

UC San Diego

UC San Diego Previously Published Works

Title

A single-nucleus transcriptomic atlas of medium spiny neurons in the rat nucleus accumbens

Permalink

<https://escholarship.org/uc/item/9vq3h6wz>

Journal

Scientific Reports, 14(1)

ISSN

2045-2322

Authors

Reiner, Benjamin C

Chehimi, Samar N

Merkel, Riley

et al.

Publication Date

2024-08-06

DOI

10.1038/s41598-024-69255-0

Peer reviewed



OPEN

A single-nucleus transcriptomic atlas of medium spiny neurons in the rat nucleus accumbens

Benjamin C. Reiner¹, Samar N. Chehimi¹, Riley Merkel^{1,2}, Sylvanus Toikumo¹, Wade H. Berrettini¹, Henry R. Kranzler^{1,3}, Sandra Sanchez-Roige^{4,5,6}, Rachel L. Kember^{1,3}, Heath D. Schmidt^{1,2,8} & Richard C. Crist^{7,8}✉

Neural processing of rewarding stimuli involves several distinct regions, including the nucleus accumbens (NAc). The majority of NAc neurons are GABAergic projection neurons known as medium spiny neurons (MSNs). MSNs are broadly defined by dopamine receptor expression, but evidence suggests that a wider array of subtypes exist. To study MSN heterogeneity, we analyzed single-nucleus RNA sequencing data from the largest available rat NAc dataset. Analysis of 48,040 NAc MSN nuclei identified major populations belonging to the striosome and matrix compartments. Integration with mouse and human data indicated consistency across species and disease-relevance scoring using genome-wide association study results revealed potentially differential roles for MSN populations in substance use disorders. Additional high-resolution clustering identified 34 transcriptomically distinct subtypes of MSNs definable by a limited number of marker genes. Together, these data demonstrate the diversity of MSNs in the NAc and provide a basis for more targeted genetic manipulation of specific populations.

Keywords Single cell transcriptomics, Medium spiny neurons, Nucleus accumbens, Rat

The nucleus accumbens (NAc) plays important roles in drug-taking and -seeking behaviors^{1–3}. The two primary output pathways of the NAc are GABAergic medium spiny neurons (MSNs) that express either dopamine type 1 receptors (D1R-expressing MSNs) or dopamine type 2 receptors (D2R-expressing MSNs)⁴. In rodents, studies show that addictive drugs, including opioids, cocaine, and methamphetamine, elicit differential responses in D1R- and D2R-expressing MSNs^{5–10}. Furthermore, these two striatal MSN populations may have opposing functional roles in reward-related behaviors: whereas activation of D1R-expressing MSNs *increases* compulsive drug intake, activation of D2R-expressing MSNs *decreases* drug reinforcement^{7,9,11,12}. For example, inactivation of D1R-expressing MSNs or activation of D2R-expressing MSNs attenuates heroin seeking, results that support bidirectional modulation of drug-mediated behaviors by distinct MSN cell populations⁸.

The NAc is broadly divided into two main subregions: the core and shell. These subregions have unique functional roles in motivated behaviors including drug taking and seeking^{13–15}. MSNs in the NAc core and shell differ both in their afferent¹⁶ and efferent^{17,18} projections, as well as their roles in motivated behaviors^{19,20}. The entire striatum, including the NAc, also contains two distinct compartments: the matrix, forming the bulk of the striatum, and the patch-like striosome interspersed throughout the matrix²¹. These two compartments are also characterized by different afferent and efferent projections²¹ and, like the core and shell, play unique roles in addiction-like behaviors^{22,23}. The prevailing evidence suggests that the relevance of MSNs to substance use disorders is not fully captured by the standard classifications of D1R/D2R expression and supports a model wherein subgroups of D1R- and D2R-expressing MSNs have distinct roles in addiction-related phenotypes.

Recent single-cell transcriptomic analyses of the NAc have identified distinct subpopulations of MSNs, including groups representing MSNs from the core, shell, matrix, and striosome. Unique gene expression profiles were

¹Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ²Department of Biobehavioral Health Sciences, School of Nursing, University of Pennsylvania, Philadelphia, PA, USA. ³Mental Illness Research, Education and Clinical Center, Crescenz VAMC, Philadelphia, PA, USA. ⁴Department of Psychiatry, University of California San Diego, La Jolla, CA, USA. ⁵Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. ⁶Institute for Genomic Medicine, University of California San Diego, San Diego, CA, USA. ⁷Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, 125 South 31st Street, Room 2207, Philadelphia, PA 19104, USA. ⁸These authors contributed equally to this work: Heath D. Schmidt and Richard C. Crist. ✉email: crist@pennmedicine.upenn.edu

discovered in MSNs from the matrix and striosome of two rhesus macaques²⁴ and a study of humans identified 10 MSN subpopulations, with the vast majority of cells falling into the primary D1R- or D2R-expressing MSN clusters²⁵. Studies of the NAc from mice and rats also identified several D1R-expressing and D2R-expressing MSN subtypes, as well as additional MSN populations with other genetic markers^{26–30}. These novel MSN subpopulations may have unique roles in addiction-like behaviors^{27–30}, similar to the functional and gene expression differences between matrix and striosomal MSNs. Similarly, a recent examination of dorsal striatum from humans and non-human primates showed the presence of D1R- and D2R-expressing MSNs, with populations of matrix and striosome neurons, and these populations had sex- and cell type-specific differences in gene expression in the context of substance use³¹. Thus, a more granular classification of MSN populations is needed to better understand individual MSN cell types and their roles in striatal function and motivated behaviors.

Large single-cell transcriptomic studies can be used to define uncommon and novel neuronal populations in nuclei throughout the brain^{26,32–34}. However, a comprehensive cell type-specific atlas of the NAc has not been generated to date. In this study, we make use of our recently-described, largest available single-cell transcriptomic NAc rat dataset ($n = 96,627$ total nuclei) to present high resolution subclustering of MSNs ($n = 48,040$ MSN nuclei)³⁰. We identified 34 transcriptomically-distinct cell populations, including previously unidentified subtypes of D1R- and D2R-expressing MSNs. We replicated these findings by identifying the same novel MSN subtypes in an independent NAc rat dataset ($n = 7641$ MSN nuclei)^{27,28}. Expression patterns for glutamate, GABA, and acetylcholine receptor genes are presented to further phenotype these subclusters. Finally, MSNs from rat, mouse²⁶, and human²⁵ were integrated together and cells scored using genome-wide association study (GWAS) summary statistics for substance use disorder (SUD) phenotypes^{35–37}, revealing potentially differential roles for MSN subpopulations in alcohol use disorder (AUD), alcohol consumption (AUDIT-C), opioid use disorder (OUD), and tobacco use disorder (TUD).

Results

Identification of five major MSN cell types in the rat nucleus accumbens

Single nucleus transcriptomic data from male, drug-naïve Brown Norway rats ($n = 11$) were aligned to the rat genome and cleaned to remove low-quality nuclei and putative doublets, yielding 96,627 nuclei (Fig. 1A). Clustering revealed the presence of a large cell population expressing high levels of *Bcl11b* and *Pde10a*, two known markers of MSNs (Fig. 1B,C). Additional marker analyses revealed that the dataset also contained populations of GABAergic interneurons, cholinergic neurons, and all expected major types of glial cells (Fig. 1D–K).

