UNIVERSITY OF CALIFORNIA,
IRVINE

Development of Novel Tools for Probing the Replicating Genome Reveals Insights Into Stem Cell Fate Regulation

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biomedical Engineering

by

Navied Akhtar

Dissertation Committee:
Associate Professor Timothy Downing, Chair
Associate Professor Elizabeth Read
Assistant Professor Fangyuan Ding

2023

DEDICATION

To

My love, Gabriela, for being by my side from the beginning,

My mother, Azam, for showing me what hard work can achieve,

My father, Farzin, for teaching me that life is meant to be lived,

My brother, Neema, for always humbling me,

and to myself, for believing in myself, and doing all this hard work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my appreciation to my committee chair, Professor Timothy Downing, who is nothing short of a savant: he persistently challenged my research and scholarship. Without his guidance, this dissertation would not have been possible.

I would like to thank my committee members, Professor Elizabeth Read and Professor Fangyuan Ding for mentoring me and guiding me along my journey. I would like to also give a very big thank you to Professor Chang Liu and Professor Michelle Digman, who helped keep me grounded and taught me that scientific pursuits are dynamic and the road is rarely ever straight.

Financial support was provided by the University of California, Irvine, and research reported in this publication was supported by the National Heart, Lung, And Blood Institute of the National Institutes of Health under Award Number T32HL116270. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

VITA

Navied Akhtar

EDUCATION

Ph.D. University of California, Irvine, Irvine, CA — May 2023
Biomedical Engineering
Advisor: Timothy L. Downing, Ph.D.

M.S. University of California, Irvine, Irvine, CA — June 2021
Biomedical Engineering

M.S. University of California, Irvine, Irvine, CA — June 2017
Biotechnology

B.A. University of California, Irvine, Irvine, CA — June 2012
Political Science

PUBLICATIONS

Published

PK Veerasubramanian, A Trinh, **N Akhtar**, W Liu, TL Downing. Biophysical and Epigenetic Regulation of Cancer Stemness, Invasiveness, and Immune Action. *Current Tissue Microenvironment Reports*, 1, 277-300 (2020). https://doi.org/10.1007/s43152-020-00021-w

N Akhtar. The 21st Century Cures Act and Its Effect on Stem Cell Research and Therapy. *UCI Stem Cell Research Center*, (2017). https://www.stemcell.uci.edu/News/journals/21_century. php

Submitted

Y Song, J Soto, SY Wong, T Hoffman, Y Wu, **N Akhtar**, S Norris, J Chu, H Park, D Kelkhoff, CE Ang, M Wernig, TL Downing, M Poo, S Li. Intermediate Matrix Stiffness Promotes Cell Reprogramming by Enhancing Histone Acetylation. (Submitted)

J Soto, Y Song, B Chen, H Park, **N Akhtar**, T Hoffman, C Ly, J Sia, S Wong, D Kelkhoff, J Chu, M Poo, T Downing, A Rowat, S Li. Reduction of Intracellular Tension and Cell Adhesion Promotes Open Chromatin Structure and Enhances Cell Reprogramming. (Submitted)

Awards

Opportunity Award Recipient at the Center for Complex Biological Systems — 2019

NIH T32 Cardiovascular Applied Research and Entrepreneurship Fellowship — 2018

ABSTRACT OF THE DISSERTATION

Development of Novel Tools for Probing the Replicating Genome Reveals Insights Into Stem Cell Fate Regulation

by

Navied Akhtar

Doctor of Philosophy in Biomedical Engineering

University of California, Irvine, 2023

Professor Timothy L. Downing, Chair

DNA replication is of paramount importance in cell proliferation, the process by which cells grow and divide to produce new cells. During this critical phase of the cell cycle, a cell duplicates its DNA, ensuring that each daughter cell receives a complete set of genetic information. Beyond mere duplication of genetic material, DNA replication also involves the crucial task of recovering the epigenome. The epigenome plays a fundamental role in regulating gene expression without changing the underlying DNA sequence. Epigenetic modifications need to be precisely and efficiently duplicated and transferred onto newly synthesized DNA post-replication, allowing the preservation of cellular identity and function across generations. It is within this context that cell fate transitions may occur. As the cell duplicates its genome and epigenome, modifications can be introduced or removed, and chromatin architecture can change, leading to alteration in gene expression that could potentially trigger a cell to leave its original state and transition into a different cell type. Therefore, DNA replication represents a critical window for the study of cell fate decisions in human embryonic stem cells and offers unique insights into the dynamic nature of the epigenome and its influence on cellular identity and development.

Due to the underexplored nature of this phase of the cell cycle, we lack the technologies to be able to reliably assay changes during and post-DNA replication. In this work, technologies were developed to provide a means to this end. First, Repli-ATAC-seq was developed to understand how the chromatin landscape is reestablished in post-replication time by assaying chromatin accessibility. This assay shed light on nucleosome dynamics after replication, bringing to light the potential timing of accessibility changes that may be reflected in biological processes such as DNA mismatch repair, transcription factor accessibility binding dynamics, and an association between increased accessibility and high gene expression. Furthermore, an S-phase fractioned single cell 10X Multiome ATAC and Gene Expression assay was produced to understand if and how chromatin architecture and gene expression are related during S-phase, and whether or not there are clues to explore cell state identities throughout S-phase. Remarkably, a neuroectodermal contributor *DACH1* revealed itself as a contributor to a potentially primed cell identity in S-phase. This work establishes techniques that could be used to identify cell cycle dynamics of chromatin accessibility and associated biological changes.

INTRODUCTION

Stem cells and cell fate transitions

Stem cells are a unique class of cells that possess the dual capability of self-renewal and differentiation. This means they can produce identical copies of themselves and also give rise to various other cell types, a crucial process for the development and maintenance of multicellular organisms. The transition from a stem cell to a more specialized cell is guided by changes in gene expression and is referred to as a cell fate transition[1]. This shift from an undifferentiated to a differentiated state is tightly regulated and controlled by a network of signals from both inside the cell and from the outside environment. The balance between self-renewal and differentiation is key to healthy organismal development and function. Disruption of this balance can lead to a range of issues, including developmental disorders, premature aging, and diseases such as cancer[2,3]. As such, understanding the mechanisms that guide stem cell behavior and cell fate transitions is a key focus of developmental biology and regenerative medicine[4].

DNA replication and epigenetic maintenance

DNA replication is a fundamental cellular process that ensures the accurate duplication of genetic material, permitting cell division and propagation. During this process, the double helix of the parent DNA molecule is unwound and each strand is used as a template to synthesize a new, complementary strand. However, beyond this basic copying of genetic information, DNA replication also involves the critical process of epigenetic maintenance[5,6]. Epigenetic modifications, which include DNA methylation and histone modifications, play vital roles in regulating gene expression without altering the underlying DNA sequence[7]. These epigenetic marks must be

accurately copied and transferred to the newly synthesized DNA to maintain cellular identity and function across generations[8]. Maintenance enzymes, such as DNA methyltransferases and histone-modifying enzymes, play an instrumental role in this process[5,9]. They ensure that the epigenetic landscape of a cell is replicated alongside the genetic material, thus preserving the cell's unique gene expression profile and functionality. However, errors in this maintenance can lead to aberrant gene expression and disease development, underscoring the importance of the precise replication of both genetic and epigenetic information[3].

DNA-protein interactions and regulation of gene expression

Interactions between DNA and proteins play an essential role in regulating gene expression. Several types of proteins, including transcription factors, chromatin remodelers, and enzymes that modify histones or DNA, interact directly with DNA to control which genes are turned on or off in a cell[10–12]. Transcription factors bind to specific DNA sequences near genes and recruit transcription machinery, thereby initiating the process of transcription. Chromatin remodelers and histone-modifying enzymes alter the structure of nucleosomes, known as chromatin, either making the DNA more accessible for transcription or compacting it to inhibit transcription. Changes in these DNA-protein interactions, driven by various internal and external signals, can lead to changes in gene expression patterns and allow for differentiation to take place.

Understanding DNA replication, cell fate transitions, and gene expression regulation requires novel tools

The process of DNA replication presents a critical period for understanding how epigenetic markers are accurately inherited, as well as identifying potential points for

cell fate determination. However, there is currently a deficit in the methodologies available to specifically probe the epigenome and DNA-protein interactions during replication[13,14]. This phase of the cell cycle offers a window into the mechanisms of epigenetic heritability, as the existing epigenetic marks must be copied and accurately transmitted to the newly synthesized DNA strands. Additionally, it presents opportunities for changes in cell fate, as adjustments to the epigenome could direct a cell toward a different developmental path. Therefore, developing methods to study the epigenome during replication is an important area for future research, as it could greatly enhance our understanding of cell differentiation, development, and disease.

SECTION 1: Repli-ATAC-seq reveals insights into the chromatin landscape in post-replication time

1.1 Introduction

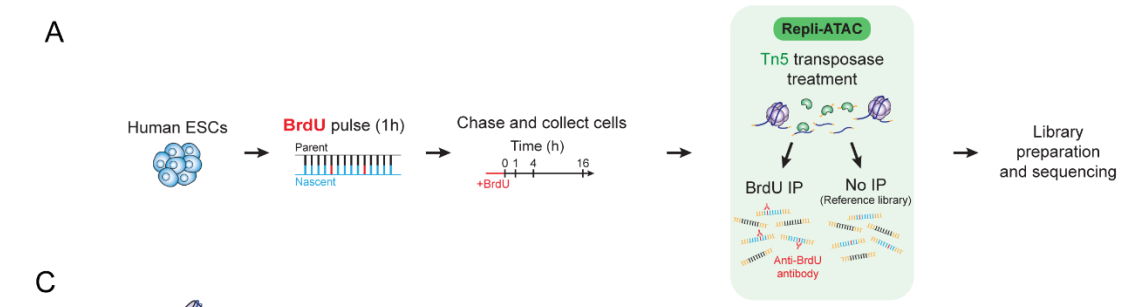
DNA replication is not only necessary for the inheritance of genetic information across cell proliferation, but also for the transmission of histones and their post-translational modifications, a process that plays an integral role in maintaining the epigenetic landscape of a cell. During DNA replication, parental histones, along with their modifications, are distributed onto the newly synthesized DNA strands and segregate in an alternating matter, helping to re-establish the original chromatin structure[15]. This segregation of histones and re-establishment of histone marks is necessary for the maintenance of gene expression patterns and cell identity across cell generations[16–18]. However, the precise mechanisms of how these histone marks are accurately replicated during DNA replication is a subject of ongoing research[19]. One technique that can be employed to probe this process is Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)[20]. ATAC-seq is a powerful tool that can provide a genome-wide resolution of chromatin accessibility by integration of a Tn5 transposase that can access euchromatin and provide TF binding information, while being inaccessible at heterochromatin. Combining ATAC-seq with a replication-associated assay method at various time points in post-replication time produced a novel assay termed replication-associated ATAC-seq (Repli-ATAC-seq). Repli-ATAC-seq can provide insights into how chromatin accessibility changes over time following DNA replication and how these changes might correspond to the distribution and modifications of histones as they pertain to gene expression and stem cell identity.

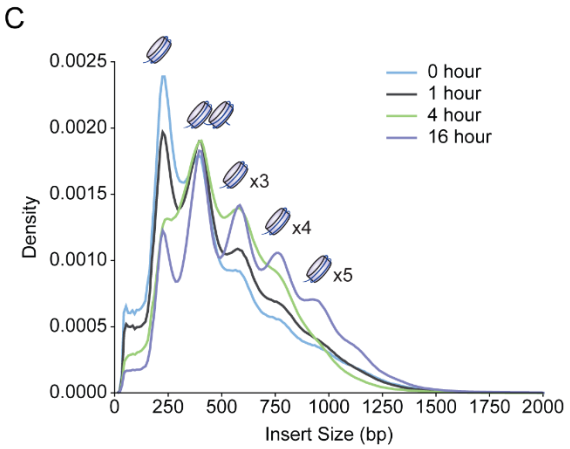
1.2 Results

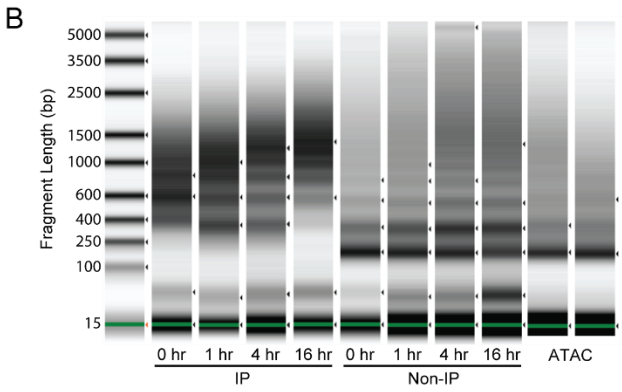
1.2.1 *Replication-Associated Assay for Transposase Accessibile Chromatin Sequencing method and initial quality metrics*

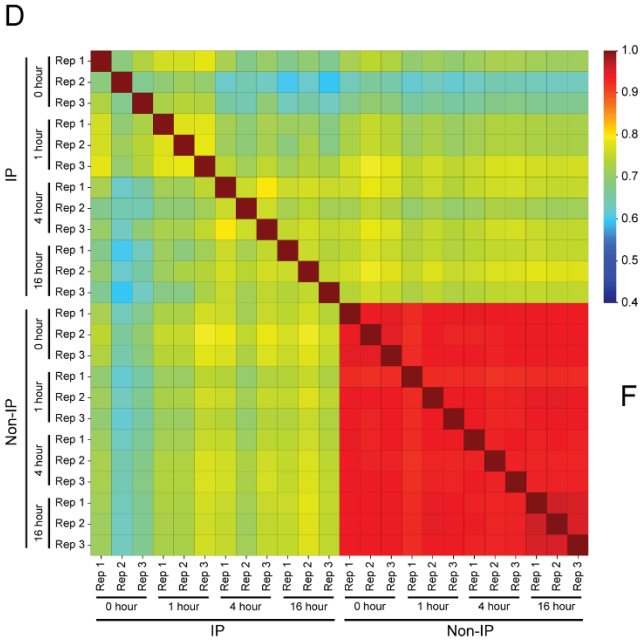
Repli-ATAC-seq of human ESCs was performed on three biological replicates by culturing cells in Bromodeoxyuridine (BrdU) containing media for 1 hour prior to removing the media and collecting cells at 0, 1, 4, and 16 hour time points (Figure 1.1A). BrdU is a synthetic thymidine analog that incorporates into newly replicating DNA and can be a target for immunoprecipitation (IP)[21,22]. The cells are lysed and Tn5 transposase enzymatic digestion is performed on 100,000 isolated nuclei per replicate and sample. It is critical to purify the DNA and perform a 5 minute 72C PCR extension step after this digestion to ensure any overhangs are converted to blunt ends for downstream PCR. After this extension step, BrdU IP is performed using anti-BrdU antibody after a 95C heat denaturation of the DNA into single strands. Anti-BrdU antibody pulls down nascent single strands of DNA, allowing for removal of fragments of DNA that were not replicating during the 1 hour BrdU pulse. It is important to keep a Non-IP reference library for comparisons, and this is done by performing an assay in parallel without the IP step. Upon completion of the BrdU IP, the DNA is purified and the samples undergo library preparation for sequencing.

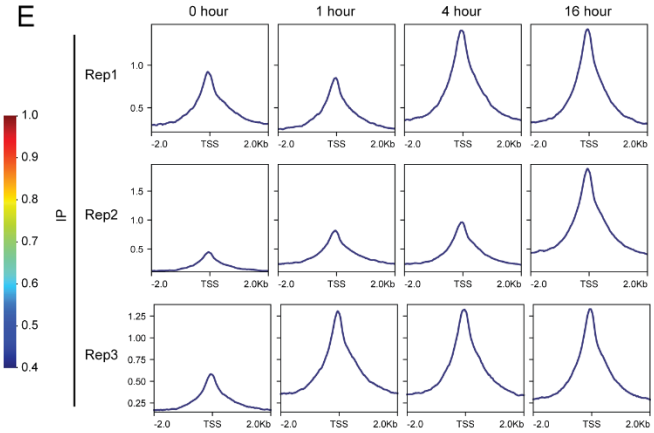
This method is similar to another published method[19], however, there are differences between the two. This method utilizes BrdU instead of EdU, and uses immunoprecipitation instead of biotin-streptavidin click chemistry. This method ends up enriching for fragments at mononucleosome lengths and larger, whereas the previously published method focuses on subnucleosomal fragments, providing unique insights into the nucleosome dynamics at play over a longer time scale.


A
Human ESCs
BrdU pulse (1h)
Parent
Nascent
Chase and collect cells
Time (h)
0 1 4 16
+BrdU
Repli-ATAC
Tn5 transposase treatment
BrdU IP
No IP
(Reference library)
Anti-BrdU antibody
Library preparation and sequencing


C
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
Density
0 hour
1 hour
4 hour
16 hour
x3
x4
x5
0 250 500 750 1000 1250 1500 1750 2000
Insert Size (bp)


B
5000
3500
2500
1500
1000
600
400
250
100
15
Fragment Length (bp)
0 hr 1 hr 4 hr 16 hr 0 hr 1 hr 4 hr 16 hr ATAC
IP
Non-IP


D
IP
Non-IP
0 hour
1 hour
4 hour
16 hour
Rep 1
Rep 2
Rep 3
1.0
0.9
0.8
0.7
0.6
0.5
0.4


E
0 hour
1 hour
4 hour
16 hour
Rep1
Rep2
Rep3
IP
-2.0 TSS 2.0Kb

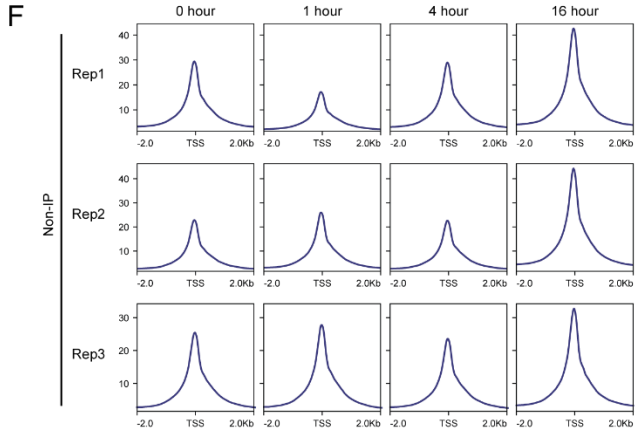

F
0 hour
1 hour
4 hour
16 hour
Rep1
Rep2
Rep3
Non-IP
-2.0 TSS 2.0Kb

Figure 1.1. Replication-Associated Assay for Transposase Accessible Chromatin (Repli-ATAC)-seq
(A) Repli-ATAC-seq protocol schematic
(B) Repli-ATAC-seq final library TapeStation for replicate 1 for 0, 1, 4, 16 hour IP and Non-IP libraries with a non-BrdU treated sample as control.
(C) Density plot of insert size distributions in base pairs for 0, 1, 4, and 16 hour Repli-ATAC-seq IP samples.
(D) Spearman correlation heatmap of the triplicate samples at each time point for Repli-ATAC-seq IP and Non-IP samples.
(E) TSS enrichment plots 2kbp up and downstream the TSS for each replicate and time point for IP and (F) Non-IP.

TapeStation analysis for the IP and Non-IP samples reveals a trend of an increase in fragment lengths for the IP samples over time, indicating that larger fragment aggregates are present at later time point of post-replication, reflecting chromatin condensation events (Figure 1.1B). Also, there was a loss of subnucleosomal fragments in the IP sample that was present in the Non-IP and regular ATAC-seq replicates. Upon sequencing, this was quantified by plotting a density curve of insert sizes across time points, which shows high mononucleosome abundance at the 0-hour time point and low abundance at the 16-hour time point (Figures 1.1C and A.1). There is a reversal of this trend across time points, most apparent in the tetra- and penta-nucleosome size regions.

1.2.2 *Comparisons of Repli-ATAC-seq biological triplicate samples*

Spearman correlations of the IP and Non-IP three biological triplicate Repli-ATAC-seq time points revealed a moderate correlation between IP samples and strong correlation between Non-IP samples (Figure 1.1D) and reflects the expected heterogeneity in accessibility during and post-DNA replication that would be absent in a bulk, full cell-cycle sample. TSS enrichment is a commonly used metric to determine ATAC-seq quality[23,24], and enrichment plots with a peak at the center

dropping off on both sides indicate an expected ATAC-seq profile for all samples and replicates (Figure 1.1E).

Peaks were called using the MACS2 algorithm[25], and the number of peaks between the IP samples increased across post-replication time ranging from 43 to 161 peaks in 0-hour replicates, 98 to 248 peaks in 1-hour replicates, 212 to 373 peaks in 4-hour replicates, and 459 to 1,288 peaks in 16-hour replicates with an average of 326.5 peaks (Figure 1.2A and Table 1.1). Although there was a slight increase in peaks called in the 16-hour Non-IP sample, all samples had greater than 40,000 peaks with a maximum peak count of 84,865 (Figure 1.2B and Table 1.1). The Non-IP 0-hour replicates had 53,687 to 69,148 peaks, the 1-hour replicates had 40,915 to 59,341 peaks, the 4-hour replicates had 49,811 to 59,230 peaks, and the 16-hour replicates had 66,309 to 84,865 peaks with an average of 71,045 peaks. The fraction of reads in peaks (FRiP) in the IP samples ranged from 0.12 to 0.28, with an average of 0.21 across all IP replicates (Figure 1.2C and Table 1.1). The FRiP in the Non-IP samples ranged from 0.19 to 0.28 across all Non-IP replicates with an average of 0.25 (Figure 1.2D and Table 1.1). Quantifying IP peaks as unique in the IP sample compared to shared with the Non-IP sample displayed how post-replication time introduces unique chromatin accessibility profiles (Figure 1.2E). To test the effect of the absence of subnucleosomal fragments observed in the IP dataset compared to the Non-IP and native ATAC libraries, all replicates and time points were filtered to insert lengths only larger than 130 base pairs (Figure 1.3A). Expectedly, peak-calling reduced some of the samples' peak counts and FRiP scores, but the overall trends remained the same for these metrics even though the absolute counts were reduced (Figures 1.3B, C, D, and E).

Table 1.1. Repli-ATAC-seq sample data

	% Mapped	# of Aligned Reads	% Unique Reads	Peak Counts	Fraction of Reads in Peaks
0 hr IP Replicate 1	99.6	239067848	3.3	43	0.16
0 hr IP Replicate 2	99.6	250253428	1.5	159	0.19
0 hr IP Replicate 3	99.6	236243748	2	161	0.14
1 hr IP Replicate 1	99.6	236753998	2.8	98	0.28
1 hr IP Replicate 2	99.6	262627049	2.5	136	0.27
1 hr IP Replicate 3	99.6	208834347	4.2	248	0.24
4 hr IP Replicate 1	99.6	248757008	3.1	373	0.2
4 hr IP Replicate 2	99.6	203180209	2.6	212	0.22
4 hr IP Replicate 3	99.5	235689404	3.1	226	0.12
16 hr IP Replicate 1	99.5	189026238	4.8	545	0.22
16 hr IP Replicate 2	99.5	186172350	6.2	1288	0.21
16 hr IP Replicate 3	99.4	180880171	4.8	459	0.21
0 hr Non-IP Replicate 1	98.7	290142490	21.3	69148	0.26
0 hr Non-IP Replicate 2	99	204597984	22.8	53678	0.19
0 hr Non-IP Replicate 3	98.9	226625836	22	65001	0.24
1 hr Non-IP Replicate 1	98.4	238460836	19.6	40915	0.22
1 hr Non-IP Replicate 2	98.3	293431780	20.5	56839	0.24
1 hr Non-IP Replicate 3	98.6	218150910	24.4	59341	0.28
4 hr Non-IP Replicate 1	98.7	290437572	19.8	59230	0.26
4 hr Non-IP Replicate 2	98.5	265773824	19.3	49811	0.23
4 hr Non-IP Replicate 3	98.2	279411302	18.3	52968	0.22
16 hr Non-IP Replicate 1	99.3	218794508	31.4	74435	0.28
16 hr Non-IP Replicate 2	99.4	253450780	30.1	84865	0.28
16 hr Non-IP Replicate 3	99.3	215400344	27.5	66309	0.27

IP replicates were also merged to determine if pooling replicates created a more robust peak-calling output (Figure 1.4). Merged IP peak-calling produced 320 peaks for 0-hour, 1530 for 1-hour, 2407 peaks for 4-hour, and 5,061 peaks for 16-hour Repli-ATAC samples (Figure 1.4A). It is interesting to note that the number of peaks are not additive in the replicates – for example, the sum of the 0-hour IP replicates independently sum to 363 instead of the 320 peaks called for 0-hour, while the sum of the 16-hour IP replicates independently would be 2,292 peaks compared to the 5,061 called peaks in the merged 16-hour sample. Furthermore, FRiP scores were 0.17, 0.27, 0.20, and 0.23, for each of the time points respectively (Figure 1.4B). Spearman correlation of the merged replicated reflected a more linear correlation between the post-replication time points (Figure 1.4C). 0-hour IP was

most correlated with 1-hour IP, while 1-hour IP was most correlated with 0-hour IP and was still highly correlated with 4-hour and 16-hour IP merged samples. 4-hour IP was most correlated with 16-hour IP and least correlated with 0-hour IP, and 16-hour IP was most correlated with 4-hour IP. Strikingly, although the merged IP peak counts decreased for 0-hour IP, this merged sample had the highest proportion of unique peaks compared to its respective Non-IP sample, supporting that immediately after replication there is a unique accessibility picture of newly replicating DNA. Finally, TSS enrichment scores of the merged IP replicates followed a clear upward trend, while Non-IP TSS enrichment scores stay constant throughout time points with a high enrichment at the 16-hour time point (Figure 1.4G).

