubstance use in the general population.
1277 *Addiction* **118**, 2026–2029 (2023).

1278 110. W. M. Compton, R. J. Valentino, R. L. DuPont, Polysubstance use in the U.S. opioid
1279 crisis. *Mol. Psychiatry* **26**, 41–50 (2021).

1280 111. B. J. Rounsaville, N. M. Petry, K. M. Carroll, Single versus multiple drug focus in
1281 substance abuse clinical trials research. *Drug Alcohol Depend.* **70**, 117–125 (2003).

1282 112. H. S. Fox, M. Niu, B. M. Morsey, B. G. Lamberty, K. Emanuel, P. Periyasamy, S. Callen,
1283 A. Acharya, G. Kubik, J. Eudy, C. Guda, S. R. Dyavar, C. V. Fletcher, S. N. Byrareddy, S.
1284 Buch, Morphine suppresses peripheral responses and transforms brain myeloid gene
1285 expression to favor neuropathogenesis in SIV infection. *Front. Immunol.* **13**, 1012884
1286 (2022).

1287 113. A. Acharya, O. A. Olwenyi, M. Thurman, K. Pandey, B. M. Morsey, B. Lamberty, N.
1288 Ferguson, S. Callen, Q. Fang, S. J. Buch, H. S. Fox, S. N. Byrareddy, Chronic morphine
1289 administration differentially modulates viral reservoirs in a simian immunodeficiency virus
1290 SIVmac251-infected rhesus macaque model. *J. Virol.* **95** (2021).

1291 114. A. L. Plaza-Jennings, A. Valada, C. O'Shea, M. Iskhakova, B. Hu, B. Javidfar, G. Ben
1292 Hutta, T. Y. Lambert, J. Murray, B. Kassim, S. Chandrasekaran, B. K. Chen, S. Morgello, H.

1293 Won, S. Akbarian, HIV integration in the human brain is linked to microglial activation and
1294 3D genome remodeling. *Mol. Cell* **82**, 4647-4663.e8 (2022).

1295 115. S. Morgello, B. B. Gelman, P. B. Kozlowski, H. V. Vinters, E. Masliah, M. Cornford, W.
1296 Cavert, C. Marra, I. Grant, E. J. Singer, The National NeuroAIDS Tissue Consortium: a new
1297 paradigm in brain banking with an emphasis on infectious disease. *Neuropathol. Appl.*
1298 *Neurobiol.* **27**, 326–335 (2001).

1299 116. S. Morgello, "Chapter 2 - HIV neuropathology" in *Handbook of Clinical Neurology*, B. J.
1300 Brew, Ed. (Elsevier, 2018)vol. 152, pp. 3–19.

1301 117. J. C. Schlachetzki, S. Gianella, Z. Ouyang, A. J. Lana, X. Yang, S. O'Brien, J. F.
1302 Challacombe, P. J. Gaskill, K. L. Jordan-Sciutto, A. Chaillon, D. Moore, C. L. Achim, R. J.
1303 Ellis, D. M. Smith, C. K. Glass, Gene expression and chromatin conformation of microglia in
1304 virally suppressed people with HIV. *Life Sci Alliance* **7** (2024).

1305 118. P. K. Riggs, A. Chaillon, G. Jiang, S. L. Letendre, Y. Tang, J. Taylor, A. Kaytes, D. M.
1306 Smith, K. Dubé, S. Gianella, Lessons for Understanding Central Nervous System HIV
1307 Reservoirs from the Last Gift Program. *Curr. HIV/AIDS Rep.* **19**, 566–579 (2022).

1308 119. D. Alvarez-Carbonell, F. Ye, N. Ramanath, Y. Garcia-Mesa, P. E. Knapp, K. F. Hauser, J.
1309 Karn, Cross-talk between microglia and neurons regulates HIV latency. *PLoS Pathog.* **15**,
1310 e1008249 (2019).

1311 120. G.-J. Wang, L. Chang, N. D. Volkow, F. Telang, J. Logan, T. Ernst, J. S. Fowler,
1312 Decreased brain dopaminergic transporters in HIV-associated dementia patients. *Brain* **127**,
1313 2452–2458 (2004).

1314 121. L. Chang, G.-J. Wang, N. D. Volkow, T. Ernst, F. Telang, J. Logan, J. S. Fowler,
1315 Decreased brain dopamine transporters are related to cognitive deficits in HIV patients with
1316 or without cocaine abuse. *Neuroimage* **42**, 869–878 (2008).

1317 122. P. M. Quizon, W.-L. Sun, Y. Yuan, N. M. Midde, C.-G. Zhan, J. Zhu, Molecular
1318 mechanism: the human dopamine transporter histidine 547 regulates basal and HIV-1 Tat
1319 protein-inhibited dopamine transport. *Sci. Rep.* **6**, 39048 (2016).

1320 123. P. J. Gaskill, D. R. Miller, J. Gamble-George, H. Yano, H. Khoshbouei, HIV, Tat and
1321 dopamine transmission. *Neurobiol. Dis.* **105**, 51–73 (2017).

1322 124. D. S. Hasin, K. D. Trautman, G. M. Miele, S. Samet, M. Smith, J. Endicott, Psychiatric
1323 Research Interview for Substance and Mental Disorders (PRISM): reliability for substance
1324 abusers. *Am. J. Psychiatry* **153**, 1195–1201 (1996).

1325 125. C. Nelson, The Composite International Diagnostic Interview (CIDI) web site. *Bull. World*
1326 *Health Organ.* **77**, 614–614 (1999).

1327 126. M. M. Jacobs, J. Murray, D. A. Byrd, Y. L. Hurd, S. Morgello, HIV-related cognitive
1328 impairment shows bi-directional association with dopamine receptor DRD1 and DRD2
1329 polymorphisms in substance-dependent and substance-independent populations. *J.*
1330 *Neurovirol.* **19**, 495–504 (2013).

1331 127. Y. Guo, J. He, S. Zhao, H. Wu, X. Zhong, Q. Sheng, D. C. Samuels, Y. Shyr, J. Long,
1332 Illumina human exome genotyping array clustering and quality control. *Nat. Protoc.* **9**, 2643–
1333 2662 (2014).

1334 128. S. Zhao, W. Jing, D. C. Samuels, Q. Sheng, Y. Shyr, Y. Guo, Strategies for processing and
1335 quality control of Illumina genotyping arrays. *Brief. Bioinform.* **19**, 765–775 (2018).

1336 129. S. Das, L. Forer, S. Schönherr, C. Sidore, A. E. Locke, A. Kwong, S. I. Vrieze, E. Y.
1337 Chew, S. Levy, M. McGue, D. Schlessinger, D. Stambolian, P.-R. Loh, W. G. Iacono, A.
1338 Swaroop, L. J. Scott, F. Cucca, F. Kronenberg, M. Boehnke, G. R. Abecasis, C. Fuchsberger,
1339 Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287
1340 (2016).

1341 130. D. Taliun, D. N. Harris, M. D. Kessler, J. Carlson, Z. A. Szpiech, R. Torres, S. A.
1342 Gagliano Taliun, A. Corvelo, S. M. Gogarten, H. M. Kang, A. N. Pitsillides, J. LeFaive, S.-B.
1343 Lee, X. Tian, B. L. Browning, S. Das, A.-K. Emde, W. E. Clarke, D. P. Loesch, A. C. Shetty,
1344 T. W. Blackwell, Q. Wong, F. Aguet, C. Albert, A. Alonso, K. G. Ardlie, S. Aslibekyan, P.
1345 L. Auer, J. Barnard, R. Graham Barr, L. C. Becker, R. L. Beer, E. J. Benjamin, L. F. Bielak,
1346 J. Blangero, M. Boehnke, D. W. Bowden, J. A. Brody, E. G. Burchard, B. E. Cade, J. F.
1347 Casella, B. Chalazan, Y.-D. I. Chen, M. H. Cho, S. H. Choi, M. K. Chung, C. B. Clish, A.
1348 Correa, J. E. Curran, B. Custer, D. Darbar, M. Daya, M. de Andrade, D. L. DeMeo, S. K.
1349 Dutcher, P. T. Ellinor, L. S. Emery, D. Fatkin, L. Forer, M. Fornage, N. Franceschini, C.
1350 Fuchsberger, S. M. Fullerton, S. Germer, M. T. Gladwin, D. J. Gottlieb, X. Guo, M. E. Hall,
1351 J. He, N. L. Heard-Costa, S. R. Heckbert, M. R. Irvin, J. M. Johnsen, A. D. Johnson, S. L. R.
1352 Kardia, T. Kelly, S. Kelly, E. E. Kenny, D. P. Kiel, R. Klemmer, B. A. Konkle, C.
1353 Kooperberg, A. Köttgen, L. A. Lange, J. Lasky-Su, D. Levy, X. Lin, K.-H. Lin, C. Liu, R. J.
1354 F. Loos, L. Garman, R. Gerszten, S. A. Lubitz, K. L. Lunetta, A. C. Y. Mak, A. Manichaikul,
1355 A. K. Manning, R. A. Mathias, D. D. McManus, S. T. McGarvey, J. B. Meigs, D. A. Meyers,
1356 J. L. Mikulla, M. A. Minear, B. Mitchell, S. Mohanty, M. E. Montasser, C. Montgomery, A.
1357 C. Morrison, J. M. Murabito, A. Natale, P. Natarajan, S. C. Nelson, K. E. North, J. R.
1358 O'Connell, N. D. Palmer, N. Pankratz, G. M. Peloso, P. A. Peyser, B. M. Psaty, D. C. Rao, S.
1359 Redline, A. P. Reiner, D. Roden, J. I. Rotter, I. Ruczinski, C. Sarnowski, S. Schoenherr, J.-S.
1360 Seo, S. Seshadri, V. A. Sheehan, M. Benjamin Shoemaker, A. V. Smith, N. L. Smith, J. A.
1361 Smith, N. Sotoodehnia, A. M. Stilp, W. Tang, K. D. Taylor, M. Telen, T. A. Thornton, R. P.
1362 Tracy, D. J. Van Den Berg, R. S. Vasan, K. A. Viaud-Martinez, S. Vrieze, D. E. Weeks, B. S.
1363 Weir, S. T. Weiss, L.-C. Weng, C. J. Willer, Y. Zhang, X. Zhao, D. K. Arnett, A. E. Ashley-
1364 Koch, K. C. Barnes, E. Boerwinkle, S. Gabriel, R. Gibbs, K. M. Rice, S. S. Rich, E.
1365 Silverman, P. Qasba, W. Gan, Trans-Omics for Precision Medicine (TOPMed) Program,
1366 TOPMed Population Genetics Working Group, G. J. Papanicolaou, D. A. Nickerson, S. R.
1367 Browning, M. C. Zody, S. Zöllner, J. G. Wilson, L. Adrienne Cupples, C. C. Laurie, C. E.
1368 Jaquish, R. D. Hernandez, T. D. O'Connor, G. R. Abecasis, Sequencing of 53,831 diverse
1369 genomes from the NHLBI TOPMed Program, *bioRxiv* (2019)p. 563866.