To identify populations of MSNs in the NAc, MSNs were subset ($n = 48,040$ MSN nuclei) and clustered at low resolution using principal components derived from variably expressed genes. Five major populations of MSNs were identified (Fig. 2A). Two clusters expressed markers of D2R-expressing MSNs (*Drd2*; Fig. 2C) but differed in the expression patterns of additional marker genes like *Scube1* and *Stk32a* (Fig. 2E and G). The remaining three clusters expressed markers of D1R-expressing MSNs (*Drd1*; Fig. 2B) but were differentiated by expression of *Htr4*, *Ebf1* and *Ppm1e*, among other markers (Fig. 2D,F and H). The identification of separate *Htr4* + and *Ebf1* + clusters

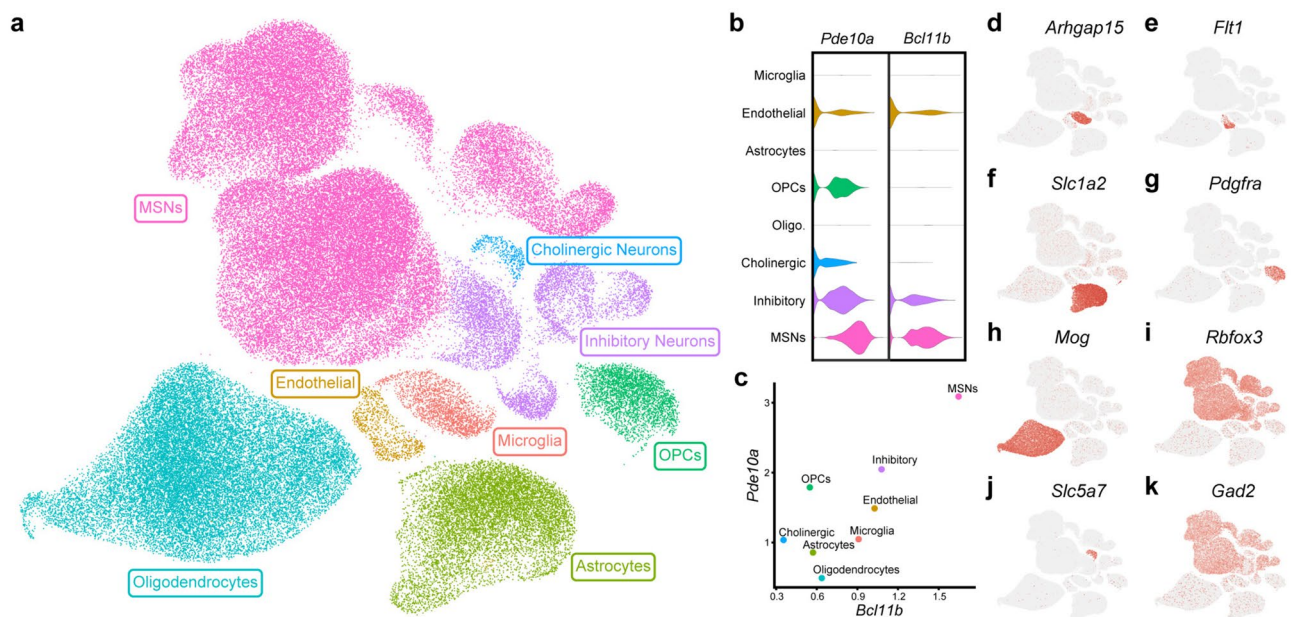


Figure 1. Clustering of snRNA-seq data. Single nucleus RNA sequencing was performed on nucleus accumbens samples from male, drug-naïve Brown Norway rats and data were clustered based on transcriptomic profile. **(a)** Uniform manifold approximation and projection (UMAP) dimension reduction plot with nuclei colored by major cell type. Normalized expression of two MSN marker genes (*Bcl11b* and *Pde10a*) is visualized by both violin plot **(b)** and scatter plot **(c)**. **(d–k)** Feature plots display expression of markers for major glial populations and non-MSN neuronal subtypes for cluster identification.

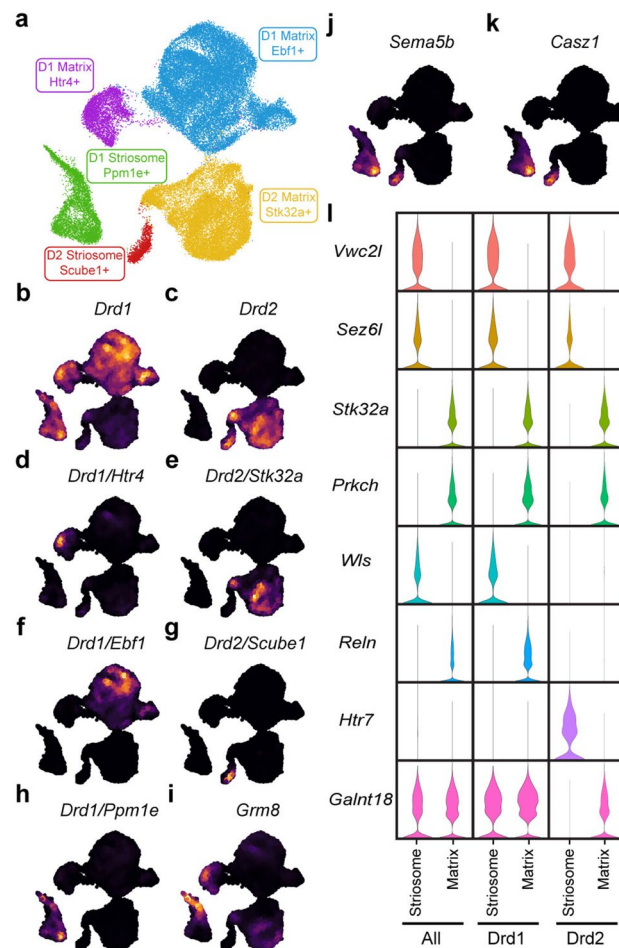


Figure 2. Low resolution clustering of MSNs. (a) UMAP dimension reduction plot of MSN clustered at low resolution. Nuclei are colored and labeled by major MSN population. Nebulosa plots visualizing expression on the MSN UMAP demonstrate (b) *Drd1* expression in three MSN populations and (c) *Drd2* expression in two populations, as well as (d–h) co-expression of major population markers with *Drd1* or *Drd2*. Expression of (i) *Grm8* and the striosomal markers (j) *Sema5b* and (k) *Cas21* are also presented. (l) Violin plots demonstrate expression differences of additional genes between all striosome or matrix MSNs, or only striosome and matrix MSNs expressing either *Drd1* or *Drd2*.

of D1R-expressing MSNs is consistent with another recent snRNA-seq analysis of rat NAc²⁸. D1R-expressing *Ppm1e* + cells have previously been labeled as *Grm8* MSNs^{28,30} due to high expression of *Grm8* (Fig. 2I). However, *Grm8* expression is also present in other D1R-expressing MSN populations (Fig. 2I), whereas *Ppm1e* expression is not (Fig. 2H). These data suggest that *Ppm1e* is the more robust marker for this major MSN subtype.

Previous single-cell analyses of mouse, macaque, and human MSNs have demonstrated that matrix and striosome neurons can be differentiated by their transcriptomic profiles^{24,26,31}. To determine whether these populations were present in our dataset, we analyzed the expression of *Sema5b* and *Cas21*, two known markers of striosome MSNs^{38,39}. Both genes were expressed in the D1R-expressing *Ppm1e* + and D2R-expressing *Scube1* + clusters (Fig. 2J,K), indicating that these are striosome cell populations. In contrast, the *Htr4* + and *Ebf1* + clusters of D1R-expressing MSNs and the *Stk32a* + cluster of D2R-expressing MSNs originate from the matrix. Comparisons of all striosome MSNs versus all matrix MSNs, D1R-expressing striosome versus D1R-expressing matrix, and D2R-expressing striosome versus D2R-expressing matrix revealed large numbers of marker genes for all groups (Fig. 2L and Tables S1–3).