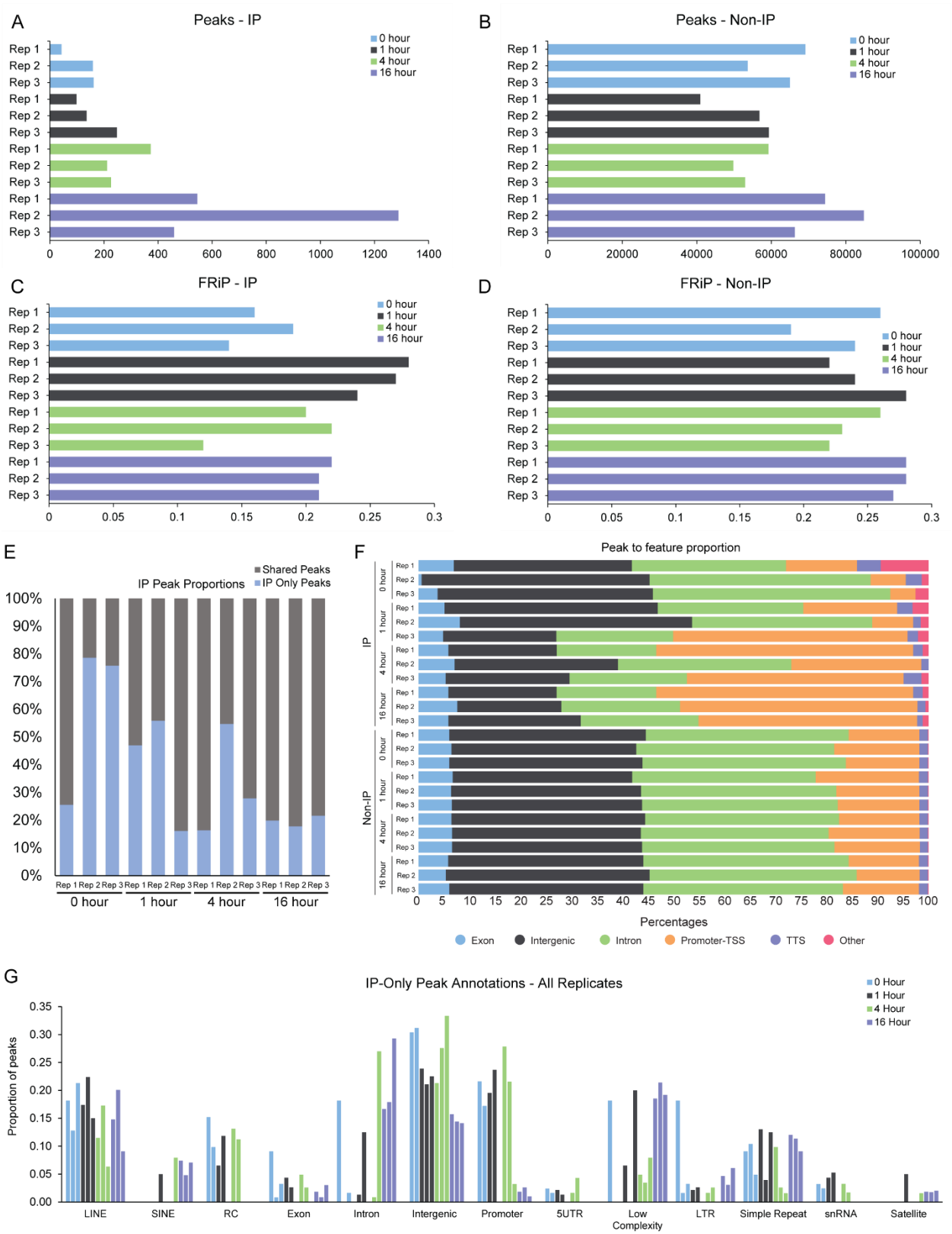

A
Peaks - IP
0 hour
1 hour
4 hour
16 hour
Rep 1
Rep 2
Rep 3
0
200
400
600
800
1000
1200
1400
B
Peaks - Non-IP
0
20000
40000
60000
80000
100000
C
FRiP - IP
0
0.05
0.1
0.15
0.2
0.25
0.3
D
FRiP - Non-IP
E
IP Peak Proportions
Shared Peaks
IP Only Peaks
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0 hour
1 hour
4 hour
16 hour
F
Peak to feature proportion
IP
Non-IP
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
Percentages
Exon
Intergenic
Intron
Promoter-TSS
TTS
Other
G
IP-Only Peak Annotations - All Replicates
0 Hour
1 Hour
4 Hour
16 Hour
Proportion of peaks
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
LINE
SINE
RC
Exon
Intron
Intergenic
Promoter
5UTR
Low Complexity
LTR
Simple Repeat
snRNA
Satellite

Figure 1.2. Triplicate features for IP and Non-IP at 0, 1, 4, and 16 hour time points.

(A) Histogram showing the distribution of peak counts for Repli-ATAC-seq IP and (B) Non-IP replicates at different time points for each replicate. Blue represents 0 hour time point, dark gray represents 1 hour time point, green represents 4 hour time point, purple represents 16 hour time point.

(C) Histogram showing the fraction of reads in peaks for Repli-ATAC-seq IP and (D) Non-IP replicates at different time points for each replicate.

(E) Stacked percentile bar chart with proportion of peaks unique in the IP sample and peaks shared between the IP and Non-IP samples for each replicate and time point. Light blue portions indicate the percentage of peaks uniquely found in each IP replicate. Grey portions indicate the percentage of peaks shared in both IP and corresponding Non-IP samples.

(F) Stacked percentile bar chart with peak to feature proportion for exons, intergenic regions, introns, promoter-TSS, TTS, and all other for Repli-ATAC-seq IP and Non-IP samples for each replicate and time point.

(G) Annotated genic region peaks from the unique IP only peaks from (E) plotted as a proportion of all unique IP only peaks for each replicate and time point.

1.2.3 *Genic feature peak enrichment in post-replication time*

To understand the biological importance of peak enrichment in the replicates and time points, the peak to genic proportion of the replicates were quantified (Figure 1.2F). There was variability in the proportion of genic features in the IP samples that were absent in the Non-IP samples. The IP samples seemed to show a higher proportion of peaks in Promoter-TSS regions of the genome compared to the Non-IP samples. The removal of insert lengths smaller than 130 base pairs did produce an effect on the proportion of genic features that peaks fell into. Peaks in exonic features almost fully disappeared, and the proportion of peaks in intergenic features increased, indicating that although there was an absence of subnucleosomal fragments in the IP samples, they were still contributing to peak-calling and were informative in genic feature enrichment (Figure 1.3F). Merging replicates showed a low proportion of exonic peaks at 0-hour that increased and maintained from 1-hour

onward (Figure 1.4F). 0-hour also had a markedly larger proportion of intergenic and intronic peaks, while the other time points had larger promoter-TSS proportions.

Plotting the individual unique peaks from the IP only triplicate samples (Figure 1.2E) and the genic feature these peaks overlap provided insights into genic feature dynamics, however, looking at the same figure in the merged replicate samples (Figure 1.4E) provides more clarity into the dynamics of what genic features these IP only peaks are falling into. LINE elements drop off dramatically in the 4-hour sample and are recovered afterward, and SINE elements decrease over the 16-hour post-replication time span. An explanation for this observation could be that chromatin remodeling is occurring at these sites, since transposable elements have been implicated in regulation of the epigenome and can function as the nucleation centers for heterchromatin[26]. Furthermore, the proportion of peaks at CpG islands spike at 4-hours post-replication, but almost completely disappear at the 16-hour time point. DNA methyltransferase (DNMT) 1 activity has been well-documented in post-replication time, and the accessibility could be attributed to the binding dynamics of DNMTs in this feature[27,28]. Low complexity and simple repeat regions of the genome contain a large proportion of peaks at 16-hours post-replication. These regions have been known to cause DNA polymerase stalling, although an explanation as to why their relative accessibility would be higher over post-replication time is elusive[29].

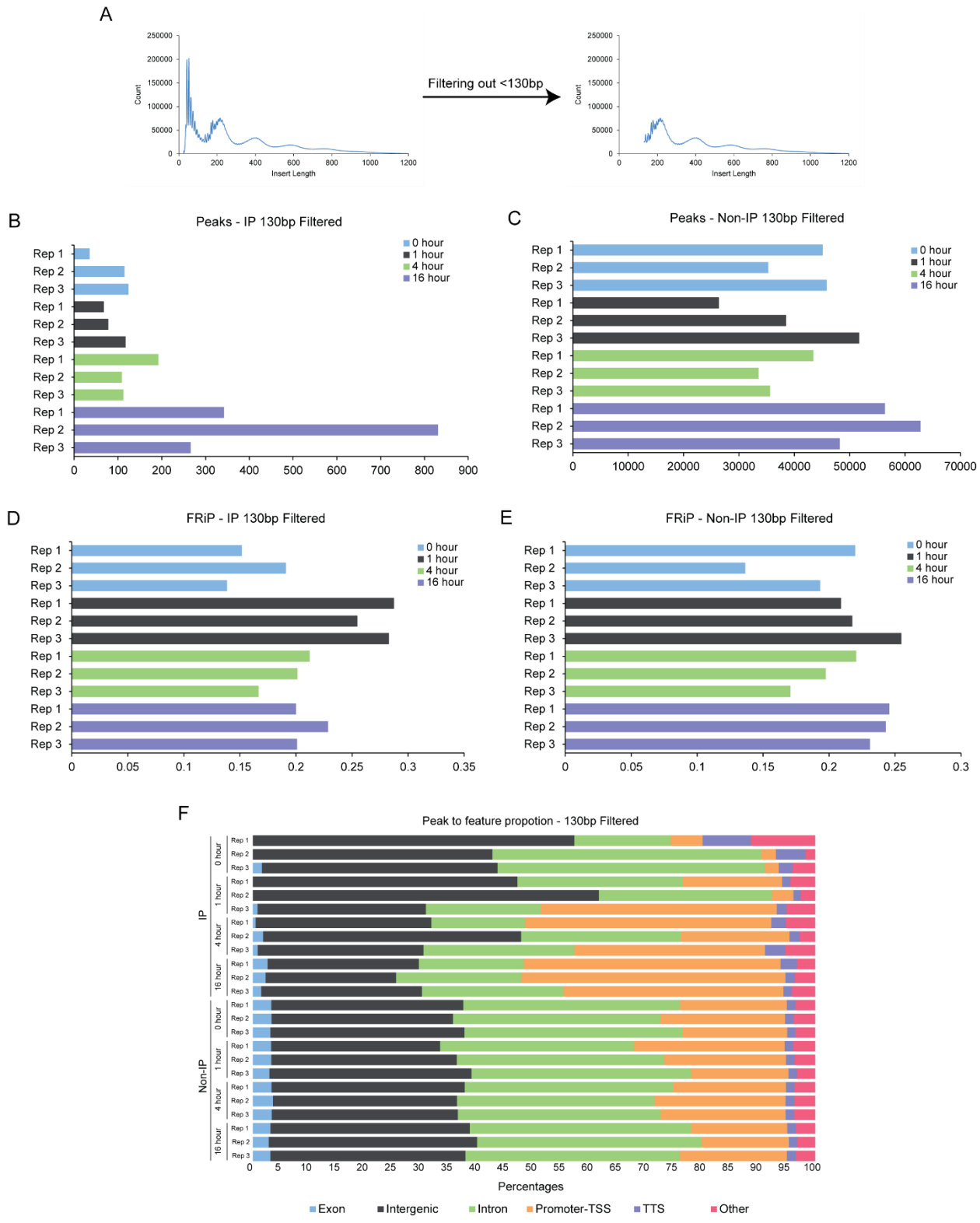


Figure 1.3. Small fragment removal conserves prior observed metrics in Repli-ATAC data.

(A) Schematic depicting removal of 130bp and smaller inserts.
(B) Histogram depicting the peak counts for 130bp filtered in Repli-ATAC-seq IP and (C) Non-IP for each replicate and time point.
(D) Histogram showing the fraction of reads in peaks for 130bp filtered in Repli-ATAC-seq IP and (E) Non-IP samples for each replicate and time point.
(F) Stacked percentile bar plot with peak to feature proportion for exons, intergenic regions, introns, promoter-TSS, TTS, and all other for 130bp filtered Repli-ATAC-seq IP and Non-IP samples for each replicate and time point.

1.2.4 *Repli-ATAC-seq shows variable probabilities in odds ratio scores at genic features*

Next, to compare the accessibility of IP compared to Non-IP Repli-ATAC-seq samples in different genic regions, odds ratios were calculated to understand accessibility likelihood. Consistent with prior trends, there was a higher likelihood of fragments to fall within introns, LINEs, SINEs, and repeats (Figure 1.5A). Remarkably, performing a similar analysis using ChromHMM[30] regions showed a higher likelihood of fragments falling at transcription elongation, weak transcription, polycomb repressed, low heterochromatin, and repetitive sites (Figure 1.5B). These are regions which would be expected to be less accessible in bulk cell samples, given these regions are generally inactive. Comparisons with histone mark TF ChIP sites, specifically at the 4-hour time point, showed a higher likelihood of accessible fragments falling at H4k20me1, H3K36me3, and H3K79me2 sites. These histone marks are documented markers for DNA mismatch repair, which could provide some resolution as to how long after replication mismatch repair takes place, and whether or not there is a separate regulatory role for this process[31–34].

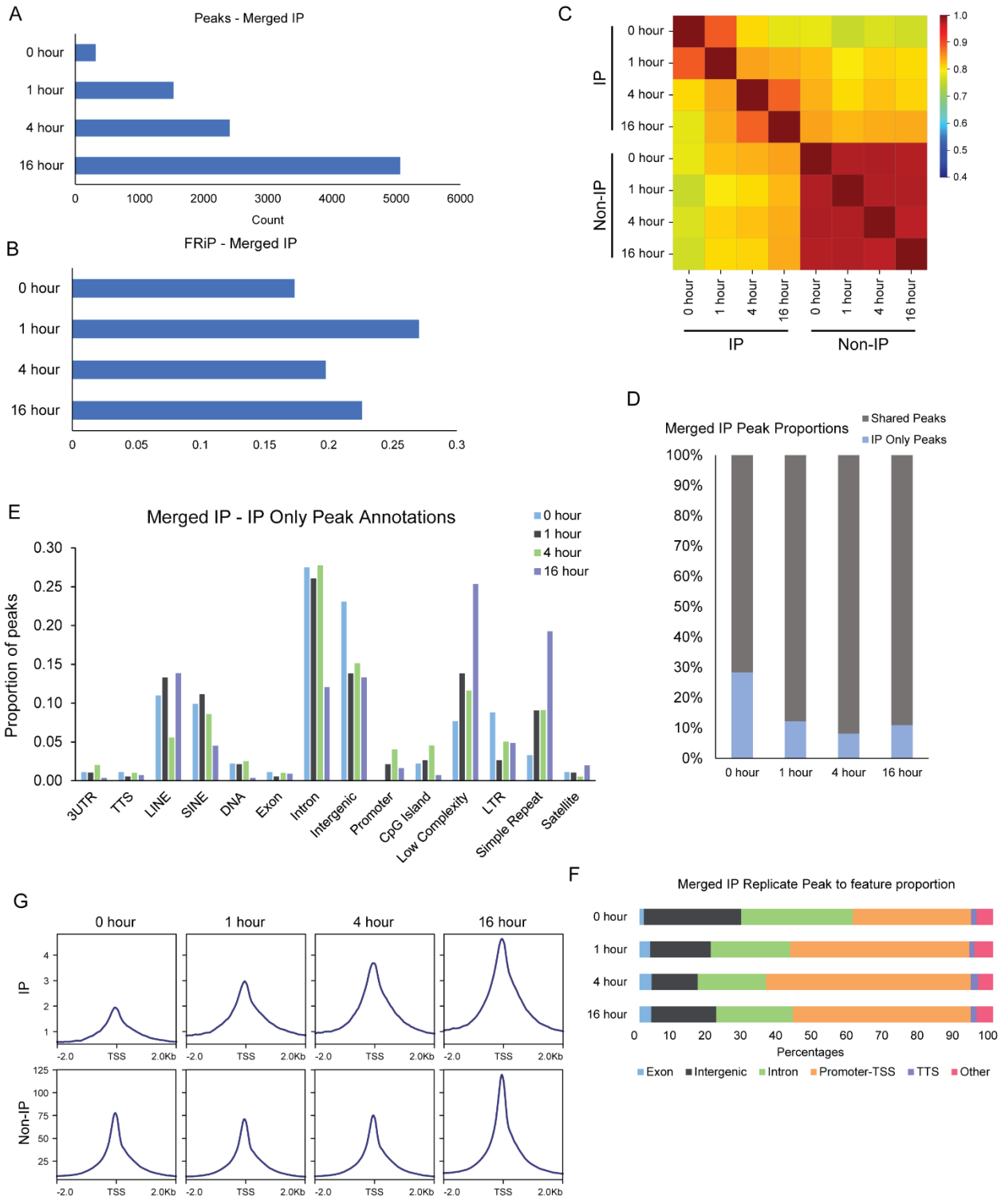


Figure 1.4. Merging replicates produces a more robust IP sample profile.

(A) Histogram depicting peak counts for merged Repli-ATAC-seq IP replicates at each time point.
(B) Histogram depicting fraction of reads in peaks for merged Repli-ATAC-seq IP replicates at each time point.
(C) Spearman correlation heatmap for merged IP replicates and merged Non-IP replicates at each time point.
(D) Stacked percentile bar plot with proportion of peaks unique in the merged Repli-ATAC-seq IP replicates and peaks shared between the merged IP replicates and merged Non-IP for each time point.
(E) Annotated genic region peaks from the unique merged Repli-ATAC-seq IP only peaks from (D) plotted as a proportion of all unique merged Repli-ATAC-seq IP only peaks for each replicate and time point.
(F) Stacked percentile bar plot with peak to feature proportion for exons, intergenic regions, introns, promoter-TSS, TTS, and all other for merged Repli-ATAC-seq IP replicates at each time point.
(G) TSS enrichment plots 2kbp up and downstream the TSS for each time point for merged replicate Repli-ATAC-seq IP and merged replicate Repli-ATAC-seq Non-IP.

1.2.5 *Repli-ATAC-seq reveals upstream accessibility in IP samples at pluripotency transcription factor binding sites and accessibility correlations with gene expression*

To understand how pluripotency maintenance may be affected in post-replication time, Repli-ATAC-seq fragments were plotted up- and downstream pluripotency TFs SOX2, OCT4, and NANOG (Figure1.6A, B, and C). Amazingly, there were striking differences between IP and Non-IP time points at all of these TFs. SOX2 binding sites showed a dramatic increase in fragment enrichment downstream of binding sites and was absent upstream of these sites and in the Non-IP samples (Figure 1.6A). OCT4 binding sites had a modest increase in fragment enrichment downstream of these binding sites in the IP sample (Figure 1.6B), which was mirrored in the NANOG IP (Figure 1.6C). These trends were absent in the Non-IP samples across all time points. This increase downstream of binding sites was present immediately post-replication, indicating that there may be transient binding of these TFs that is affecting accessibility at these pluripotency maintenance binding sites.

This observation is similar to another cell cycle TF-related dynamic, mitotic bookmarking, the process in which TFs stay bound to key regions of the genome during mitosis so that fast transcriptional activation occurs upon entry into G1[35,36]. When the nearest genes downstream of these binding sites were identified and gene ontology (GO) performed, sites that only had SOX2 binding sites enriched for genes involved in stem cell maintenance (Figure 1.6D), concordant with reports of a role for mitotic bookmarking of SOX2 in pluripotency and differentiation[37]. SOX2/OCT4 shared sites, OCT4 sites, and NANOG only sites lacked stem cell maintenance GO biological process terms (Figures 1.6E and F). To determine if this enrichment of downstream IP accessibility was a technical artefact from the Repli-ATAC-seq method or the computational processing, IP and Non-IP enrichment were plotted for 30 other TFs in H1-ESCs (Figures A.2, A.3, and A.4). An increase in downstream Non-IP accessibility was not detected for any TF other than BCL11A and SP2, and downstream accessibility in IP samples were detected for BCL11A, FOSL1, NRSF, P300, RXRA, SIX5, and SP2.

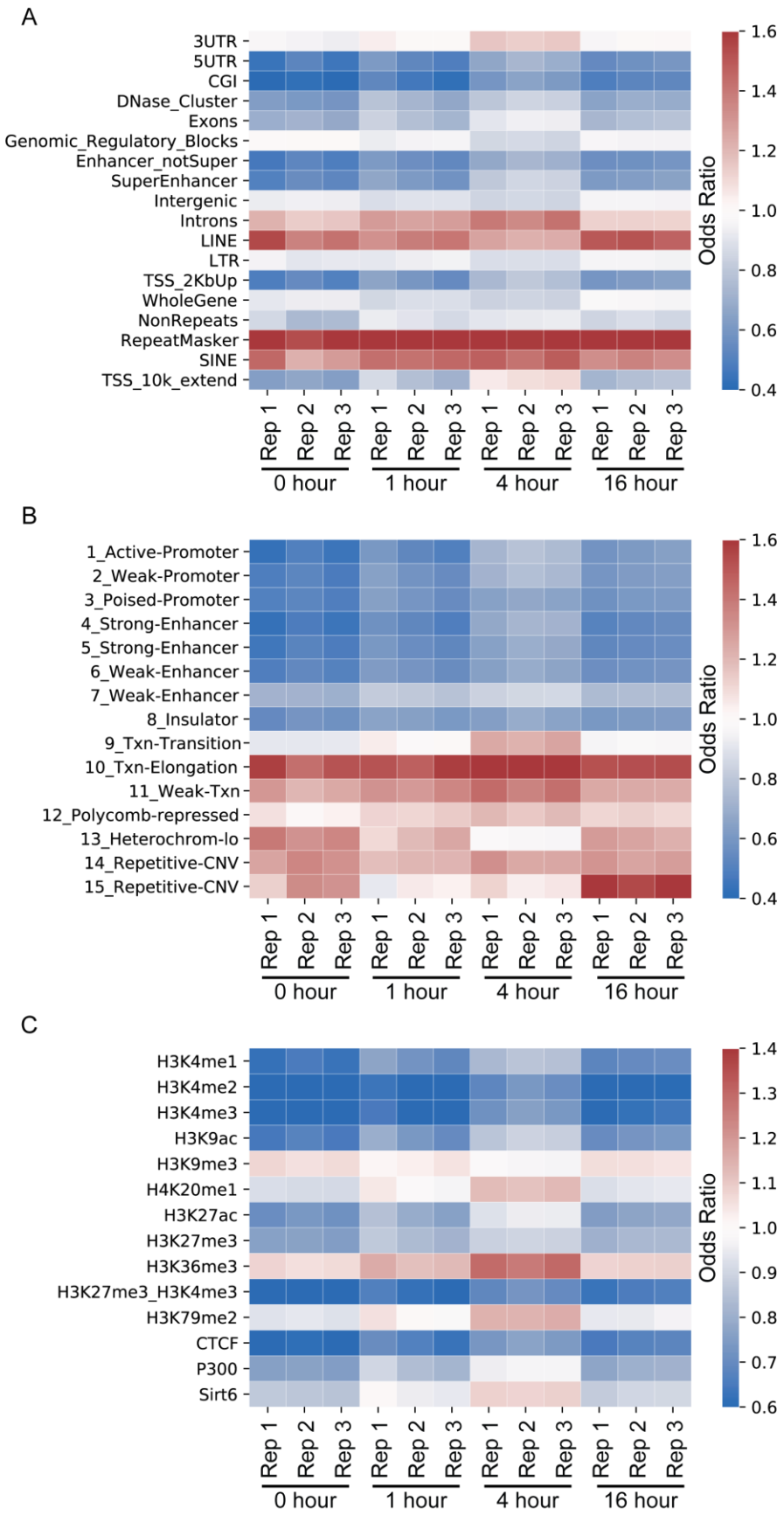

A
3UTR
5UTR
CGI
DNase_Cluster
Exons
Genomic_Regulatory_Blocks
Enhancer_notSuper
SuperEnhancer
Intergenic
Introns
LINE
LTR
TSS_2KbUp
WholeGene
NonRepeats
RepeatMasker
SINE
TSS_10k_extend
Odds Ratio
1.6
1.4
1.2
1.0
0.8
0.6
0.4
Rep 1
Rep 2
Rep 3
0 hour
1 hour
4 hour
16 hour
B
1_Active-Promoter
2_Weak-Promoter
3_Poised-Promoter
4_Strong-Enhancer
5_Strong-Enhancer
6_Weak-Enhancer
7_Weak-Enhancer
8_Insulator
9_Txn-Transition
10_Txn-Elongation
11_Weak-Txn
12_Polycomb-repressed
13_Heterochrom-lo
14_Repetitive-CNV
15_Repetitive-CNV
C
H3K4me1
H3K4me2
H3K4me3
H3K9ac
H3K9me3
H4K20me1
H3K27ac
H3K27me3
H3K36me3
H3K27me3_H3K4me3
H3K79me2
CTCF
P300
Sirt6
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6

Figure 1.5. Odds ratios reveal unique post-DNA replication-associated chromatin accessibility relationships with genomic features.