1370 131. C. Fuchsberger, G. R. Abecasis, D. A. Hinds, minimac2: faster genotype imputation.
1371 *Bioinformatics* **31**, 782–784 (2015).

1372 132. P.-R. Loh, P. Danecek, P. F. Palamara, C. Fuchsberger, Y. A Reshef, H. K Finucane, S.
1373 Schoenherr, L. Forer, S. McCarthy, G. R. Abecasis, R. Durbin, A. L Price, Reference-based
1374 phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448
1375 (2016).

133. P.-R. Loh, P. F. Palamara, A. L. Price, Fast and accurate long-range phasing in a UK Biobank cohort. *Nat. Genet.* **48**, 811–816 (2016).

134. P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, H. Li, Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021).

135. Interpreting Intronic and Antisense Reads in 10x Genomics Single Cell Gene Expression Data - Official 10x Genomics Support, *10x Genomics*. https://www.10xgenomics.com/support/single-cell-gene-expression/documentation/steps/sequencing/interpreting-intronic-and-antisense-reads-in-10-x-genomics-single-cell-gene-expression-data.

136. Interpreting Single Cell Gene Expression Data With and Without Intronic Reads - Official 10x Genomics Support, *10x Genomics*. https://www.10xgenomics.com/support/single-cell-gene-expression/documentation/steps/sequencing/interpreting-single-cell-gene-expression-data-with-and-without-intronic-reads.

137. R. León-Rivera, B. Morsey, M. Niu, H. S. Fox, J. W. Berman, Interactions of Monocytes, HIV, and ART Identified by an Innovative scRNAseq Pipeline: Pathways to Reservoirs and HIV-Associated Comorbidities. *MBio* **11** (2020).

138. S. J. Fleming, M. D. Chaffin, A. Arduini, A.-D. Akkad, E. Banks, J. C. Marioni, A. A. Philippakis, P. T. Ellinor, M. Babadi, Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender, *bioRxiv* (2022)p. 791699.

139. FireCloud. https://portal.firecloud.org/#methods/cellbender/remove-background/13.

140. Terra. https://app.terra.bio/.

141. K. Voss, J. Gentry, G. Van der Auwera, Full-stack genomics pipelining with GATK4 + WDL + Cromwell (2017). https://doi.org/10.7490/f1000research.1114631.1.

142. I. Virshup, S. Rybakov, F. J. Theis, P. Angerer, F. Alexander Wolf, anndata: Annotated data, *bioRxiv* (2021)p. 2021.12.16.473007.

143. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281-291.e9 (2019).

144. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 1–5 (2018).

145. L. McInnes, J. Healy, N. Saul, L. Großberger, UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018).

146. E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I. W. H. Kwok, L. G. Ng, F. Ginhoux, E. W. Newell, Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.*, doi: 10.1038/nbt.4314 (2018).

147. I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P.-R. Loh, S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

1414 148. A. Ianevski, A. K. Giri, T. Aittokallio, Fully-automated and ultra-fast cell-type
1415 identification using specific marker combinations from single-cell transcriptomic data. *Nat.*
1416 *Commun.* **13**, 1246 (2022).

1417 149. S. Puvogel, A. Alsema, L. Kracht, M. J. Webster, C. S. Weickert, I. E. C. Sommer, B. J. L.
1418 Eggen, Single-nucleus RNA sequencing of midbrain blood-brain barrier cells in
1419 schizophrenia reveals subtle transcriptional changes with overall preservation of cellular
1420 proportions and phenotypes. *Mol. Psychiatry* **27**, 4731–4740 (2022).

1421 150. K. Tiklová, Å. K. Björklund, L. Lahti, A. Fiorenzano, S. Nolbrant, L. Gillberg, N.
1422 Volakakis, C. Yokota, M. M. Hilscher, T. Hauling, F. Holmström, E. Joodmardi, M. Nilsson,
1423 M. Parmar, T. Perlmann, Single-cell RNA sequencing reveals midbrain dopamine neuron
1424 diversity emerging during mouse brain development. *Nat. Commun.* **10**, 581 (2019).

1425 151. Z. Fang, X. Liu, G. Peltz, GSEApy: a comprehensive package for performing gene set
1426 enrichment analysis in Python. *Bioinformatics* **39** (2023).

1427 152. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis,
1428 K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis,
1429 S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene
1430 ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**,
1431 25–29 (2000).

1432 153. Gene Ontology Consortium, S. A. Aleksander, J. Balhoff, S. Carbon, J. M. Cherry, H. J.
1433 Drabkin, D. Ebert, M. Feuermann, P. Gaudet, N. L. Harris, D. P. Hill, R. Lee, H. Mi, S.
1434 Moxon, C. J. Mungall, A. Muruganugan, T. Mushayahama, P. W. Sternberg, P. D. Thomas,
1435 K. Van Auken, J. Ramsey, D. A. Siegele, R. L. Chisholm, P. Fey, M. C. Aspromonte, M. V.
1436 Nugnes, F. Quaglia, S. Tosatto, M. Giglio, S. Nadendla, G. Antonazzo, H. Attrill, G. Dos
1437 Santos, S. Marygold, V. Strelets, C. J. Tabone, J. Thurmond, P. Zhou, S. H. Ahmed, P.
1438 Asanitthong, D. Luna Buitrago, M. N. Erdol, M. C. Gage, M. Ali Kadhum, K. Y. C. Li, M.
1439 Long, A. Michalak, A. Pesala, A. Pritazahra, S. C. C. Saverimuttu, R. Su, K. E. Thurlow, R.
1440 C. Lovering, C. Logie, S. Oliferenko, J. Blake, K. Christie, L. Corbani, M. E. Dolan, H. J.
1441 Drabkin, D. P. Hill, L. Ni, D. Sitnikov, C. Smith, A. Cuzick, J. Seager, L. Cooper, J. Elser, P.
1442 Jaiswal, P. Gupta, P. Jaiswal, S. Naithani, M. Lera-Ramirez, K. Rutherford, V. Wood, J. L.
1443 De Pons, M. R. Dwinell, G. T. Hayman, M. L. Kaldunski, A. E. Kwitek, S. J. F.
1444 Laulederkind, M. A. Tutaj, M. Vedi, S.-J. Wang, P. D'Eustachio, L. Aimo, K. Axelsen, A.
1445 Bridge, N. Hyka-Nouspikel, A. Morgat, S. A. Aleksander, J. M. Cherry, S. R. Engel, K.
1446 Karra, S. R. Miyasato, R. S. Nash, M. S. Skrzypek, S. Weng, E. D. Wong, E. Bakker, T. Z.
1447 Berardini, L. Reiser, A. Auchincloss, K. Axelsen, G. Argoud-Puy, M.-C. Blatter, E. Boutet,
1448 L. Breuza, A. Bridge, C. Casals-Casas, E. Coudert, A. Estreicher, M. Livia Famiglietti, M.
1449 Feuermann, A. Gos, N. Gruaz-Gumowski, C. Hulo, N. Hyka-Nouspikel, F. Jungo, P. Le
1450 Mercier, D. Lieberherr, P. Masson, A. Morgat, I. Pedruzzi, L. Pourcel, S. Poux, C. Rivoire, S.
1451 Sundaram, A. Bateman, E. Bowler-Barnett, H. Bye-A-Jee, P. Denny, A. Ignatchenko, R.
1452 Ishtiaq, A. Lock, Y. Lussi, M. Magrane, M. J. Martin, S. Orchard, P. Raposo, E. Speretta, N.
1453 Tyagi, K. Warner, R. Zaru, A. D. Diehl, R. Lee, J. Chan, S. Diamantakis, D. Raciti, M.
1454 Zarowiecki, M. Fisher, C. James-Zorn, V. Ponferrada, A. Zorn, S. Ramachandran, L.
1455 Ruzicka, M. Westerfield, The Gene Ontology knowledgebase in 2023. *Genetics* **224** (2023).

1456 154. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A.
1457 Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment
1458 analysis: a knowledge-based approach for interpreting genome-wide expression profiles.
1459 *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

155. V. K. Mootha, C. M. Lindgren, K.-F. Eriksson, A. Subramanian, S. Sihag, J. Lehar, P. Puigserver, E. Carlsson, M. Ridderstråle, E. Laurila, N. Houstis, M. J. Daly, N. Patterson, J. P. Mesirov, T. R. Golub, P. Tamayo, B. Spiegelman, E. S. Lander, J. N. Hirschhorn, D. Altshuler, L. C. Groop, PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).

156. D. Agarwal, C. Sandor, V. Volpato, T. M. Caffrey, J. Monzón-Sandoval, R. Bowden, J. Alegre-Abarrategui, R. Wade-Martins, C. Webber, A single-cell atlas of the human substantia nigra reveals cell-specific pathways associated with neurological disorders. *Nat. Commun.* **11**, 4183 (2020).

Acknowledgments

General: We thank the staff and patients of the Manhattan HIV Brain Bank; this research would not be possible without their valuable contributions. We also thank all members of the Single Cell Opioid Responses in the Context of HIV (SCORCH) consortium for stimulating and helpful discussions, and Roger Pique-Regi for guidance on demuxlet software. This work was supported in part through the computational and data resources provided by the Scientific Computing and Data Center at Icahn School of Medicine at Mount Sinai. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Funding: This work was supported by
National Institutes of Health grant U01DA053600 (SA, SM)
National Institutes of Health grant R61DA048207 (SA, SM)
National Institutes of Health grant R01NS108801 (SM)
National Institutes of Health grant RFAG060961 (SM)
National Institutes of Health grant U24MH100931 (SM)
National Institutes of Health contract 75N95023C00015 (SM)
National Institutes of Health Office of Research Infrastructure grant S10OD026880 (Patricia Kovach, ISMMS)
National Institutes of Health Office of Research Infrastructure grant S10OD030463 (Patricia Kovach, ISMMS)
National Center for Advancing Translational Sciences Clinical and Translational Science Awards grant UL1TR004419 (Rosalind Wright, ISMMS)