Cross-species conservation of major MSN cell types

To determine whether the major MSN populations are conserved across species, we integrated our rat data with MSNs from previously published drug-naïve mouse and human NAc datasets^{25,26}. A UMAP of the integrated datasets revealed MSN clusters with contributions from all three species (Fig. 3A). These clusters map to our groupings in the rat of D1R striosome (D1R *Ppm1e* +), D1R matrix (D1R *Htr4* + and D1R *Ebf1* +), D2R striosome (D2R *Scube1* +), and D2R matrix (D2R *Stk32a* +) and the integrated data identify the analogous MSN populations in mouse and human NAc (Fig. 3B–E). These findings are consistent with the previous mouse study that identified atypical, “patch-like” or striosome MSN populations expressing either *Drd1* (D1_3, D1_8) or *Drd2*

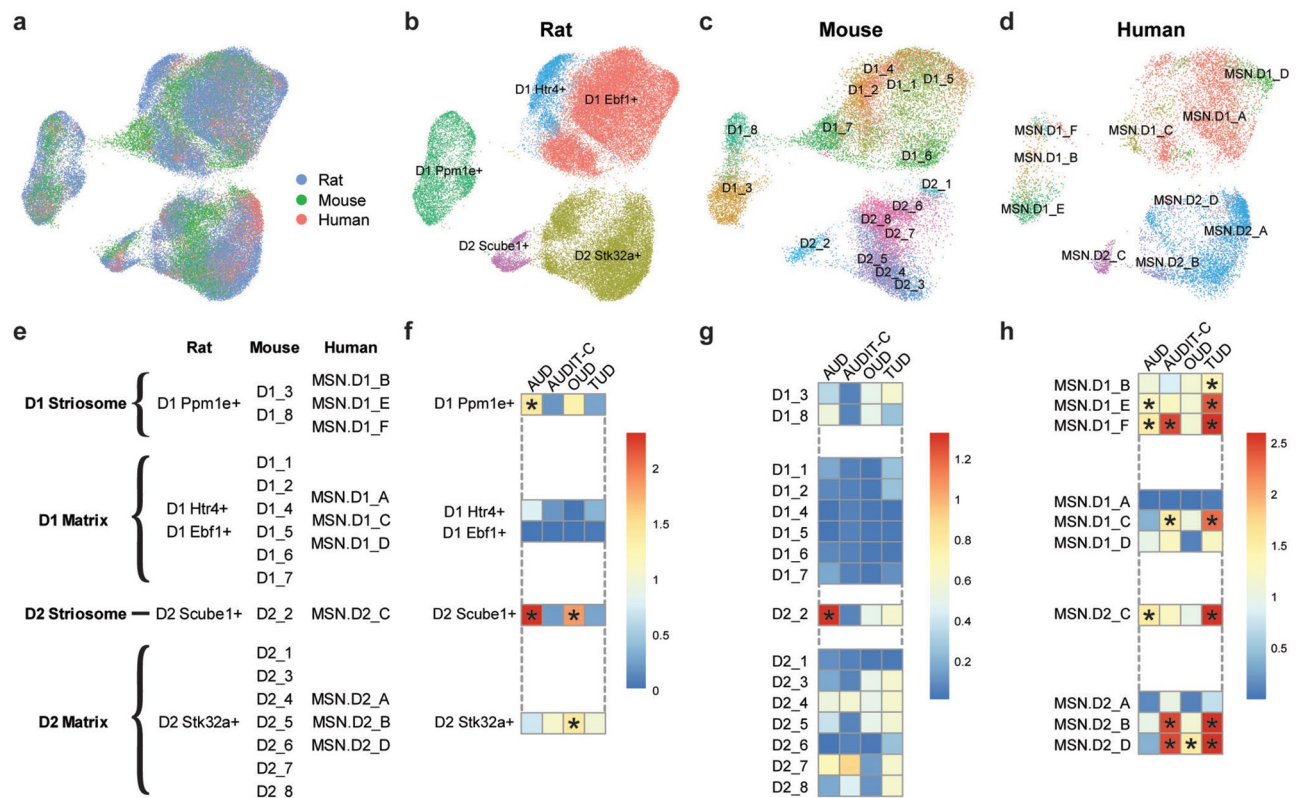


Figure 3. Species comparison and GWAS enrichment. MSN data from the present study were integrated with previously published MSN snRNA-seq data from mouse and human NAc. (a) A combined UMAP shows successful integration of MSNs from all three species. The MSN nuclei from each species are subsequently displayed separately and color-coded based on the cluster identifications from the current study for (b) rat and the original publications for (c) mouse and (d) human. (e) Based on the integrated data, clusters for all species are mapped to one of four major MSN populations: D1 striosome, D1 matrix, D2 striosome, and D2 matrix. Genome-wide association study summary statistics were used to assess MSN clusters in (f) rat, (g) mouse, and (h) human for enriched expression of genes associated with alcohol use disorder (AUD), alcohol consumption (AUDIT-C), opioid use disorder (OUD), and tobacco use disorder (TUD). Heatmap colors indicate the $-\log_{10}$ (p-value) from Monte Carlo tests after correction for multiple testing with a false discovery rate of 0.05. Significant cell type associations are indicated by asterisks.

(D2_2)²⁶ and demonstrate that consistent transcriptomic separation of striosome and matrix MSN cell types is possible across species.

Identification of MSN subtypes associated with substance use disorder phenotypes

Integrating single-cell transcriptomic data with GWAS summary statistics can implicate cell populations in phenotypes of interest and provide targets for downstream functional analyses. To identify MSN cell populations with potential relevance to SUDs, we used scDRS⁴⁰ to calculate disease-relevance scores for each individual cell in the rat, mouse, and human datasets using gene-level GWAS results for AUD³⁷, AUDIT-C³⁷, OUD³⁶, and TUD³⁵. The most consistent finding was a significant enrichment of genes associated with AUD in the D2R-expressing striosome clusters in all three species (Fig. 3F–H). AUD-associated genes were also enriched in additional rat (D1 *Ppm1e*+) and human (MSN.D1_E, MSN.D1_F) MSN clusters, further implicating the striosome compartment in AUD (Fig. 3F–H). Clusters enriched for AUD or AUDIT-C-associated genes were almost entirely non-overlapping, consistent with prior literature showing minimal genetic overlap between these two related phenotypes (Fig. 3F–H)^{37,41}. D2R-expressing matrix MSNs in humans and rats, but not mice, were also significantly enriched for OUD-associated genes (Fig. 3F–H). Genes associated with TUD were also enriched in 7 of the 10 human MSN clusters but no clusters in the rat or mouse MSNs (Fig. 3F–H).

Discovery of novel MSN subtypes in the rat nucleus accumbens

To identify novel MSN subtypes, principal components were derived for each of the five major populations of MSNs. Each population was then subclustered separately and the resulting datasets were merged, which resulted in 34 MSN subclusters (Fig. 4A). The relationships between these new subclusters are shown in the UMAP and constellation plots in Fig. 4A and B, respectively. Marker analyses of the 34 individual subclusters show that these populations are transcriptomically distinct (Fig. 4C–G and Table S4). Although clustered at high-resolution, all subclusters can be separated and defined by the expression patterns of a relatively small number of genes

