Cell type-specific expression of the molecular players in mouse prefrontal cortex during cocaine addiction

Eric Teng

tengeric1595@gmail.com

11th Grade / Kent School

Abstract. Cocaine addiction is an issue that affects more than 5 million people in America per year. Although there has been much research into the genes and chemicals responsible for cocaine addiction, there are many specific questions left unanswered. Our experiment attempts to further previous research into certain molecular players. We follow up with their use of single-cell RNA sequencing on the prefrontal cortex cells of mice undergoing cocaine intravenous self-administration. Data of 12 samples from both saline and cocaine treated mice which are found on the Gene Expression Omnibus public database were retrieved. Using the Seurat function of RStudio, the data was merged into objects, normalized, clustered, and labeled into one of eight cell types. What resulted was a detailed UMAP plot displaying the clusters, their gene expression level, expression frequency, and their cell type. With this plot, we were able to determine the specific cell types that express the genes encoding the pre-established molecular players (ΔFosB, MeCP2, and BDNF). When the analysis was expanded to a cell-type specific level, it was discovered some of these genes were selectively expressed in excitatory neurons and non-neuronal cells. Going further into the analysis, we determined the 6 genes with the most varied gene expression over the 3 stages of cocaine addiction for each of the 8 cell types. Overall, our computational analysis of publicly available transcriptome datasets from mouse addiction model provides new insights into the molecular basis of cocaine addiction.

Keywords: Cocaine Addiction, single-cell RNA sequencing, gene expression.

1. Introduction

The recurring surge of cocaine-related drug overdoses presents pressing public health complications. Nearly 5 million Americans reported use in 2017, 19 thousand were hospitalized from abuse in 2016, and 15 thousand died from an overdose in 2016 [1]. Human studies and preclinical studies have suggested prefrontal cortex (PFC) plays a crucial role in driving cocaine addiction-related behavior[2,3]. PFC is a significant brain region regulating behavior [4]; it drives cognitive functions such as learning and memory, decision-making, emotion, and social behavior through the dynamic integration of sensory-stimulus-driven inputs from multiple brain regions [5]. The PFC weaves together the complex relationships between cause and effect, allowing the brain to create goals and anticipate the actions that would lead to their fulfillment. Exposure to cocaine results in widespread transcriptional and epigenetic alterations in the brain. It is observed that the introduction of cocaine can modify histones and methylate DNA, leading to the regulation of gene expression [6]. The study also mentions how humans undergo a similar transcriptional change when maturing, whether it be postnatal or during the teenage years, but the introduction of cocaine can counteract the PFC's development. However, cocaine can still have an impact in later years, since the PFC has a high level of plasticity compared to other brain regions.

Classifying all cells in the PFC into 8 cell types, a recent study uses single-cell RNA sequencing (scRNA-seq) to document cell type-specific transcriptional adaptations across the PFC during different stages of cocaine addiction [7]. This study laid the foundation for the transcriptional analysis of molecular players and pathways in each different cell type. The 3 main molecular players this research covers are ΔFosB, MeCP2, and BDNF, all of which are observed to have vital roles in cocaine addiction. ΔFosB is a transcriptional factor that is built up in the nuclear accumbens after substance abuse. It regulates gene expression dynamics, heightening the rewarding effects of cocaine and increasing the subject's desire of the drug [8,9]. MeCP2 modifies chromatin and functions along

with histones. It modulates DNA methylation and interacts with microRNA to control an individual's motivation to consume cocaine [10]. The final player, BDNF, is important for synaptic plasticity. Long-term exposure to cocaine increases BDNF levels in the ventral tegmental area (VTA) and nuclear accumbens [11]. Like ΔFosB, BDNF will increase cocaine's rewarding effect and increases the chance of relapse [12].

To better understand the cellular mechanisms involved in cocaine addiction, we ask whether addiction-related genes referred to in previous studies such as ΔFosB, MeCP2, and BDNF exhibit a cell type-specific expression in PFC. Via further examining the computational analysis of publicly available transcriptome datasets to a cell-type specific level, we attempted to further study the cell-type specific expression pattern of these molecular players using bioinformatic tools.

The cocaine intravenous self-administration (IVSA) model is a gold standard in addiction studies[13]. The maintenance and withdrawal of IVSA mimic the binge use of drugs and drug abstinence, which are key stages in the cycle of addiction. Here we aimed to identify the top differentially expressed genes in each characterized cell type in PFC in cocaine IVSA maintenance, short-term withdrawal (48 hours), and long-term withdrawal (15 days) using differential expression analysis methods.

Collectively, our study extends the previous reports by reevaluating the transcriptional dynamics of addiction-relevant genes in cell-type specific levels in PFC. Instead of a general analysis of the effects cocaine has on gene expression, we specifically analyzed the 6 highest variable genes in each of the 8 cell types across 3 different stages of cocaine addiction. Suppose a specific gene was among the most variable in multiple cells. In that case, it could mean that it is a molecular player in cocaine addiction, as the more a gene's expression is altered over time of withdrawal, the more likely it is to be a factor in addiction. Our newly identified differentially expressed genes in each cell type will provide potential new insights into the molecular basis of cocaine addiction.

2. Materials and Methods

2.1. List of materials

Raw single-cell RNA-sequencing data generated by 10X Genomics platform (10X Genomics, Inc., Pleasanton, CA) was downloaded from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE124952 and a copy of dataset was stored in my personal computer.

R Studio codes used in the original article provided by Bhattacherjee, A. et al [12].

Personal computer, RStudio, R Packages (Seurat, Tidyverse, Rcolorbrewer)

Internet access for downloading datasets and code.

2.2. Experimental Procedures

ScRNA datasets retrieving. Single cell expression matrix generated by Cell Ranger software from mouse samples treated with cocaine or saline will be obtained from the NCBI-GEO portal with the GEO accession GSE124952. https://www.ncbi.nlm.nih.gov/geo/. GEO, or Gene Expression Omnibus, is a public data repository containing genomic datasets. GSE124952 specifically is the repository containing the genome of the PFC samples from the 6 saline-treated mice and 6 cocaine-treated mice during maintenance, 48-hour withdrawal, and 15 days of withdrawal. 12 single-cell gene expression matrice datasets were downloaded. Each PFC sample contained barcodes, genes, and matrix files which were then sorted into specific files.