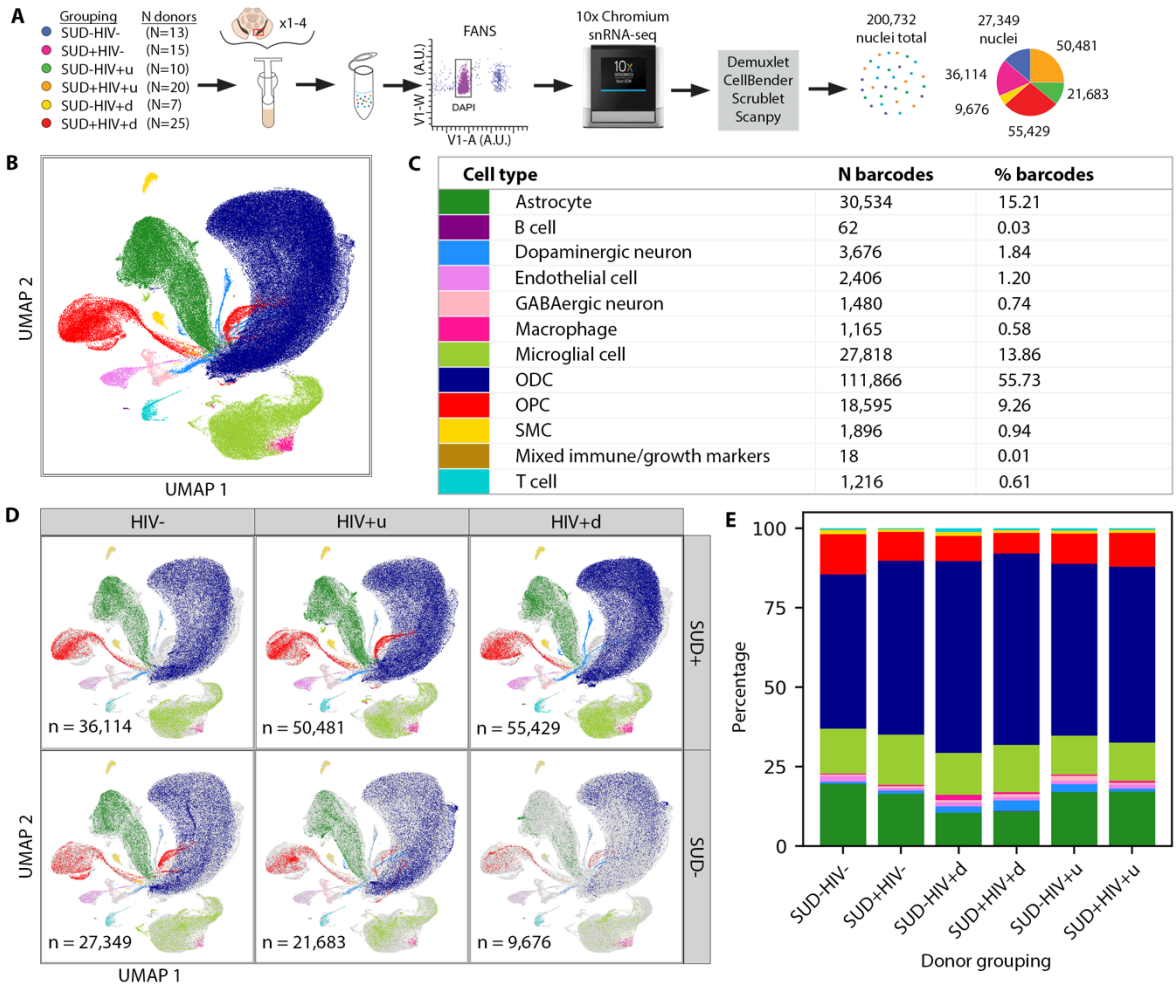
Author contributions:
Conceptualization: SA, SM
Methodology: AW, MJ, SA, SM
Software: AW
Validation: AW, MJ, AV, TL, SA, SM
Formal analysis: AW
Investigation: AW, MJ, GM, EG, JM, AV, TL, SM
Resources: SA, SM
Data curation: AW, MJ, AV, TL, SM
Writing—original draft: AW
Writing—review & editing: AW, MJ, SA, SM
Visualization: AW
Supervision: AW, MJ, GM, SA, SM
Project administration: SA, SM

Funding acquisition: SA, SM

Competing interests:
The authors declare that they have no competing interests.

Data and materials availability:
All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Data presented in this study were produced as part of the Single Cell Opioid Response in the Context of HIV consortium (SCORCH: RRID:SCR_022600). De-identified snRNA-seq data, donor SNP arrays, and associated metadata will be deposited at dbGaP (https://www.ncbi.nlm.nih.gov/gap/) and made available prior to publication. The data at dbGaP is controlled-access and requires special handling; please follow dbGaP guidelines and contact dbGaP for support. Mapped, demultiplexed snRNA-seq data will also be made publicly available at the NEMO (Neuroscience Multi-omic) Archive under identifier nemo:dat-zngw2c8 (https://assets.nemoarchive.org/dat-zngw2c8) prior to publication. Code used to perform the analyses in this paper will be made publicly available on Github and Zenodo prior to publication.

1527 **Figures and Tables**



1528
1529 **Fig. 1. snRNA-seq data generation and cell typing results. (A)** Schematic of study
1530 design and snRNA-seq data acquisition procedure (showing at left the number of
1531 donors per study group and at right, with the same color-coding, each group's final
1532 number of QC-passing nuclei). "V1-W" and "V1-A": width and area, respectively,
1533 of violet-channel fluorescence signal used for detecting intact nuclei with
1534 fluorescence-activated nuclear sorting (FANS); "DAPI": 4′,6-diamidino-2-
1535 phenylindole. **(B)** 2D UMAP showing the cell types we identified, with cell type
1536 colors as shown in **(C)**. **(C)** Numbers and fractions of nuclei of each cell type. **(D)**
1537 UMAPs showing nuclei by cell type for each donor group (colors as in **(C)**). **(E)**
1538 Cell type proportions for each donor group (see also Table S5), with colors as
1539 shown in **(C)**: red, oligodendrocyte precursor cells (OPC); dark blue,
1540 oligodendrocytes (ODC); light green, microglia; dark green, astrocytes; and less-
1541 frequent cell types including in light blue, DA neurons and rose, GABA neurons.

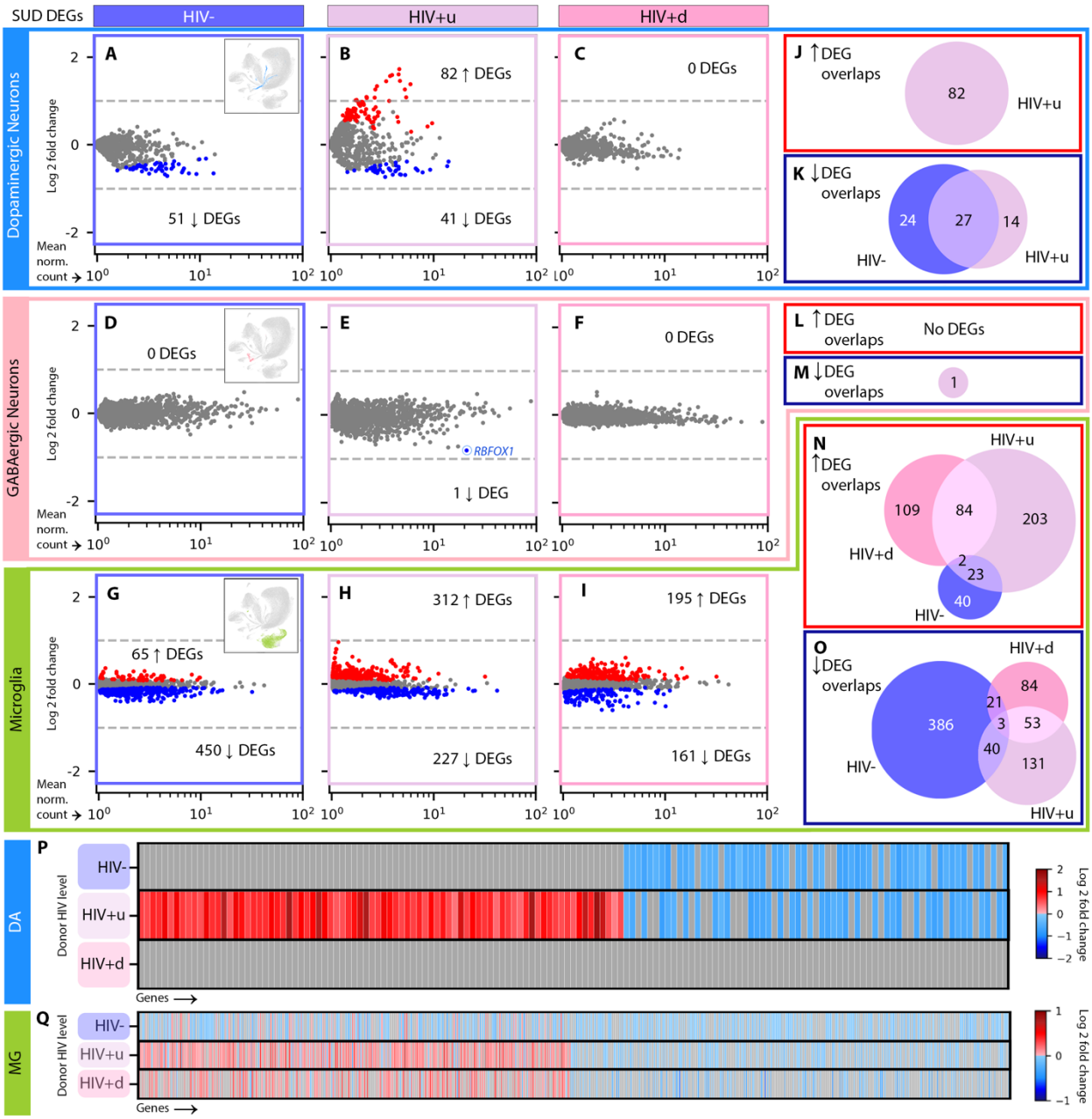


Fig. 2. SUD-attributed DEGs by cell type. (A)-(C) SUD DEGs in DA neurons of **(A)** HIV- (dark purple), **(B)** HIV+u (light purple), and **(C)** HIV+d (pink) donors. Red and blue dots: significantly up- and down- regulated DEGs, respectively ($p < 0.05$). Numbers of up- and down- DEGs are shown. **(D)-(F):** SUD DEGs in GABA neurons for **(D)** HIV-, **(E)** HIV+u, and **(F)** HIV+d donors. In **(E)**, the DEG *RBFOX1* is circled. **(G)-(I)** SUD DEGs in microglia for **(G)** HIV-, **(H)** HIV+u, and **(I)** HIV+d donors. **(J)-(O)** Venn diagrams showing the extent of overlap among SUD DEGs across HIV levels for **(J)** DA neuron up- and **(K)** down- DEGs; **(L)** GABA neuron up- and **(M)** down- DEGs; and **(N)** microglia up- and **(O)** down-DEGs. **(P)-(Q)** Gene-by-gene comparisons of differential transcription across HIV levels for **(P)** DA neurons and **(Q)** microglia. Columns in **(P)**, **(Q)** contain all genes arising as SUD DEGs at any HIV level; rows correspond to HIV levels. Genes per HIV level (row) are colored by log2 fold change (l2fc) if they are DEGs (see color bar) or gray if not. DA neuron, GABA neuron, and microglial nuclei are shown in the UMAP insets in **(A)**, **(D)**, and **(G)**, respectively.