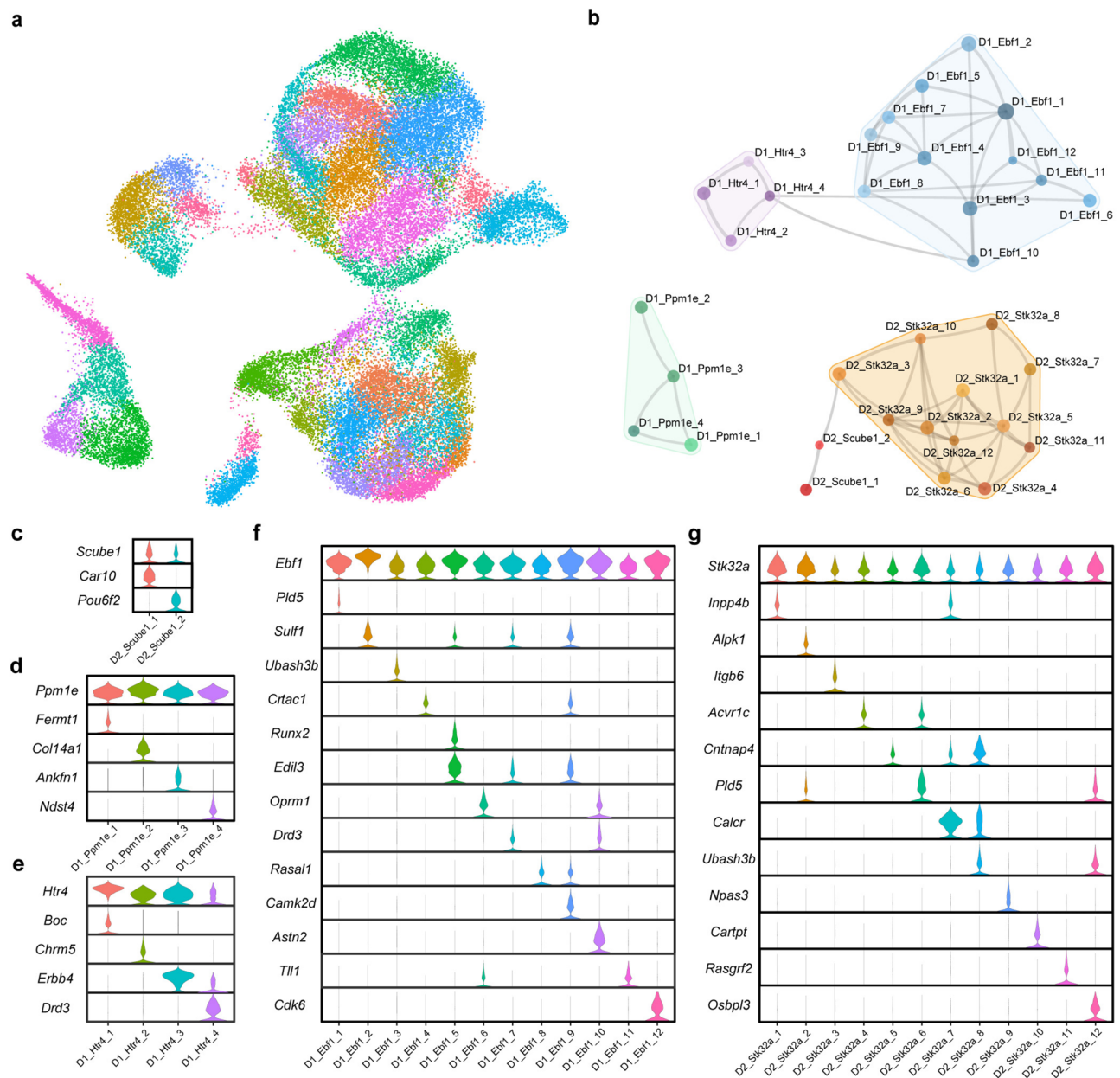


Figure 4. High resolution subclustering of MSNs. Each major population of MSNs was separately subclustered and then remerged. The relationships between the final 34 subclusters are displayed as both (a) a UMAP dimension reduction plot and (b) a constellation plot. Violin plots show normalized expression of marker genes for subclusters within the (c) D2 *Scube1*+, (d) D1 *Ppm1e*+, (e) D1 *Htr4*+, (f) D1 *Ebf1*+, and (g) D2 *Stk32a*+ major MSN populations.

(Fig. 4C–G and Table S5). The subclusters produced by our high-resolution clustering include many previously undescribed subtypes of MSNs. For example, two of the novel MSN cell types, the D1R-expressing *Ppm1e*+ cells expressing *Fermt1* (D1_Ppm1e_1) or *Col14a1* (D1_Ppm1e_2), are highlighted in Fig. 5C and E, respectively. These novel subtypes each make up ~2–3% of all MSNs in the rat NAc. The subclustering analysis also clarified the expression patterns of prior MSN markers. We observed expression of *Chst9*, recently identified as a marker of Grm8 MSNs⁴², in 2 of the 4 D1R *Ppm1e*+ subclusters (Table S4). These data indicate that *Chst9*+ MSNs represent a distinct subset of striosomal D1R-expressing MSNs.

To determine if this more granular classification of MSNs revealed additional associations with SUD phenotypes, the scDRS analysis was performed on the 34 MSN subclusters. The two D2R-expressing *Scube1*+ cell populations were significantly enriched for expression of genes associated with AUD, whereas two D2R-expressing *Stk32a*+ subclusters were enriched for genes associated with AUDIT-C score (Fig. S1). No MSN cell populations were significantly associated with OUD or TUD (Fig. S1).

Higher resolution clustering increases the possibility that some clusters may be driven by sample-specific variation in gene expression rather than variation between cell types, such that the clusters may not represent

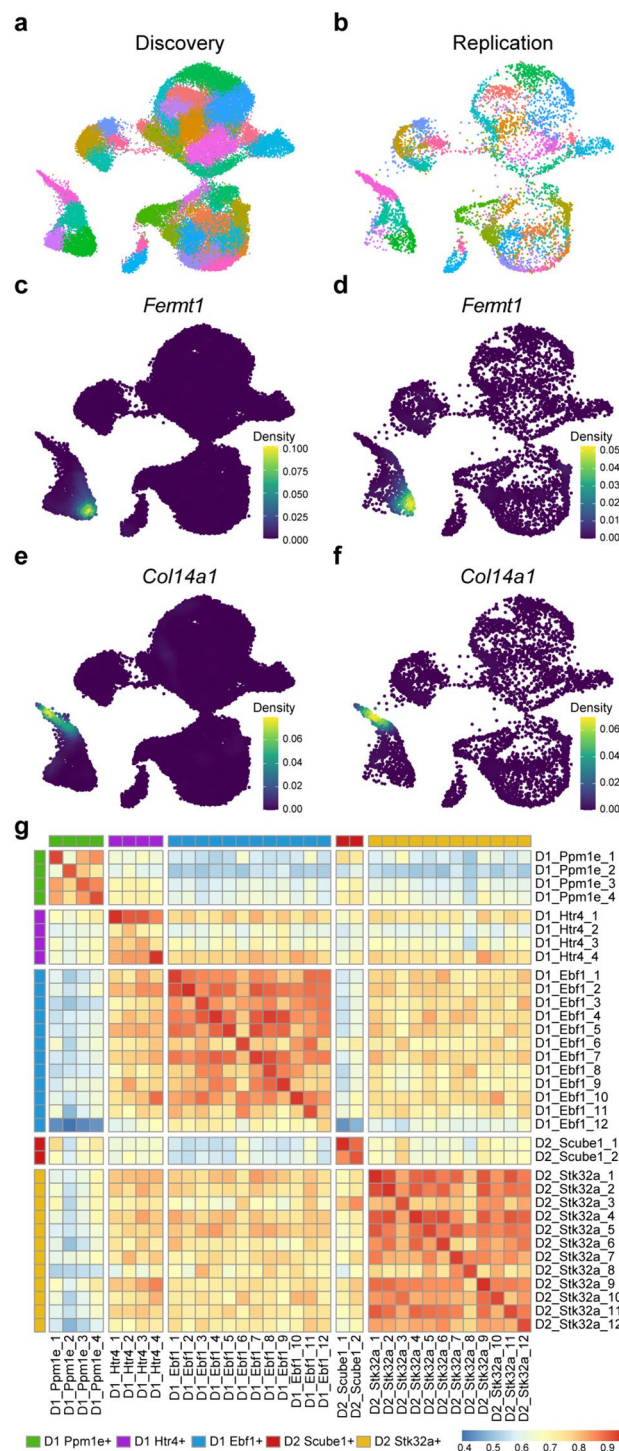


Figure 5. Replication of MSN subclustering in an independent snRNA-seq dataset. **(a)** UMAP dimension reduction plot of the current MSN dataset ($n = 48,040$) color-coded by subcluster. **(b)** Rat NAc MSN nuclei ($n = 7641$) from previously published work^{27,28} mapped to the discovery dataset UMAP. The replication nuclei were assigned to subclusters defined in the discovery dataset based on transcriptomic similarity and colored based on those subcluster assignments. Nebulosa plots show the expression profiles of *Fermt1* and *Col14a1* in the discovery (c,e) and replication datasets (d,f). **(g)** Heatmap representing Pearson correlation coefficients for gene expression between the 34 MSN subclusters in the discovery dataset and the assigned subclusters in the replication dataset.

true cell populations. All 34 MSN subclusters contained nuclei from all samples and no sample contributed more than 25% of the nuclei in any one subcluster (Table S6), suggesting that the subclusters defined here represent distinct MSN populations in the rat NAc rather than bioinformatic artifacts.