Replicating data preprocessing and cell clustering. Extracted from the public GitHub repository, we followed the codes provided by Bhattacherjee, A. et al [12]. and methods up and through clustering to replicate data pre-processing and quality control. Since some of the functions from previous code were unavailable in the current version of Seurat, we revised it so it could run with the newer versions of R and Seurat. The R package "Seurat" was used in processing single-cell RNA sequences. Data from a total of 29,864 single cells from 12 independent biological samples with various gene numbers and counts were loaded (gene numbers and counts are visualized in Supplementary Figure 1a). After

loading Samples 1-12 into their respective categories (cocaine or saline), we merged each category into its own Seurat object. Normalizing the Seurat objects filtered the cells by removing cells with particularly high mitochondrial mRNA (>10%, it implies that the cell is dying) or possible double droplet cells. The batch effect will be removed with a canonical correlation analysis (CCA) by selecting variable genes of at least two standard deviations from the expected dispersion. The top 2000 of the genes within the cocaine and saline samples will be used in Principal Component Analysis, t-SNE dimensionality reduction, and UMAP dimensionality reduction. With CCA, the Seurat objects were merged back into one. Source of heterogeneity in cells and genes was visualized by Dimension reduction heatmap (Supplementary Figure 1b). Preliminary clusters were created using the Seurat function "FindClusters".

Identifying each cluster by cell marker genes. From here, all code is original. 59 subclusters were identified using resolution 1.6. UMAP reduction had a much more distinct and comprehensive graph than t-SNE, so we focused on the UMAP dimensional reduction graph. Preliminary clusters in this graph that expressed the same maker genes were combined. Using the package "SummarizedExperiment", we were able to graph a dot plot for each different cell-type specific gene marker, displaying both their expression rate and expression level in each of the 59 clusters. Firstly, clusters will be annotated using the expression of the known marker genes. By looking at unique expressions of specific genes in the 59 clusters, we were able to categorize the 59 clusters into the 8 different cell types: astrocytes, excitatory neurons, inhibitory neurons, endothelial cells, oligodendrocyte cells, microglial cells, oligodendrocyte precursor cells (OPC), and NF Oligo cells. Some clusters were designated as double droplets and were therefore removed.

Gene candidates' expression allocation. Then, the expression levels of candidates such as ΔFosB, MeCP2, and BDNF, will be measured the 8 identified cell clusters. To visualize the expression pattern of our gene candidate in broad cell clusters, we will plot heatmaps using "FeaturePlot" function.

To search for highly differentially expressed genes in each cell type, we identified the marker genes in three stages of cocaine IVSA from 6 samples from cocaine-treated groups. Dot plots were made to show the top 6 differentially expressed gene levels across three groups of samples from cocaine intravenous self-administration maintenance, withdrawal of 48 hours, and withdrawal of 15 days across cocaine IVSA maintenance.

3. Results

3.1. Cell types classification

According to the source paper, 29,864 single cells obtained from 12 biological samples from both saline controls and cocaine IVSA models were sequenced. To prepare the samples for sequencing, PFC samples from acute coronal brain sections of P60 mice were extracted. The tissue was then dissociated into a single-cell suspension, and the cells were captured using the 10X Chromium platform from 10X Genomics and used to construct cDNA libraries for sequencing. First, we filtered the cells by removing cells with particularly high mitochondrial mRNA (>10%). Then, low quality neuronal and non-neuronal cells (<800 genes for non-neuronal and <1500 genes for neuronal cells) were filtered out. Upon dimension reduction with tSNE and UMAP shown in Figure 1, it is very clear that the UMAP graph is much more precise in displaying clusters. tSNE is too spread out and some clusters overlaps which implies while UMAP can better separate each cluster. Therefore, UMAP was selected to display our data thereafter.

Cell type-specific signature markers such as *Snap25* for global neuronal cells, *Slc17a7* for excitatory neurons, *Gad2* for inhibitory neurons, *Gja1* for astrocytes, *Aspa* for oligodendrocyte, *Bmp4* for newly formed oligodendrocytes, *Pdgfra* for oligodendrocyte precursors, *C1qa* for microglia, and *Flt1* for endothelial cells were examined in each of the 59 clusters. We generated a dot plot displays the 59 clusters' gene expression levels and frequencies for the multiple types of gene markers (Figure 2). On the earlier UMAP and tSNE plots, each cell type is seen to be genetically similar to one other since they are all relatively close to one another on the graphs. The double droplets were removed

from the plot. Using the knowledge of the unique gene expressions for each of the 8 cell types and the dot plots, we categorized the 59 clusters into their respective cell types. After assigning each cluster to a cell type and making a new dot plot displaying various cell markers which are already characterized, we can clearly see the specific gene patterns that are unique to each cell type (Figure 3), which confirms the validity of our cell type classification. Eight broad clusters of PFC cell types based on signature markers was visualized in UMAP plot (Figure 4).