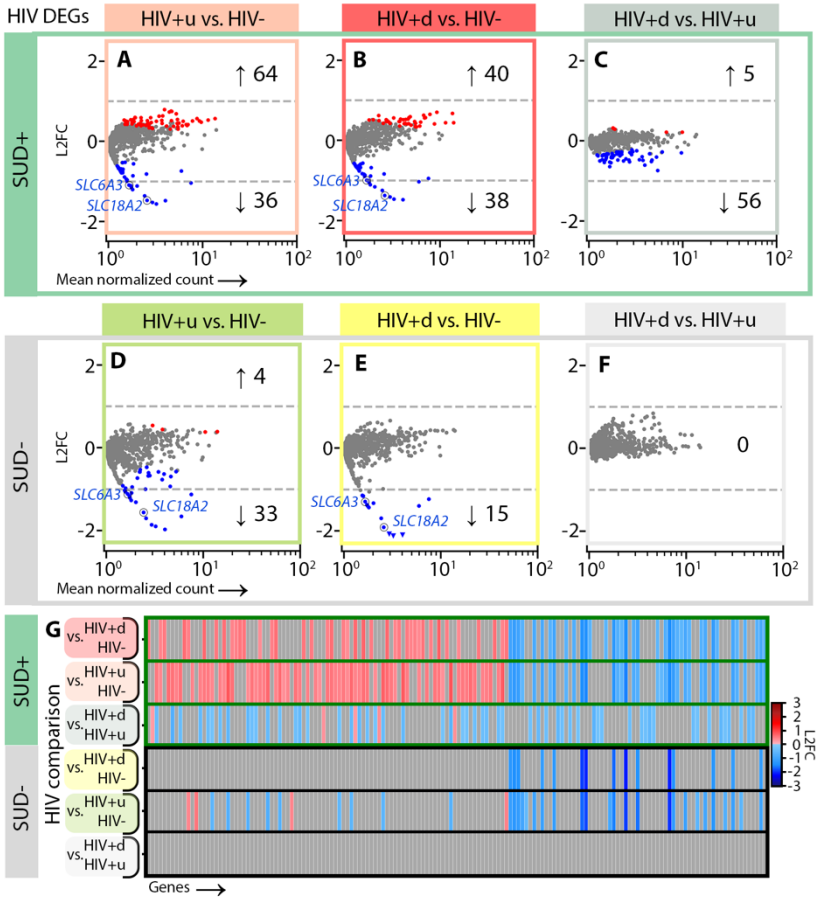


Fig. 3. HIV-attributed DEGs for DA neurons. (A)-(C) HIV DEGs occurring in SUD+ individuals ($p < 0.05$) for DA neurons, for DEAs of **(A)** SUD+HIV+u vs. SUD+HIV- (peach), **(B)** SUD+HIV+d vs. SUD+HIV- (red), and **(C)** SUD+HIV+d vs. SUD+HIV+u (gray). **(D)-(F)** HIV DEGs occurring in SUD- individuals for DEAs of **(D)** SUD-HIV+u vs. SUD-HIV- (green), **(E)** SUD-HIV+d vs. SUD-HIV- (yellow), and **(F)** SUD-HIV+d vs. SUD-HIV+u (gray). **(G)** Gene-by-gene comparisons of differential transcription across HIV comparisons for (top 3 rows) SUD+ and (bottom 3 rows) SUD-, similar to **Figs. 2P**, **Q**. In **(A)**, **(B)**, **(D)**, and **(E)**, monoamine transporter DEGs *SLC6A3* and *SLC18A2* are circled.

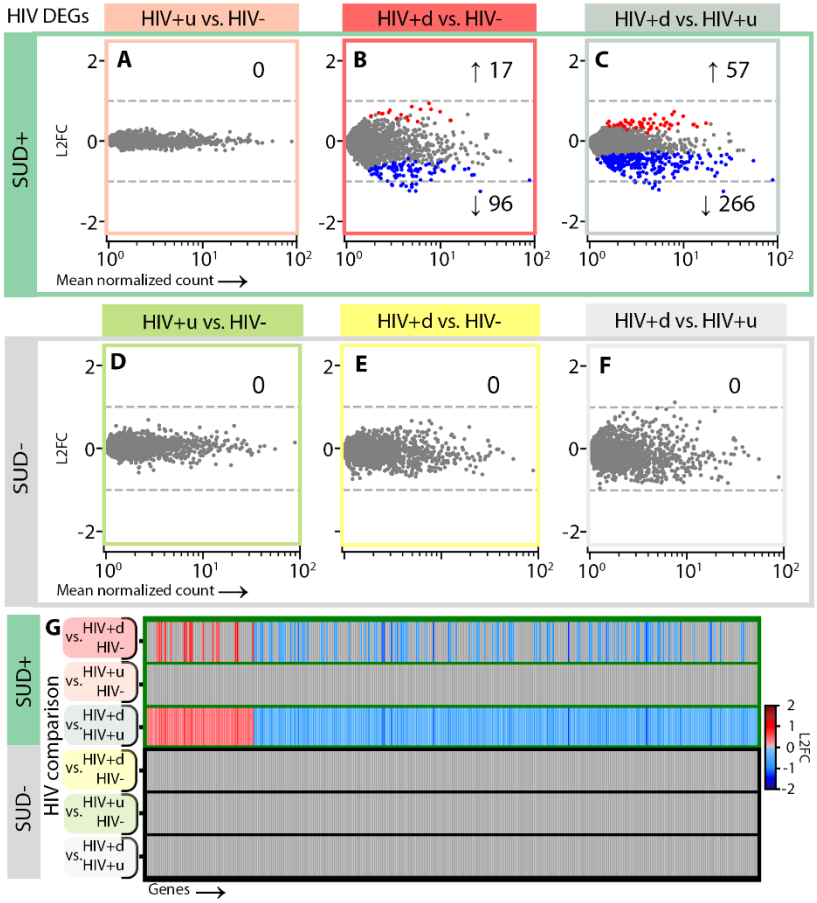


Fig. 4. HIV-attributed DEGs for GABA neurons. (A)-(C) HIV DEGs ($p < 0.05$) for GABA neurons in SUD+ donors, for DEAs of **(A)** SUD+HIV+u vs. SUD+HIV- (peach), **(B)** SUD+HIV+d vs. SUD+HIV- (red), and **(C)** SUD+HIV+d vs. SUD+HIV+u (gray). **(D)-(F)** HIV DEGs in SUD- donors for **(D)** SUD-HIV+u vs. SUD-HIV- (green), **(E)** SUD-HIV+d vs. SUD-HIV- (yellow), and **(F)** SUD-HIV+d vs. SUD-HIV+u (gray). **(G)** Gene-by-gene comparisons of differential transcription across HIV comparisons for SUD+ and SUD- donors.

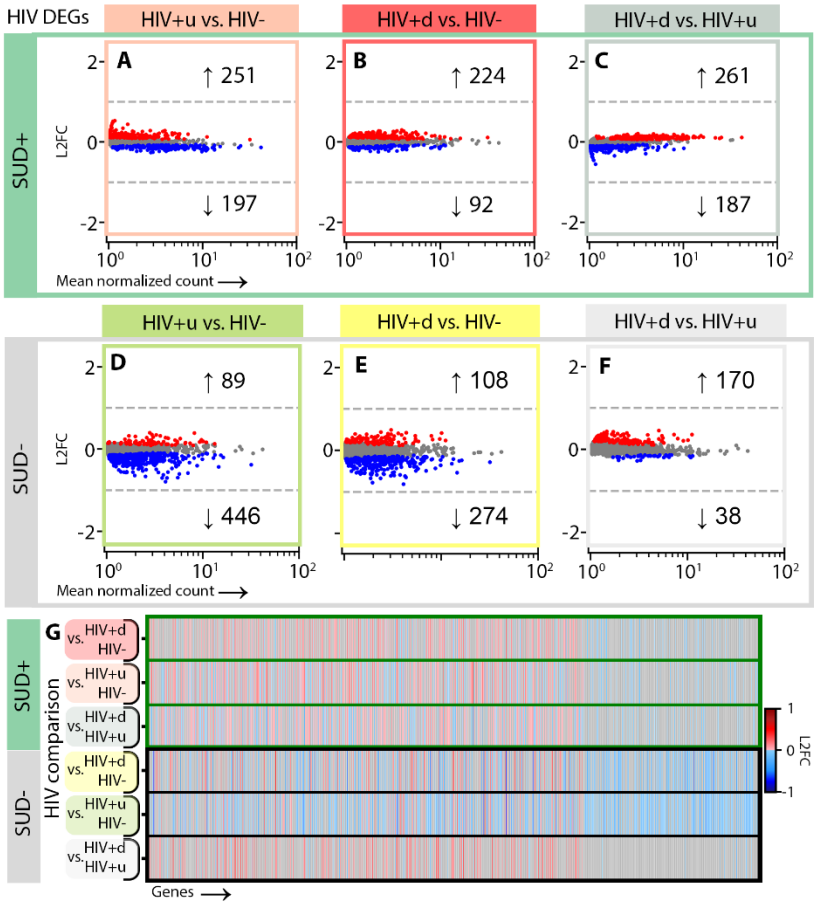


Fig. 5. HIV-attributed DEGs for microglia. (A)-(C) HIV DEGs ($p < 0.05$) for microglia in SUD+ donors for DEAs of **(A)** SUD+HIV+u vs. SUD+HIV- (peach), **(B)** SUD+HIV+d vs. SUD+HIV- (red), and **(C)** SUD+HIV+d vs. SUD+HIV+u (gray). **(D)-(F)** HIV DEGs in SUD- donors for DEAs of **(D)** SUD-HIV+u vs. SUD-HIV- (green), **(E)** SUD-HIV+d vs. SUD-HIV- (yellow), and **(F)** SUD-HIV+d vs. SUD-HIV+u (gray). **(G)** Gene-by-gene comparisons of differential transcription across HIV comparisons for SUD+ and SUD- donors.