(A) Heatmap of odds ratios of fragments present in various genic regions for different replicates and time points. A value of 1.0 (white) indicates no difference between Repli-ATAC-seq IP and Non-IP sample, a value greater than 1.0 (red) indicates higher odds of the outcome occurring in the IP sample, and a value less than 1.0 (blue) indicates higher odds of the outcome occurring in the Repli-ATAC-seq Non-IP sample. 0.95 CI for all odds ratios < ± 0.05.

(B) Same as (A) but for ChromHMM annotated regions

(C) Same as (A) but for histone mark and TF ChIPs.

To determine if there was a general genome-wide correlation at TSS sites between gene expression and chromatin accessibility in post-replication time, LOESS plots of 2kbp up- and downstream TSS Repli-ATAC-seq accessibility scores were plotted against binned mean expression levels of a WT huESC dataset (Figure 1.6G). Mean expression and Repli-ATAC-seq accessibility were correlated, with high mean expression corresponding with the highest accessibility and low mean expression corresponding with the lowest accessibility at TSSs across all time points. Strikingly, the accessibility enrichment and separation between expression levels increased in post-replication time, indicating there is a correlation between gene expression levels and post-replication accessibility recovery.

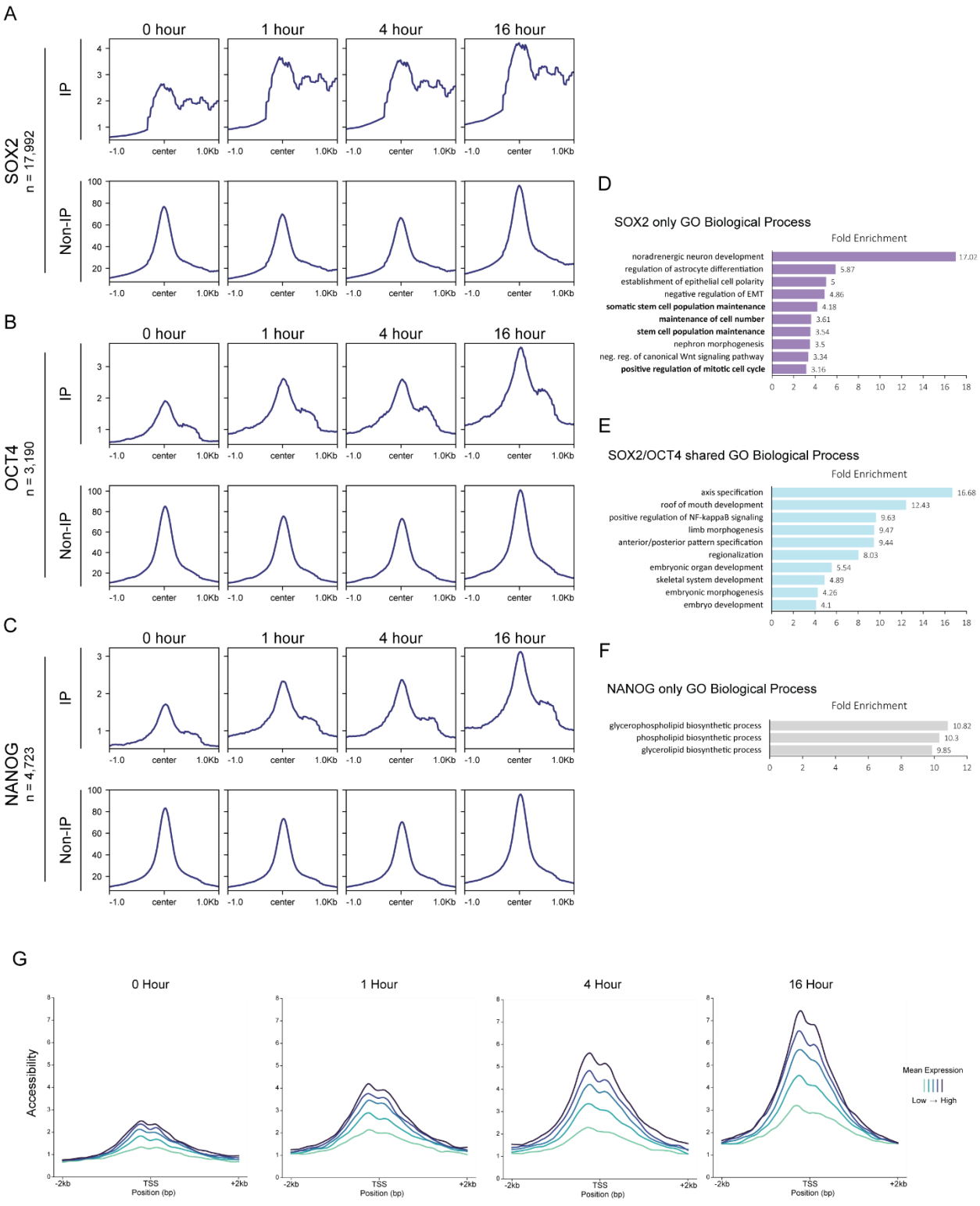

A
SOX2
n = 17,992
IP
Non-IP
0 hour
1 hour
4 hour
16 hour
-1.0
center
1.0Kb
B
OCT4
n = 3,190
C
NANOG
n = 4,723
D
SOX2 only GO Biological Process
Fold Enrichment
noradrenergic neuron development 17.02
regulation of astrocyte differentiation 5.87
establishment of epithelial cell polarity 5
negative regulation of EMT 4.86
somatic stem cell population maintenance 4.18
maintenance of cell number 3.61
stem cell population maintenance 3.54
nephron morphogenesis 3.5
neg. reg. of canonical Wnt signaling pathway 3.34
positive regulation of mitotic cell cycle 3.16
E
SOX2/OCT4 shared GO Biological Process
Fold Enrichment
axis specification 16.68
roof of mouth development 12.43
positive regulation of NF-kappaB signaling 9.63
limb morphogenesis 9.47
anterior/posterior pattern specification 9.44
regionalization 8.03
embryonic organ development 5.54
skeletal system development 4.89
embryonic morphogenesis 4.26
embryo development 4.1
F
NANOG only GO Biological Process
Fold Enrichment
glycerophospholipid biosynthetic process 10.82
phospholipid biosynthetic process 10.3
glycerolipid biosynthetic process 9.85
G
0 Hour
1 Hour
4 Hour
16 Hour
Accessibility
TSS
Position (bp)
-2kb
+2kb
Mean Expression
Low → High

Figure 1.6. Repli-ATAC-seq reveals upstream accessibility in IP samples at pluripotency transcription factor binding sites and accessibility correlations with gene expression.

(A) SOX2 binding region enrichment for Repli-ATAC-seq IP and Non-IP samples for each time point.
(B) Same as (A), but for OCT4.
(C) Same as (A), but for NANOG.
(D) 5kbp downstream unique SOX2 binding region Gene Ontology biological process terms fold enrichment. Bold terms correspond to stem cell maintenance processes.
(E) Same as (D), but for 5kbp downstream of shared SOX2/OCT4 binding regions.
(F) Same as (D), but for 5kbp downstream of NANOG binding regions.
(G) Accessibility plots at promoter-TSS regions of Repli-ATAC-seq IP time points for genes in 20 percentile bins of mean gene expression.

1.3 Discussion

Here, a novel assay, Repli-ATAC-seq, was developed producing a biological triplicate dataset of post-replication chromatin accessibility in hESCs. The first observation of this dataset was that upon comparing IP samples to Non-IP samples, chromatin condensation over a 16-hour post-replication time span was revealed. Although many of the quality control metrics, including spearman correlations between samples, TSS enrichment scores, peak counts, and FRiP scores were positive, an improvement of these metrics were seen upon merging the biological triplicate samples. Replicate merging lead to a spearman correlation organized through post-replication time and produced a more digestible peak to genic region enrichment. Removing subnucleosomal fragments had a minimal effect on these outcomes.

Another observation in the Repli-ATAC-seq dataset was that there is a large degree of variability in between IP and Non-IP, as well as between time points. IP fragments had a higher likelihood of falling within introns, LINEs, SINES, and repeats. Furthermore, ChromHMM regions with an expectedly low accessibility were more

likely to be accessible in the IP compared to the Non-IP. Taken together, these findings indicate there is a higher accessibility in regions canonically less accessible in bulk samples, indicating the presence of heterogeneity in post-replication time may have a distinct role in chromatin architecture. Remarkably, there was a higher likelihood of accessible fragments at DNA mismatch repair histone marks H4K20me1, H3K36me3, and H2K79me2 at 4 hours post-replication. Sirt6, another TF involved in DNA mismatch repair[38], also shared a higher likelihood of accessibility at 4 hours post-replication, indicating this time point may have a unique state in post-replication time.

The next observation involved identifying how accessibility is affected in post-replication time at pluripotency maintenance TFs SOX2, OCT4, and NANOG. SOX2 showed a high downstream accessibility in the IP sample that was lacking in the Non-IP. Albeit to a lesser extent, downstream accessibility from OCT4 and NANOG sites in the IP samples were also higher than Non-IP samples. When performing the same analysis with 30 other TFs, FOSL1, NRSF, P300, RXRA, and SIX5 showed downstream accessibility in the IP sample only, indicating there exists an inability for nucleosome compaction that happens after DNA is replicated in specific, TF-associated contexts. This type of comparison of IP and Non-IP allows for identification of asymmetry in the replicating genome upstream of TF binding sites.

Finally, correlating mean gene expression and accessibility in post-replication time at TSSs revealed the stunning revelation that post-replication accessibility increases with expression levels. This finding is critical to pursue, and, ideally, performing single-cell sequencing to identify the biological importance of this finding would prove fruitful.

1.4 Methods

Repli-ATAC-seq protocol for HUES64 cells

BrdU treating and preparing the cells for transposase reaction:

1) Aspirate media from well(s). Note: This protocol was originally done at 0, 1, 4, and 16 hour timepoints, so 4 wells from a 6-well were used for one set of biological replicates.
2) Add 2mL mTeSR media with 50 mM BrdU to plates with cells for 1 hour.
3) After 1 hour, aspirate the BrdU containing media and wash with 1mL non-BrdU containing mTeSR media two times.
4) a. Aspirate the media in the 0 hour time point well and add 1 mL Accutase. Place back in incubator for 10 minutes.
   b. For the non-0 hour time points, add 2mL of non-BrdU containing mTeSR and place back in the incubator.
5) Quench the 0 hour well with 1mL mTesR, collect the cells in a 15 mL conical tube, spin down at 200 xg for 5 mins at 4C, aspirate the supernatant, and resuspend in 1 mL cold PBS buffer.
6) Collect 100k cells in a 1.5 mL Eppendorf tube and spin down at 200 xg for 5 mins at 4C using a bucket rotor. Note: Using a non-bucket rotor will create more cell number loss than desirable.
7) Aspirate the supernatant and gently resuspend the cell pellet in 50 uL of cold lysis buffer. Spin down immediately at 500 xg for 10 mins at 4C.
8) Carefully discard supernatant and immediately continue to transposition reaction.

Transposition reaction:

1) Make sure the cell pellet is set on ice

2) Make the transposition reaction mix:

    a. 50 uL TD (2x reaction buffer)

    b. 5 uL TDE1 (Nextera Tn5 Transposase)

    c. 45 uL Nuclease Free H2O

3) Resuspend nuclei in the transposition reaction mix

4) Incubate the transposition reaction at 37C for 30 mins (Thermomixer 37C)

5) Immediately following transposition, purify using a Qiagen MinElute PCR Purification Kit

6) Elute transposed DNA in 20 uL Elution Buffer (10 mM Tris buffer, pH 8) – should be included in kit.

7) To amplify transposed DNA fragments, combine the following in a 0.2 ml PCR tube:

    a. 20 µl Transposed DNA

    b. 5 ul Nuclease Free H20

    c. 25 µl NEBNext High-Fidelity 2x PCR Master Mix

8) Thermal cycle for ONLY 1 cycle of 72C for 5 minutes. This first 5 minute extension at 72C is critical to allow extension of both ends of the primer after transposition, which generates amplifiable fragments.

9) Immediately following this extension step, purify using a Qiagen MinElute PCR Purification Kit and elute in 20 uL Elution Buffer.

10) This purified DNA can be kept at 4C until the 1 , 4, and 16 hour time points are brought to this step of the protocol as well. From this point forward, the different time point samples can be done in parallel.

BrdU Immunoprecipitation:

1) Take 10 uL of the sample (non-IP sample) and place at 4C for future steps. The other 10 uL will be used for BrdU immunoprecipitation (IP sample)
2) Add 490 uL 1x TE buffer to the IP sample and incubate the sample at 95C (done on Thermomixer) for 5 mins to heat-denature the DNA.
3) Cool the sample on ice-water for 2 mins
4) Add 60 uL of 10x IP buffer to a new 1.5 mL tube and add the denatured DNA to this tube.
5) Add 40 uL of 12.5 ug/mL anti-BrdU antibody (11.25 uL stock in 450 uL 1x IP – feel free to adjust this total amount based on how much you will end up using on the day so as to not waste antibody).
6) Incubate for 1 hour at room temp with constant rotation – cover tubes with box to keep them in the dark
7) Add 20 ug of rabbit anti-mouse IgG (10 ul of stock)
8) Incubate for 1 hour at room temperature with constant rotations – cover tubes with box to keep them in the dark
9) Centrifuge at 17,000xg for 5 mins at 4C (This will be done on a normal microcentrifuge, not the large swing bucket centrifuge)
10) Remove the supernatant completely. If the pellet becomes loose, then briefly centrifuge the sample again in order to completely remove the supernatant without disturbing the pellet. Several centrifugations may be necessary to completely remove the supernatant.
11) Add 750 uL ice cold 1x IP buffer and centrifuge at 17,000 xg for 5 mins at 4C.

12) Remove supernatant completely as in step 10

13) Resuspend the pellet in 200 uL digestion buffer with freshly added 0.25 mg/ml proteinase K. Incubate overnight at 37C (Thermomixer)

14) Add 100 ul fresh digestion buffer with freshly added 0.25 mg/ml proteinase K.

15) Incubate samples for 60 mins at 56C

16) Purify the sample using AMPure XP Beads (2x bead volume, 300 uL in this case) and resuspend in 20 uL Elution Buffer

PCR Amplification: (This section is taken from Buenrostro et al 2016)

1) To amplify transposed DNA fragments, combine the following in a 0.2 ml PCR tube:
    a. 20 µl Transposed DNA
    b. 2.5 µl 25 µM Custom Nextera PCR Primer 1
    c. 2.5 µl 25 µM Custom Nextera PCR Primer 2 (Contains Barcode)
    d. 25 µl NEBNext High-Fidelity 2x PCR Master Mix
2) Thermal cycle as follows:
    a. 1 cycle of 72°C for 5 min, 98°C for 30 sec
    b. 5 cycles of 98°C for 10 sec, 63°C for 30 sec, 72°C for 1 min
3) To reduce GC and size bias in PCR, the appropriate number of PCR cycles is determined using qPCR allowing us to stop amplification prior to saturation. To run a qPCR side reaction, combine the following in qPCR compatible consumables:
    a. 5 µl of previously PCR amplified DNA
    b. 4.41 µl Nuclease Free H_2O

c. 0.25 µl 25 µM Customized Nextera PCR Primer 1
d. 0.25 µl 25 µM Customized Nextera PCR Primer 2
e. 0.09 µl 100x SYBR Green I
f. 5 µl NEBNext High-Fidelity 2x PCR Master Mix

4) Using a qPCR instrument, cycle as follows:
    a. 1 cycle of 98°C for 30 sec
    b. 20 cycles of 98°C for 10 sec, 63°C for 30 sec, 72°C for 1 min
5) To calculate the additional number of cycles needed, plot linear Rn versus cycle and determine the cycle number that corresponds to ¼ of maximum fluorescent intensity.
6) Run the remaining 45 µl PCR reaction to the cycle number determined by qPCR. Cycle as follows:
    a. 1 cycle of 98°C for 30 sec
    b. N cycles of 98°C for 10 sec, 63°C for 30 sec, 72°C for 1 min
        i. Cycle for an additional N cycles, where N is determined using qPCR.
7) Purify amplified library using Qiagen MinElute PCR Purification Kit. Elute the purified library in 20 µl Elution Buffer (10 mM Tris Buffer, pH 8). Be sure to dry the column before adding elution buffer.
8) Perform Qubit and tapestation HSD5000.

Materials:

mTesR1 media, BrdU (BD Biosciences), Phosphate Buffered Saline (PBS) Molecular biology grade IGEPAL CA-630, Proteinase K, 1x TE buffer, 2x TD (2x reaction buffer, Illumina Cat #FC-121-1030), TDE1 (Nextera Tn5 Transposase,

Illumina Cat #FC-121-1030), Qiagen MinElute PCR Purification Kit, Elution Buffer (10 mM Tris buffer, pH 8), NEBNext High-Fidelity 2x PCR Master Mix (New England Labs Cat #M0541), 25 uM Custom Nextera PCR Primer 1, 25 uM Custom Nextera PCR Primer 2, 100x SYBR Green I (Invitrogen Cat #S-7563), 0.2-ml PCR tubes, 1.5 ml Eppendorf tubes, 15 ml conical tubes, PCR Thermal cycler, anti-BrdU antibody (BD Pharmagen), rabbit anti-mouse IgG (BD Biosciences), Ampure XP beads (Beckman Coulter)

Lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% IGEPAL CA-630)

Digestion Buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, and 0.5% SDS): To prepare 50 ml, combine 44 ml of autoclaved ddH2O, 2.5 ml of 1 M Tris-HCl (pH8.0), 1 ml of 0.5 M EDTA and 2.5 ml of 10% (wt/vol) SDS. Store at room temperature.

IP buffer (10×): To prepare 50 ml, combine 28.5 ml of ddH2O, 5 ml of 1 M sodium phosphate (pH 7.0), 14 ml of 5 M NaCl and 2.5 ml of 10% (wt/vol) Triton X-100. Store at room temperature.

IP buffer (1×): To prepare 50 ml, add 5 ml of 10× IP buffer to 45 ml of autoclaved ddH2O. Store at room temperature.

ATAC-seq processing pipeline

FASTQ files were trimmed using Trim Galore! and files were aligned using BWA. Duplicates were marked using picard and filtering was done using SAMtools to remove mitochondrial DNA, duplicate reads, and unmapped reads. Further filtering was performed using BAMtools, removing reads containing greater than four mismatches and reads with an insert size greater than 2kb. Peak calling was

performed using MACS2 using the –nomodel parameter. BEDtools intersect was used for detecting genomic overlaps with peaks or with fragments.

Insert size calculations and filtering

Picard was used to quantify insert sizes using the *CollectInsertSizeMetrics* function with M=0.5 for each bam file. The histogram counts were merged across replicates and plotted as a density of all reads in the sample as normalization. Insert sizes were filtered for less than 130 bp using awk 'substr($0,1,1)=="@" || ($9>= 130) || ($9<=-130)' on BAM files.

Correlation heatmaps

bigWig files were produced from BAM files using deepTools *bamCoverage* function. Next, deepTools *multiBigwigSummary bins* was used to create an array with an average score across 10kbp bins and a spearman correlation was calculated using deepTools *plotCorrelation*. Heatmap plotting was performed using Seaborn in Python3.

Enrichment score calculations

deepTools *computeMatrix* function was used with bigWig files of the Repli-ATAC samples and BED files for regions of interest. Enrichment Score is calculated by taking the average read depth in the 100 bps at each flanking region of the distribution and calculating a fold change at each position over that average read depth. --referencePoint was set to center and enrichment panels were created using deepTools *plotHeatmap* function. H1-hESC TF ChIPs were found from the ENCODE project under GSE32465.

Odds ratio calculations

Odds ratios were calculated by taking the number of fragments inside the feature and in the IP sample "A", the number of fragments outside the feature and in the IP sample "B", the number of fragments inside the feature and in the Non-IP sample "C", and the number of fragments outside the feature and in the Non-IP sample "D". These numbers were then taken to create an odds ratio using the equation: A*D/B*C. The values were organized in an array and turned into a heatmap using Seaborn in Python3.

Gene Ontology biological process identification

Genomic Regions Enrichment of Annotations Tool (GREAT) was used with corresponding TF ChIP-seq BED files to find gene associations 5kbp downstream of TF binding sites. These genes were then used with PANTHER to identify enriched gene ontology biological process terms.

Accessibility plots for mean gene expression around TSS

Mean expression values were calculated using a publicly available WT ESC dataset (SRP299892) and processed in Seurat. Genes were binned in 20 percentile rankings from low to high mean expression. A 4 kbp window around the TSS for each gene was taken and accessibility from bigwig files were cut into single base pair values at these sites. These were averaged across each expression level and loess was used to smooth the values.

SECTION 2: Understanding 10X Single Cell Multiome ATAC and Gene Expresssion Sequencing in human samples

2.1 Introduction

Understanding the complex biology of human cells requires the integration of genomic, epigenomic, transcriptomic, and proteomic data, oftentimes referred to as multiome sequencing. A specific type of assay, 10X Single Cell Multiome ATAC + GEX sequencing, leverages droplet based single-cell sequencing to simultaneously profile chromatin accessibility and gene expression in the same single cell. Here we present a summary of papers of this specific modality of multiome sequencing from human samples, given that the limitations of human samples make it difficult to produce studies with numerous cell types and mutations compared to other model organisms such as *mus musculus*. One of the papers analyzed the central nervous system and found that certain genes persist in a primed accessibility state across multiple cell types, ready for cues to activate gene expression by other chromatin-related mechanisms.[39] Another paper explored skin fibroblasts from patients with cutaneous systemic sclerosis and identified AP-1 circuitry as a key regulator of healthy to pathogenic sclerosis-associated fibroblasts.[40] Separate single-cell ATAC-seq and RNA-seq experiments performed on cerebral cortex samples revealed that integration of the datasets accurately reflects gene activity with distal gene cis-regulatory element accessibility linking to gene expression, which was verified by multiome sequencing.[41] Additionally, chromosomal rearrangements in lineage-ambiguous leukemias causing deregulation of a master transcription factor, BCL11B, were identified by analyzing accessibility compared to normal hematopoietic chromatin signatures.[42] The results implicated BCL11B as a progenitor cell fate driver. Other

papers analyzed cell types such as hematopoietic stem and progenitor cells, cancer-associated fibroblasts, breast tissue, and oocytes. The data revealed cell-type-specific transcription factor regulatory circuitry, where differential chromatin accessibility signals distinguished cell subsets between healthy and diseased patients.