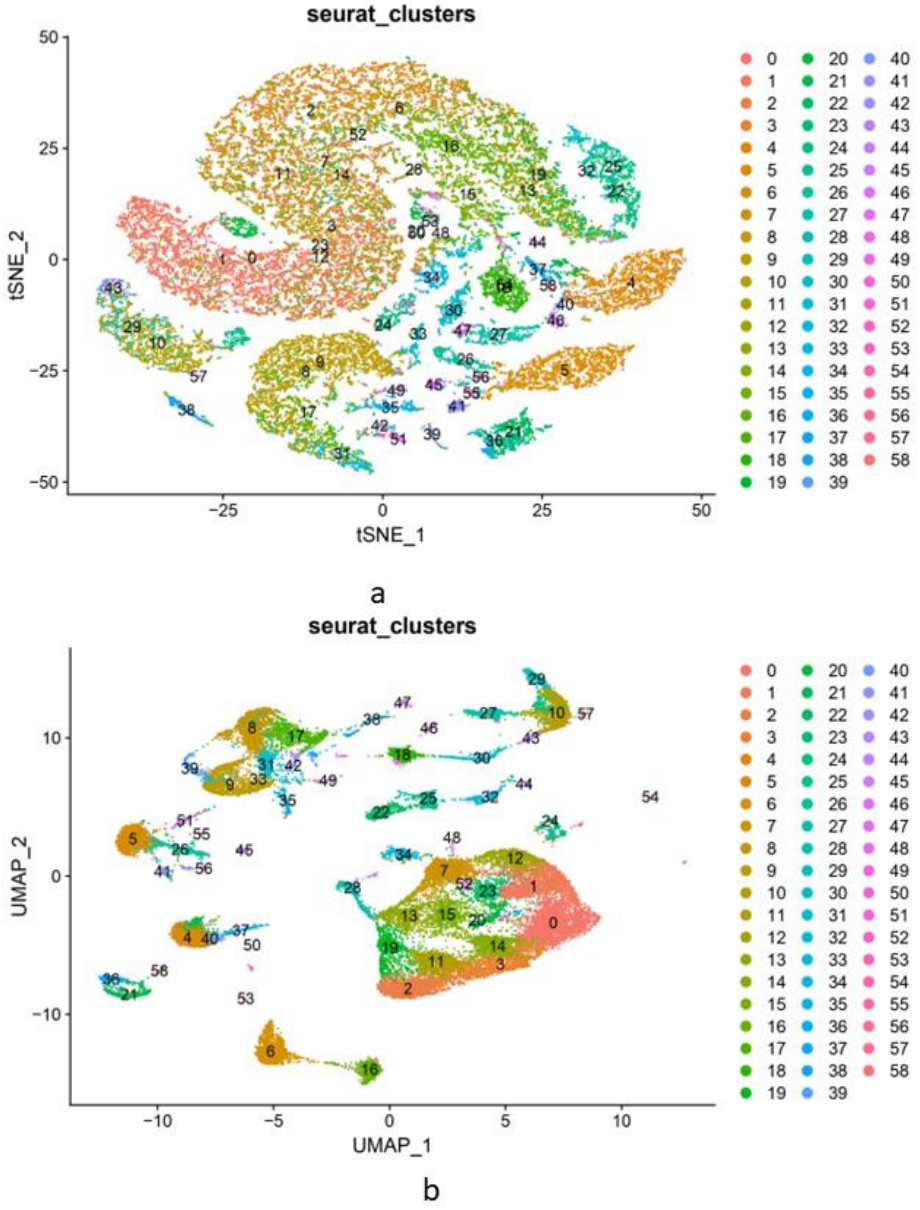


Figure 1. a) t-SNE plot visualizing cluster assignments of cells. b) plot visualizing cluster assignments of cells.

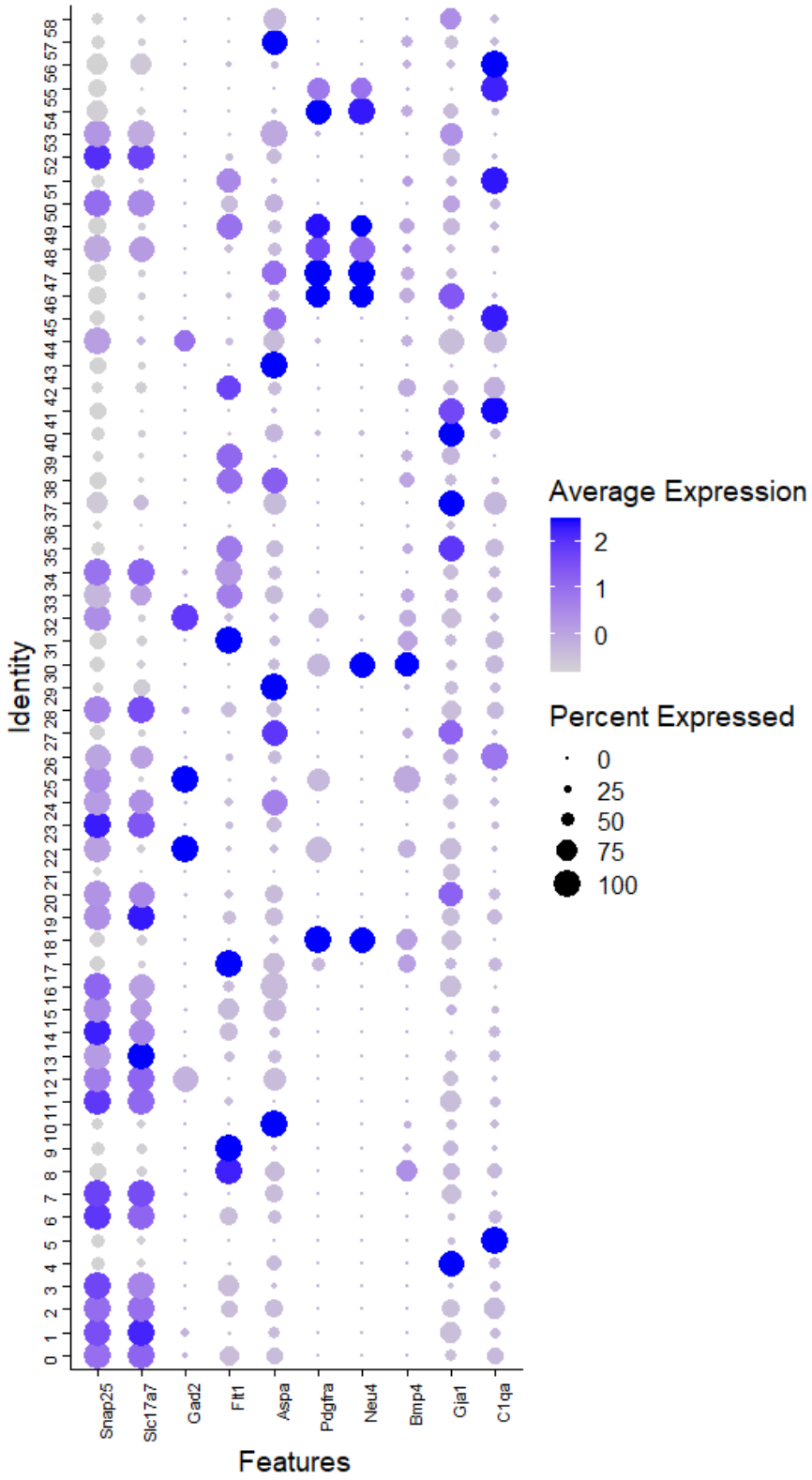


Figure 2. Frequency and Expression Levels of signature cell type-specific markers in these 59 preliminary Clusters.