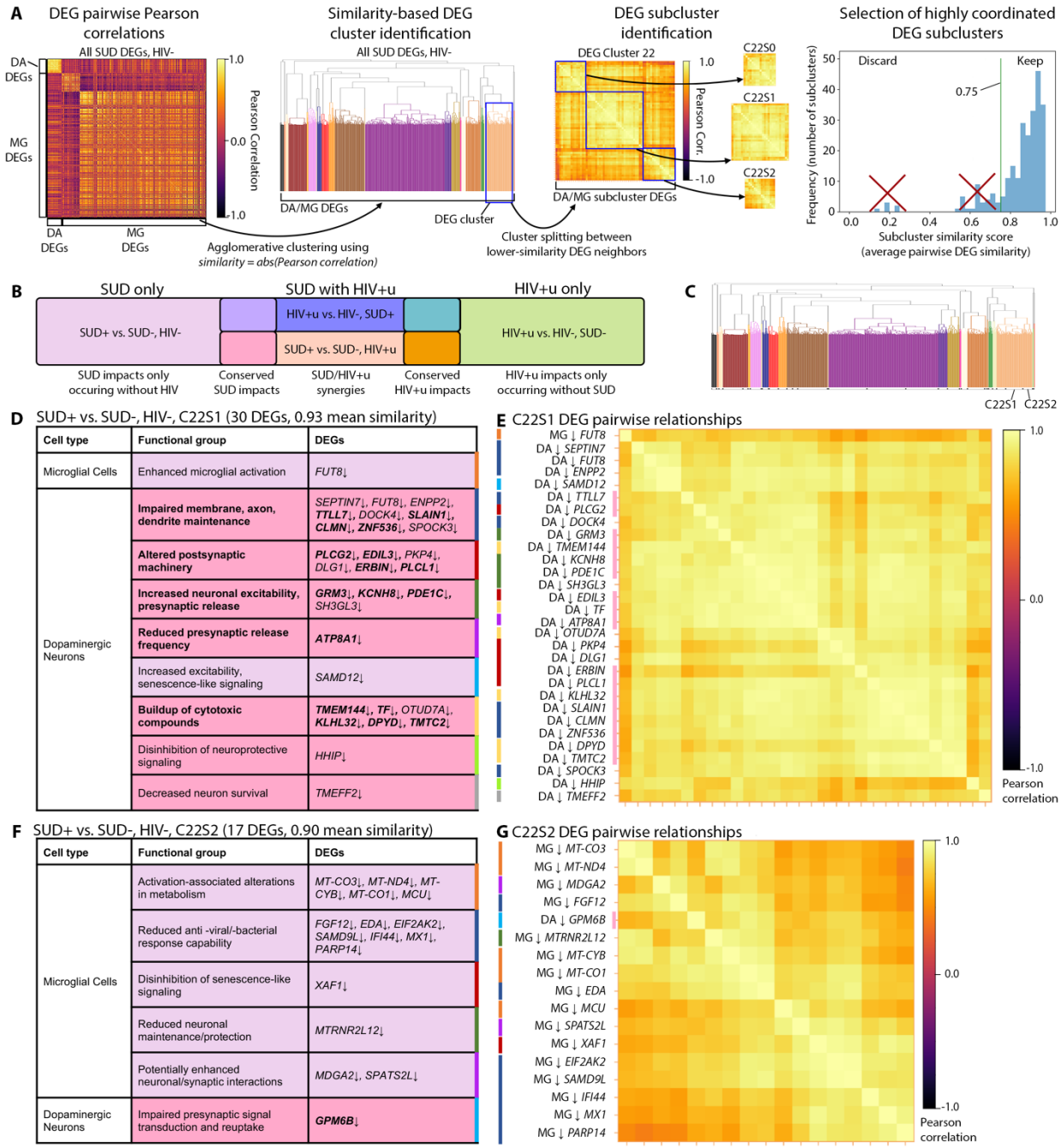


Fig. 6. DEG subclustering approach and DA neuron/microglial DEG subclusters showing SUD impacts. (A) Schematic of our approach for identifying DEG subclusters, for DA neuron (DA) and microglial (MG) DEGs, for the DEA comparing SUD+HIV- vs. SUD-HIV- (Materials and Methods). **(B)** Venn diagram showing interpretations for DEG subcluster functions based on the DEAs they arise in, for the 4 DEAs assessing impacts of SUD (left, **Figs. 2A, G**), HIV+u (right, **Figs. 3D, 5D**), or SUD+HIV+u (middle, **Figs. 2B, H**, **3A, 5A**). **(C)** Dendrogram for DEGs from SUD+HIV- vs. SUD-HIV-, showing subclusters for SUD impacts. Labels denote subclusters shown in **(D)-(G)**. **(D)** Subcluster C22S1 DEGs, organized by functional group (indexed by color at right). Fill color follows the legend in **(B)**; bold denotes DEGs recurring as SUD impacts in SUD+HIV+u donors. **(E)** Heatmap showing pairwise DEG correlations for subcluster C22S1, with DEG names colored by functional group to left and pink to right if recurring in SUD+HIV+u. **(F)**, **(G)** Similar to **(D)**, **(E)**, but for subcluster C22S2.

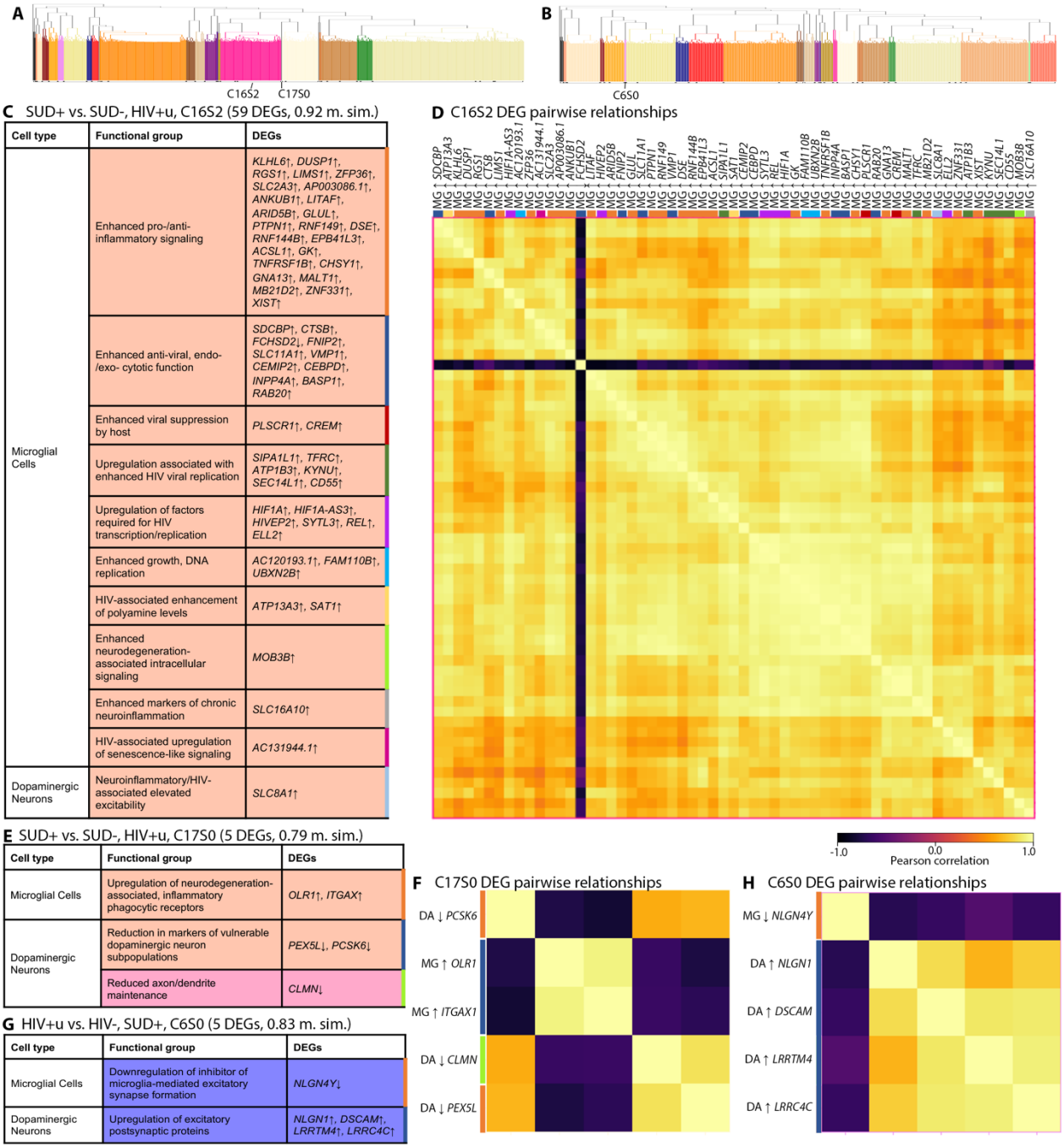


Fig. 7. DA neuron/microglial DEG subclusters showing added functional impacts for SUD+HIV+u donors. (A) Dendrogram relating DA/MG DEGs for the comparison of SUD+HIV+u vs. SUD-HIV+u (**Figs. 2B**, **H**), with labels indicating subclusters shown in **(C)**-**(F)**. **(B)** Dendrogram relating DEGs for SUD+HIV+u vs. SUD+HIV- (**Figs. 3A**, **5A**), with labels indicating subcluster in **(G)**, **(H)**. **(C)** Subcluster C16S2 DEGs organized by functional group (indexed by color at right); fill color follows the legend in **Fig. 6B** (pink: SUD impacts; orange: impacts specific to SUD+HIV+u dual diagnosis). "m. sim": subcluster mean pairwise similarity. **(D)** Heatmap of pairwise DEG correlations for subcluster C16S2, with DEG names colored by functional group. **(E)**, **(F)** Similar to **(C)**, **(D)** but for subcluster C17S0. **(F)**, **(H)** Similar to **(C)**, **(D)** but for subcluster C6S0.