2.2 Summaries of Findings

2.2.1 *Multiome in the Brain and Nervous System*

Meijer et al. performed multiome sequencing on gray matter from two healthy individuals and cell types of the central nervous system (CNS) were identified.[39] They were able to identify astrocytes, ENDO, ESTRO, excitatory neurons, inhibitory neurons, microglia and related cells (MIGL), oligodendrocytes, and oligodendrocyte precursor cells (OPCs). They found that the promoters of five MHC-I pathway genes were accessible not only in MIGL, but also in OPCs, mature oligodendrocytes, astrocytes, inhibitory neurons, and excitatory neurons. Despite the accessibility at these promoters, expression of these genes lacked concordance with accessibility. Two of five of those genes were only expressed in MIGL and at very low levels, while one of five of those genes was expressed in multiple cell types, leaving two of five genes unexpressed. MHC-II pathway genes showed higher specificity, with only MIGL, OPCs, and excitatory neurons having accessible chromatin at the promotor for *HLA-DMB*, and expression only at MIGL. Also, for *HLA-DQB1* and *HLA-DRB1*, MIGL and neurons showed accessibility at the promoter but expression of the respective genes only in MIGL. Next, the researchers investigated whether or not multiple sclerosis susceptibility SNPs overlapped with chromatin accessibility and gene expression. The group found that the promoter and SNP for *HLA-B* was accessible in all identified cell types. Transcript levels were high in OPC and MIGL only however.

The same promoter and SNP accessibility profile was found for *ZHX3* but transcript levels were high for all cell types except for MIGL. The researchers concluded that certain genes persist in a primed accessibility state, ready for cues to activate gene expression by other chromatin-related mechanisms.

Montefiori et. al investigated lineage-ambiguous leukemias and identified chromosomal rearrangements causing deregulation of the expression of a master transcription factor, *BCL11B*.[42] Multiome sequencing was performed on two *BCL11B* group samples and accessibility was compared to different normal hematopoietic chromatin signatures. Both samples were highly enriched for long-term hematopoietic stem and progenitor cell (HSPC) signatures and for activated HSPC signatures. Furthermore, *BCL11B* expression correlated with this chromatin accessibility. When enriching for BCL11B binding using ChromVAR, activated HSPC signatures were highest in the BCL11B leukemia samples, implicating BCL11B as a progenitor cell fate driver in this leukemia subtype.

To understand the gene regulatory networks (GRNs) involved in human brain development, Fleck et al. utilized multiome sequencing and produced a GRN called Pando to assess transcription, chromatin accessibility, TF binding, and *cis*-regulatory elements (CREs) of nine human iPSC and ESC lines along with a GLI3 KO line across brain organoid differentiation.[43] Initially, CCA was performed on singleome scRNA-seq and scATAC-seq samples across differentiation time, and the integration of this CCA was validated by measuring a strong correlation to a multiome dataset of the same time point. As organoid generation has been reported to be variable depending on cell lines, the group gauged the self-patterning variation across the nine stem cell lines using single-cell multiome sequencing. They detected transcriptional clusters

that organized across an anterior to posterior axis, and these markers were consistent between the numerous cell lines. The proportion of cells in clusters varied between the lines however, and this similarity extended to CREs of patterning-related genes. Although there is variation in patterning between the cell lines, they posited a conserved GRN involved in brain patterning. Next, the group investigated the developmental context of a previously reported cell fate commitment TF, GLI3. The group produced *GLI3* KO lines and performed multiome sequencing on 3 week old brain organoids. The analysis revealed comparable cell clustering profiles between the KO and WT cells, however, there was both differential expression and accessibility in the telencephalic progenitor population. These genes included *HES1*, *HES4*, *HES5*, and *EMX2*. The group was able to elucidate GRNs involved in cell fate specification of developing brain organoids and validate a role for GLI3 in telencephalic cell-fate specification.

Ma et al. performed multiome sequencing on human dorsolateral prefrontal cortex (dlPFC) of five donors to understand regulatory mechanisms in human-specific evolutionary differences of *FOXP2* expression.[44] From scRNA-seq comparisons between species, *FOXP2* microglia-expressed enrichment was higher in human samples and of particular interest, as mutations of *FOXP2* are present in developmental verbal dyspraxia and various neuropsychiatric disorders. The analysis characterized CREs interacting with and proximal to *FOXP2* in microglia and intratelencephalic excitatory neurons. GRNs highlighted NFIA, a *FOXP2* target, as having a regulatory role in non-intratelencephalic excitatory neurons. Furthermore, the group identified the *FOXP2* regulon as the top regulon specific to

intratelencephalic neurons. Taken together, this indicated FOXP2 as a critical regulator for intratelencephalic excitatory neurons.

2.2.2 *Multiome in Cancer and the Immune System*

Another research group characterized tumor infiltrating lymphocyte (TIL) heterogeneity by first identifying five major groups from scRNA-seq from four different high grade serous ovarian carcinomas.[45] These groups were validated through a multiomic approach, showing concomitant nuclear gene expression and accessible chromatin at the loci for the genes identifying each of these five groups. This approach unveiled numerous co-regulated genes in these groups, including genes associated with T cell stemness, T cell effector activity, T cell proliferation, and T cell exhaustion. Furthermore, novel markers were identified in specific tissue resident memory (TRM) subsets. An increase of *ETV1* during differentiation of TRM effector to TRM proliferative to TRM exhausted cells was identified, and *GZMK* expression levels were at their peak in TRM stem-like cells and decreased throughout the differentiation process to exhaustion. Finally, an enrichment of zinc finger domains in TRM stem-like cells was identified, while POU domains were enriched as differentiation progressed, with FOX and HOX binding domains showing higher accessibility in TRM exhausted cells.

Boukhaled et al. performed multiome sequencing on two healthy, four PBMC melanoma patient samples selected for high IFN-I response capacity (IRC), and four PBMC melanoma patient samples selected for low IRC.[46] Clustering was performed and CD4+ T effector cells were analyzed further. Similarity fusion analysis indicated a high similarity in chromatin accessibility for high IRC patients, while both sequencing modalities had high similarity in the low IRC population. This indicated

that IRC groups were defined by more than IFN-I response, and gene expression and chromatin accessibility profiles define these groups at the basal state. Gene set enrichment analysis (GSEA) showed enrichment of pathways correlated to CD4 T cell dysfunction in the high IRC group, while enrichment of genes associated with T cell memory formation was identified in the low IRC group. Next, SCENIC was used to compare TF binding to regulons. TF networks associated with T follicular helper and T cell dysfunction were identified in the high IRC population. Interestingly, FLI1 and RUNX1 were highlighted as top networks activated in this population, which have been correlated with CD8 T cell exhaustion and restrict CD8 T cell memory differentiation. In the low IRC population, AP-1 TF regulons, such as FOS, JUN, and TAF7, were highlighted. Ingenuity pathway analysis (IPA) identified NFATC2 as a regulator of gene expression in the high IRC population. Expression of NFAT genes, *FLI1*, and *RUNX1* were increased in the high IRC group, while *JUN* and *FOS* expression was higher in the low IRC population, with a decrease in NFAT family genes. This indicated that a divergent gene expression program exists upon T cell receptor binding, with IFN-I hyper-responsiveness associating with a dysfunctional T cell state. Network analysis showed enhancement of genes downstream of T cell activation, migration, and cell adhesion in the high IRC population, which has also been shown in exhausted T cells. The high IRC population also showed upregulation of epigenetic pathways compared to low IRC. These genes encode for enzymes involved in transcriptional repression, transcriptional activation, chromatin stability, and T cell exhaustion enzymes such as EZH2 and DNMT3A. Differentially accessible region (DAR) analysis showed a higher number of accessible regions in the low IRC group compared to the high IRC group, indicating an increased transcriptional potential in

the low IRC population and a restriction of chromatin accessibility and transcriptional repression in the low IRC group. Genomic Regions Enrichment Analysis Tool (GREAT) was performed and pathways were enriched for immune regulation and inflammation in both IRC groups compared to healthy controls. The high IRC population showed a reduction in pathways involved in negative regulation of IFN-I, pointing to epigenetic repression of IFN-I signaling inhibition. There was a distinct lack of expression of IFN-I stimulated genes (ISGs), so the group posited that responsiveness to IFN-I may be a primed epigenetic state from interactions prior to IFN-I exposure. ISGs were similarly accessible between the high and low IRC groups, however, accessibility of STAT and IRF binding motifs that activate ISG expression were enriched in the high IRC group, with T cell exhaustion TF binding motif enrichment also presenting in the high IRC group. These were absent in the low IRC group, and the research group concluded that IFN-I responsiveness between the groups are differentially poised states that exist.

Another group showed that multiome sequencing of a CD34+ population with a large transcriptional flux recapitulated a B-lymphoid/plasmacytoid dendritic cells bifurcation at DARs for these lines that were accessible at regions for both of these cell types.[47] This indicated a large fluctuation in gene expression dynamics is accompanied by accessibility at regions that are accessible separately between the two cell types.

Ors et al. developed Topic Inference of Transcriptionally Associated Networks (TITAN) for multiomic analysis of the heterogeneity of estrogen response across time in estrogen receptor alpha (ER) positive breast cancer cell lines.[48] After performing scRNA-seq on a number of lines, single-cell multiome sequencing was performed on

MCF-7 and T-47D cells to understand if changes in chromatin states were responsible for changes in transcription upon 48h estrogen exposure in these lines. ESR1 and FOXM1 signatures were highlighted as different groups in both modalities of the multiome analysis, correlating differentially accessible regions (DARs) to these RNA topics. These signatures were recapitulated using DEG, TF enrichment, and TF accessibility analysis, indicating a divergent pathway in breast cancer development in response to estrogen. ATAC profiles of highly expressed genes indicated differences in TSS, gene bodies, and TTS accessibility, confirming chromatin state differences in the two groups.

Another group investigated the gene expression and chromatin accessibility profiles of cancer associated fibroblasts (CAFs) in multiple solid tumor types.[49] They initially identified three populations for CAFs in mouse tumors that had distinct gene expression and chromatin accessibility profiles: steady state-like, mechanoresponsive, and immunuomodulatory. To relate these findings in human cell lines, multiomic analysis was performed on three human breast tumors. CAF clusters were recapitulated in the human cell lines, confirming that these CAF subpopulations are conserved across species.

2.2.3 *Multiome in Other Cell Types*

Gur et al. performed multiome sequencing on skin fibroblasts from nine face and diffuse cutaneous systemic sclerosis (dSSc) and nine healthy patients.[40] Label transferring for cells was done from a prior RNA cellular reference map, and single-cell clustering was performed based on these labels. Clustering of the ATAC signal reproduced these results, indicating chromatin accessibility signals exist defining these subsets. Differential peak analysis and motif enrichment analysis unveiled cell

subset-type specific transcription factor regulatory circuitry. There were limited differences at promoters, but distal chromatin accessible regions showed significant differences between healthy and diseased patients. These differentially accessible distal regions enriched for genes identified by the group as perturbed in the sclerosis-associated fibroblasts (ScAFs): *LGR5*, *CDKN2A*, and *SLPI/MATN4*. There was enrichment of DNA motifs for ATF3, JUN, and CREB1 in ScAFs and a reduction in enrichment for TWIST1, HAND2, MCI1, NEUROD1, PRRX1, and PRRX2 motifs. Furthermore, there were enriched motifs shared in ScAFs and other SSc fibroblast subsets indicating global TF binding perturbations. The researchers concluded that they have identified AP-1 circuitry as a key regulator of healthy to pathogenic ScAFs.

Finally, Wang et al. performed multiome sequencing on eight post-mortem retinas from four healthy human subjects.[50] The group produced predicted target genes by searching for SNPs in accessibility peaks that correlated with expression of a nearby gene. This method produced a 32.1% hit rate of ATAC peaks and 202 SNPs were identified in promoter peaks that had predicted target genes. One particular SNP, rs4821699, was found to be most accessible in retinal ganglion cells, and the peak with this SNP was predicted to target *TRIOBP*, a gene implicated in glaucoma. This produced a novel hypothesis that rs48221699 mutant *TRIOBP* expression in retinal ganglion cells could play a role in glaucoma. Another SNP, rs17421627, is implicated in Macular Telangiectasia and had a predicted target gene with SNP accessibility Müller glia and astrocytes. Feature linkage connected this SNP to LINC00461, and deletion of the loci where this SNP resides downregulates

LINC00461. Taken together, this multiome resource provides insight into disease mechanisms of noncoding risk variants in the human retina.

2.3 Conclusion

Multiome sequencing is essential to understanding human cellular biology, and analyzing the chromatin accessibility of specific cell types can yield insights into their functional potential. Certain multiome analyses on cell lines recapitulate transcriptomic information whereas others do not, and chromatin accessibility does not always correlate with gene expression, but transcription factor binding motifs seem to provide more insight. Furthermore, correlations with SNP sites is a worthwhile analysis method and can produce testable, novel hypotheses. The future of multiome sequencing research will likely focus on uncovering novel cellular regulatory networks and identifying potential therapeutic targets.

SECTION 3: S-phase fractioned 10X Multiome sequencing reveals a neuroectoderm-primed cell subset present unique to DNA replication

3.1 Introduction

Changes in gene expression are central to cellular processes, and these are often intimately associated with alterations in chromatin accessibility[51]. However, our understanding of these dynamics, particularly during the S-phase of the cell cycle when DNA replication occurs, remains incomplete[52]. The S-phase represents a unique stage where the chromatin landscape requires recovery to a steady state, potentially allowing for shifts in gene expression profiles and, consequently, cell fate decisions[53,54]. Despite the significance of this phase, the correlation between gene expression changes and chromatin accessibility during S-phase is not well characterized, owing in part to technical challenges. However, advancements in technology and the development of novel experimental approaches hold promise in unraveling these complexities. Techniques such as scRNA-seq and scATAC-seq enabled the exploration of cell-to-cell variability and chromatin accessibility on an unprecedented scale. Such methodologies can offer novel insights into the dynamics of gene regulation during S-phase, particularly in stem cells where cell fate decisions are fundamental. A recent method, 10X Multiome ATAC + Gene Expression integrates both of these single-cell modalities by assaying gene expression and chromatin accessibility in the same single-cell. This powerful tool can provide novel insights into the dynamics of gene regulation during S-phase when paired with S-phase sorting and fractionation, unveiling clues about cell fate decision making during this integral phase of the cell cycle.

3.2 Results

3.2.1 *Optimization of Single Cell 10X Multiome ATAC + Gene Expression protocol with S-phase fractionation*

Performing an established protocol and introducing large perturbations to the workflow requires identification of the optimal method for preserving the state of the cell input into the final phases of the protocol. This process involves choosing metrics to be able to understand whether protocol changes are producing a meaningful improvement over prior iterations. Running at least two iterations allows for the comparative analysis of outcomes and enables ensuring the validity of subsequent analyses. The process entails isolating the nuclei of human ESCs, staining with the DNA stain 7-Aminoactinomycin D (7AAD), performing FACS on 4 separate S-phase fractions, and performing Tn5 transposase treatment and ATAC/GEX library preparation (Figure 3.1C).

The first replicate of the 10X Multiome sequencing involved initially using a time-course for nuclei isolation and testing for percent viability for each sample. This time-course with nuclei lysis buffer ranged from 4-10 minutes at 1 minute intervals, and samples were checked using a cell Countess system with nuclei-cell lysates stained with Trypan Blue. Upon reaching 5% viability at 7 minutes, there were diminishing returns at the 8, 9, and 10 minute mark. Having low cell viability is important, because 7AAD requires either cell permeabilization or nuclei isolation to be able to access and intercalate DNA, and Trypan Blue stains lysed cells. Above 5% viability, many cells would still have intact cell membranes, which would reduce FACS efficiency downstream in the protocol.

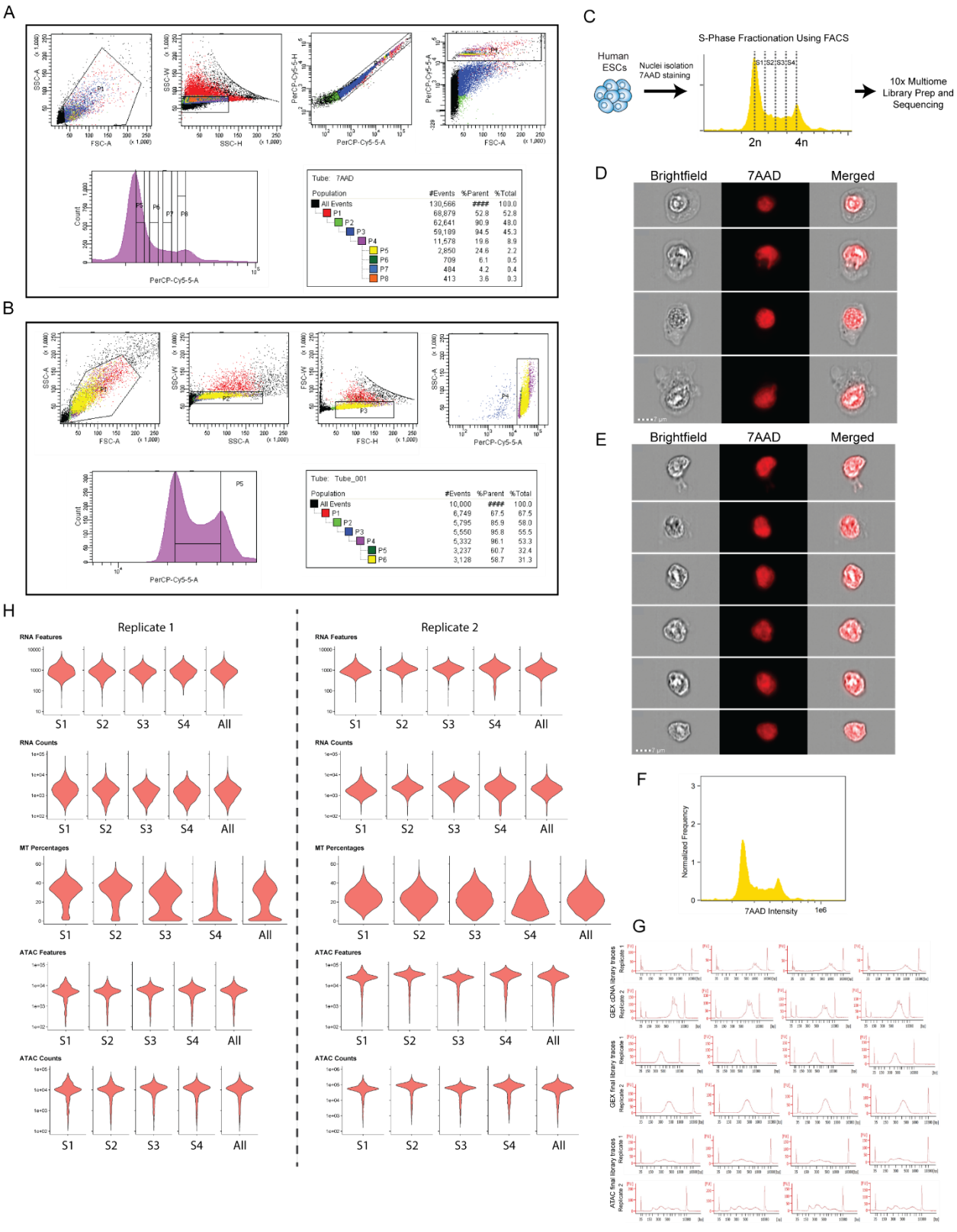

A
B
C
S-Phase Fractionation Using FACS
Human ESCs
Nuclei isolation 7AAD staining
10x Multiome Library Prep and Sequencing
2n
4n
D
Brightfield
7AAD
Merged
E
F
Normalized Frequency
7AAD Intensity
G
H
Replicate 1
Replicate 2
RNA Features
RNA Counts
MT Percentages
ATAC Features
ATAC Counts
S1
S2
S3
S4
All

Figure 3.1. 10x Multiome S-phase fractionation replicates 1 and 2 method optimization reveals differences in sorting outcomes and library quality.

(A) Replicate 1 FACS plots with total cell counts in each gate
(B) Replicate 2 FACS plots with total cell counts in each gate
(C) S-phase fractionation strategy schematic
(D) ImageStream images of underlysed cells
(E) ImageStream images of lysed cells
(F) ImageStream 7AAD intensity readouts from nuclei input
(G) BioAnalyzer traces of 10X Multiome libraries for replicates 1 and 2. Top – GEX cDNA library traces. Middle – GEX final library traces. Bottom – ATAC final library traces.
(H) Violin plots of RNA features, RNA counts, mitochondrial percentages, ATAC features, and ATAC counts for S-phase fractioned samples of replicates 1 and 2.

Using cells with 5% viability treated with 7AAD, FACS was performed (Figure 3.1A) and cells were isolated for 4 different S-phase fractions: S1, S2, S3, and S4 (Figure 3.1A bottom left, and Figure 3.1C, middle). The percentage of cells that were 7AAD positive that fell into the cell cycle curve was 8.9% of the total population. At this point, it was unknown whether or not this metric was optimal or not, so the nuclei were moved forward in the protocol and 10X Multiome library preparation and sequencing was performed. The number of cells collected for each of the 4 phases ranged from 1773-2907 cells (Table A.3.1). The first quality control metric that indicated there may have been issues with this first replicate was the ATAC TSS enrichment score, a common ATAC-seq control metrics, which ranged from 3.63-4.05. A TSS enrichment score of 5 is considered optimal, however, given this was a new protocol and that these fragments were transitioning from 2n to 4n copy number, it was unclear whether or not these values indicated a deficiency in the dataset.

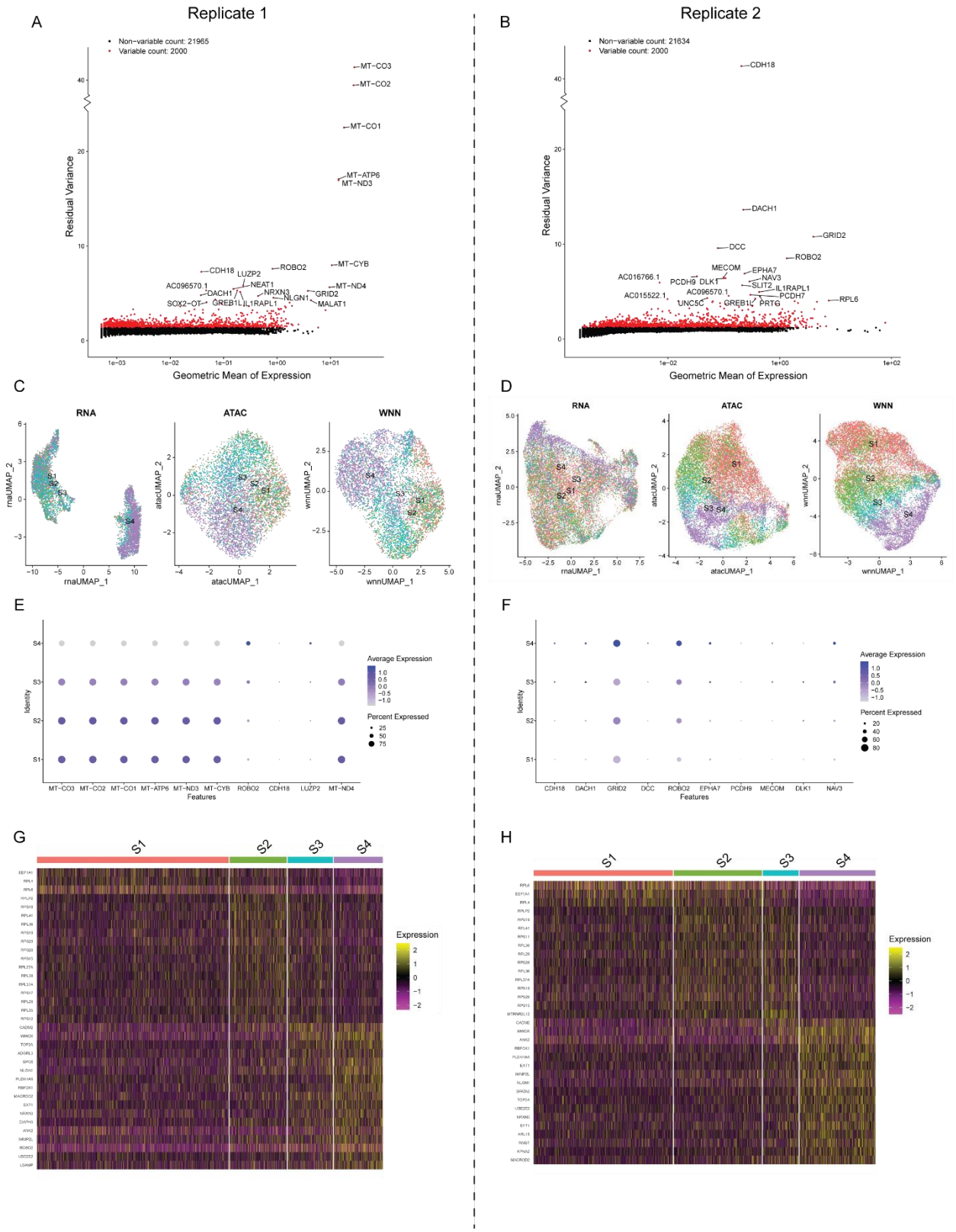

A
Replicate 1
Non−variable count: 21965
Variable count: 2000
Residual Variance
Geometric Mean of Expression
MT−CO3
MT−CO2
MT−CO1
MT−ATP6
MT−ND3
MT−CYB
ROBO2
CDH18
LUZP2
NEAT1
AC096570.1
MT−ND4
DACH1
NRXN3
GRID2
NLGN1
SOX2−OT
GREB1L
LIL1RAPL1
MALAT1
B
Replicate 2
Non−variable count: 21634
Variable count: 2000
CDH18
DACH1
GRID2
DCC
ROBO2
MECOM
EPHA7
NAV3
AC016766.1
PCDH9
DLK1
SLIT2
IL1RAPL1
AC096570.1
PCDH7
AC015522.1
UNC5C
GREB1L
PRTG
RPL6
C
RNA
ATAC
WNN
rnaUMAP_1
rnaUMAP_2
atacUMAP_1
atacUMAP_2
wnnUMAP_1
wnnUMAP_2
S1
S2
S3
S4
D
E
Identity
Features
Average Expression
Percent Expressed
MT−CO3
MT−CO2
MT−CO1
MT−ATP6
MT−ND3
MT−CYB
ROBO2
CDH18
LUZP2
MT−ND4
F
CDH18
DACH1
GRID2
DCC
ROBO2
EPHA7
PCDH9
MECOM
DLK1
NAV3
G
H
Expression

Figure 3.2. S-phase fractioned replicates show a difference in highly variable features at mitochondrial DNA.