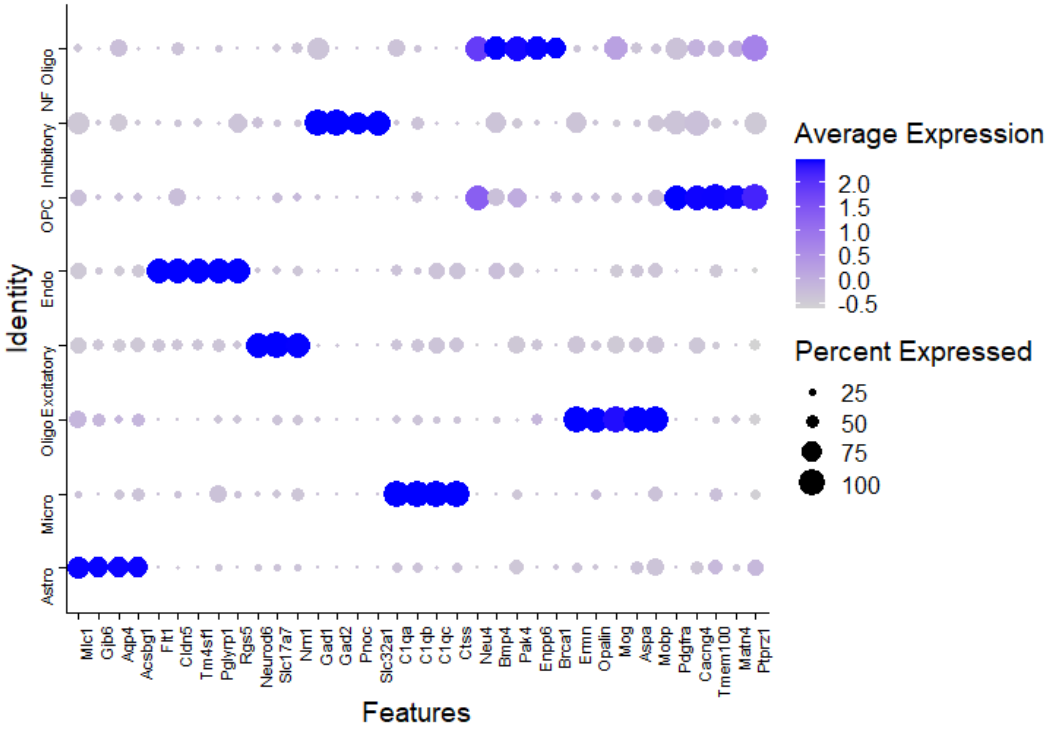


Figure 3. Dot plot showing broader cell-type specific gene markers in 59 preliminary clusters.

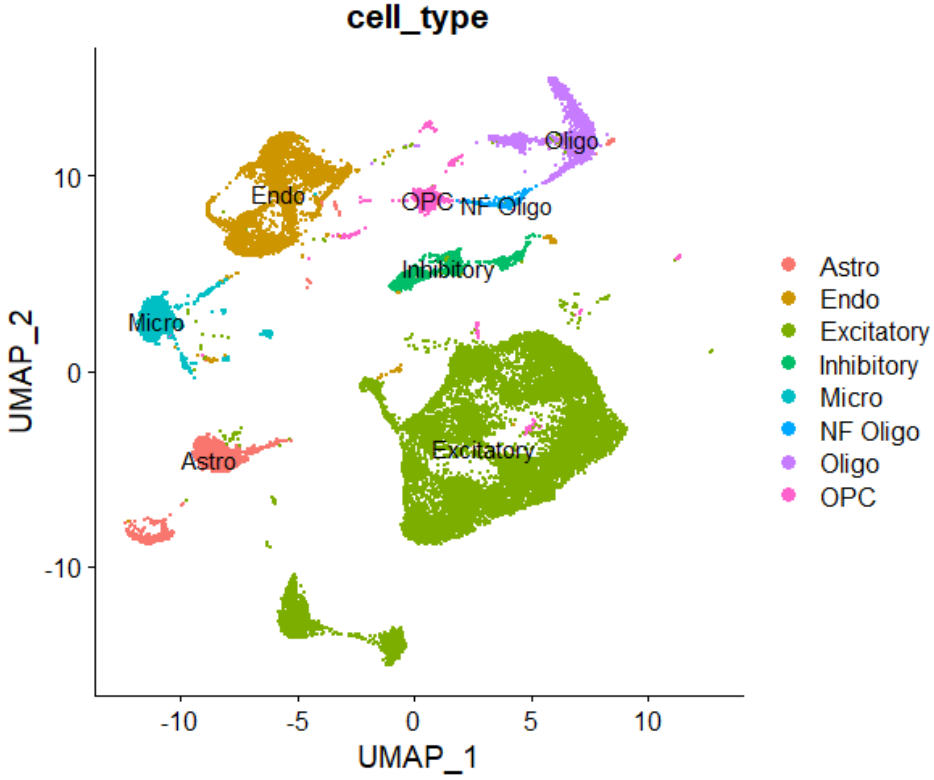


Figure 4. UMAP plot showing the broad clustering of PFC cell types.

3.2. Cell type-specific expression of ΔFosB, MeCP2, and BDNF

We investigated the cell type selective expression of previously reported addiction related gene *Fosb*, *Mecp2*, and *Bdnf* using present single-cell RNA-seq data of the PFC. We examined these gene expression on the dataset in cocaine treated samples. Based on UMAP dimension reduction, we could identify the expression of these genes in the different cell type clusters (Figure 5). *Mecp2* and *Bdnf*

were almost exclusively expressed in excitatory neurons (Figure 5b-c). However, *Fosb* was mainly found in non-neuronal cells, such as Endo and Micro (Fig. 5a).