Profiling glutamate, GABA, and acetylcholine receptor expression in MSN subtypes

Activity of MSNs in the NAc is regulated, in part, by glutamate, GABA and acetylcholine signaling, but the expression patterns of these receptors in MSN subtypes are still not well defined. To address this knowledge gap, we analyzed the expression of genes encoding glutamate, GABA, and acetylcholine receptors in our MSN dataset (Fig. S2–3). *Grik2* was the most highly expressed of all ionotropic glutamate receptor genes, with the highest expression occurring in D1R-expressing matrix MSNs, whereas *Grik3* was expressed exclusively in other MSN populations (Fig. S2A). In contrast, *Grik1* had lower overall expression and was limited to a select but diverse group of MSN subclusters (Fig. S2A). Expression of the metabotropic glutamate receptor gene *Grm8* also was highly variable across MSN subclusters (Fig. S2B). Genes encoding GABA receptor subunits were generally less variable across the MSN cell populations, with *Gabrg3*, *Gabrb3*, and *Gabrb1* having notably higher expression than other family members (Fig. S2C). *Chrna7* and *Chrng* were the only nicotinic acetylcholine receptors with notable expression in the dataset (Fig. S3A). Several muscarinic cholinergic receptors were also expressed, with *Chrm3* showing highly variable expression and *Chrm4* only present in D1R-expressing MSNs (Fig. S3B).

Replication of novel MSN subtypes in an independent dataset

To support the validity of the subclusters identified in our analysis, we examined MSNs from an independent snRNA-seq dataset generated from drug-naïve rat NAc ($n = 7641$ MSN nuclei)^{27,28}. Mapping this replication dataset onto our discovery dataset ($n = 48,040$ MSN nuclei) yielded cells from all 34 MSN subclusters (Fig. 5A,B). To demonstrate consistency between the discovery and replication datasets, we analyzed expression of specific marker genes that distinguish two closely related MSN subclusters: *Fermt1* (D1_Ppm1e_1) and *Col14a1* (D1_Ppm1e_2). The novel subtypes expressing these marker genes were present in the replication sample (Fig. 5C–F) with little to no background expression of either gene observed, matching the expression pattern in the discovery dataset. More broadly, Pearson correlation analysis indicated strong overlap in gene expression between the matched subclusters in the discovery and replication datasets (Fig. 5G). These data support the reproducibility of our findings and highlight uncommon and previously undescribed MSN cell types and the specificity of their respective primary marker genes.

Discussion

Single-cell transcriptomic datasets are generally small due to the substantial costs of these experiments and often lack adequate numbers of cells from uncommon populations to effectively define those populations during clustering. Overcoming the sample size limitations can reveal novel cell subtypes⁴³. An analysis of ~1.3 million cells from across the mouse cortex and hippocampal formation by the Allen Institute for Brain Science identified 364 neuronal populations, many of which were previously unidentified³³. Likewise, the HypoMap atlas of hypothalamus single-cell data also defined 130 subtypes from data on ~220,000 neurons³². Integration of large-scale, multimodal single-cell data, as in the BRAIN Initiative Cell Census Network's analysis of the motor cortex, can also help to identify large numbers of neuronal populations³⁴. Our study contributes to these large-scale neural phenotyping efforts by examining the transcriptome profile of ~79,000 NAc MSNs across rats, mice, and humans. By focusing on MSNs in the neuroanatomically restricted region of the NAc, we identified cell populations that are conserved between species and share associations with human phenotypes, suggesting that these MSN subtypes have high potential for translational studies.

We used single-cell transcriptomic data from rat NAc tissue to perform high-resolution subclustering and phenotyping of MSNs. The cell atlas described here provides marker combinations that define each of the 34 MSN populations, including novel subtypes, and distinct expression patterns for receptor genes known to be involved in MSN function. Previously, the largest study of the NAc included ~20,000 MSNs from drug-naïve mice²⁶. That study focused primarily on clustering results with 16 MSN populations and demonstrated that these MSN subtypes had distinct spatial patterns within the NAc²⁶. The study also included a secondary, higher-resolution subclustering of MSNs. This analysis identified 57 cell populations, most of them previously undescribed in mice, and showed clearly that significant heterogeneity exists in MSNs of the NAc²⁶. Our clustering of ~48,000 rat MSNs similarly supports the heterogeneous nature of MSNs and revealed multiple novel cell subtypes in the NAc that could be defined by expression of a relatively small number of marker genes. These subclusters were not driven by individual samples and replication analysis in an independent rat NAc dataset^{27,28} supported the validity of these subclusters. Conservation of individual markers between our MSN subclusters and those from the prior mouse work is unclear. For example, *Tac2* expression is limited to specific D1R-expressing cells in mouse NAc²⁶ and these MSN subtypes regulate cocaine reward⁴⁴, but *Tac2* was not identified as a marker gene for any subclusters in our rat data. In contrast, *Gpr149* and *Nfix* are expressed in non-overlapping subsets of D1R-expressing striosomal MSNs in both species²⁶. This mixed concordance could be caused by inter-species differences or technical artifacts resulting from the methodological differences between the studies (e.g. different generations of 10x Chromium chemistry used for library preparation). Despite this, our data demonstrate that NAc MSNs from the striosome and matrix can be distinguished based on transcriptomic profile, consistent with data from mice and macaques^{24,26}, and that this distinction is also evident in human NAc MSN data. The replication analysis also suggests the MSN subclusters are stable across demographic variables, given the sex and strain differences between the discovery cohort (male, Brown Norway) and replication cohort (male and female, Sprague Dawley). In total, our results highlight the significant benefit of larger sample sizes for improving cell-type discovery efforts.

Although we identified 34 MSN subclusters in this study, there is no a priori way to know how many truly distinct cell populations are present in a given tissue. Decisions on clustering parameters are therefore driven by a combination of prior biological knowledge of cellular diversity within the tissue, the number of cells or nuclei in the dataset, and the ability to define distinct sets of cluster markers at the chosen resolution. Historically, our knowledge of uncommon neuronal subtypes in the NAc, like most brain regions, is lacking and thus cannot provide guideposts for higher-resolution analyses. Although we identified 34 MSN subclusters, there are likely additional MSN populations yet to be defined given the greater numbers of neuronal subtypes identified in significantly larger datasets of the hypothalamus, cortex, and hippocampus^{32,33}. Increases in the number of cells or nuclei from control brain tissue, potential integration of existing drug-naïve sample datasets to increase sample sizes, and collection of multi-modal data should be prioritized for all brain regions, as it will enable deep phenotyping of neural cell populations and facilitate downstream functional analyses and translational research.

Gene expression profiles vary between different cell types. For example, expression of D1R- and D2R-expressing MSN marker genes are largely confined to their respective groups in our rat NAc data, agreeing with a prior study of the mouse NAc²⁶. Notably, we did not find any 'hybrid' MSNs, defined by expression of both D1R and D2R, which had previously been observed in the dorsal striatum of non-human primates and humans^{24,31}. Together, these data suggest that hybrid D1R- and D2R-expressing MSNs may not be present in the ventral striatum or that this population of cells is limited to higher order primates. In addition to expression differences between cell types, individual populations of cells may also have wide transcriptomic variability associated with altered states. Activated neurons represent a classic example of this phenomenon, showing different transcriptomic patterns than inactive neurons due to the expression of immediate early genes associated with the active state. The 34 MSN subclusters defined in this study are "transcriptomically distinct", a term that encompasses both differences in cell type and differences in cell state, and it is difficult to parse these two differences without further functional experiments. An argument can be made that two populations differing only in state would be expected to form a gradient of expression differences rather than yielding definitive markers and RNA velocity analyses may help to identify these relations between the individual subclusters. Evidence for state changes captured in snRNA-seq data could guide future work, including studies of induction of MSN plasticity.