(A) Variable feature plot for replicate 1. Top 2000 genes highlighted in red and annotated with the top 20 variable genes.
(B) S-phase fraction labeled RNA, ATAC, and WNN UMAPs for replicate 1.
(C) S-phase fraction labeled RNA, ATAC, and WNN UMAPs for replicate 2.
(D) Dot plot of the top 10 variable genes in replicate 1 labeled by S-phase identity.
(E) Dot plot of the top 10 variable genes in replicate 2 labeled by S-phase identity.
(F) Heatmap of the gene expression of top 20 marker genes in each S-phase fraction for replicate 1.
(G) Heatmap of the gene expression of top 20 marker genes in each S-phase fraction for replicate 2.

For replicate 2, a different method was used to optimize cell lysis and nuclei isolation. Cell Countess and Trypan Blue staining may give false positive hits by staining debris aggregates, and visually detecting the difference between debris and nuclei is a limitation of the Countess the machine. Instead, a multi-checkpoint method was used to detect nuclei lysis efficiency. First, the time-course was repeated, but instead of staining, nuclei were placed on a microscope slide and visually assessed at 60X magnification. At this resolution, nuclei morphology and cell membranes were optically detectable. Interestingly, at 4 minutes, there was not only complete cell membrane lysis, but also complete nuclei lysis, indicating over-lysis was occurring. Bringing the lysis time down to 1.5 minutes yielded promising results, where nuclei membranes were still intact while most cells had membrane lysis. The next checkpoint was using the same input sample that was to be used in FACS. A small amount of the FACS input was taken and used on ImageStream (Figures 3.1D and E). ImageStream provided a comprehensive look at the various states of the nuclei. Figure 3.1D shows examples of cell membranes still intact, with some nuclei lysis occurring (Figure 3.1D rows 2 and 4). Figure 3.1E shows isolated nuclei with various

nuclei morphology. Row 1 shows an example of nuclei lysis and chromatin leakage into solution, while row 6 shows an example of nuclei blebbing. Combining ImageStream with our optic validation at 60X magnification allowed us to feel much more confident in our FACS sample input, and allowed for a "go no-go" approach for moving forward with the sample preparation depending on these quality control decisions.

Replicate 2 FACS produced a much more robust sorting outcome, with 53.3% of the total population being 7AAD positive and falling into the cell cycle curve (Figure 3.1B, bottom left). Furthermore, the ImageStream cell cycle curve recapitulated a similar 7AAD intensity (Figure 3.1F). The number of cells collected for each of the 4 phases ranged from 2044-7512 (Table A.3.1), a marked improvement over replicate 1. Furthermore, ATAC TSS enrichment score ranged from 4.65-5.35, another improvement over replicate 1.

BioAnalyzer traces from each of the replicates further solidified the improvement of replicate 2 over replicate 1 (Figure 3.1G). GEX cDNA library traces were more pronounced in replicate 2 (Figure 3.1G, top two rows), although GEX final library traces looked similar (Figure 3.1G, middle two rows). Interestingly, ATAC final library traces between the two replicates were considerably different (Figure 3.1G, bottom two rows). Replicate 2 samples produced an entirely extra peak at higher fragment sizes that were absent in replicate 1, supporting the notion that nuclei over-lysis and chromatin spillage into solution was occurring in replicate 1.

When quantifying Seurat quality control metrics, we further see disparities between the two datasets (Figure 3.1H). Although the number of RNA features and counts are comparable between the two datasets, we observe a tighter distribution

of the metrics in replicate 2 compare to replicate 1, indicating more consistency between the RNA quality of the cells (Figure 3.1H, top two rows). Another are where we see differences between the two replicates is the mitochondrial percentages of each sample (Figure 3.1H, middle row). Replicate 1 has markedly higher mitochondrial read content in the samples compared to replicate 2. High mitochondrial reads in scRNA-seq samples has been associated with cell stress or death during sample processing [55]. Furthermore, the tighter distribution in replicate 1 may suggest mitochondrial spillage in the lysate, indicated by a more consistent mitochondrial read percentage in each individual cell. The ATAC features and counts (Figure 3.1H, bottom two rows) show similar dispersion profiles, and then differences in counts and features between the two replicates could be attributed to sequencing depth, as the ATAC samples for both replicates were not sequenced to saturation due to costs and computational time. Taken together, these findings suggest the protocol developed for replicate 2 is more robust, however it is important to validate this notion by exploring the sequenced datasets of each replicate.

3.2.2 *Variable feature detection and single-cell weighted nearest neighbor clustering resolves differences between 10X Multiome ATAC + Gene Expression with S-phase fractionation replicates and validates the second replicate as an optimized method*

To further explore differences between the two replicates, highly variable features were determined using a mean variability plot (Figure 3.2A and B). For replicate 1, 7 of the 10 top variable genes were mitochondrial genes (*MT-CO3, MT-CO2, MT-CO1, MT-ATP6, MT-ND3, MT-CYB, MT-ND4*), while replicate 2 had no mitochondrial reads contributing to the variability in the dataset. Interestingly, of the

top 20 variable genes annotated, 6 genes are shared between the datasets (*CDH18, DACH1, GRID2, ROBO2, GREB1L, ILRAPL1*), indicating that although there are overlaps in positive hits in the datasets, high mitochondrial read counts are a confounding signal present in replicate 1.

UMAP reduction and clustering was performed for the RNA and ATAC datasets and was integrated using a weighted nearest neighbors (WNN) approach[56] (Figures 3.2C and D). UMAP reduction of the RNA with S-phase fraction identity labeling shows a distinct separation in replicate 1 for S4, with a uniform cluster combining S1, S2, and S3 (Figure 3.2C, left). This is likely due to the high mitochondrial read count existing in S1, S2, and S3 for replicate 1, which is less prevalent in S4 of the same replicate (Figure 3.1H, middle row, left and Figure 3.2E). The same observation was not recapitulated in the RNA UMAP for replicate 2, although the cells do seem to have some S-phase separation (Figure 3.2D, left). Variable features were similarly expressed across each replicate's S-phase identities (Figures 3.1E and F). Heatmaps for the top 10 differentially expressed genes in S-phase identities for replicates 1 and 2 fail to help resolve differences between the RNA profiles between the replicates (Figures 3.2G and H).

UMAP reduction of the ATAC datasets shows S-phase separation in both of the replicates, however, there is a clearer aggregation of S-phase labels in replicate 2 compared to replicate 1 (Figure 3.2C and D, middle). WNN integration of the two modalities helps resolve overlaps in S-phase fractions, with replicate 2 producing a clear gradient from S1 to S2 to S3 to S4, while replicate 1 has an S1/S2 mix to S3 to S4 (Figure 3.2C and D, right), indicating replicate 2 reveals more defined differences between the S-phases.

Finding clusters at a resolution of 0.7 (Figures 3.3A and B) and identifying cluster hallmarks (Figures 3.3C and D) reveals contributing factors to S-phase identities. The differentially expressed gene heatmap from replicate 1 highlights 8 clusters, with clusters 1, 2, 3, and 4 as clusters containing cells with highly expressed mitochondrial genes, explaining why the S1, S2, and S3 identities are intermixed in the RNA UMAP. Furthermore, clusters 0, 5, and 6 have a reduced expression of mitochondrial genes, corresponding to the S4 label, explaining the separation in the replicate 1 RNA UMAP. The top 10 variable features from replicate 1 exhibit minimally variable average expression across the cluster identities (Figure 3.3E). The same resolution produces 12 clusters in replicate 2 and shows a much more dynamic and variable contribution to cluster identities (Figures 3.3D and F). Mitochondrial gene expression is still exhibited in replicate 2 cluster hallmarks, however, mitochondrial genes are not the sole contributor to the differential expression hallmarks of any of the clusters as they are in replicate 1. Due to these findings, replicate 2 was chosen as the dataset to perform further downstream analysis moving forward.

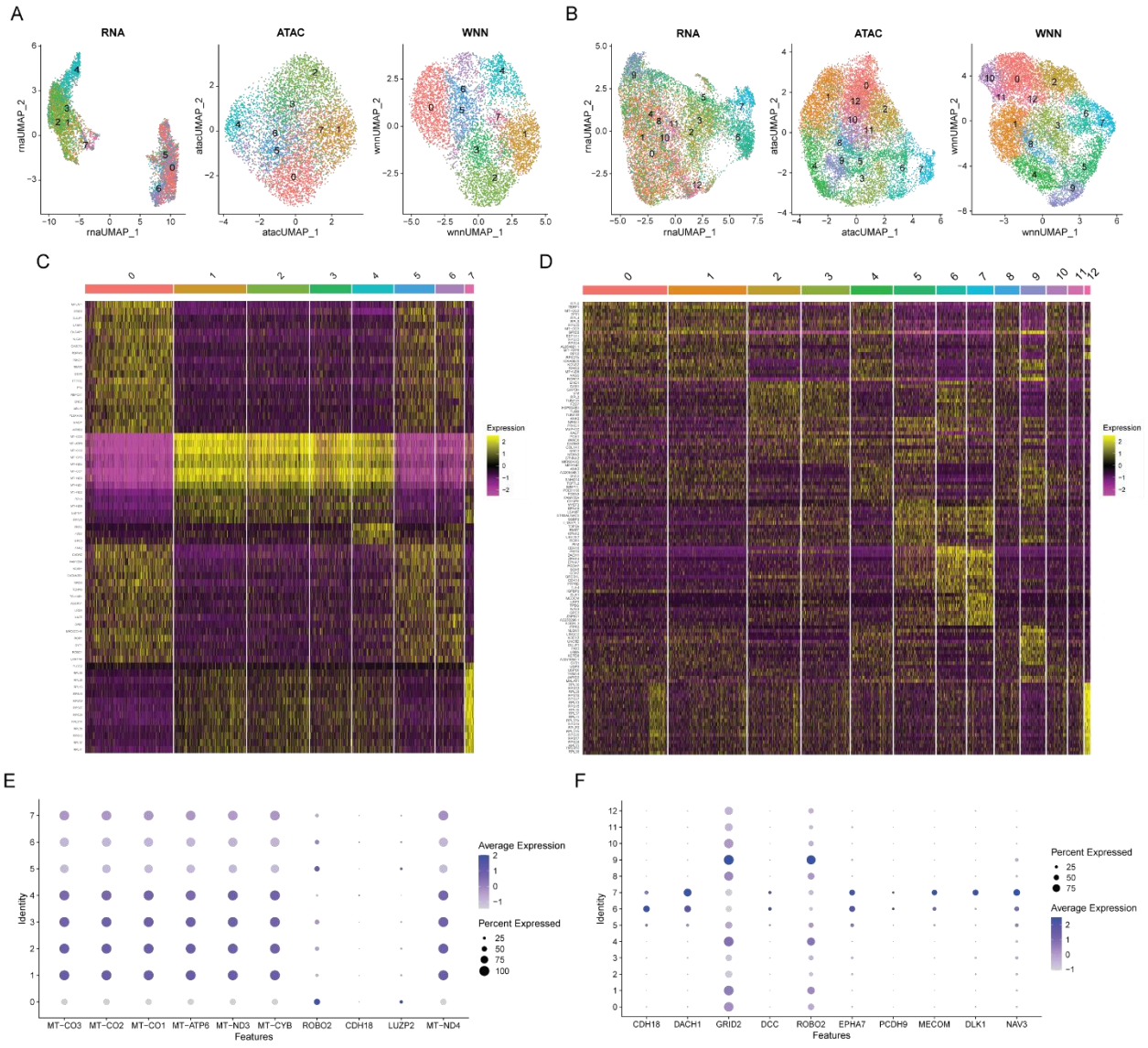
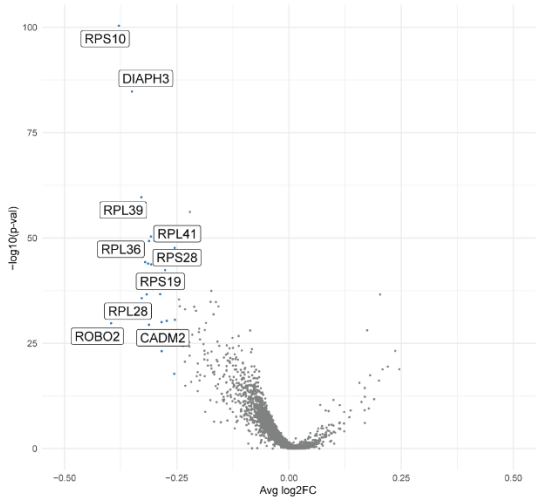


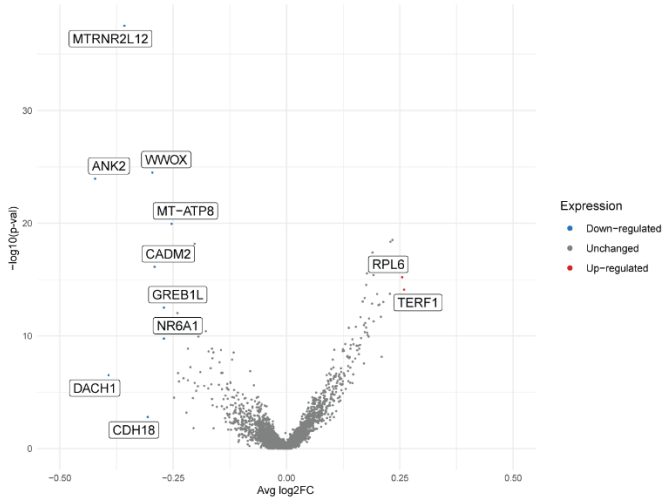
Figure 3.3. Clustering of multiome replicates shows a distinct difference in cluster identity gene hallmarks.

(A) RNA, ATAC, and WNN UMAPs with resolution = 0.7 for replicate 1.
(B) RNA, ATAC, and WNN UMAPs with resolution = 0.7 for replicate 2.
(C) Heatmap of the gene expression of top 20 variable genes in each cluster for replicate 1.
(D) Heatmap of the gene expression of top 20 variable genes in each cluster for replicate 2.
(E) Dot plot of the top 10 variable genes in replicate 1 labeled by cluster identity.
(F) Dot plot of the top 10 variable genes in replicate 2 labeled by cluster identity.

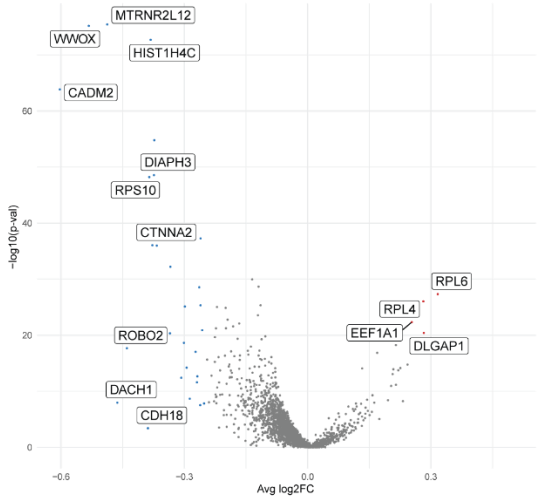
S1vS2



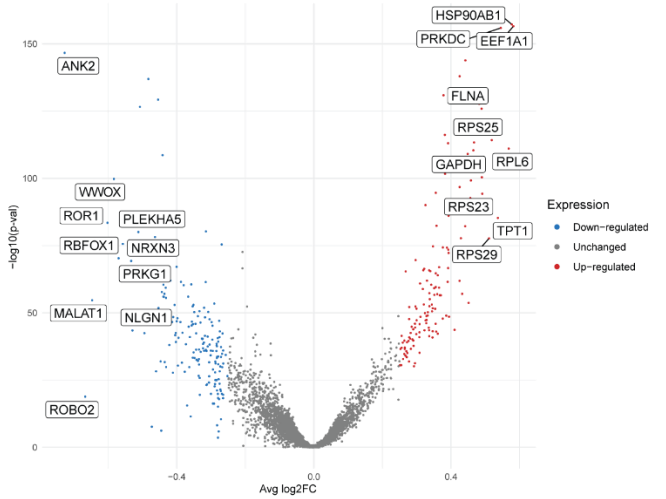
S2vS3



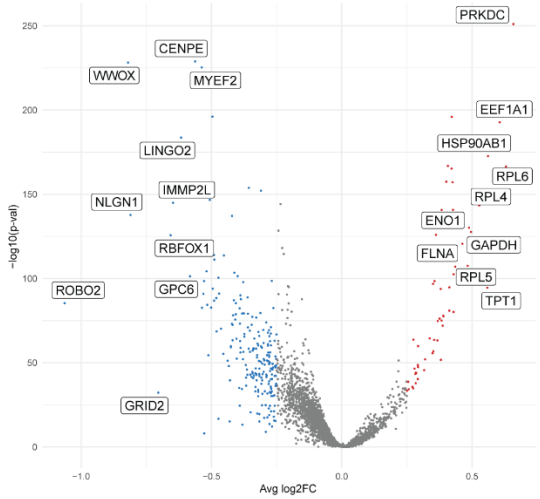
S1vS3



S2vS4



S1vS4



S3vS4

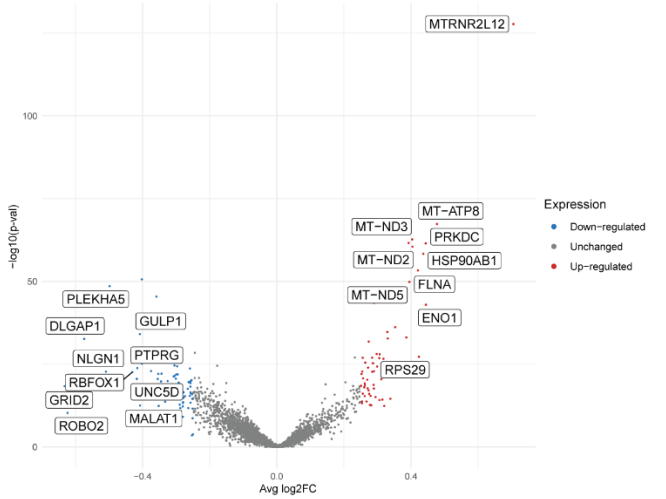


Figure 3.4. Differential gene expression reveals insights into S-phase fractions and highlights nervous system related genes in each fraction.

S-phase identity comparisons from replicate 2 are labeled by corresponding sample comparisons above each volcano plot. Up- and down-regulated genes were classified by having an average log2FC of greater than 0.25 or less than -0.25. Up-regulated genes are labeled in red and down-regulated are labeled in blue. The top 10 genes from each expression level are annotated.

3.2.3 *Differentially expressed genes between S-phase fractions are involved in nervous system development*

Comparing up- and downregulated differentially expressed genes between S-phases reveals RPL genes, RPS genes, and genes primarily related to nervous system development (Figure 3.4). *ROBO2*, a gene implicated in axon development[57], and CADM2, a gene that has been shown to play a part in brain development[58], are downregulated in S1 versus S2. When comparing differentially expressed genes between S2 and S3, *WWOX*, *ANK2*, *CADM2*, *DACH1*, and *CDH18* are downregulated while *TERF1* is upregulated. *WWOX* has been found to be involved in neuronal migration and early cortical development[59,60], *ANK2* has been shown to regulate embryonic neurodevelopment and axonal branching[61,62], *DACH1* has been shown to regulate neural crest cell migration during embryonic development[63], and *CDH18* is expressed in the central nervous system, specifically the brain[64]. *TERF1* has been implicated in telomere maintenance and cancer[65,66]. Differentially expressed genes in S3 versus S4 fractions enrich for similar nervous system gene types. *DLGAP1*, *PLEKHA5*, *NLGN1*, *GRID2*, and *UNC5D* were downregulated, while *FLNA* was upregulated. *DLGAP* family genes, *PLEKHA5*, *NLGN1*, *GRID2*, and *UNC5D* are expressed in the brain and are important for cortical and neuronal cell survival[67–71], and *FLNA* regulates neuronal migration[72]. Although there is overlap in the genes that are up- and downregulated between the S-phase fractions, it is interesting to note

the differentially expressed genes vary between numerous nervous system genes, indicating there may be a neural lineage priming effect occurring during S-phase.

3.2.4 *Chromatin accessibility and gene expression relationships add resolution to S-phase fraction specific cell types.*

To further resolve potential primed states throughout S-phase fractions, coverage plots for a common pluripotency factor, *POU5F1*, and two of the highest variably expressed genes in the dataset, *DACH1* and *CDH18*, were produced to detect correlations between accessibility and gene expression (Figure 3.5A, B, and C). *POU5F1* was not a differentially expressed gene in any of the S-phases, and this information was recapitulated in the expression levels and accessibility landscape across all S-phase fractions (Figure 3.5A, top right and top left). Although S2 showed downregulation of *DACH1* and *CDH18*, which is slightly apparent in the gene expression violin plots, the accessibility landscape looks similar across the S-phase fractions of both genes (Figure 3.5B and C, top right and top left). To determine if there was a general genome-wide correlation at TSS sites between gene expression and chromatin accessibility, LOESS plots of 2kbp up- and downstream TSS S-phase fraction accessibility scores were plotted against binned mean expression levels of a WT huESC dataset (Figure 3.5D). Mean expression and S-phase fraction accessibility were correlated, with high mean expression corresponding with the highest accessibility and low mean expression corresponding with the lowest accessibility at TSSs. This trend held true across all S-phase fractions, indicating that although gene expression and accessibility are correlated, there lacked a unique accessibility signature at the TSS of genes in S-phase fractions.

To incorporate the accessibility modality of the dataset, chromVAR was used to identify TF motifs associated with each S-phase (Figure 3.6, red plots). Correlations between gene expression of specific TFs and the accessibility of the corresponding TF binding motif were quantified using Presto[73] (Figure 3.6). This allowed for detecting highly expressed TFs and highly accessible binding TFs in each of the S-phases using a Wilcoxon ranking method. SOX2 and SOX4, both of which are involved in maintenance of pluripotency as well as neural development[74–77], were the only statistically significant hits in S1 (Figure 3.6A), supporting the notion that a neural priming effect could be beginning in early S-phase. S2 contained numerous hits with SREBF1, MGA, ZNF682, and NFKB2 being the top four hits (Figure 3.6B). SREBF1 has been identified as a proneural and neurogenesis TF[78]. Intriguingly, MGA has been shown to have numerous roles in development, including safeguarding ESCs from differentiating into endoderm, being indispensable for cell survival during peri-implantation, and negatively regulating the oncogenic/pluripotency MYC pathway[79–81]. Zinc finger proteins are expressed in many neural lineages and are involved in neuro-related diseases[82], and NF-kB TFs are abundant in the brain and central nervous system[83]. Furthermore, MYCN and POU5F1 were also contained in S2 (Table A.3.4), suggesting there may be a balance in this S-phase state for pluripotency maintenance with an opportunity for cell fate changes to occur. Interestingly, S3 returned no positive hits, potentially suggesting this phase may be a steady state for TF expression and motif accessibility with low enrichment for any particular TF expression and motif combination. Top S4 hits enriched for FOX family genes (FOXP1, FOXN3, FOXO3, FOXK1, and FOXG1), as well as ELF2 and E2F3 (Figure 3.6C and Table A.3.4). FOX family genes have been well-documented in a neural

developmental context, including development of the ectoderm, neurons, and brain and regulation of neural stem cell homeostasis[84–90]. ELF2 expression is present in brain tissue, is involved in cell cycle progression, and has been implicated in angiogenesis[91,92]. E2F3 is a target in neural precursor proliferation and cell cycle exit, has a regulatory effect on SOX2 for differentiation in the developing cortex, and has been established as a neurogenic factor in neural precursor cells[93–95]. Taken together, integration of gene expression for TFs and their corresponding enrichment of accessible binding motifs helps unveil a delicate balance in the expression and activity of different neural progenitor genes and pluripotency maintenance genes in different S-phases.