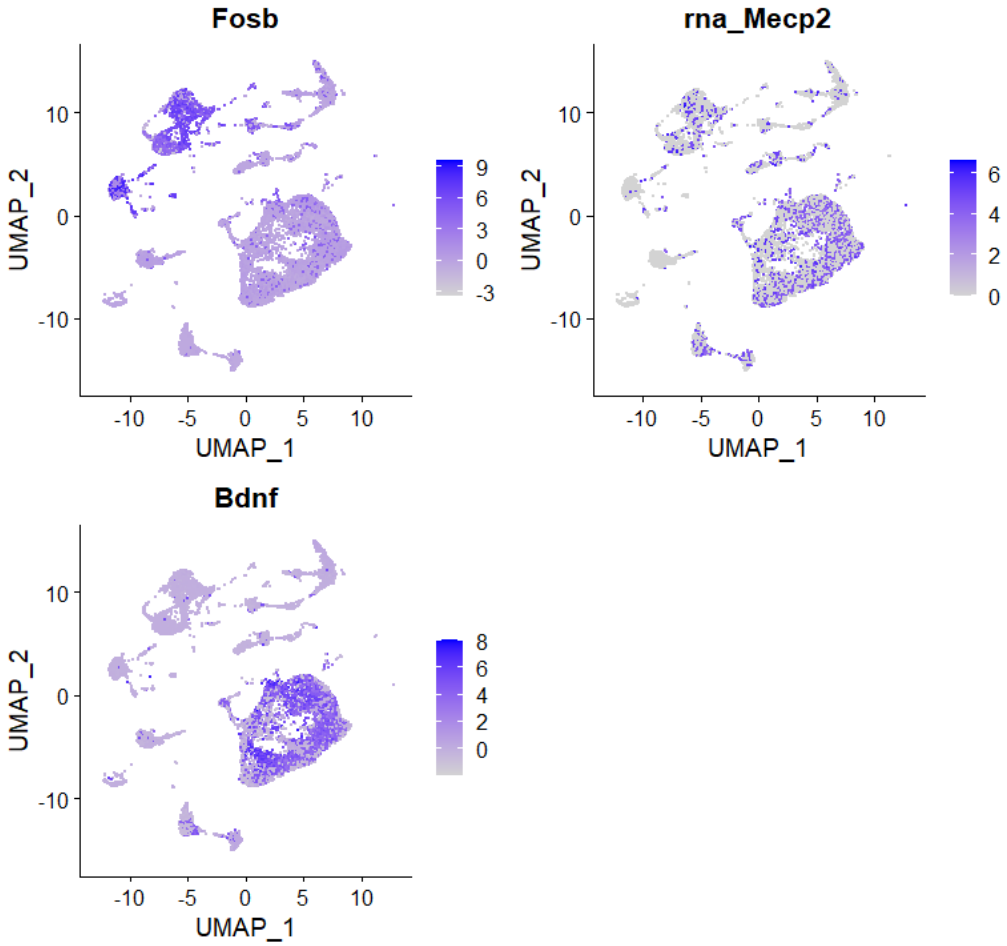
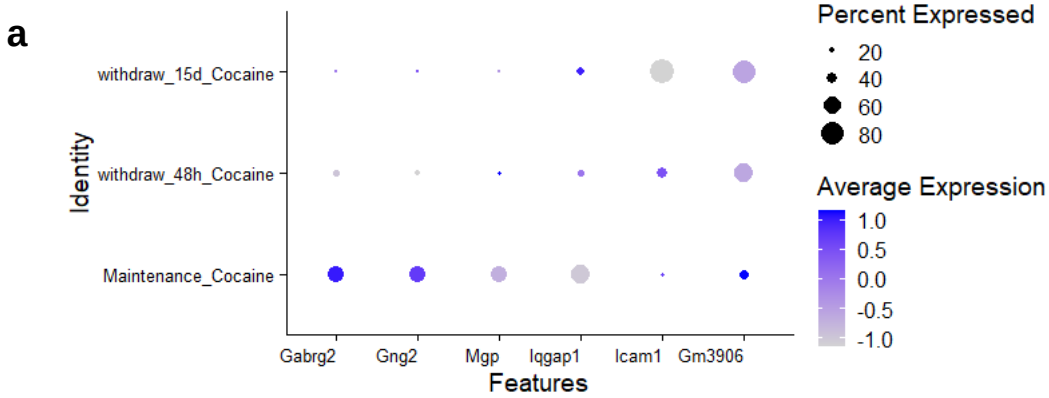
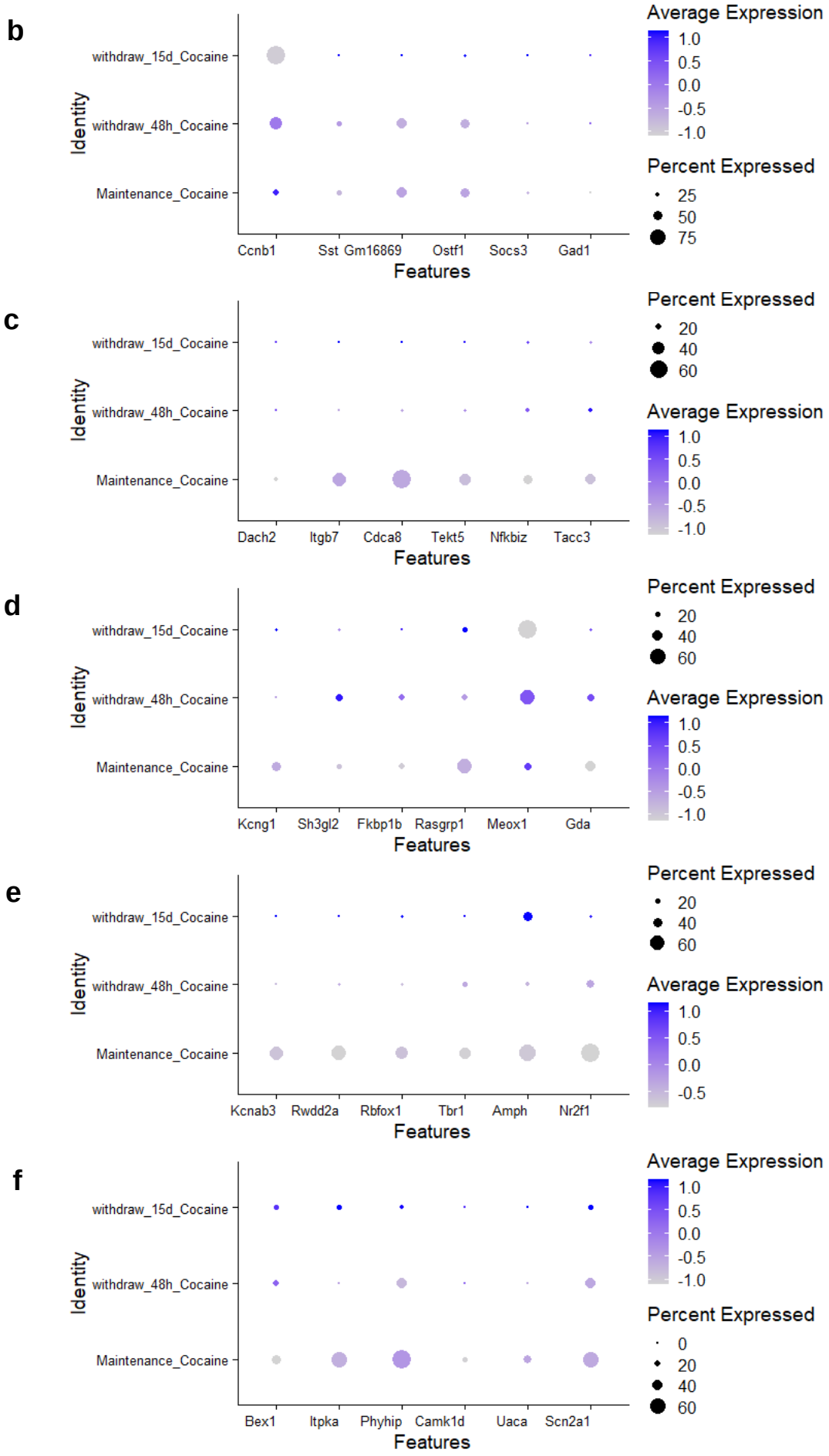


Figure 5. Cell type specific expression of the relevant genes in mouse PFC.