In addition to phenotyping MSNs, analysis of our single-cell data and gene-level GWAS data highlighted cell populations enriched for the expression of genes associated with substance use-related phenotypes. The implicated MSN subtypes are potentially relevant for understanding the pathophysiology of substance use and misuse. The most significant finding is an association between the D2R-expressing *Scube1* + MSNs and AUD. In this analysis, analogous clusters in both mouse and human MSN data were also associated with AUD, supporting the potential translational relevance of these cells in AUD. Based on the expression of known markers, these cells represent D2R-expressing MSNs originating from the striosome. Prior work supports the hypothesis that striosome MSNs are involved in reward and reinforcement learning^{45,46} and MSNs in the striosome project to areas of the brain involved in addiction-like behaviors (e.g. ventral tegmental area)⁴⁷. Furthermore, manipulation of circuits targeting the striosome or matrix compartments of the striatum had differential effects on cost-benefit decision making⁴⁸. Mice were more likely to choose a high-reward, high-cost option over a low-reward, low-cost option after inhibition of inputs to the striosome⁴⁸. Despite these compelling connections, the role of the striosome in the effects of chronic alcohol exposure is unknown. Functional phenotyping of D2R-expressing striosome MSNs by other, non-sequencing methods should be prioritized to determine the role that these cells play in the relationship between the NAc and risk of AUD. Similarly, populations of D2R-expressing matrix MSNs were significantly enriched for OUD-associated genes in both humans and rats, and warrant functional follow-up studies. In contrast, enrichment for TUD-associated genes were observed in the majority of human MSN subtypes but not in rodent populations. These data could reflect notable species differences in the expression patterns of the top genes identified in the TUD GWAS and make it difficult to hypothesize whether TUD-related circuits in the NAc are conserved.

The identification of 34 MSN subclusters raises questions about the spatial location, circuit involvement, and overall function of these cell populations. The major D1R-expressing *Htr4* + and *Ebf1* + MSN populations observed in our analysis were previously seen in the replication dataset and shown to have significant spatial biases towards the NAc shell and core, respectively²⁸. Analyses in mice similarly showed subregion expression differences between MSN clusters²⁶. These findings suggest that the subclusters identified in the current study are also likely to have non-uniform distributions in the rat NAc. Future experiments with highly multiplexed fluorescent in situ hybridization (FISH) would help define the spatial profiles for each MSN subtype. Circuit tracing could also be used to define spatial distributions and study functional relevance, the ultimate goal of the type of neuronal phenotyping presented here. It is well established that MSNs in the NAc core and shell have different functions^{13–15}. Similar differences exist between neurons in the matrix and those in the striosome^{22,23}. These findings, however, are typically based on the generic classification of MSNs within those regions as either D1R- or D2R-expressing cells. Our results indicate that manipulation of all D1R- or D2R-expressing MSNs, even if restricted to only one of these NAc subregions or compartments, will inevitably affect multiple distinct MSN subtypes. It is therefore impossible to know which populations are truly relevant to any observed phenotypic effects. Cell atlases like the one presented here make it possible to define subclusters of cells and their respective marker genes, allowing more targeted genetic manipulation and functional dissection of behavior.

Materials and methods

Medium spiny neuron clustering

Single nucleus RNA sequencing (snRNA-seq) data from adult male, drug-naïve Brown Norway rats ($n = 11$) were obtained from our prior study of the NAc³⁰. Sequencing reads were aligned to the rn7.2.110 version of the rat genome (downloaded 09.13.2023) and count matrices were generated using Cell Ranger v7.1.0⁴⁹. Seurat objects

were created for all samples using *Seurat* v4.3⁵⁰. Nuclei with ≤ 200 genes detected or $\geq 5\%$ of reads from the mitochondrial genome were removed. After initial clustering, *SoupX* v1.6.2 was used to correct the count matrix for each sample for the presence of cell-free mRNA⁵¹. Expression data were normalized using SCTransform while regressing out the effect of the number of unique molecular identifiers (UMIs) per nucleus, and integration was used to combine all samples. Doublets were identified with *scDblFinder* v1.12.0⁵². All doublets and all clusters with a majority of nuclei labeled as doublets were removed. Additional putative doublet clusters with mixed cell type markers and clusters with low average UMIs were also removed. Data for all nuclei belonging to MSN clusters were subset and used for low-resolution clustering using the first 20 principal components (PCs) and a resolution of 0.05 to identify major cell populations. To identify MSN subtypes, each major MSN population was subclustered independently with newly calculated PCs. Subclusters with poor markers were removed and the remaining 34 subclusters were merged to create the final dataset of 48,040 MSNs. A constellation plot was created using modified versions of published code⁵³ and the *scrattch.hicat* R package by the Allen Institute for Brain Science (<https://github.com/AllenInstitute/scrattch>). All other plots were generated with the *Seurat* and *ggplot2* packages.

Cross-species comparison

Seurat objects were generated from processed single nucleus/cell transcriptomic datasets for drug-naïve human²⁵ and mouse²⁶ NAc. Both datasets were then subset to MSNs using cluster identifications from their respective publications. Human gene symbols were converted to mouse orthologs using the *nichenetr* R package⁵⁴. Rat, mouse, and human MSN datasets were integrated using *Seurat* v4.3 and analogous MSN populations were identified by UMAP.

Single cell disease-relevance score (scDRS) analysis

We performed gene-based association testing for 4 well-powered SUD GWAS human datasets [i.e. alcohol use disorder (AUD)³⁷, alcohol consumption as measured by AUDIT-C³⁷, opioid use disorder (OUD)³⁶, and tobacco use disorder (TUD)³⁵] in FUMA v1.5.4⁵⁵ using MAGMA v1.06⁵⁶, which employs multiple regression models to detect multiple marker effects that account for SNP *p*-values and linkage disequilibrium (LD) between markers. All GWAS results were based on European ancestry, and the European 1000 Genomes Project phase 3 panel was used as the LD reference. A modified version of the *scDRS* program (<https://github.com/martinjzhang/scDRS>) was created to convert between human and rat orthologs (https://github.com/CristLab/scDRS_Rat) and used to generate disease-relevance scores for each nucleus based on gene expression profiles and Z-scores from the MAGMA analyses⁴⁰. Significant enrichment of expression of disease-associated genes was assessed for each major cluster in the rat, mouse, and human MSN datasets, as well as the rat MSN subclusters, by Monte Carlo analysis using the *scDRS* standard analytical pipeline⁴⁰. *P*-values were corrected for multiple testing using a False Discovery Rate of 0.05.

Replication analysis

Count matrices from an independent set of previously analyzed drug-naïve rat NAc snRNA-seq samples were obtained from NCBI GEO (accession numbers: GSE137763, GSE222418)^{27,28} and analyzed with *Seurat* v4.3. Nuclei with < 500 genes or $> 5\%$ mitochondrial transcripts were removed as low quality, whereas nuclei with high numbers of UMI were removed as putative multiplets. After QC, data from each sample were merged into a single *Seurat* object, and then normalized, scaled, and clustered. MSN clusters were identified by high expression of *Bcl11b* and *Pde10a*, known markers of MSNs⁵⁷ and extracted from the full dataset. The replication MSN data were mapped to our discovery MSN data UMAP using the *Seurat* MapQuery function and predicted cluster identities were assigned to each nucleus in the replication dataset. Pearson correlation coefficients of gene expression between clusters in the discovery and replication datasets were calculated in *clustifyr*⁵⁸ and visualized with *heatmap*. Nebulosa plots were generated using the *Nebulosa* R package⁵⁹.

Data availability

snRNA-seq count matrices for our discovery dataset are available at the NCBI Gene Expression Omnibus (GEO) under accession number GSE263307. The cleaned *Seurat* objects for the full dataset and the final subclustered MSNs are also available upon reasonable request. Please contact the corresponding author. The raw count matrices used in the replication analysis are available at GEO accession numbers GSE137763 and GSE222418. Cleaned data from mouse and human NAc can be obtained at GEO accession number GSE118020 and https://github.com/LieberInstitute/10xPilot_snRNAseq-human, respectively. Gene-level summary statistics from SUD GWAS are available from their respective groups^{35–37}.

Code availability

All R code used for these analyses is available at https://github.com/CristLab/msn_subclustering

Received: 16 May 2024; Accepted: 2 August 2024

Published online: 06 August 2024

References

- Schmidt, H. D. & Pierce, R. C. Cocaine-induced neuroadaptations in glutamate transmission: Potential therapeutic targets for craving and addiction. *Ann. N. Y. Acad. Sci.* **1187**, 35–75. <https://doi.org/10.1111/j.1749-6632.2009.05144.x> (2010).