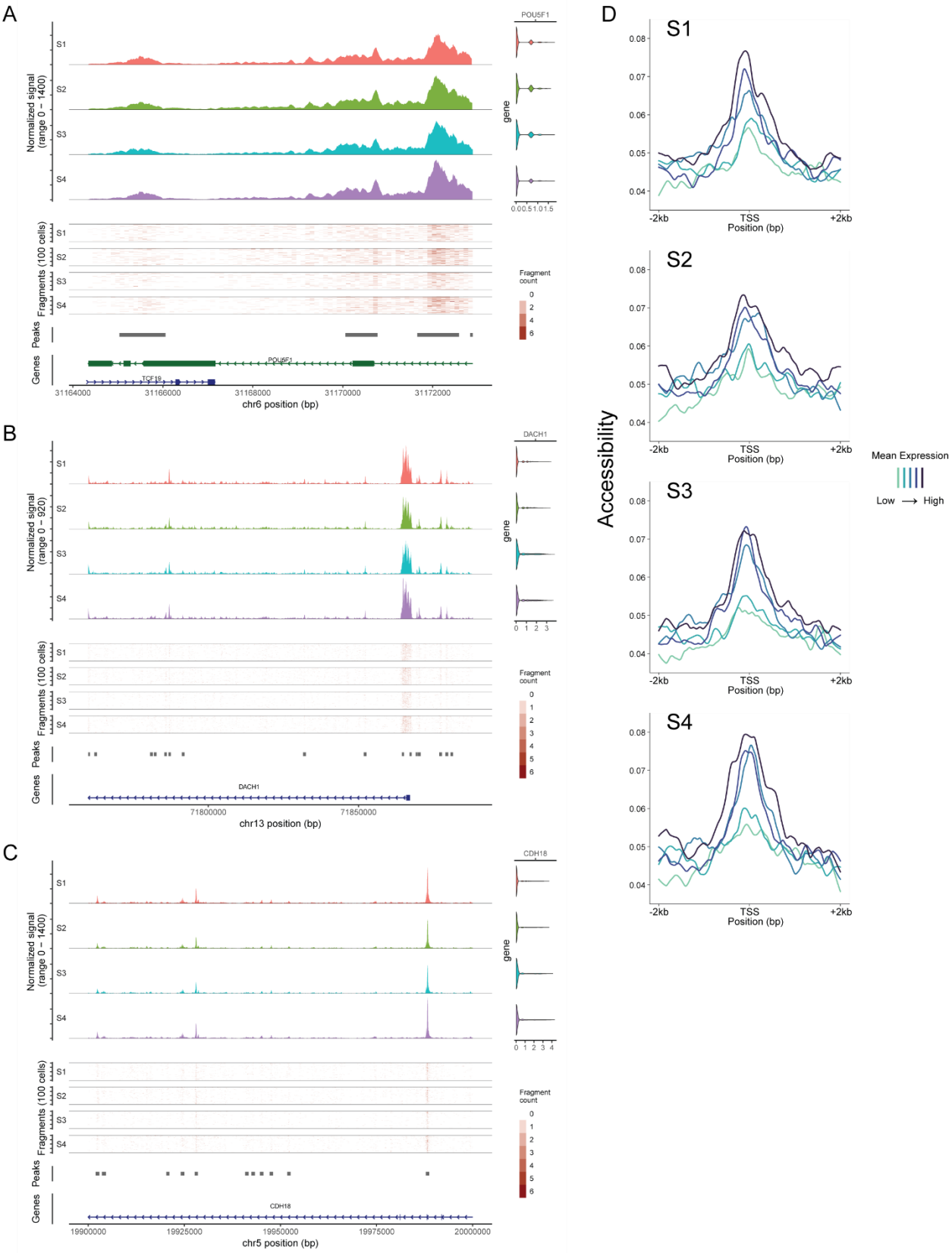

A
Normalized signal (range 0 – 1400)
S1
S2
S3
S4
Fragments (100 cells)
Peaks
Genes
POU5F1
TCF19
31164000
31166000
31168000
31170000
31172000
chr6 position (bp)
gene
Fragment count
B
Normalized signal (range 0 – 920)
DACH1
71800000
71850000
chr13 position (bp)
C
CDH18
19900000
19925000
19950000
19975000
20000000
chr5 position (bp)
D
Accessibility
S1
S2
S3
S4
-2kb
TSS
+2kb
Position (bp)
Mean Expression
Low → High

Figure 3.5. Gene expression and accessibility across the TSS of genes is steady among S-phase fractions.

(A) Coverage plot for *POU5F1* in S-phase fractions with normalized accessibility signal (top left), expression level (top right), down-sampled (100 cells) fragment count (middle left), and peaks (middle bottom).
(B) Same as (A) but for *DACH1*.
(C) Same as (A) but for *CDH18*.
(D) Accessibility plots at promoter regions of S-phase fractions for genes in 20 percentile bins of mean gene expression.

3.2.5 *S-phase fraction-agnostic expression levels reveal DACH1 as a unique cluster identity marker antagonistic to POU5F1*

Observing gene expression across clusters within each S-phase fraction provides an opportunity to understand temporal S-phase changes in gene expression (Figure 3.7A and B). Strikingly, gene expression of pluripotency markers *POU5F1* and *SOX2* show a generally S-phase fraction-agnostic expression profile (Figure 3.7A and B, top two rows). Cluster 12 has a uniquely high expression of *POU5F1*, and this is recapitulated across all four S-phase fractions, with increased expression across clusters 0, 2, 9, 10, and 11 in S2, clusters 0 and 1 in S3, and clusters 1 and 10 in S4. *SOX2* expression is high in all clusters, except cluster 4, 5, and 9, and although there are some increases in these clusters in different S-phases, the expression levels between clusters in S-phase fractions remain consistent.

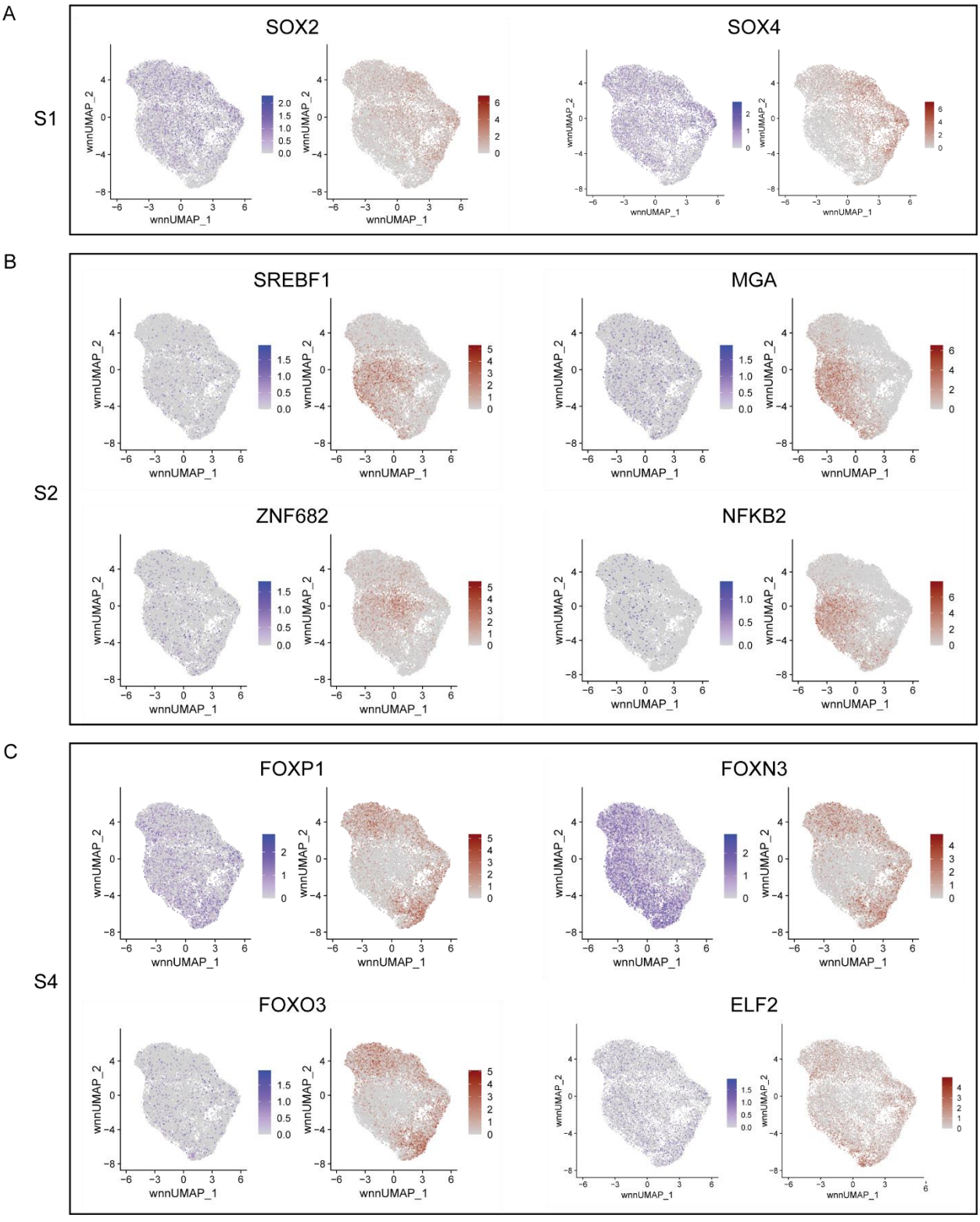

A
S1
SOX2
SOX4
B
S2
SREBF1
MGA
ZNF682
NFKB2
C
S4
FOXP1
FOXN3
FOXO3
ELF2
wnnUMAP_1
wnnUMAP_2

Figure 3.6. Presto gene expression and TF motif enrichment reveals unique gene expression-TF accessibility enrichments displayed in the WNN UMAP across S-phase fractions.

(A) Gene expression and motif plots for top presto hits (SOX2 and SOX4) in S1.
(B) Gene expression and motif plots for top presto hits (SREBF1, MGA, ZNF682, and NFKB2) in S2.
(C) Gene expression and motif plots for top presto hits (FOXP1, FOXN3, FOXO3, and ELF2) in S4.

When looking at the top two variable features in the dataset, *DACH1* and *CDH18*, we see their expression almost exclusively in clusters 6 and 7 (Figure 3.7A, bottom two rows), and these clusters were the most separated from other clusters in the RNA UMAP (Figure 3.3B, left panel). This high expression is recapitulated in an S-phase fraction-agnostic manner, with the highest expression being prevalent in clusters 6 and 7 throughout all S-phase fractions (Figure 3.7B, bottom two rows). Remarkably, although S-phase fractions lacked variability in gene expression and accessibility at *POU5F1* (Figure3.5A), clusters 6 and 7 show a reduction in *POU5F1* accessibility peaks and a slightly reduced expression level (Figure 3.7C). Integrating TF expression and binding motif accessibility enrichment of POU5F1 using presto shows how *DACH1* and *CDH18* expression occupy a region with low expression of *POU5F1* and a distinct absence of POU5F1 binding motif enrichment (Figure 3.7D). To understand if this was a uniquely S-phase phenomenon, a WT hESC scRNA-seq dataset was used to determine gene expression levels of *POU5F1*, *SOX2*, *DACH1*, and *CDH18* (Figure3.7E and F). *POU5F1* and *SOX2* were highly expressed across all clusters of the WT dataset, while *DACH1* and *CDH18* had close to no expression across any of the clusters, indicating these $DACH1^{+}/CDH18^{high}$ and $DACH1^{high}$ clusters are unique to S-phase.

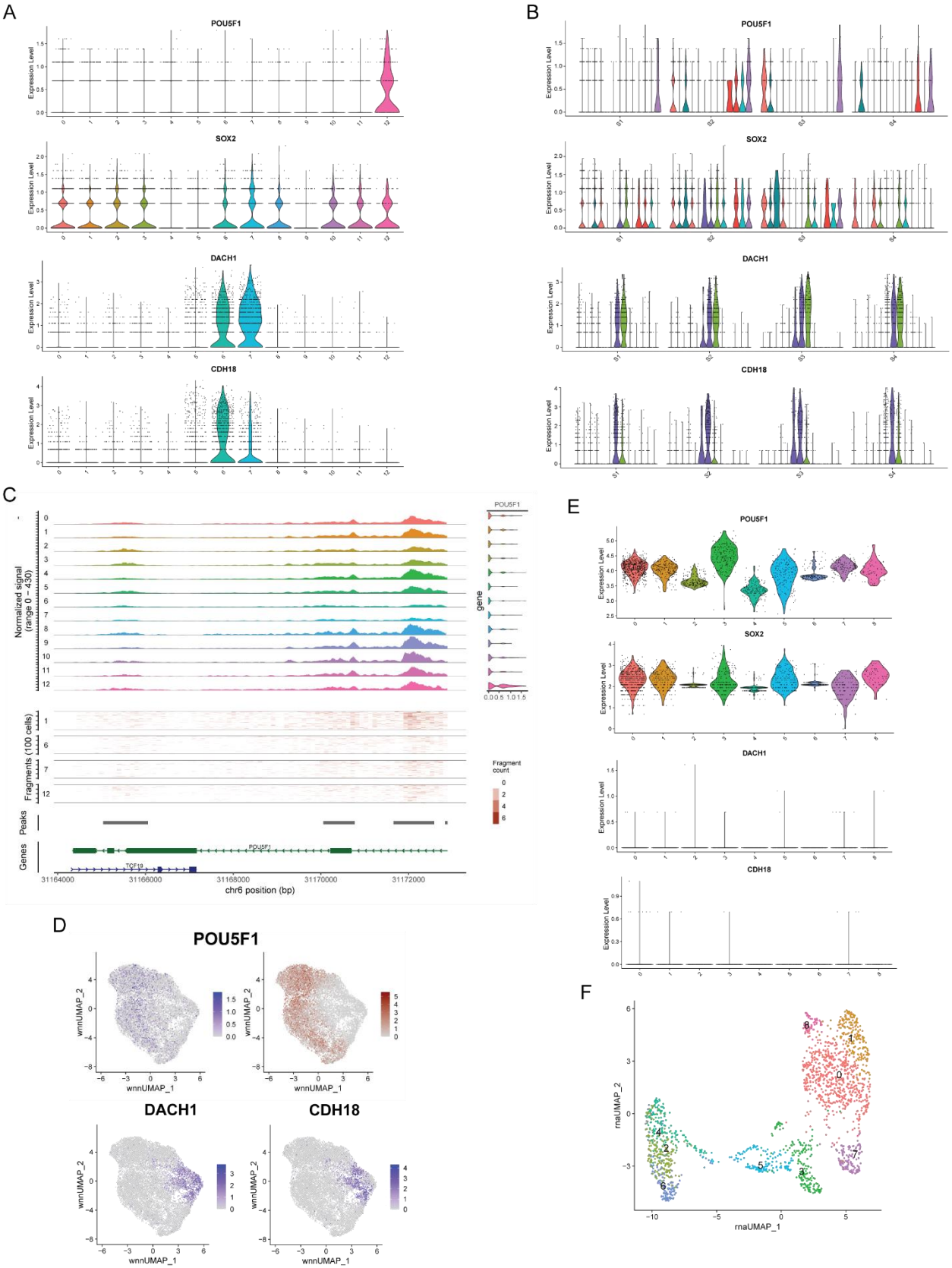

A
POU5F1
SOX2
DACH1
CDH18
Expression Level
B
S1
S2
S3
S4
C
Normalized signal
(range 0 – 430)
Fragments (100 cells)
Peaks
Genes
gene
Fragment count
chr6 position (bp)
31164000
31166000
31168000
31170000
31172000
D
POU5F1
DACH1
CDH18
wnnUMAP_1
wnnUMAP_2
E
F
rnaUMAP_1
rnaUMAP_2

Figure 3.7. S-phase fraction agnostic expression levels reveal *DACH1* as a unique cluster identity marker antagonistic to *POU5F1*.

(A) Violin plots of gene expression based on cluster identities for *POU5F1*, *SOX2*, *DACH1*, and *CDH18*.
(B) Violin plots of gene expression from (A) based on cluster identities in each S-phase fraction.
(C) Coverage plot for *POU5F1* in clusters with normalized accessibility signal (top left), expression level (top right), down-sampled (100 cells) fragment count (middle left), and peaks (middle bottom).
(D) Gene expression and motif plots for *POU5F1* and gene expression plots for *DACH1* and *CDH18* displayed in the WNN UMAP.
(E) Violin plots of gene expression from (A) in clusters of WT ESCs.
(F) Corresponding UMAP at resolution = 0.7 for gene expression data in (E).

3.2.6 *Gene expression and TF motif enrichment of DACH1^{+}/CDH18^{high} and DACH1^{high} and clusters reveal neural stem cell (NSC)-primed cell subsets in S-phase.*

DACH1 has been shown to have an indispensable role in neural crest migration and neural development during embryonic development[63,96]. Interestingly, it has also been reported to repress *SOX2*, *NANOG*, and *KLF4* signatures with DACH1 binding to the promoters of these pluripotency feature[97]. Furthermore, *DACH1* overexpression stimulated tumor organoid formation through BMP signaling suppression, a pathway that, when inhibited, establishes neuroectoderm[98,99]. *CDH18* expression has been shown to exist in the brain and can act as a tumor suppressor in glioma cells, and type II cadherins are involved in the organization of neurons into layers[64,100].

To help determine cell identities of *DACH1^{+}/CDH18^{high}* and *DACH1^{high}* clusters, presto was once again used to identify top TF expression and motif accessibility hits in these clusters (Figure 3.8A and Table A.3.5). Although there was a lack of variability in accessibility at the *CDH18* and *DACH1* genes (Figure 3.8B and C), presto revealed SOX4, MEIS2, and LHX5 shared between the two clusters. SOX4 was

previously discussed as a neural development gene, and MEIS2 is indispensable for cranial neural crest development and mandibular arch patterning from neural crest cells[101,102]. LHX5 has been implicated in forebrain specification, morphogenesis of Purkinje cells and Cajal-Retzius cells, dendritogenesis, and a regulator of mamillary body differentiation[103–108].

RFX4, PAX3, and MEIS family genes were top hits for the *DACH1*$^{+}$/*CDH18*^{high} cluster. RFX4 is considered a neural progenitor marker, is essential for ventricle formation in the forebrain and dorsoventral patterning, is a regulator of neural stem and progenitor cells, regulates sonic hedgehog signaling, and is implicated in the development of glia and oligodendrocyte progenitor cells[109–114]. PAX3 is expressed during CNS development, neural crest cell development, and is involved in neural tube closure and patterning[115–121]. MEIS family genes play a critical role in the brain and are expressed in neural crest cells[122–124]. The *DACH1*^{high} cluster contained RFX3 and EMX2 as top hits. RFX3 directs cilia development in glia and regulates neuron patterning required for development[125–129]. EMX2 is implicated in cortical specification and organization from neural stem cells, is a transcriptional regulator in neural progenitor cells, and regulates stem cell proliferation of adult central nervous system[130–136]. Taken together, these findings suggest these clusters are primed to become neural stem precursor cells, most likely progressing toward neural crest cells.

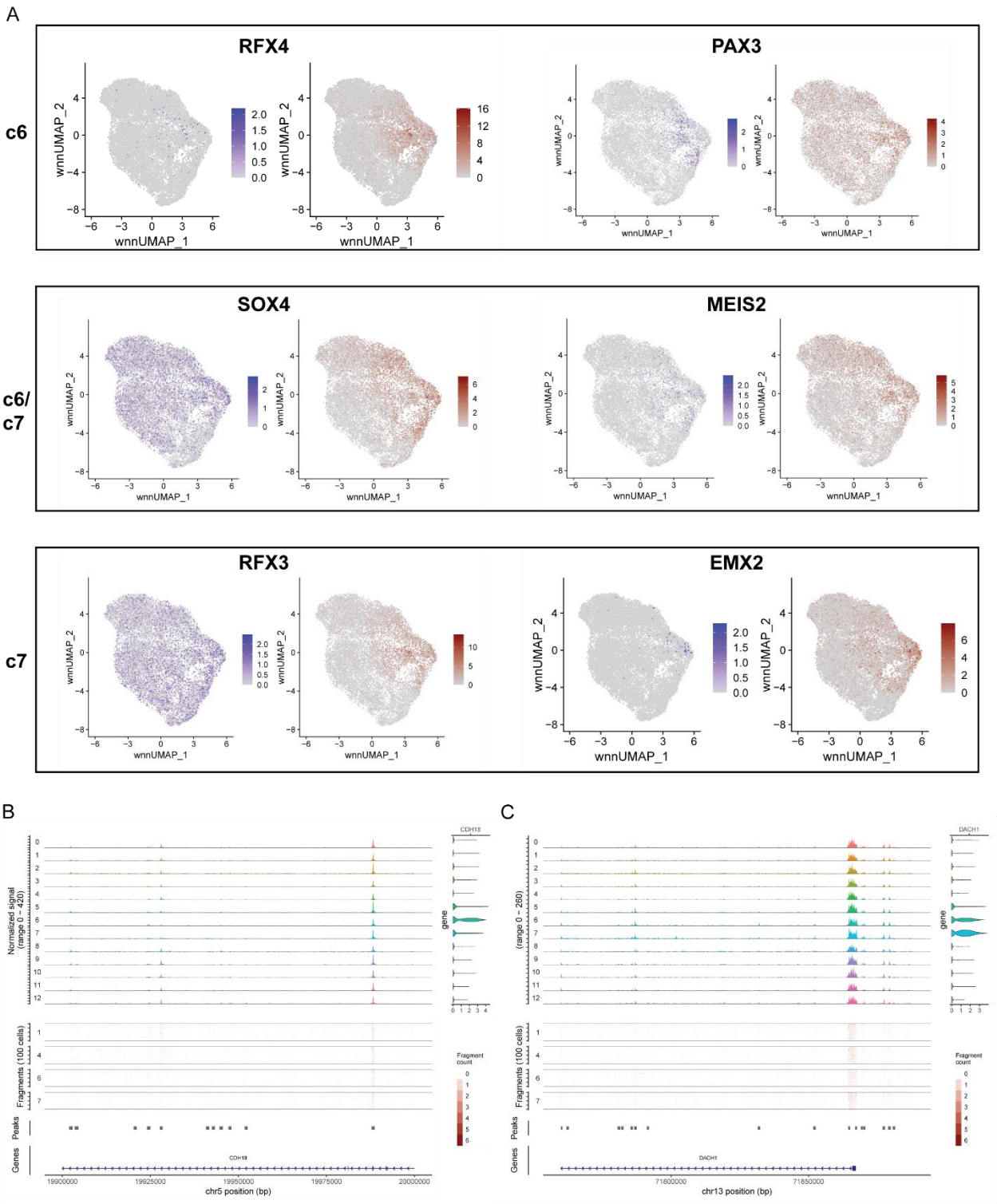


Figure 3.8. Presto gene expression and TF motif enrichment correlations of *DACH1*$^{+}$/*CDH18*^{high} and *DACH1*^{high} clusters reveal NSC-primed cells in S-phase.

(A) Gene expression and motif plots of top presto hits for cluster 6 (*RFX4* and *PAX3*), shared in cluster 6 and 7 (*SOX4* and *MEIS2*), and cluster 7 (*RFX3* and *EMX2*).