3.3. Differentially expressed genes in each cell type

We examined the top differentially expressed genes in each characterized cell type in PFC in cocaine IVSA maintenance, short-term withdrawal (48 hours), and long-term withdrawal (15 days) using differential expression analysis methods. Top 6 most varied genes over the course of 3 stages of addiction for each cell type were identified (Figure 6), which is summarized in the Table 1.




b
withdraw_15d_Cocaine
withdraw_48h_Cocaine
Maintenance_Cocaine
Identity
Ccnb1 Sst Gm16869 Ostf1 Socs3 Gad1
Features
Average Expression
1.0
0.5
0.0
-0.5
-1.0
Percent Expressed
25
50
75
c
withdraw_15d_Cocaine
withdraw_48h_Cocaine
Maintenance_Cocaine
Identity
Dach2 Itgb7 Cdca8 Tekt5 Nfkbiz Tacc3
Features
Percent Expressed
20
40
60
Average Expression
1.0
0.5
0.0
-0.5
-1.0
d
withdraw_15d_Cocaine
withdraw_48h_Cocaine
Maintenance_Cocaine
Identity
Kcng1 Sh3gl2 Fkbp1b Rasgrp1 Meox1 Gda
Features
Percent Expressed
20
40
60
Average Expression
1.0
0.5
0.0
-0.5
-1.0
e
withdraw_15d_Cocaine
withdraw_48h_Cocaine
Maintenance_Cocaine
Identity
Kcnab3 Rwdd2a Rbfox1 Tbr1 Amph Nr2f1
Features
Percent Expressed
20
40
60
Average Expression
1.0
0.5
0.0
-0.5
f
withdraw_15d_Cocaine
withdraw_48h_Cocaine
Maintenance_Cocaine
Identity
Bex1 Itpka Phyhip Camk1d Uaca Scn2a1
Features
Average Expression
1.0
0.5
0.0
-0.5
-1.0
Percent Expressed
0
20
40
60

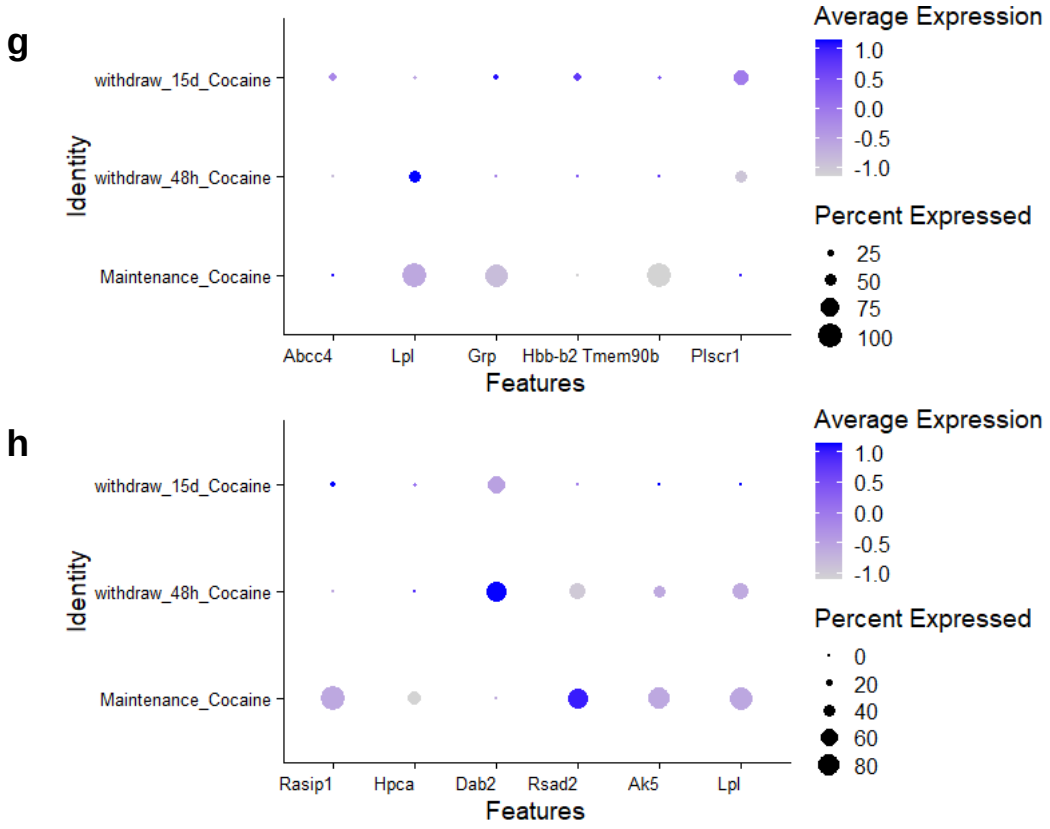


Figure 6. The gene expression and frequency of the 6 most variable genes over 3 periods of withdrawal for each of the 8 cell types: a) Astro b) Excitatory c) Inhibitory d) Endo e) Micro f) Oligo g) OPC h)NF Oligo

Table 1. Top 6 most varied genes for each cell type.