2. Schmidt, H. D., Anderson, S. M. & Pierce, R. C. Stimulation of D1-like or D2 dopamine receptors in the shell, but not the core, of the nucleus accumbens reinstates cocaine-seeking behaviour in the rat. *Eur. J. Neurosci.* **23**, 219–228. <https://doi.org/10.1111/j.1460-9568.2005.04524.x> (2006).
3. Gibson, G. D., Millan, E. Z. & McNally, G. P. The nucleus accumbens shell in reinstatement and extinction of drug seeking. *Eur. J. Neurosci.* **50**, 2014–2022. <https://doi.org/10.1111/ejn.14084> (2019).
4. Kravitz, A. V., Tye, L. D. & Kreitzer, A. C. Distinct roles for direct and indirect pathway striatal neurons in reinforcement. *Nat. Neurosci.* **15**, 816–818. <https://doi.org/10.1038/nn.3100> (2012).
5. Enoksson, T., Bertran-Gonzalez, J. & Christie, M. J. Nucleus accumbens D2- and D1-receptor expressing medium spiny neurons are selectively activated by morphine withdrawal and acute morphine, respectively. *Neuropharmacology* **62**, 2463–2471. <https://doi.org/10.1016/j.neuropharm.2012.02.020> (2012).
6. O'Neal, T. J., Bernstein, M. X., MacDougall, D. J. & Ferguson, S. M. A conditioned place preference for heroin is signaled by increased dopamine and direct pathway activity and decreased indirect pathway activity in the nucleus accumbens. *J. Neurosci.* **42**, 2011–2024. <https://doi.org/10.1523/JNEUROSCI.1451-21.2021> (2022).
7. Lobo, M. K. *et al.* Cell type-specific loss of BDNF signaling mimics optogenetic control of cocaine reward. *Science* **330**, 385–390. <https://doi.org/10.1126/science.1188472> (2010).
8. O'Neal, T. J., Nooney, M. N., Thien, K. & Ferguson, S. M. Chemogenetic modulation of accumbens direct or indirect pathways bidirectionally alters reinstatement of heroin-seeking in high- but not low-risk rats. *Neuropsychopharmacology* **45**, 1251–1262. <https://doi.org/10.1038/s41386-019-0571-9> (2020).
9. Beutler, L. R. *et al.* Balanced NMDA receptor activity in dopamine D1 receptor (D1R)- and D2R-expressing medium spiny neurons is required for amphetamine sensitization. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 4206–4211. <https://doi.org/10.1073/pnas.1101424108> (2011).
10. James, A. S. *et al.* Opioid self-administration results in cell-type specific adaptations of striatal medium spiny neurons. *Behav. Brain Res.* **256**, 279–283. <https://doi.org/10.1016/j.bbr.2013.08.009> (2013).
11. Calipari, E. S. *et al.* In vivo imaging identifies temporal signature of D1 and D2 medium spiny neurons in cocaine reward. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 2726–2731. <https://doi.org/10.1073/pnas.1521238113> (2016).
12. Flanigan, M. & LeClair, K. Shared motivational functions of ventral striatum D1 and D2 Medium Spiny Neurons. *J. Neurosci.* **37**, 6177–6179. <https://doi.org/10.1523/JNEUROSCI.0882-17.2017> (2017).
13. Floresco, S. B. The nucleus accumbens: An interface between cognition, emotion, and action. *Annu. Rev. Psychol.* **66**, 25–52. <https://doi.org/10.1146/annurev-psych-010213-115159> (2015).
14. Di Chiara, G. Nucleus accumbens shell and core dopamine: Differential role in behavior and addiction. *Behav. Brain Res.* **137**, 75–114. [https://doi.org/10.1016/s0166-4328\(02\)00286-3](https://doi.org/10.1016/s0166-4328(02)00286-3) (2002).
15. Zahm, D. S. Functional-anatomical implications of the nucleus accumbens core and shell subterritories. *Ann. N. Y. Acad. Sci.* **877**, 113–128. <https://doi.org/10.1111/j.1749-6632.1999.tb09264.x> (1999).
16. Li, Z. *et al.* Cell-type-specific afferent innervation of the nucleus accumbens core and shell. *Front. Neuroanat.* **12**, 84. <https://doi.org/10.3389/fnana.2018.00084> (2018).
17. Smith, R. J., Lobo, M. K., Spencer, S. & Kalivas, P. W. Cocaine-induced adaptations in D1 and D2 accumbens projection neurons (a dichotomy not necessarily synonymous with direct and indirect pathways). *Curr. Opin. Neurobiol.* **23**, 546–552. <https://doi.org/10.1016/j.conb.2013.01.026> (2013).
18. Soares-Cunha, C. *et al.* Nucleus accumbens medium spiny neurons subtypes signal both reward and aversion. *Mol. Psychiatry* **25**, 3241–3255. <https://doi.org/10.1038/s41380-019-0484-3> (2020).
19. Rossi, L. M. *et al.* Role of nucleus accumbens core but not shell in incubation of methamphetamine craving after voluntary abstinence. *Neuropsychopharmacology* **45**, 256–265. <https://doi.org/10.1038/s41386-019-0479-4> (2020).
20. Durst, M., Könczöl, K., Balázs, T., Eyre, M. D. & Tóth, Z. E. Reward-representing D1-type neurons in the medial shell of the accumbens nucleus regulate palatable food intake. *Int. J. Obes. (Lond.)* **43**, 917–927. <https://doi.org/10.1038/s41366-018-0133-y> (2019).
21. Brimblecombe, K. R. & Cragg, S. J. The striosome and matrix compartments of the striatum: A path through the labyrinth from neurochemistry toward function. *ACS Chem. Neurosci.* **8**, 235–242. <https://doi.org/10.1021/acschemneuro.6b00333> (2017).
22. Cui, Y. *et al.* Targeted expression of μ -opioid receptors in a subset of striatal direct-pathway neurons restores opiate reward. *Nat. Neurosci.* **17**, 254–261. <https://doi.org/10.1038/nn.3622> (2014).
23. Crittenden, J. R. & Graybiel, A. M. Basal Ganglia disorders associated with imbalances in the striatal striosome and matrix compartments. *Front. Neuroanat.* **5**, 59. <https://doi.org/10.3389/fnana.2011.00059> (2011).
24. He, J. *et al.* Transcriptional and anatomical diversity of medium spiny neurons in the primate striatum. *Curr. Biol.* **31**, 5473–5486. <https://doi.org/10.1016/j.cub.2021.10.015> (2021).
25. Tran, M. N. *et al.* Single-nucleus transcriptome analysis reveals cell-type-specific molecular signatures across reward circuitry in the human brain. *Neuron* **109**, 3088–3103.e3085. <https://doi.org/10.1016/j.neuron.2021.09.001> (2021).
26. Chen, R. *et al.* Decoding molecular and cellular heterogeneity of mouse nucleus accumbens. *Nat. Neurosci.* **24**, 1757–1771. <https://doi.org/10.1038/s41593-021-00938-x> (2021).
27. Savell, K. E. *et al.* A dopamine-induced gene expression signature regulates neuronal function and cocaine response. *Sci. Adv.* **6**, eaba4221. <https://doi.org/10.1126/sciadv.aba4221> (2020).
28. Phillips, R. A. *et al.* Distinct subpopulations of D1 medium spiny neurons exhibit unique transcriptional responsiveness to cocaine. *Mol. Cell. Neurosci.* **125**, 103849. <https://doi.org/10.1016/j.mcn.2023.103849> (2023).
29. Avey, D. *et al.* Single-cell RNA-seq uncovers a robust transcriptional response to morphine by glia. *Cell Rep.* **24**, 3619–3629.e3614. <https://doi.org/10.1016/j.celrep.2018.08.080> (2018).
30. Reiner, B. C. *et al.* Single nucleus transcriptomic analysis of rat nucleus accumbens reveals cell type-specific patterns of gene expression associated with volitional morphine intake. *Transl. Psychiatry* **12**, 374. <https://doi.org/10.1038/s41398-022-02135-1> (2022).
31. Phan, B. N. *et al.* Single nuclei transcriptomics in human and non-human primate striatum in opioid use disorder. *Nat. Commun.* **15**, 878. <https://doi.org/10.1038/s41467-024-45165-7> (2024).
32. Steuernagel, L. *et al.* HypoMap—a unified single-cell gene expression atlas of the murine hypothalamus. *Nat. Metab.* **4**, 1402–1419. <https://doi.org/10.1038/s42255-022-00657-y> (2022).
33. Yao, Z. *et al.* A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222–3241.e3226. <https://doi.org/10.1016/j.cell.2021.04.021> (2021).
34. BICCN. A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86–102. <https://doi.org/10.1038/s41586-021-03950-0> (2021).
35. Toikumo, S. *et al.* Multi-ancestry meta-analysis of tobacco use disorder prioritizes novel candidate risk genes and reveals associations with numerous health outcomes. *medRxiv* <https://doi.org/10.1101/2023.03.27.23287713> (2023).
36. Kember, R. L. *et al.* Cross-ancestry meta-analysis of opioid use disorder uncovers novel loci with predominant effects in brain regions associated with addiction. *Nat. Neurosci.* **25**, 1279–1287. <https://doi.org/10.1038/s41593-022-01160-z> (2022).
37. Kember, R. L. *et al.* Genetic underpinnings of the transition from alcohol consumption to alcohol use disorder: Shared and unique genetic architectures in a cross-ancestry sample. *Am. J. Psychiatry* **180**, 584–593. <https://doi.org/10.1176/appi.ajp.21090892> (2023).