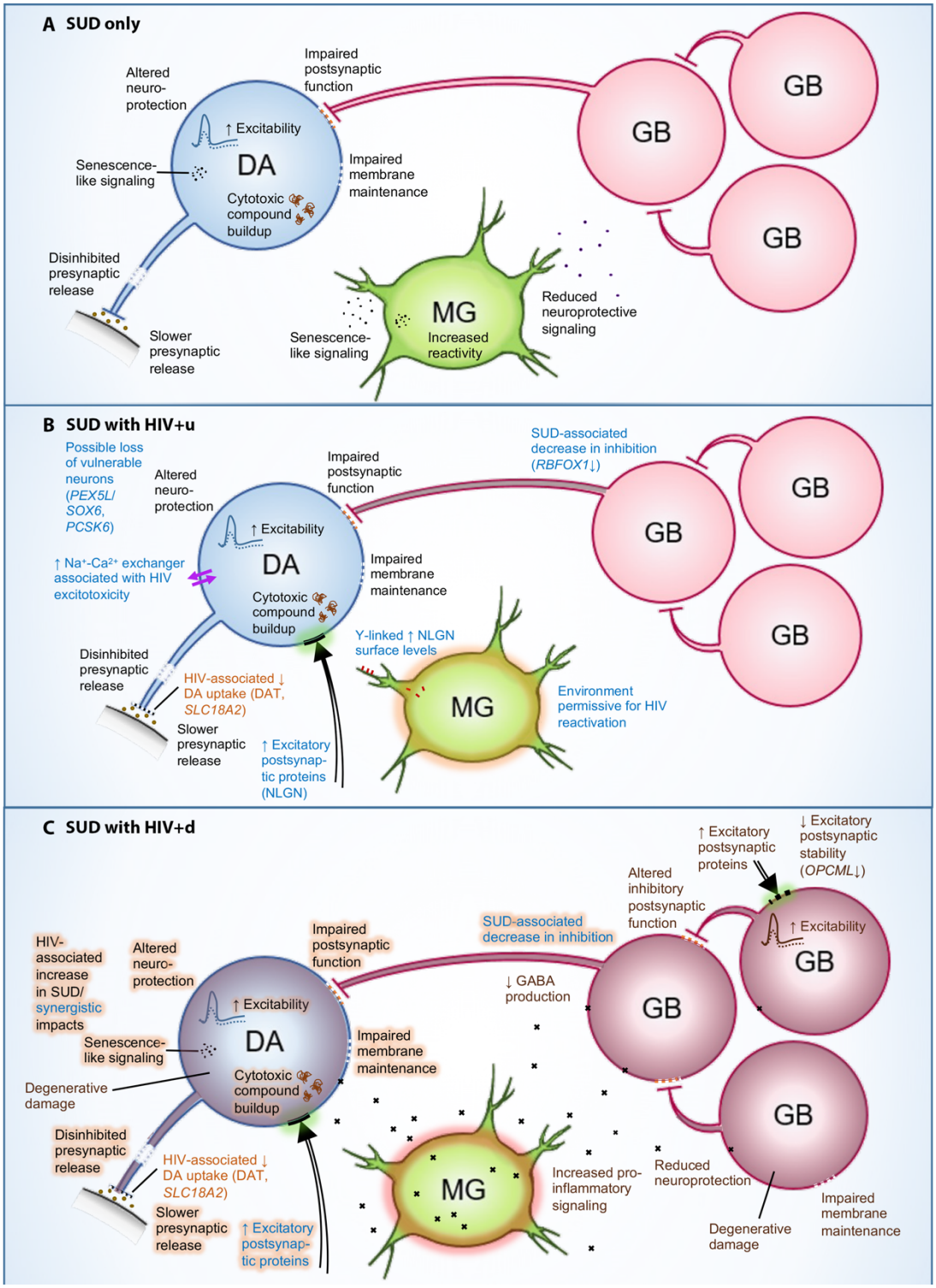


Fig. 8. Summary of functional impacts identified for SUD and HIV. Schematics summarizing the biological impacts identified for **(A)** SUD only, **(B)** SUD with HIV+u, and **(C)** SUD with HIV+d, on DA neurons (blue cells; "DA"), GABA neurons (pink cells; "GB"), and microglia (green cells; "MG"). Descriptions indicate biological impacts, with font color showing DEA-based association with HIV and/or SUD (black: SUD impacts; light brown: HIV+u impacts; blue: HIV+u/SUD+ synergistic impacts; dark brown: HIV+d impacts), and with light brown highlights indicating HIV+d-associated enhancement of impacts.

1628 **Table 1. Demographic characteristics of donor groups.**

	Entire Sample	SUD- HIV- (n=13)	SUD+ HIV- (n=15)	SUD- HIV+u (n=10)	SUD+ HIV+u (n=20)	SUD- HIV+d (n=7)	SUD+ HIV+d (n=25)
Mean (standard deviation) age [years]	54.0 (11.4)	47.5 (2.9)	54.3 (2.7)	60.1 (3.3)	60.8 (2.3)	54.0 (4.0)	49.4 (2.1)
Number (percent) male sex	51 (57%)	7 (54%)	12 (80%)	5 (50%)	10 (50%)	4 (57%)	13 (52%)
Race/ethnicity by number (percent)	34 (38%) Non-Hisp Black 30 (33%) Hisp. 20 (22%) Non-Hisp White 6 (7%) Other	2 (15%) Non-Hisp Black 6(46%) Hisp 3 (23%) Non-Hisp White 2 (15%) Other	4 (27%) Non-Hisp Black 5(33%) Hisp 6 (40%) Non-Hisp White 0 (0%) Other	2 (20%) Non-Hisp Black 2(20%) Hisp 6 (60%) Non-Hisp White 0 (0%) Other	9 (45%) Non-Hisp Black 8(40%) Hisp 1 (5%) Non-Hisp White 2 (10%) Other	3 (43%) Non-Hisp Black 1(14%) Hisp 1 (14%) Non-Hisp White 2 (29%) Other	14 (56%) Non-Hisp Black 8(32%) Hisp 3 (12%) Non-Hisp White 0 (0%) Other
Median (IQR) CD4 T cell count	165 (71, 425)	N/A	N/A	378 (208, 648)	269 (135, 421)	242 (185, 400)	69 (15, 102)
Median (IQR) nadir CD4 T cell count	54 (12, 139)	N/A	N/A	113 (80, 270)	60 (13, 172)	93 (17, 185)	22 (6, 58)
Opioid/ cocaine SUD diagnoses by number	Opioid and cocaine 28 Opioid only 14 Cocaine only 18	N/A	Opioid and cocaine 10 Opioid only 4 Cocaine only 1	N/A	Opioid and cocaine 10 Opioid only 4 Cocaine only 6	N/A	Opioid and cocaine 8 Opioid only 6 Cocaine only 11
Number (percent) with exposure to other substances	67 (74%)	3 (23%)	10(67%)	8 (80%)	19 (95%)	6 (86%)	21 (84%)

1629 Statistical comparisons of donor characteristics across donor groups are provided in Table S2.
1630 Numbers of nuclei by donor group are provided in Table S3.
1631
1632
1633

Supplementary Materials for

Transcriptional impacts of substance use disorder and HIV on human ventral midbrain neurons and microglia

Alyssa Wilson* *et al.*

*Corresponding author. Email: alyssa.wilson@mssm.edu

This PDF file includes:

Figs. S1 to S5
Tables S3, S6, S8

Other Supplementary Materials for this manuscript include the following:

Tables S1, S2, S4, S5, S7
Data S1 to S9

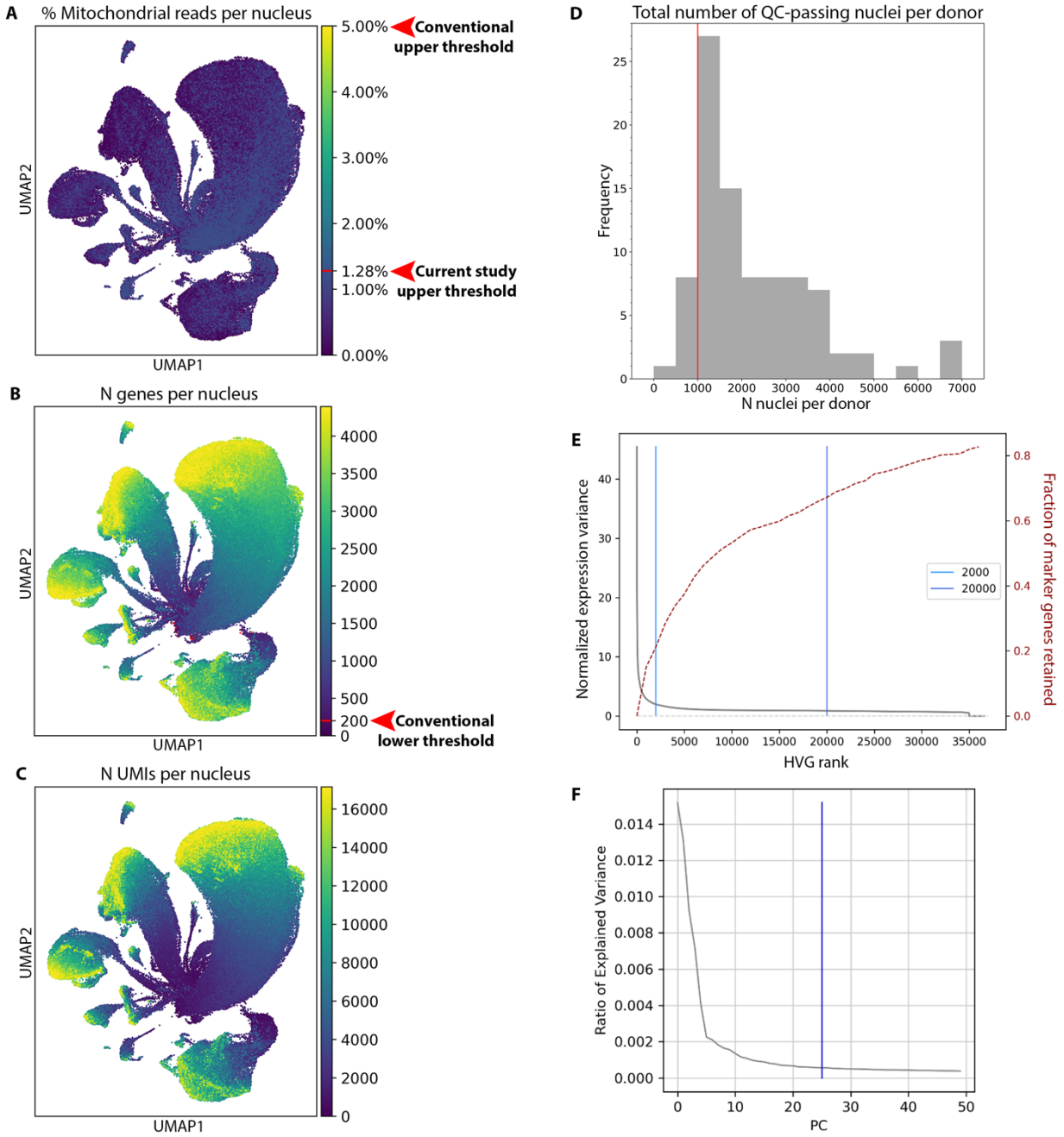


Fig. S1. Additional information about snRNA-seq QC and pre-processing.

(A)-(C) UMAPs showing values of QC heuristics, only for nuclei passing all QC nuclei. **(A)** Percent mitochondrial reads/nucleus. Red line/arrowhead at 1.28% shows the maximum value in our QC-passing nuclei; line at 5% shows a conventional filtering maximum. **(B)** Number of genes/nucleus; red line/arrowhead shows a conventional lower threshold of 200; QC-passing nuclei with <200 genes are colored red (see Fig. S2). **(C)** Number of UMIs/nucleus. **(D)** Distribution of the number of QC-passing nuclei/donor. Red line: the 1,000-nuclei threshold exceeded by 90% of donors. **(E)** Parameters used to determine the number of HVGs retained in analysis. Genes are ranked left to right by decreasing variability; black line (left y-axis) shows each gene's normalized expression variance; red line (right y-axis) shows the fraction of all genes ≤ that rank in our marker gene database. Blue lines show (right) our cutoff (20,000 HVGs) and (left) a conventional one (2,000 HVGs). **(F)** Explained variances of principle component analysis components ("PCs"); blue line shows the maximum component we retained, the 25th PC.