Cell type	**Top 6 Most Varied Genes Over the Course of 3 Stages of Addiction for Each Cell Type**					
Astro:	*Gabrg2*	*Gng2*	*Mgp*	*Iqgap1*	*Icam1*	*Gm3906*
Excitatory:	*Ccnb1*	*Sst*	*Gm16869*	*Ostf1*	*Socs3*	*Gad1*
Inhibitory:	*Dach2*	*Itgb7*	*Cdca8*	*Tekt5*	*Nfkbiz*	*Tacc3*
Endo:	*Kcng1*	*Sh3gl2*	*Fkbp1b*	*Rasgrp1*	*Meox1*	*Gda*
Micro:	*Kcnab3*	*Rwdd2a*	*Rbfox1*	*Tbr1*	*Amph*	*Nr2f1*
Oligo:	*Bex1*	*Itpka*	*Phyhip*	*Camk1d*	*Uaca*	*Scn2a1*
OPC:	*Abcc4*	*Lpl*	*Grp*	*Hbb-b2*	*Tmem90b*	*Plscr1*
NF Oligo	*Rasip1*	*Hpca*	*Dab2*	*Rsad2*	*Ak5*	*Lpl*

4. Discussion

By conducting computational analysis of publicly accessible single-cell RNA sequencing datasets in a mouse model of cocaine addiction, this study found that ΔFosB is predominantly expressed in non-neuronal cells, especially in endothelial and microglial cells, while MeCP2 and BDNF are expressed abundantly in excitatory neurons. Furthermore, we identified the genes that exhibited the most significant differential expression during different stages of cocaine IVSA, providing insight into the genes that are most affected by cocaine addiction.

Why did we choose to study the specific genes in our experiment? To better understand the cellular mechanisms underlying addiction, we ask whether previously identified addiction-related genes demonstrate cell type-specific expression in the PFC. To answer that, gene dynamics will take place in the same cell type because we know cocaine addiction is a chronic disease and presumably affects

only the expression level of certain genes. Addiction does not alter already present genes to express something else entirely. Therefore, we can assume that each cell type will have the same type of effect as one other since they are still expressing the same genes.

Why did we choose to study the specific genes in our experiment? As mentioned in the introduction, the genes express for the molecular players ΔFosB, MeCP2, and BDNF, each have vital roles that stand out more than other genes when it comes to the development of cocaine addiction. ΔFosB builds up in the nuclear accumbens after addiction. Since it is a type of transcription factor and lasts quite long in the neurons, it is likely that ΔFosB could alter gene expression while and after addiction. MeCP2 works with chromatins and histones, its purpose to methylate DNA. Incorrect methylation could cause unwanted gene expression changes. In this case, MeCP2 could increase a consumer's desire to cocaine. Finally, BDNF has a role in synaptic plasticity. BDNF will increase the rewarding effect of cocaine and potential for relapse. Furthermore, many studies have already shown that these are impactful molecular players in addiction and their importance is confirmed.

Why did we pick the most variable genes over the 3 stages of addictions for each specific cell type as candidates for molecular players? Cocaine-related genes are related to the cell types that express them. The top genes with the most variability across the 3 stages of addiction are almost certain to have an impact on how each cell type functions in addiction. However, that is not to say that other genes with less variation may also contribute to the development of cocaine addiction. For example, although we have already confirmed that ΔFosB, MeCP2, and BDNF are molecular players in cocaine addiction, none are within the top 6 most variable genes of any of the 8 cell types. This is because difference in expression level does not correlate with the importance that the gene plays in cocaine addiction: it correlates with the degree that gene is related to addiction. Therefore, it could be very possible that another undiscovered molecular player might be related to a gene that is within the top of the most variable genes but not the top 6.

There are many methods to further study cocaine addiction-related genes identified in this paper:

1) Although differential expression analysis was performed in this experiment, whether any gene pathways are involved is still unknown. In future studies, we could use gene pathway analysis in each cluster to hopefully gain more insight into the gene expressions of each cluster.

2) We can look at gene allocation in the subcluster of excitatory neurons. We see that most of the molecular players were exhibited in the excitatory neurons. If we are able to specify the clusters into smaller groups, we could be provided with more specific information on the gene expression and frequencies of the cell type. Subclustering is both a simple and practical procedure that could supply many benefits.

3) We can enlarge the sample size to make receive more precise results and possibly even find other genes that may be involved in cocaine addiction. As mentioned in the questions, perhaps there are certain genes that do play a role in addiction but do not change much in expression over the course of addiction. Since we only looked at the genes with the largest change in expression levels, future studies could determine if such genes exist and if there is a way to find them.

4) With the top 6 most differential genes from each specific cell type, we can further examine what molecular players are related to each gene when the subject is addicted to cocaine. Many of these genes have already been proven to be linked to drug addiction. For example, *Sst* expression in excitatory neurons is highly variable across cocaine IVSA development and withdrawal. It is also reported somatostatin (SST)-expressing interneurons regulates behavioral responses to cocaine [14]. However, while many of the high-variance genes have already been proven to be linked to drug addiction, some do not possess studies that relate them to drug addiction.

5. Conclusion

Cocaine addiction leads to many epigenetic and transcriptional alterations in the PFC, suppressing or increasing the expression levels of certain genes by methylating histones. Many of these changes can be related to certain molecular players. The most important of these molecular players are ΔFosB,

MeCP2, and BDNF. To better understand the cellular mechanisms involved in cocaine addiction, we asked whether candidate genes *Fosb*, *Mecp2*, and *Bdnf* exhibit a cell type-specific expression in the PFC. In our analysis, we discovered that *Fosb* is mostly expressed in non-neuronal cells, especially in endothelial and microglial cells; *Mecp2* and *Bdnf* are abundantly expressed in excitatory neurons. Also, by finding the genes that were most differentially expressed in various stages of cocaine IVSA, we figured out which genes were most affected by cocaine addiction and therefore the specific molecular players that were responsible for the development of addiction. The specific cell type expression of the addiction gene signature based on the computational analysis might be relevant to design potential new pharmaceutical approaches to tackle addiction, and high-variance genes might be a new target in therapy development.

Bibliography

[1] 2018 Annual Surveillance Report of Drug-related Risks and Outcomes | National Rural Health Resource Center. https://www.ruralcenter.org/resource-library/2018-annual-surveillance-report-of-drug-related-risks-and-outcomes.

[2] Goldstein, R. Z. & Volkow, N. D. Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. Nat. Rev. Neurosci. 12, 652–669 (2011).

[3] Chen, B. T. et al. Rescuing cocaine-induced prefrontal cortex hypoactivity prevents compulsive cocaine seeking. Nature 496, 359–362 (2013).

[4] Miller, E. K., Freedman, D. J. & Wallis, J. D. The prefrontal cortex: categories, concepts and cognition. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 357, 1123–1136 (2002).

[5] Funahashi, S. & Andreau, J. M. Prefrontal cortex and neural mechanisms of executive function. J. Physiol. Paris 107, 471–482 (2013).