(B) Coverage plot for *CDH18* and (C) *DACH1*

3.3 Discussion

Developing and utilizing new technologies requires a deep consideration for the intricacies and the effects of changing variables in a protocol. Here, the 10X Single Cell Multiome ATAC and GEX protocol was heavily modified by including S-phase fractionation into the workflow, producing a one-of-a-kind dataset of replicating hESCs. Optimizing the workflow by implementing checkpoints to determine nuclei viability proved fruitful, with higher nuclei recovery, faster throughput, and lower mitochondrial read counts, allowing for clarity in the dataset. Computational processing provided another lens for this lever of clarity. UMAP representations improved separation between S-phase fractions and provided more dynamic and informative gene expression between replicates. When reducing dimensions of RNA datasets, replicate 1 produced a separation for the S4 fraction, which disappeared in replicate 2. This could be explained by the prominence of mitochondrial genes present in S1, S2, and S3 fractions of the first replicate. Furthermore, the ATAC UMAP representation showed a stronger separation between S-phase fractions in replicate 2 compared to replicate 1, and integrating the two modalities using a WNN approach solidified a separation between each S-phase fraction in replicate 2 that was absent in replicate 1. This optimized method proved to be superior to the original method, and this optimized dataset was used moving forward for downstream analysis.

Variable feature detection informed the initial idea that neuronal priming was potentially occurring and presented itself when observing changes in S-phase gene expression changes. Differentially expressed genes between S-phase fractions highlighted neuronal migratory, cortical development, axonal branching, and cell

survival genes, and these differentially expressed genes were unique to each fraction, indicating a dynamic neural priming effect throughout S-phase.

When observing chromatin accessibility with gene expression information, pluripotency gene *POU5F1* expression and accessibility was similar across all S-phase fractions. Interestingly, the two most variable gene features in the dataset, *DACH1* and *CDH18*, also exhibited similar gene expression and accessibility across all S-phase fractions. Searching for a correlation between accessibility between gene expression and accessibility of S-phase fractions across TSSs genome-wide confirmed this lack of accessibility variability. However, integrating TF gene expression and TF binding motif accessibility showed a different correlation. S-phase fractions had a host of different hits for this integration method, with a strong prevalence of neural progenitor genes and pluripotency maintenance genes in different S-phase fractions, suggesting there may be a delicate balance at play between neural differentiation and stem cell maintenance.

When clustering using an unsupervised method, gene expression of pluripotency markers *POU5F1* and *SOX2* showed small differences across cluster expression in different S-phase fractions. This S-phase fraction-agnostic expression was also detected in *DACH1* and *CDH18*, and these two genes were detected in two distinct clusters. Remarkably, $DACH1^{+}/CDH18^{high}$ and $DACH1^{high}$ clusters showed a unique reduction in *POU5F1* promoter accessibility. Furthermore, integrating the gene expression and TF binding motif accessibility of *POU5F1* showed that *DACH1* and *CDH18* expression occupies regions where *POU5F1* TF binding motif accessibility lacks enrichment. Comparing *POU5F1*, *SOX2*, *DACH1*, and *CDH18* gene expression levels in a scRNA-seq WT hESC dataset indicated high expression levels across all clusters

for *POU5F1* and *SOX2*, and little to no expression of *DACH1* and *CDH18* in any cluster, indicating this *DACH1*$^{+}$/*CDH18*^{high} and *DACH1*^{high} expression was a uniquely S-phase related phenomenon.

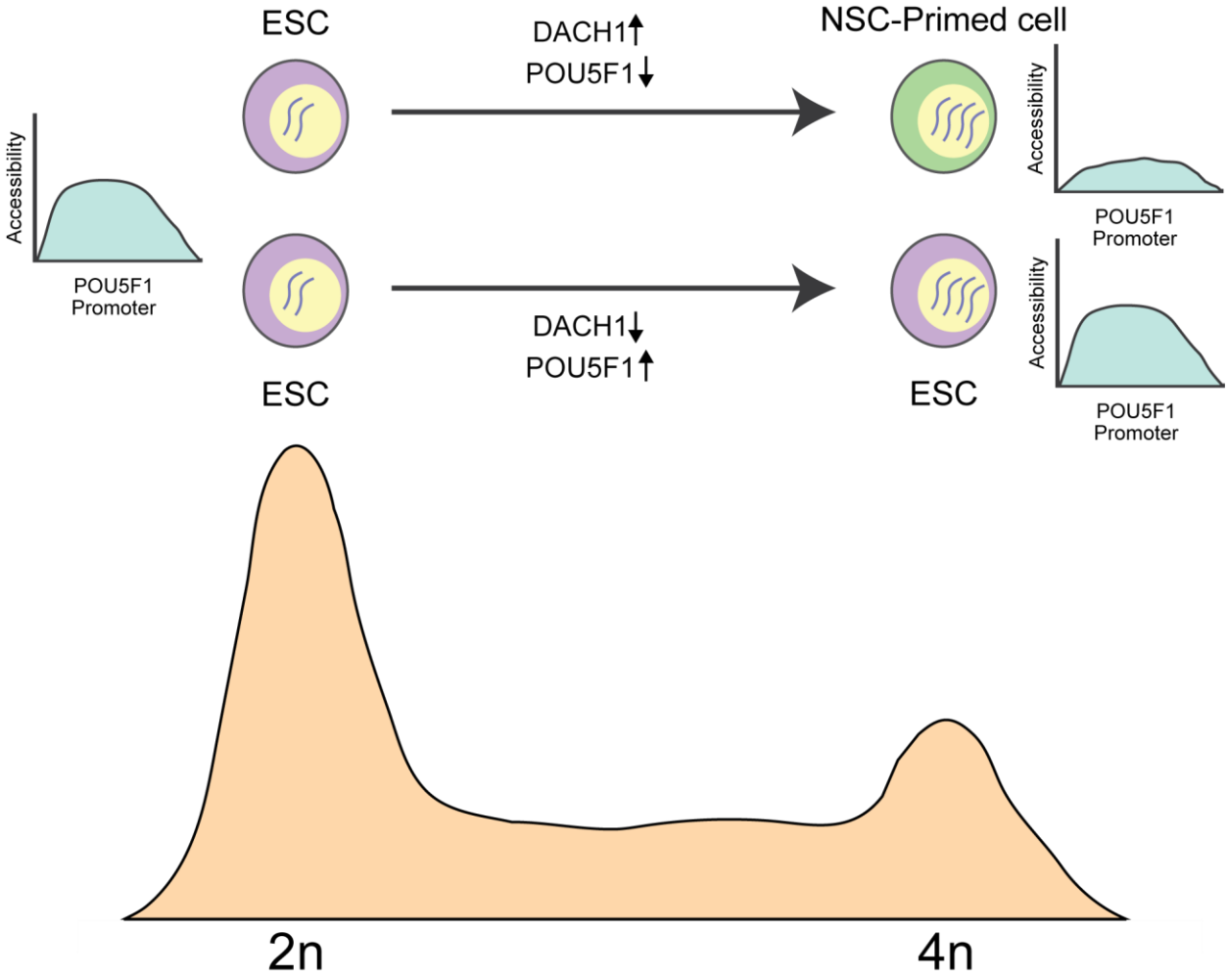


Figure 3.9. Schematic depicting a model of *DACH1* and *POU5F1* gene expression and accessibility interactions during S-phase.

DACH1 has been shown to repress pluripotency signatures by binding to the promoters of these features and has been implicated in establishing neuroectoderm. *DACH1*$^{+}$/*CDH18*^{high} and *DACH1*^{high} clusters showed a prevalence of gene expression of SOX, MEIS, LHX, PAX, and RFX family genes and an enrichment of their

corresponding TF binding motif accessibility. These TFs have been implicated in cortical development, neural stem cell establishment, and neural cell regulation and patterning, implicating these *DACH1*$^{+}$/*CDH18*^{high} and *DACH1*^{high} clusters as a potential primed neural stem precursor cell subset.

Given these findings, it is interesting to posit a model of *DACH1* and *POU5F1* interactions with *DACH1* expression antagonistic to *POU5F1* expression and accessibility in S-phase (Figure 3.9). It is possible that other mechanisms responsible for initiating *DACH1* expression in a small subset of cells trigger during early S-phase and allow an opportunity that limits *POU5F1* expression during this phase of the cell cycle. This reduction in *POU5F1* expression would cause weakness in stem cell maintenance and allow for priming of a neural stem cell state, and should other factors be present to fully trigger this cell state transition, this window of the cell cycle could allow this cell identity change to occur. This dataset provides unique insights into S-phase in the context of stem cell maintenance and identity, and it should facilitate advances in our understanding of stem cell fate changes and the identification of novel gene targets involved in development.

Methods

S-phase sorted 10X multiome sample preparation for sequencing of HUES64 cells

HUES64 were cultured in mTeSR1 media for maintenance.

Nuclei isolation and 7AAD staining:

1) For one run, the amount of HUES64 cells from two wells from a 6-well dish at >80% confluency are sufficient.
2) Aspirate media and add 1 mL of Accutase to each well
3) Incubate for 18-20 minutes – gently shake/swirl plate at 10 mins so that cells singularize properly
    a. Note: It is extremely important that cells are fully singularized at this step but not left in Accutase for too long. Not leaving cells in Accutase for long enough creates too many multicellular clumps, which causes the need for more aggressive pellet resuspension, but leaving cells in Accutase for too long causes lysis and debris aggregates which causes cells to clump to one another.
4) Add 1 mL of mTeSR1 media to quench the reaction and collect cells in a 50 mL conical tube
5) Spin down cells at 200 xg at 4C using the swing bucket rotor centrifuge.
6) Aspirate supernatant, and gently resuspend cells in 1 mL ice cold 3% BSA in PBS.
7) Pre-rinse a 40 um strainer with PBS and filter the sample through the pre-rinsed strainer into a 50 mL conical tube.

a. This is not done in the original 10X protocol, but this step isolates more singularized cells which reduces the amount of agitation necessary of resuspending the pellet, preserving nuclei membranes.

8) Spin down cells at 200 xg at 4C using the swing bucket rotor centrifuge.
9) Remove supernatant gently using a 1 mL hand pipettor.
10) Resuspend pellet in NP40 Lysis Buffer and transfer to a 1.5 mL eppendorf DNA Lobind tube for 1.5 minutes.
    a. Note: This process should be performed very gently using **wide-bore 1 mL pipette tips** at first, and once the tips are too wide to continue this resuspension (clumps are too small for the wide bore tip to make smaller), switch to a normal-bore 1 mL pipette tip to singularize cells. This pipetting process should be done in the 1.5 mL eppendorf tube, not the 50 mL conical.
    b. **The cells should only be in the NP40 Lysis Buffer for 1.5 minutes**, but this is ample time to resuspend the pellet nearly fully. A few very small clumps may remain, but do not exceed 1.5 minutes in NP40 lysis buffer.
11) Centrifuge at 300 xg for 5 minutes at 4C using a swing bucket rotor centrifuge.
12) Remove supernatant and add 1 mL PBS + 3% BSA + 1 U/ul RNase Inhibitor. Incubate on ice for 3 minutes.
13) Pipette in the same manner as step 10a, using a wide-bore pipette tip, and then a normal-bore tip.

14) Centrifuge at 300 xg for 5 minutes at 4C using a swing bucket rotor centrifuge.
15) Pre-rinse a FACS tube strainer with PBS.
16) Remove supernatant, resuspend pellet in 1 mL PBS + 3% BSA + 1 U/ul RNase Inhibitor in the same manner as step 10a, using a wide-bore pipette tip, and then a normal-bore tip.
17) Add 10 ul 7AAD and pipette gently to mix (~5 times), and add the solution to a FACS tube with the pre-rinsed strainer on top.

FACS Sorting:

1) Bring 4 1.5 mL eppendorf Lobind tubes with 200 ul PBS + 3% BSA + 2 U/ul RNase Inhibitor
    a. Note: The extra RNase inhibitor is so that the concentration still remains high when diluted by the addition of nuclei in sheath buffer from the sorter
2) Sort on a BD FACSAria Fusion machine.
3) Cells were sorted using a 100 um nozzle with voltage parameters as follows:
    a. FSC – 242 V
    b. SSC – 330 V
    c. PerCP-Cy5-5 – 467 V
4) FSC threshold values were set at 50,000
5) Laser Window Extension was set to 2.0 and FSC Area Scaling was set to 0.50
6) Make sure the adaptor to keep the tubes cooled is used.

Note: Save 50 uL of input sample for imagestream analysis as QC.

Nuclei permeabilization:

1) Centrifuge sorted tubes at 500xg for 5 minutes at 4C using a swing bucket rotor centrifuge.
2) Aspirate supernatant and resuspend in 100 uL 0.1X Lysis Buffer and pipette mix 5 times.
3) Incubate on ice for 2 minutes
4) Add 1 mL Wash Buffer and pipette mix 5 times.
5) Centrifuge tubes at 500xg for 5 minutes at 4C using a swing bucket rotor centrifuge.
6) Remove supernatant, resuspend in 5 uL Diluted Nuclei Buffer, and proceed to 10X Multiome library preparation.
    a. Note: Normally, there would be a cell count to find final nuclei concentration prior to the ATAC-seq portion, however, it is unlikely for any sample to have > 50k nuclei in total, so the concentration will never be > 10k nuclei/uL which is the maximum concentration for the assay. To determine nuclei counts, a light sequencing of the GEX libraries should be performed if the BioAnalyzer traces look as expected.

Sorting and ImageStream

FACS was performed at the UCI Institute for Immunology Flow Core on a BD FACSAria Fusion Sorter using a 100 micron nozzle. Voltage parameters were set at 242 for FSC, 330 for SSC, and 467 for PerCP-Cy5-5 (for 7AAD detection). FSC threshold was set at 50,000, and FSC Area Scaling was set to 0.50. ImageStream MKII was used to attain brightfield and 7AAD nuclear stains using standard settings and a 100 micron nozzle.

10x Multiome computational processing

Cell Ranger ARC v2.0.1 was used for replicate 1 and v2.0.2 was used for replicate 2. BCL files were converted to FASTQ using *mkfastq* and FASTQ was aligned to hg38, filtered, barcode counted, and peak called using *count*. The filtered feature matrix h5 files were analyzed using Seurat and Signac. Peaks in the samples of each replicate were merged to form a common peak set, and cell cycle annotations were added to each sample according to their S-phase fraction (S1, S2, S3, or S4). The replicates were filtered to keep cells with ATAC counts less than $1x10^6$ and greater than 3000, keep cells with RNA counts less than 9000 and greater than 500, and remove cells with mitochondrial percentages greater than 35%. SCTransform normalization was performed on the RNA, and term-frequency inverse-document frequency was calculated and singular value decomposition was computed before calling a UMAP with LSI reduction. A weighted nearest neighbors graph was calculated using a weighted combination of the RNA and ATAC using *FindMultiModelNeighbors* in Seurat with dimensions 1:50 of the RNA and 2:10 of the ATAC. A WNN UMAP was produced and clusters were identified using the original Louvain algorithm with a resolution of 0.7. Variable feature detection was performed using the *FindVariableFeatures* function in Seurat with the vst selection method. Gene expression markers for identity classes (S-phase fraction identity or cluster identity) were done using a Wilcoxon Rank Sum test with a log2FC threshold of 0.25.

Accessibility plots for mean gene expression around TSS

Mean expression values were calculated using a publicly available WT ESC dataset (SRP299892) and processed in Seurat similar to the 10X multiome processing method. Genes were binned in 20 percentile rankings from low to high mean

expression. A 4 kbp window around the TSS for each gene was taken and accessibility from bigwig files were cut into single base pair values at these sites. These were averaged across each expression level and loess was used to smooth the values.

Motif accessibility analysis using chromVAR and presto

chromVAR was used to calculate a per-cell accessibility score for known motifs and a chromvar assay was added to the Seurat object. To find gene and corresponding gene TF binding motif enrichment pairs, the presto package was used to calculate a p-value based on the Wilcox rank sum test of both the RNA and chromvar assay data. From this, a ranked list is produced for the top hits for paired RNA expression and corresponding gene TF accessibility.

SECTION 4: Immunolocalized Micrococcal Nuclease Sequencing

4.1 Introduction

Various proteins, including transcription factors and chromatin-modifying enzymes, interact with DNA to regulate gene expression, shaping a cell's phenotype[10]. In the context of stem cells, specific sets of transcription factors are known to sustain the stem cell state by promoting the expression of self-renewal genes and inhibiting differentiation genes[137]. Moreover, chromatin-modifying proteins alter the chromatin landscape, affecting the accessibility of DNA to the transcriptional machinery and, thus, influencing gene expression[138]. Therefore, precise DNA-protein interactions are fundamental for preserving stem cell identity and ensuring proper balance between self-renewal and differentiation.

To map and understand these crucial DNA-protein interactions, techniques such as Chromatin Immunoprecipitation Sequencing (ChIP-seq) and Micrococcal Nuclease Sequencing (MNase-seq) are employed. ChIP-seq enables the identification of the genomic binding sites of DNA-associated proteins[139]. It involves the use of antibodies to immunoprecipitate a protein of interest along with its bound DNA. The co-precipitated DNA is then sequenced and mapped back to the genome to identify the binding sites of the protein. On the other hand, MNase-seq involves the use of micrococcal nuclease, an enzyme that preferentially digests accessible DNA, leaving behind DNA that is protected by bound proteins such as histones[140]. Sequencing the remaining DNA provides information on the locations of these protein-bound regions, or nucleosomes, offering insights into chromatin structure. Here, we attempt to develop Immunolocalized-MNase (IL-MNase)-seq, a method to localize MNase to a protein of interest and initiating MNase digestion to enrich for DNA bound to the

protein of interest. Immunolocalization will allow for isolating colonies or sections of colonies on a dish for not only enriching DNA-protein interactions of interest, but also enriching for cells in a spatial modality using a technique such as laser-capture microdissection.

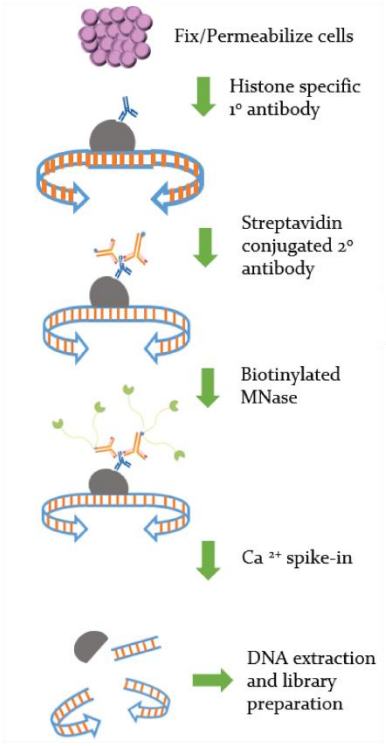
4.2 Results

4.2.1 *Immunolocalized Micrococcal Nuclease Sequencing*

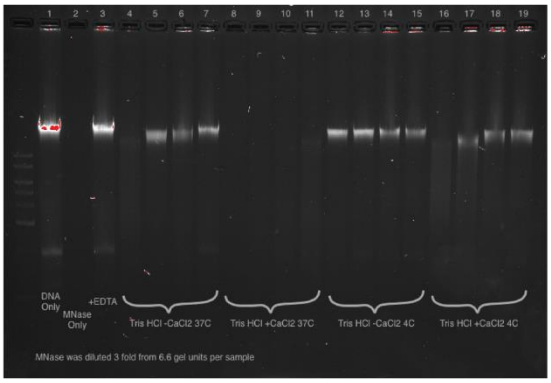
IL-MNase-seq is performed by using a streptavidin conjugated secondary antibody and biotinylated MNase similar to how immunofluorescence staining is performed (Figure 4.1A). First, cells are fixed and permeabilized and a primary antibody of interest is incubated overnight with the cells on a dish. Next, a secondary streptavidin conjugate antibody is incubated with the cells for one hour. Biotinylated MNase is added to the dish with PBS lacking calcium, as calcium is needed for MNase activation and digestion. A calcium spike-in is performed, the reaction is quenched after 30 minutes, DNA purified, and library preparation is performed.

First, MNase digestion was optimized. MNase was incubated at 37C and 4C with and without $CaCl_2$ (Figure 4.1B). There was mild digestion at 37C and no digestion at 4C for the -$CaCl_2$ treatment. Adding in $CaCl_2$ caused complete digestion at 37C and mild digestion at 4C, indicating MNase was temperature sensitive and its activity can be controlled in a temperature-mediated manner. Next, the effects of a calcium chelator, EDTA, was examined (Figure 4.1C). EDTA showed complete inactivity of MNase. The effects of biotinylation of MNase were unknown, so digestion of DNA using streptavidin-bead bound biotinylated MNase was performed at 37C with $CaCl_2$ in both Tris-HCL and PBS (Figure 4.1D).

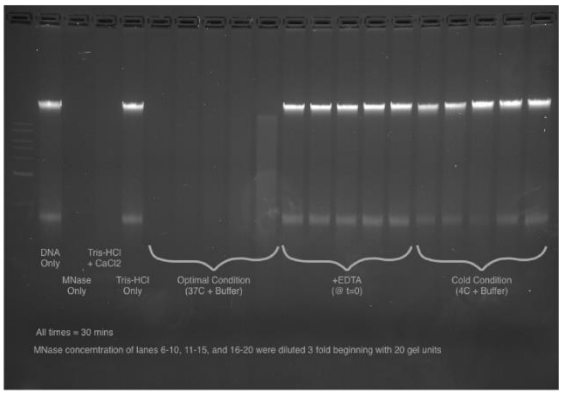
A


Fix/Permeabilize cells
Histone specific 1° antibody
Streptavidin conjugated 2° antibody
Biotinylated MNase
Ca ²⁺ spike-in
DNA extraction and library preparation

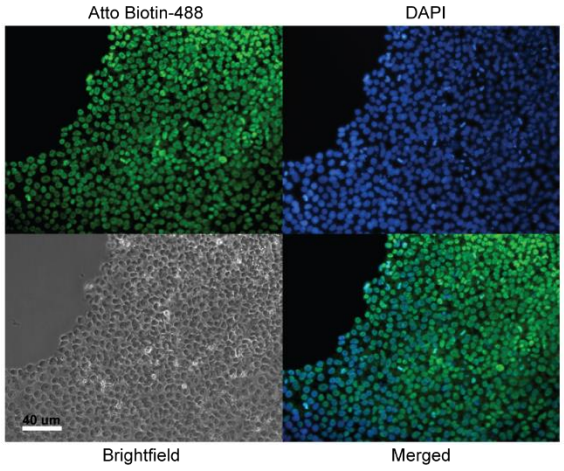
B


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
DNA Only
MNase Only
+EDTA
Tris HCl -CaCl2 37C
Tris HCl +CaCl2 37C
Tris HCl -CaCl2 4C
Tris HCl +CaCl2 4C
MNase was diluted 3 fold from 6.6 gel units per sample

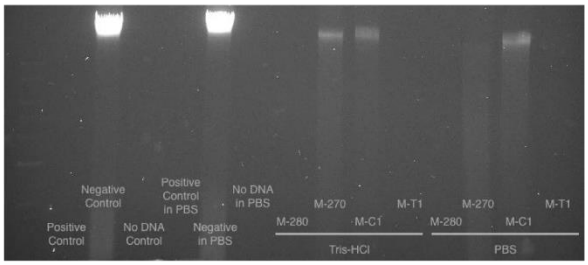
C


DNA Only
Tris-HCl + CaCl2
MNase Only
Tris-HCl Only
Optimal Condition (37C + Buffer)
+EDTA (@ t=0)
Cold Condition (4C + Buffer)
All times = 30 mins
MNase concentration of lanes 6-10, 11-15, and 16-20 were diluted 3 fold beginning with 20 gel units

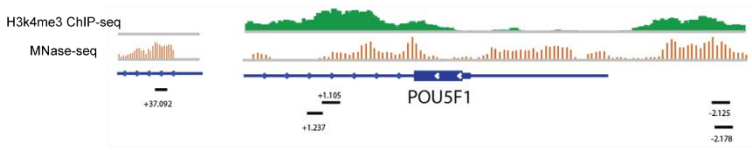
E


Atto Biotin-488
DAPI
40 um
Brightfield
Merged

D


Negative Control
Positive Control in PBS
No DNA in PBS
M-270
M-T1
M-270
M-T1
Positive Control
No DNA Control
Negative in PBS
M-280
M-C1
M-280
M-C1
Tris-HCl
PBS

F


H3k4me3 ChIP-seq
MNase-seq
+37.092
+1.105
+1.237
POU5F1
-2.125
-2.178

G

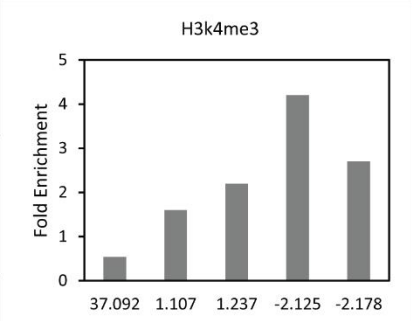

H3k4me3
Fold Enrichment
0
1
2
3
4
5
37.092
1.107
1.237
-2.125
-2.178

Figure 4.1. Immunolocalized Micrococcal Nuclease-sequencing
(A) IL-MNase-Seq protocol schematic
(B) Gel of digested DNA by different concentrations of MNase in lanes at 37C and 4C, and with $CaCl_2$ and without $CaCl_2$.
(C) Gel of digested DNA by different concentrations of MNase in lanes at 37C with optimal buffer, with EDTA, and at 4C with optimal buffer.
(D) Gel of digested DNA by biotinylated MNase bound to streptavidin conjugated beads in Tris-HCL and PBS.
(E) Immunofluorescence staining of streptavidin conjugated H3k4me3 antibody with Atto Biotin-488.
(F) Primer locations for PCR primers for anti-H3k4me3 *POU5F1* IL-MNase PCR enrichment.
(G) Fold enrichment at locations up- and downstream the *POU5F1* promoter using IL-MNase-Seq.