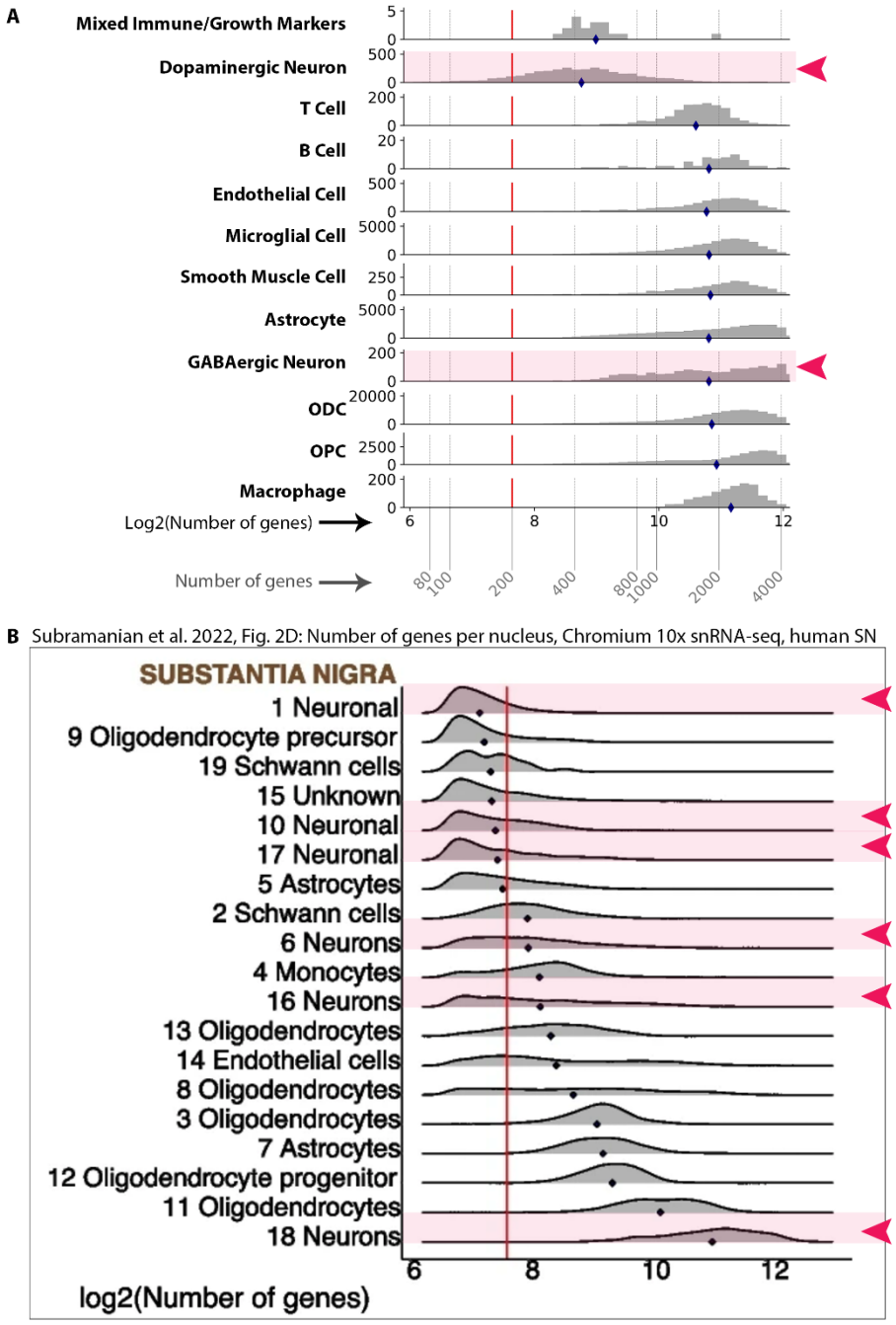


Fig. S2. Comparison of N genes/nucleus, by cell type, to published SN snRNA-seq data. **(A)** Histograms showing the log-transformed number of genes/nucleus, by cell type, for our QC-passing nuclei (**Figs. 1B**, **C**). Gray dashed lines show corresponding raw (non-log-transformed) gene counts. **(B)** Histograms showing log-transformed number of genes/nucleus for each cluster identified in previously published human SN snRNA-seq data, prepared similarly to ours (via Chromium 10x (*31*)). For **(A)** and **(B)**, neuronal cell types are marked with pink highlights and arrowheads, distribution means are indicated by blue diamonds, and red lines show the conventional lower threshold of 200 genes/nucleus.

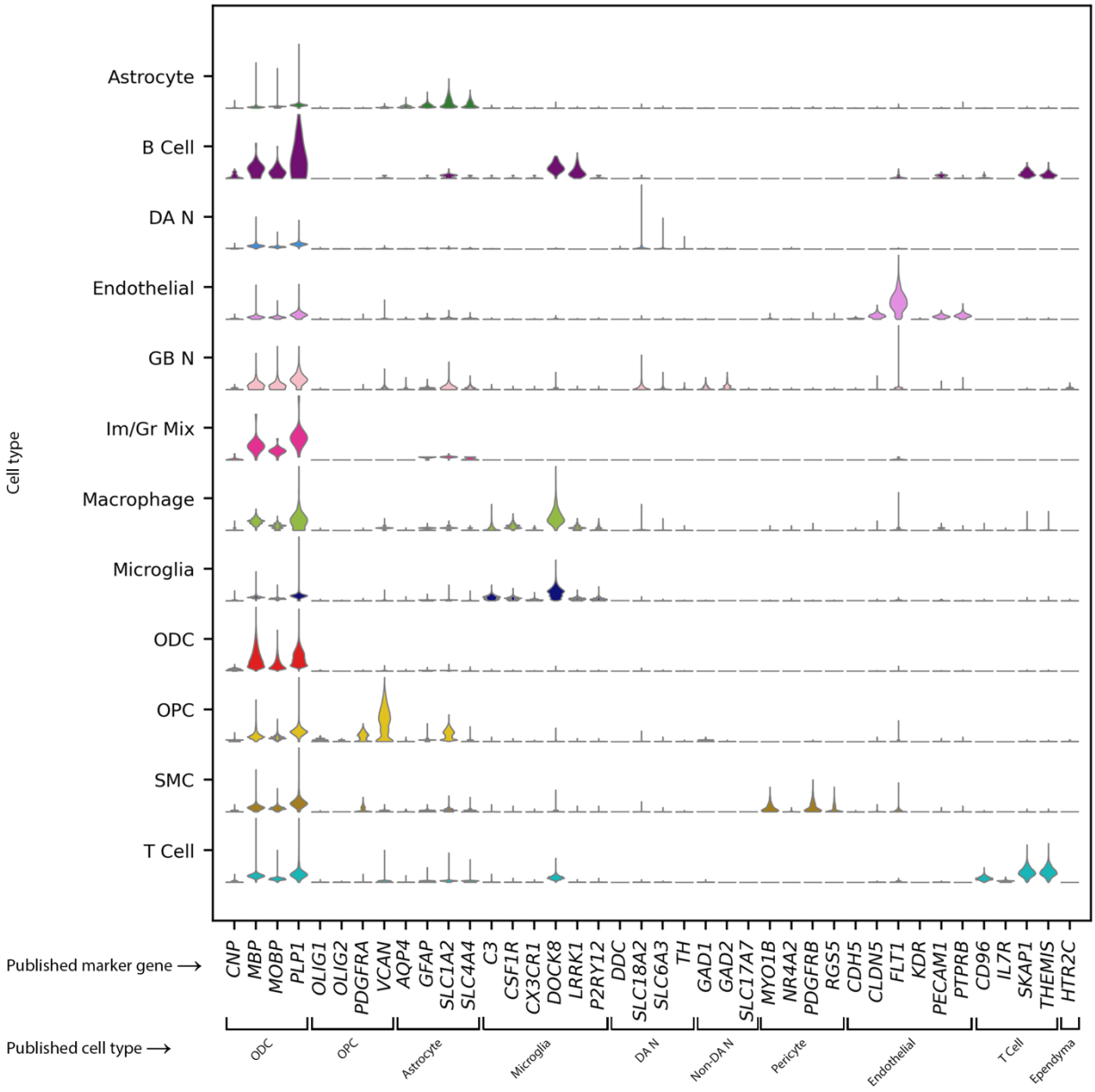


Fig. S3. Expression of previously published SN cell type marker genes, by cell type. Expression in each of our cell types (rows) of previously identified (Wei et al. 2023 (*34*)) marker genes for ventral midbrain cell types (columns).

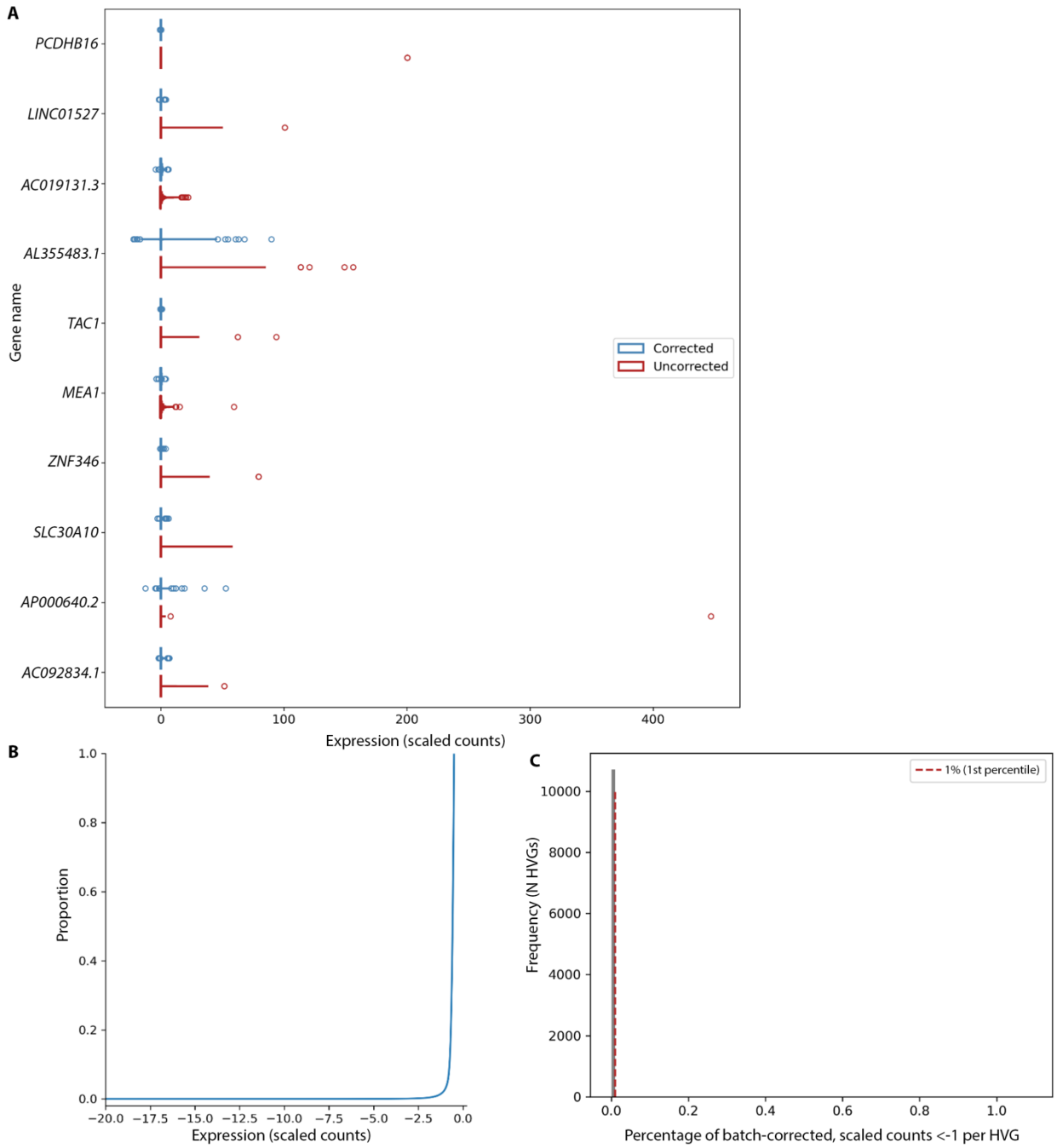


Fig. S4. Impacts of batch correction on gene expression distributions for QCed data. **(A)** Gene expression distributions before (red) vs. after (blue) batch correction, for 10 example HVGs selected randomly from our 20,000. **(B)** An empirical cumulative distribution showing the values of all batch-corrected, still-scaled read counts with values ≤ -0.5 (that would round to non-negligible negative counts ≤-1), highlighting a vanishing number of counts shifted to values <-2 by batch correction. **(C)** Distribution showing the percentages of each HVG's batch-corrected counts with values <-1. Red line denotes 1% (corresponding to the lowest 1st percentile of an HVG's distribution).