[6] Nestler, E. J. & Lüscher, C. The Molecular Basis of Drug Addiction: Linking Epigenetic to Synaptic and Circuit Mechanisms. Neuron 102, 48–59 (2019).

[7] Bhattacherjee, A. et al. Cell type-specific transcriptional programs in mouse prefrontal cortex during adolescence and addiction. Nat. Commun. 10, 4169 (2019).

[8] Nestler, E. J., Barrot, M. & Self, D. W. ΔFosB: A sustained molecular switch for addiction. Proc. Natl. Acad. Sci. 98, 11042–11046 (2001).

[9] Kelz, M. B. et al. Expression of the transcription factor ΔFosB in the brain controls sensitivity to cocaine. Nature 401, 272–276 (1999).

[10] Im, H.-I., Hollander, J. A., Bali, P. & Kenny, P. J. MeCP2 controls BDNF expression and cocaine intake through homeostatic interactions with microRNA-212. Nat. Neurosci. 13, 1120–1127 (2010).

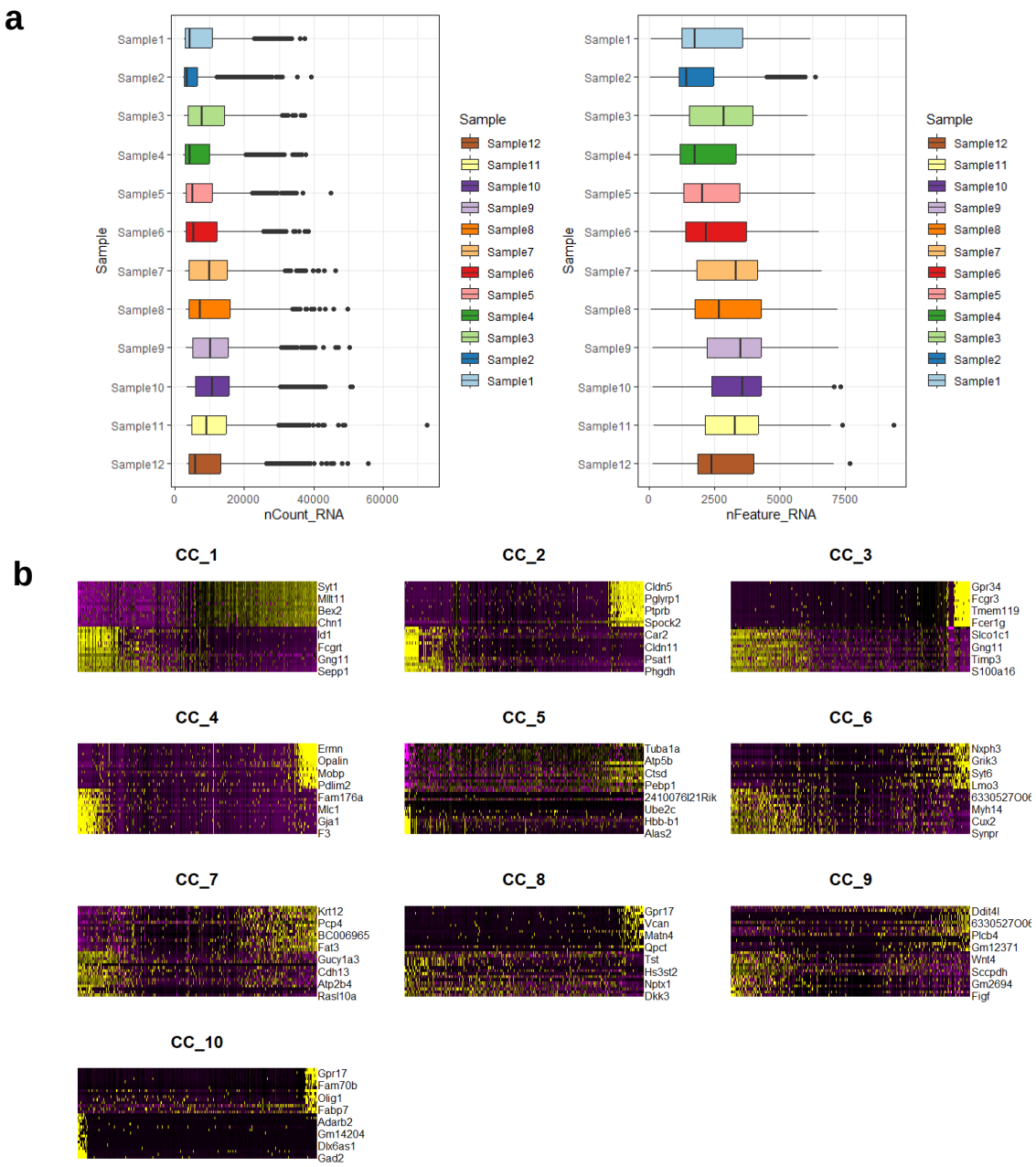
[11] Lu, L., Dempsey, J., Liu, S. Y., Bossert, J. M. & Shaham, Y. A Single Infusion of Brain-Derived Neurotrophic Factor into the Ventral Tegmental Area Induces Long-Lasting Potentiation of Cocaine Seeking after Withdrawal. J. Neurosci. 24, 1604–1611 (2004).

[12] Schoenbaum, G., Stalnaker, T. A. & Shaham, Y. A role for BDNF in cocaine reward and relapse. Nat. Neurosci. 10, 935–936 (2007).

[13] Kmiotek, E. K., Baimel, C. & Gill, K. J. Methods for Intravenous Self Administration in a Mouse Model. J. Vis. Exp. e3739 (2012) doi:10.3791/3739.

[14] Ribeiro, E. A. et al. Transcriptional and physiological adaptations in nucleus accumbens somatostatin interneurons that regulate behavioral responses to cocaine. Nat. Commun. 9, 3149 (2018).

Appendix



Appendix Fig 1. a. boxplots showing the distribution of the number of genes (nFeature_RNA, right) detected in each of the 12 samples and corresponding gene counts (nCounts_RNA, Left); b. the heatmap of top differentially expressed genes in each principal component identified by canonical correlation analysis. Both cells and genes are sorted by their principal component scores (shows how similar gene expression in each cell is compared to each other).

R code availability:

R code used to conduct this study is saved in a shared repository (https://github.com/tengeric1595/Cell-type-specific-expression-of-the-molecular-players-in-mouse-prefrontal-cortex-during-cocaine-add.git)