38. Martín, A. *et al.* A spatiomolecular map of the striatum. *Cell Rep.* **29**, 4320–4333.e4325. <https://doi.org/10.1016/j.celrep.2019.11.096> (2019).
39. Saunders, A. *et al.* Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015–1030.e1016. <https://doi.org/10.1016/j.cell.2018.07.028> (2018).
40. Zhang, M. J. *et al.* Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *Nat. Genet.* **54**, 1572–1580. <https://doi.org/10.1038/s41588-022-01167-z> (2022).
41. Kranzler, H. R. *et al.* Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nat. Commun.* **10**, 1499. <https://doi.org/10.1038/s41467-019-09480-8> (2019).
42. Andraka, E., Phillips, R. A., Brida, K. L. & Day, J. J. Chst9 marks a spatially and transcriptionally unique population of oprm1-expressing neurons in the nucleus accumbens. *bioRxiv* <https://doi.org/10.1101/2023.10.16.562623> (2023).
43. Siletti, K. *et al.* Transcriptomic diversity of cell types across the adult human brain. *Science* **382**, eadd7046. <https://doi.org/10.1126/science.add7046> (2023).
44. Zhao, Z. D. *et al.* A molecularly defined D1 medium spiny neuron subtype negatively regulates cocaine addiction. *Sci. Adv.* **8**, eabn3552. <https://doi.org/10.1126/sciadv.abn3552> (2022).
45. Hong, S. *et al.* Predominant striatal input to the lateral Habenula in Macaques comes from striosomes. *Curr. Biol.* **29**, 51–61.e55. <https://doi.org/10.1016/j.cub.2018.11.008> (2019).
46. Evans, R. C. *et al.* Functional dissection of basal ganglia inhibitory inputs onto Substantia nigra dopaminergic neurons. *Cell Rep.* **32**, 108156. <https://doi.org/10.1016/j.celrep.2020.108156> (2020).
47. Watabe-Uchida, M., Zhu, L., Ogawa, S. K., Vamanrao, A. & Uchida, N. Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* **74**, 858–873. <https://doi.org/10.1016/j.neuron.2012.03.017> (2012).
48. Friedman, A. *et al.* A corticostriatal path targeting striosomes controls decision-making under conflict. *Cell* **161**, 1320–1333. <https://doi.org/10.1016/j.cell.2015.04.049> (2015).
49. Zheng, G. X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049. <https://doi.org/10.1038/ncomms14049> (2017).
50. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e3529. <https://doi.org/10.1016/j.cell.2021.04.048> (2021).
51. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* <https://doi.org/10.1093/gigascience/giaa151> (2020).
52. Germain, P. L., Lun, A., Garcia Meixide, C., Macnair, W. & Robinson, M. D. Doublet identification in single-cell sequencing data using. *F1000Research* **10**, 979. <https://doi.org/10.12688/f1000research.73600.2> (2021).
53. Bhaduri, A. *et al.* An atlas of cortical arealization identifies dynamic molecular signatures. *Nature* **598**, 200–204. <https://doi.org/10.1038/s41586-021-03910-8> (2021).
54. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: Modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162. <https://doi.org/10.1038/s41592-019-0667-5> (2020).
55. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826. <https://doi.org/10.1038/s41467-017-01261-5> (2017).
56. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219. <https://doi.org/10.1371/journal.pcbi.1004219> (2015).
57. Arlotta, P., Molyneaux, B. J., Jabaudon, D., Yoshida, Y. & Macklis, J. D. Ctip2 controls the differentiation of medium spiny neurons and the establishment of the cellular architecture of the striatum. *J. Neurosci.* **28**, 622–632. <https://doi.org/10.1523/JNEUROSCI.2986-07.2008> (2008).
58. Fu, R. *et al.* clustifyr: An R package for automated single-cell RNA sequencing cluster classification. *F1000Research* **9**, 223. <https://doi.org/10.12688/f1000research.22969.2> (2020).
59. Alquicira-Hernandez, J. & Powell, J. E. Nebulosa recovers single-cell gene expression signals by kernel density estimation. *Bioinformatics* **37**, 2485–2487. <https://doi.org/10.1093/bioinformatics/btab003> (2021).

Acknowledgements

The authors would like to thank Dr. Yafang Zhang, Kael Ragnini, Jennifer Ben Nathan, and Mariela Jennings for their technical contributions to this project. This work was supported in part by a State of Pennsylvania Department of Health Nonformula Tobacco Settlement Act Grant, Pharmacogenetics of Opioid Use Disorder (W.H.B.); National Institutes of Health grants R01 DA037897 (H.D.S.), R21 DA045792 (H.D.S.), R21 DA 057458 (H.D.S. & R.C.C.), R21 DA 055846 (B.C.R.), NIH/NIDA DP1DA054394 (S.S.-R.), K01 AA028292 (R.L.K.), and R01 AA030056 (H.R.K. & R.L.K.); Tobacco-Related Disease Research Program (TRDRP) Grant Number T32IR5226 (S.S.-R.); and Department of Veterans Affairs grant I01 BX004820 (H.R.K.). The funding sources had no role in the conceptualization of this work, analysis of the data, or preparation of the manuscript.

Author contributions

B.C.R., H.D.S., and R.C.C. conceptualized the project. B.C.R., W.H.B., H.R.K., S.S.-R., R.L.K., H.D.S., and R.C.C. obtained funding for the project. B.C.R., H.R.K., R.L.K., H.D.S., and R.C.C. designed the analyses. B.C.R., S.N.M., R.M., H.D.S., and R.C.C. interpreted the analyses and outcomes. B.C.R. and R.C.C. performed the bioinformatic analyses. S.T., H.R.K., S.S.-R., and R.L.K. conceptualized the human genome-wide association studies, provided genetic data and performed genetic analyses. R.C.C., B.C.R., and H.D.S. wrote the original manuscript. All authors participated in reviewing and editing the manuscript for publication.

Competing interests

B.C.R. receives research funding from Novo Nordisk and Boehringer Ingelheim that was not used in support of these studies. H.R.K. is a member of advisory boards for Dicerna Pharmaceuticals, Sophrosyne Pharmaceuticals, Enthion Pharmaceuticals, and Clearmind Medicine; a consultant to Sobrera Pharmaceuticals; the recipient of research funding and medication supplies for an investigator-initiated study from Alkermes; a member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative, which was supported in the last 3 years by Alkermes, Dicerna, Ethypharm, Lundbeck, Mitsubishi, Otsuka, and Pear Therapeutics; and a holder of U.S. patent 10,900,082 titled: "Genotype-guided dosing of opioid agonists", issued 26 January 2021. All other authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-69255-0>.

Correspondence and requests for materials should be addressed to R.C.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024