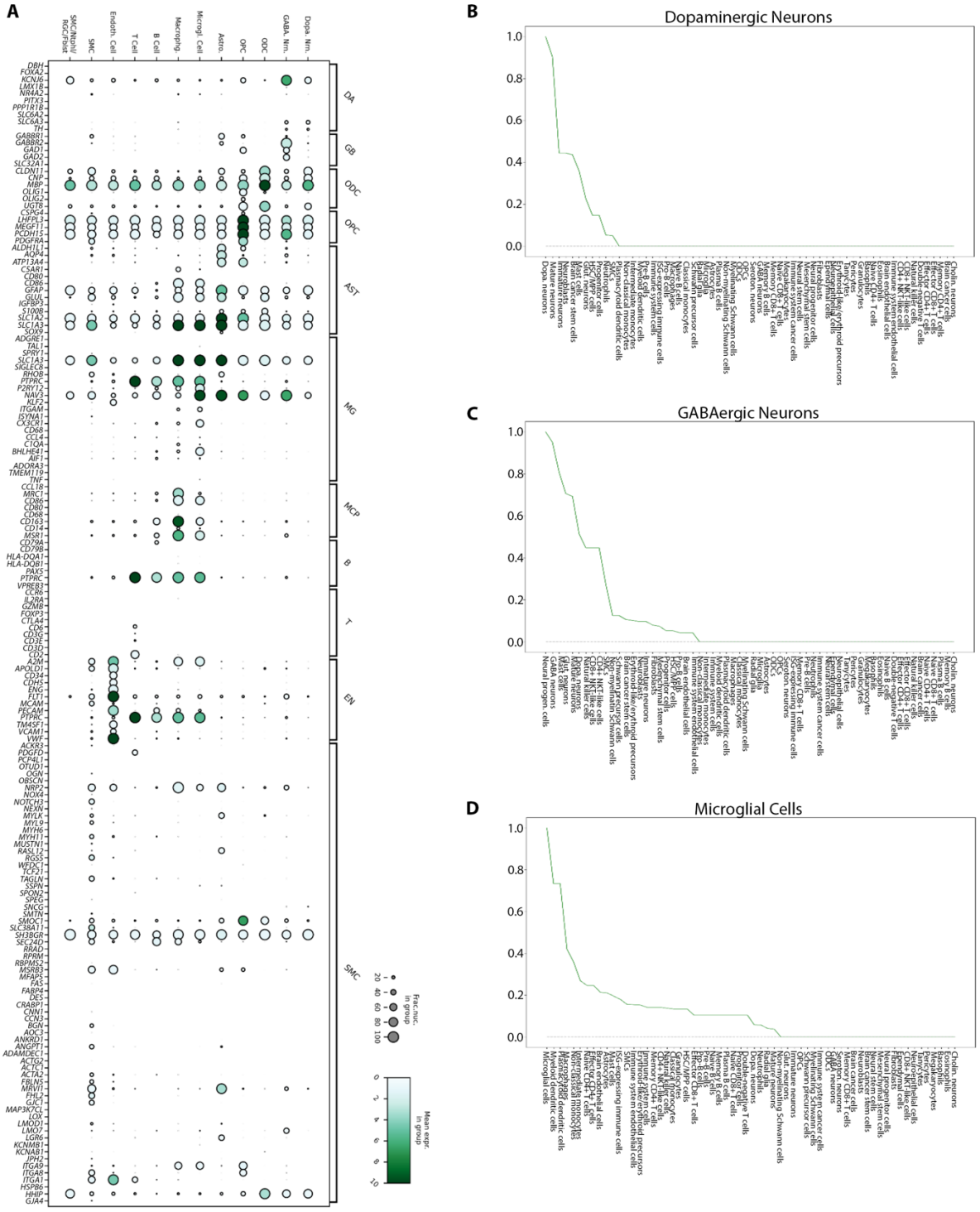


Fig. S5. Further information about likelihood-based cell typing of clusters.

(A) Dot plot showing, for each cell type in our dataset (top labels), expression of various expected SN cell type marker genes (left labels: expected marker genes, right labels: expected SN cell types). **(B)-(D)** Possible database (Data S6) cell types, ordered left to right by decreasing log-likelihood (and normalized to a maximum value of 1), for our **(B)** DA neuron, **(C)** GABA neuron, and **(D)** microglia clusters.

1712 **Table S1. (separate file)**
1713 Donor characteristics.
1714
1715 **Table S2. (separate file)**
1716 Statistical comparisons of donor characteristics across donor groups. Bold text with asterisks
1717 indicates $p<0.05$. Double dashes indicate repeated comparisons or not-relevant self-comparisons.
1718
1719 **Table S3.**
1720 Numbers of donors and nuclei per SUD drug class, by HIV status.

1721 **Table S4. (separate file)**
1722 snRNA-seq nuclear yields by donor and cell type.

1723 **Table S5. (separate file)**
1724 Statistical comparisons of cell type proportions across donor groups. Multiple-comparison-
1725 adjusted p-values are shown for Kolmogorov-Smirnov tests; bold and asterisks indicate $p<0.05$.
1726 The cell type proportions per donor group used in these comparisons are shown below this table.
1727
1728 **Table S6.**
1729 DA neuron (DA)/microglia (MG) DEG subcluster compositions, for SUD+/HIV+u DEAs.
1730
1731 **Table S7. (separate file)**
1732 Comparison of gene set enrichment results produced using Enrichr vs. GSEApy.
1733
1734 **Table S8.**
1735 Fractions of significantly correlated DEG pairs in subclustering analyses.
1736
1737

1738 **Table S3. Numbers of donors and nuclei per SUD drug class, by HIV status.**

HIV status	N donors N nuclei cocaine SUD	N donors N nuclei opioid SUD	N donors N nuclei cocaine + opioid SUDs	N donors N nuclei neither SUD	Total N for HIV status
HIV-	1 2,083	4 5,421	10 28,610	13 27,349	28 63,463
HIV+u	6 16,593	4 10,607	10 23,281	10 21,683	30 72,164
HIV+d	11 20,867	6 19,229	8 15,333	7 9,676	32 65,105
Total N for SUD drug class	18 39,543	14 35,257	28 67,224	30 58,708	90 200,732

1739
1740

1741 **Table S6. DA neuron (DA)/microglia (MG) DEG subcluster compositions, for SUD+/HIV+u**
1742 **DEAs.**

DEA (effect)	Total N DEG subclusters	Median (IQR) N DEGs per subcluster	Median (IQR) subcluster mean similarity	N mixed DA/MG DEG subclusters	Median (IQR) N DEGs per mixed DA/MG subcluster	Median (IQR) mixed DA/MG subcluster mean similarity
SUD+HIV- vs. SUD-HIV- **(SUD impacts without HIV)**	28	13 (7, 23)	0.88 (0.83, 0.94)	2	24 (20, 27)	0.91 (0.91, 0.92)
SUD-HIV+u vs. SUD-HIV- **(HIV+u impacts without SUD)**	34	8 (5, 20)	0.94 (0.89, 0.95)	6	10 (7, 14)	0.90 (0.87, 0.92)
SUD+HIV+u vs. SUD-HIV+u **(SUD impacts with HIV+u)**	28	12 (7, 21)	0.89 (0.85, 0.92)	6	16 (8, 38)	0.87 (0.80, 0.91)
SUD+HIV+u vs. SUD+HIV- **(HIV+u impacts with SUD)**	27	9 (5, 22)	0.89 (0.85, 0.92)	3	9 (7, 23)	0.90 (0.87, 0.91)
All	117	10 (6, 21)	0.91 (0.86, 0.95)	17	12 (6, 20)	0.90 (0.86, 0.92)

1743
1744
1745

1746 **Table S8. Fractions of significantly correlated DEG pairs in subclustering analyses.**

Cell type mix	N DEAs analyzed	Median (IQR) fraction significantly correlated DEG pairs per DEA, across DEAs
DA neurons, microglia	4	0.30 (0.28, 0.30)
GABA neurons, microglia	2	0.27 (0.27, 0.28)
DA neurons, GABA neurons	2	0.46 (0.44, 0.49)
All	8	0.30 (0.28, 0.34)

1747
1748

Data S1. (separate file)

DEG lists, organized by cell type (DA neuron, GABA neuron, or microglia) and DEA type (impacts of HIV or SUD).

Data S2. (separate file)

Lists of Enrichr gene sets enriched by DEGs per DEA, cell type, and direction of dysregulation (up/down). Significantly enriched gene sets (with adjusted p-value <0.05; column E) are highlighted in green.

Data S3. (separate file)

Basic annotations for DEG subclusters, organized by cell type.

Data S4. (separate file)

Focused annotations for DEG subclusters, with references.

Data S5. (separate file)

UMAPs comparing expression data before (left) vs. after (right) batch correction, showing nuclei from one donor across pooled libraries ("batches"). Each plot title lists the donor's ID, and each library is shown in a different color.

Data S6. (separate file)

Cell type marker gene database used for cell typing.

Data S7. (separate file)

Heatmaps of DEG subclusters showing significant DEG pairwise Pearson correlations and related underlying data.

Data S8. (separate file)

Heatmaps showing mean normalized counts for each subcluster's DEGs, per case group donor, with donors labeled by SUD drug class (opioid, cocaine, or opioid+cocaine) and organized by their cross-DEG expression similarities (shown with a dendrogram above each heatmap).

Data S9. (separate file)

Zip file containing lists of GSEApy prerank gene sets enriched by each DEA's DEGs, organized by cell type. Files with names ending in "__all" include all gene sets having any DEGs; files ending with "__sig" include only significantly enriched gene sets.