Biotinylation of MNase had little effect on the digestion capability of the enzyme. Next, we used a biotinylated dye, Atto Biotin-488, to determine if streptavidin conjugation of the secondary antibody had a negative effect on primary-secondary antibody binding (Figure 4.1E). Streptavidin conjugation seemed to have no effect on secondary antibody binding, with a strong overlap of the fluorescence dye with DNA dye DAPI.

Confident that the independent pieces of the assay were functioning with the added chemical modifications, IL-MNase-seq was performed on hESCs enriching for H3k4me3, an abundant histone marker for active promoters[141]. PCR was performed at sites enriched for H3k4me3 and MNase-seq (Figure 4.1F). Fold enrichment revealed an enrichment for H3k4me3 in a manner similar to the H3k4me3 ChIP-seq and MNase-seq tracks, suggesting the assay was working as intended.

4.3 Discussion

Here, a novel assay, IL-MNase-seq was developed to be used for assaying DNA-protein interactions in a unique manner. This assay was repeated 9 more times, with little-to-no reproducibility, and a new and improved version of this assay was

produced shortly after developing this, CUT&RUN[142], removing the need to further develop this assay.

Methods

Immunolocalized MNase-Seq Protocol

IMPORTANT All washes should be done carefully to not disturb cell adhesion to plate. Take care to pipette against the wall of wells, and never use a vacuum aspirator.

Cell fixation, chromatin digestion and extraction

1) Fix cells (on culture substrate) with 1% PFA for 10 min at RT.
2) Quench reaction with 125 mM glycine.
3) Aspirate and wash 2x with ice-cold PBS (Ca-; Mg-).
4) Permeabilize cell membrane with ice-cold 0.2% Triton X-100 and 3% BSA for 20 mins with gentle rocking at RT.
5) Aspirate and wash 2x with ice-cold PBS (Ca-; Mg-) (Leave washes in for 5 mins 4°C with gentle rocking)
6) Incubate with 1:500 Primary antibody in 3% BSA overnight with gentle rocking at 4°C
7) Aspirate and wash 2x with ice-cold PBS (Ca-; Mg-) (Leave washes in for 5 mins 4°C with gentle rocking)
8) Incubate with 1:250 secondary streptavidin-conjugated antibody for 1.5 hrs at RT
9) Aspirate and wash 3x with PBS (Ca-; Mg-) (Leave washes in for 5 mins 4°C with gentle rocking)

Chromatin digestion

10) Incubate with 1:30 Biotinylated MNase (5 U/ul original stock) in PBS for 1.5 hrs with gentle rocking at RT

11) Aspirate and wash 2x with PBS (Ca-; Mg-) (Leave washes in for 5 mins 4°C with gentle rocking)

12) Incubate with 20 mM $CaCl_2$ to initiate MNase digestion for 30 mins with gentle rocking at 37°C

13) Gently remove supernatant without disturbing cells on culture substrate

14) Quench MNase activity by adding 95 ul 50 mM EDTA/EGTA

15) Scrape cells from well and place all supernatant in 1.5 ul tube for each sample

16) Add 5 ul 20% SDS, incubate for 10 mins @ 65°C

17) Add 16 ul Reverse cross linking mixture to sample, mix well, seal with parafilm, and incubate overnight (or 5 hours) at 65°C

DNA Purification: AMPure XP Bead Purification

18) Gently shake bottle to resuspend beads and equilibrate 2X AMPure XP Beads to RT

19) Add 200 ul beads to reaction and mix by pipetting 10 times, incubate for 5 mins at RT and place on magnet until clear

20) Aspirate clear solution and discard

21) Wash beads with 200 ul freshly prepared 70% ethanol and incubate for 1 min, aspirate ethanol and discard

22) Repeat once more

23) Quick spin beads, place beads back on magnet, and pipette and discard residual ethanol

24) Leave the tubes open and all the beads to dry

25) Remove beads from magnetic stand and add 52 ul 1x TE buffer, gently mix bead and solution by pipette 20 times

26) Incubate for 5 mins at RT and place on magnetic stand

27) Transfer clear solution to new tube

28) Use 2 ul for Qubit analysis (Skip this step for low input, if skipping this step, only elute in 50 ul 1x TE Buffer in prior step)

29) Proceed to library preparation.

Reverse crosslinking mixture: 250 mM Tris-HCl, pH 6.5, 62.5 mM EDTA pH 8.0, 1.25 M NaCl, 5 mg/ml Proteinase K, 62.5 ng/ul RNase A.

Immunofluorescence staining

Cells were fixed in 1% PFA for 10 mins at RT and the membrane permeabilized with 0.2% Triton X-100 and 3% BSA in PBS for 20 mins with gentle rocking at RT. 1:500 primary antibody in 3% BSA in PBS was added and left overnight with gentle rocking at 4C. 1:250 secondary streptavidin conjugated antibody was added for 1.5 hours at RT with gentle rocking. Atto Biotin-488 was added for 1 hour at RT with gentle rocking. 300nM DAPI was added for 10 minutes at RT with gentle rocking.

MNase biotinylation and antibody-streptavidin conjugation

MNase biotinylation was done using Thermo Scientific EZ-Link Micro NHS-PEG4-Biotinylation Kit per the manufacturer's instructions. Streptavidin conjugation to secondary antibody was performed using abcam's Streptavidin COnjugation Kit per the manufacturer's instructions.

SECTION 5: Conclusions and Future Work

5.1 Conclusions

In this body of work, three separate techniques were developed to probe how DNA-protein interactions affect gene expression during or post- DNA replication in human embryonic stem cells. First, Repli-ATAC-seq was developed to assay how the chromatin landscape develops and changes in post-replication time. This assay provided information regarding the heterogeneity in post-replication time and showed how DNA mismatch repair mechanisms, mitotic bookmarking, and gene expression levels may be involved in the post-replication landscape.

Next, a brief summary of the literature of 10X Single Cell Multiome ATAC and Gene Expression in human samples was performed to understand the tools and ideas available in the scientific community for this nascent technique. This technique was applied in the context of DNA replication in a novel way by incorporating S-phase fractionation into the protocol – a never before documented feat. This unique method proved to be fruitful, providing information suggesting neuroectoderm fate priming in a subset of S-phase cells and highlighting DACH1 as a potential POU5F1 antagonist.

Finally, IL-MNase-seq proved to be irreproducible, but provided valuable experience in understanding how to develop protocols and methods that built the foundation for being able to produce Repli-ATAC-seq and S-phase fraction 10X Multiome sequencing.

5.2 Future Work

The assays in this dissertation deliver only preliminary findings in their respective datasets. Future research about Repli-ATAC-seq will likely need higher resolution by including more time points. Furthermore, it would be prudent to reduce the pulse time of BrdU so that less time has passed since a region has replicated before chase. Furthermore, the analysis of the S-phase fractioned multiome dataset has only scratched the surface. All 12 clusters have the potential to be identified as a cell subtype. Also, the inclusion of the G1 and G2/M peaks would provide increased detail and give a complete view of the cell cycle. This would create a resource that would be used for decades to come to analyze stem cell fate and identity during development.

REFERENCES

1. Gifford, C. A. *et al.* Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells. *Cell* **153**, 1149–1163 (2013).
2. Heyn, H. *et al.* Distinct DNA methylomes of newborns and centenarians. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 10522–10527 (2012).
3. Sharma, S., Kelly, T. K. & Jones, P. A. Epigenetics in cancer. *Carcinogenesis* **31**, 27–36 (2010).
4. Yamanaka, S. Pluripotent Stem Cell-Based Cell Therapy-Promise and Challenges. *Cell Stem Cell* **27**, 523–531 (2020).
5. Almouzni, G. & Cedar, H. Maintenance of Epigenetic Information. *Cold Spring Harb. Perspect. Biol.* **8**, (2016).
6. Budhavarapu, V. N., Chavez, M. & Tyler, J. K. How is epigenetic information maintained through DNA replication? *Epigenetics Chromatin* **6**, 32 (2013).
7. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17**, 487–500 (2016).
8. Ducasse, M. & Brown, M. A. Epigenetic aberrations and cancer. *Mol. Cancer* **5**, 60 (2006).
9. Lu, D. Epigenetic modification enzymes: catalytic mechanisms and inhibitors. *Acta Pharm. Sin. B* **3**, 141–149 (2013).
10. Wang, G. *et al.* Understanding Transcription Factor Regulation by Integrating Gene Expression and DNase I Hypersensitive Sites. *Biomed Res. Int.* **2015**, 757530 (2015).
11. Tyagi, M., Imam, N., Verma, K. & Patel, A. K. Chromatin remodelers: We are the drivers!! *Nucleus* **7**, 388–404 (2016).

12. Dong, X. & Weng, Z. The correlation between histone modifications and gene expression. *Epigenomics* vol. 5 113–116 at https://doi.org/10.2217/epi.13.13 (2013).
13. Ferraz, R. A. C. *et al.* DNA–protein interaction studies: a historical and comparative analysis. *Plant Methods* **17**, 82 (2021).
14. Cozzolino, F., Iacobucci, I., Monaco, V. & Monti, M. Protein-DNA/RNA Interactions: An Overview of Investigation Methods in the -Omics Era. *J. Proteome Res.* **20**, 3018–3030 (2021).
15. Escobar, T. M. *et al.* Active and Repressed Chromatin Domains Exhibit Distinct Nucleosome Segregation during DNA Replication. *Cell* **179**, 953-963.e11 (2019).
16. Shukla, V., Vaissière, T. & Herceg, Z. Histone acetylation and chromatin signature in stem cell identity and cancer. *Mutat. Res. Mol. Mech. Mutagen.* **637**, 1–15 (2008).
17. Buszczak, M. & Spradling, A. C. Searching Chromatin for Stem Cell Identity. *Cell* **125**, 233–236 (2006).
18. Guo, C. & Morris, S. A. Engineering cell identity: establishing new gene regulatory and chromatin landscapes. *Curr. Opin. Genet. Dev.* **46**, 50–57 (2017).
19. Stewart-Morgan, K. R., Reverón-Gómez, N. & Groth, A. Transcription Restart Establishes Chromatin Accessibility after DNA Replication. *Mol. Cell* **75**, 284-297.e6 (2019).
20. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol.*

Biol. **109**, 21.29.1-21.29.9 (2015).

21. Tice, R., Schneider, E. L. & Rary, J. M. The utilization of bromodeoxyuridine incorporation into DNA for the analysis of cellular kinetics. *Exp. Cell Res.* **102**, 232–236 (1976).
22. Marchal, C. *et al.* Genome-wide analysis of replication timing by next-generation sequencing with E/L Repli-seq. *Nat. Protoc.* **13**, 819–839 (2018).
23. Yan, F., Powell, D. R., Curtis, D. J. & Wong, N. C. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biol.* **21**, 22 (2020).
24. Zhang, H. *et al.* Extensive evaluation of ATAC-seq protocols for native or formaldehyde-fixed nuclei. *BMC Genomics* **23**, 214 (2022).
25. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
26. Slotkin, R. K. & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **8**, 272–285 (2007).
27. Charlton, J. *et al.* Global delay in nascent strand DNA methylation. *Nat. Struct. Mol. Biol.* **25**, 327–332 (2018).
28. Charlton, J. *et al.* TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. *Nat. Genet.* **52**, 819–827 (2020).
29. Murat, P., Guilbaud, G. & Sale, J. E. DNA polymerase stalling at structured DNA constrains the expansion of short tandem repeats. *Genome Biol.* **21**, 209 (2020).
30. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **12**, 2478–2492 (2017).

31. Ljungman, M., Parks, L., Hulbatte, R. & Bedi, K. The role of H3K79 methylation in transcription and the DNA damage response. *Mutat. Res. Mutat. Res.* **780**, 48–54 (2019).
32. Huang, Y. & Li, G.-M. DNA mismatch repair in the context of chromatin. *Cell Biosci.* **10**, 10 (2020).
33. Corvalan, A. Z. & Coller, H. A. Methylation of histone 4's lysine 20: a critical analysis of the state of the field. *Physiol. Genomics* **53**, 22–32 (2020).
34. Li, F. *et al.* The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSα. *Cell* **153**, 590–600 (2013).
35. Liu, Y. *et al.* Widespread Mitotic Bookmarking by Histone Marks and Transcription Factors in Pluripotent Stem Cells. *Cell Rep.* **19**, 1283–1293 (2017).
36. Bellec, M. *et al.* The control of transcriptional memory by stable mitotic bookmarking. *Nat. Commun.* **13**, 1176 (2022).
37. Deluz, C. *et al.* A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. *Genes Dev.* **30**, 2538–2550 (2016).
38. Tian, X. *et al.* SIRT6 Is Responsible for More Efficient DNA Double-Strand Break Repair in Long-Lived Species. *Cell* **177**, 622-638.e22 (2019).
39. Meijer, M. *et al.* Epigenomic priming of immune genes implicates oligodendroglia in multiple sclerosis susceptibility. *Neuron* **110**, 1193-1210.e13 (2022).
40. Gur, C. *et al.* LGR5 expressing skin fibroblasts define a major cellular hub perturbed in scleroderma. *Cell* **185**, 1373-1388.e20 (2022).
41. Trevino, A. E. *et al.* Chromatin and gene-regulatory dynamics of the

developing human cerebral cortex at single-cell resolution. *Cell* **184**, 5053-5069.e23 (2021).

42. Montefiori, L. E. *et al.* Enhancer Hijacking Drives Oncogenic BCL11B Expression in Lineage-Ambiguous Stem Cell Leukemia. *Cancer Discov.* **11**, 2846–2867 (2021).
43. Fleck, J. S. *et al.* Inferring and perturbing cell fate regulomes in human brain organoids. *Nature* (2022) doi:10.1038/s41586-022-05279-8.
44. Ma, S. *et al.* Molecular and cellular evolution of the primate dorsolateral prefrontal cortex. *Science (80-.).* **377**, eabo7257 (2023).
45. Anadon, C. M. *et al.* Ovarian cancer immunogenicity is governed by a narrow subset of progenitor tissue-resident memory T cells. *Cancer Cell* **40**, 545-557.e13 (2022).
46. Boukhaled, G. M. *et al.* Pre-encoded responsiveness to type I interferon in the peripheral immune system defines outcome of PD1 blockade therapy. *Nat. Immunol.* **23**, 1273–1283 (2022).
47. Nagaharu, K. *et al.* A bifurcation concept for B-lymphoid/plasmacytoid dendritic cells with largely fluctuating transcriptome dynamics. *Cell Rep.* **40**, 111260 (2022).
48. Ors, A. *et al.* Estrogen regulates divergent transcriptional and epigenetic cell states in breast cancer. *Nucleic Acids Res.* **50**, 11492–11508 (2022).
49. Foster, D. S. *et al.* Multiomic analysis reveals conservation of cancer-associated fibroblast phenotypes across species and tissue of origin. *Cancer Cell* **40**, 1392-1406.e7 (2022).
50. Wang, S. K. *et al.* Single-cell multiome of the human retina and deep learning

nominate causal variants in complex eye diseases. *Cell Genomics* **2**, 100164 (2022).

51. Tsompana, M. & Buck, M. J. Chromatin accessibility: a window into the genome. *Epigenetics Chromatin* **7**, 33 (2014).
52. Friman, E. T. *et al.* Dynamic regulation of chromatin accessibility by pluripotency transcription factors across the cell cycle. *Elife* **8**, e50087 (2019).
53. Groth, A., Rocha, W., Verreault, A. & Almouzni, G. Chromatin Challenges during DNA Replication and Repair. *Cell* **128**, 721–733 (2007).
54. MacAlpine, D. M. & Almouzni, G. Chromatin and DNA replication. *Cold Spring Harb. Perspect. Biol.* **5**, a010207 (2013).
55. Ilicic, T. *et al.* Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* **17**, 29 (2016).
56. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
57. Fouquet, C. *et al.* Robo1 and robo2 control the development of the lateral olfactory tract. *J. Neurosci. Off. J. Soc. Neurosci.* **27**, 3037–3045 (2007).
58. Mori, M., Rikitake, Y., Mandai, K. & Takai, Y. Roles of Nectins and Nectin-Like Molecules in the Nervous System BT - Cell Adhesion Molecules: Implications in Neurological Diseases. in (eds. Berezin, V. & Walmod, P. S.) 91–116 (Springer New York, 2014). doi:10.1007/978-1-4614-8090-7_5.
59. Iacomino M, Baldassari S, Tochigi Y, Kośla K, Buffelli F, Torella A, Severino M, Paladini D, Mandarà L, Riva A, Scala M, Balagura G, Accogli A, Nigro V, Minetti C, Fulcheri E, Zara F, Bednarek AK, Striano P, S. H. and S. V. Loss of Wwox Perturbs Neuronal Migration and Impairs Early Cortical Development. *Front.*

Neurosci. **14**, (2020).

60. Cheng, Y.-Y. *et al.* Wwox deficiency leads to neurodevelopmental and degenerative neuropathies and glycogen synthase kinase 3β-mediated epileptic seizure activity in mice. *Acta Neuropathol. Commun.* **8**, 6 (2020).
61. Yang, R. *et al.* ANK2 autism mutation targeting giant ankyrin-B promotes axon branching and ectopic connectivity. *Proc. Natl. Acad. Sci.* **116**, 15262–15271 (2019).
62. Kawano, S. *et al.* Autism-associated ANK2 regulates embryonic neurodevelopment. *Biochem. Biophys. Res. Commun.* **605**, 45–50 (2022).
63. Kim, Y.-K., Lee, H., Ismail, T., Kim, Y. & Lee, H.-S. Dach1 regulates neural crest migration during embryonic development. *Biochem. Biophys. Res. Commun.* **527**, 896–901 (2020).
64. Bai, Y.-H. *et al.* A Novel Tumor-Suppressor, CDH18, Inhibits Glioma Cell Invasiveness Via UQCRC2 and Correlates with the Prognosis of Glioma Patients. *Cell. Physiol. Biochem. Int. J. Exp. Cell. Physiol. Biochem. Pharmacol.* **48**, 1755–1770 (2018).
65. Derevyanko, A. *et al.* Gene therapy with the TRF1 telomere gene rescues decreased TRF1 levels with aging and prolongs mouse health span. *Aging Cell* **16**, 1353–1368 (2017).
66. Martínez, P. *et al.* Increased telomere fragility and fusions resulting from TRF1 deficiency lead to degenerative pathologies and increased cancer in mice. *Genes Dev.* **23**, 2060–2075 (2009).
67. Rasmussen, A. H., Rasmussen, H. B. & Silahtaroglu, A. The DLGAP family: neuronal expression, function and role in brain disorders. *Mol. Brain* **10**, 43

(2017).

68. Yamada, K., Nomura, N., Yamano, A., Yamada, Y. & Wakamatsu, N. Identification and characterization of splicing variants of PLEKHA5 (Plekha5) during brain development. *Gene* **492**, 270–275 (2012).

69. Camporesi, E. *et al.* Neuroligin-1 in brain and CSF of neurodegenerative disorders: investigation for synaptic biomarkers. *Acta Neuropathol. Commun.* **9**, 19 (2021).

70. Selimi, F. *et al.* Lurcher GRID2-induced death and depolarization can be dissociated in cerebellar Purkinje cells. *Neuron* **37**, 813–819 (2003).

71. Takemoto, M. *et al.* Laminar and areal expression of unc5d and its role in cortical cell survival. *Cereb. Cortex* **21**, 1925–1934 (2011).

72. Zhang, J. *et al.* Filamin A regulates neuronal migration through brefeldin A-inhibited guanine exchange factor 2-dependent Arf1 activation. *J. Neurosci. Off. J. Soc. Neurosci.* **33**, 15735–15746 (2013).

73. Sarpe, I. & Vandin, F. PRESTO: Simple and Scalable Sampling Techniques for the Rigorous Approximation of Temporal Motif Counts. in *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)* 145–153 doi:10.1137/1.9781611976700.17.

74. Bergsland, M. *et al.* Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev.* **25**, 2453–2464 (2011).

75. Avilion, A. A. *et al.* Multipotent cell lineages in early mouse development on SOX2 function. *Genes Dev.* **17**, 126–140 (2003).

76. Dy, P. *et al.* The three SoxC proteins--Sox4, Sox11 and Sox12--exhibit overlapping expression patterns and molecular properties. *Nucleic Acids Res.*

36, 3101–3117 (2008).

77. Taranova, O. V *et al.* SOX2 is a dose-dependent regulator of retinal neural progenitor competence. *Genes Dev.* **20**, 1187–1202 (2006).
78. Toledo, E. M. *et al.* Srebf1 Controls Midbrain Dopaminergic Neurogenesis. *Cell Rep.* **31**, 107601 (2020).
79. Qin, J. *et al.* Mga safeguards embryonic stem cells from acquiring extraembryonic endoderm fates. *Sci. Adv.* **7**, (2021).
80. Washkowitz, A. J. *et al.* Mga is essential for the survival of pluripotent cells during peri-implantation development. *Development* **142**, 31–40 (2015).
81. Llabata, P. *et al.* Multi-Omics Analysis Identifies MGA as a Negative Regulator of the MYC Pathway in Lung Adenocarcinoma. *Mol. Cancer Res.* **18**, 574–584 (2020).
82. Bu, S., Lv, Y., Liu, Y., Qiao, S. & Wang, H. Zinc Finger Proteins in Neuro-Related Diseases Progression. *Front. Neurosci.* **15**, 760567 (2021).
83. Shih, R.-H., Wang, C.-Y. & Yang, C.-M. NF-kappaB Signaling Pathways in Neurological Inflammation: A Mini Review. *Front. Mol. Neurosci.* **8**, 77 (2015).
84. Pearson, C. A. *et al.* Foxp1 Regulates Neural Stem Cell Self-Renewal and Bias Toward Deep Layer Cortical Fates. *Cell Rep.* **30**, 1964-1981.e3 (2020).
85. Schuff, M. *et al.* FoxN3 is required for craniofacial and eye development of Xenopus laevis. *Dev. Dyn.* **236**, 226–239 (2007).
86. Ahlenius, H. *et al.* FoxO3 regulates neuronal reprogramming of cells from postnatal and aging mice. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 8514–8519 (2016).
87. Renault, V. M. *et al.* FoxO3 regulates neural stem cell homeostasis. *Cell Stem*

Cell **5**, 527–539 (2009).

88. Coronel-Córdoba, P. *et al.* FoxK1 is Required for Ectodermal Cell Differentiation During Planarian Regeneration. *Front. cell Dev. Biol.* **10**, 808045 (2022).
89. Hou, P.-S., hAilín, D. Ó., Vogel, T. & Hanashima, C. Transcription and Beyond: Delineating FOXG1 Function in Cortical Development and Disorders. *Front. Cell. Neurosci.* **14**, 35 (2020).
90. Cargnin, F. *et al.* FOXG1 Orchestrates Neocortical Organization and Cortico-Cortical Connections. *Neuron* **100**, 1083-1096.e5 (2018).
91. Oettgen, P. *et al.* Characterization of NERF, a novel transcription factor related to the Ets factor ELF-1. *Mol. Cell. Biol.* **16**, 5091–5106 (1996).
92. Guan, F. H. X. *et al.* The antiproliferative ELF2 isoform, ELF2B, induces apoptosis in vitro and perturbs early lymphocytic development in vivo. *J. Hematol. Oncol.* **10**, 75 (2017).
93. McClellan, K. A. *et al.* Unique requirement for Rb/E2F3 in neuronal migration: evidence for cell cycle-independent functions. *Mol. Cell. Biol.* **27**, 4825–4843 (2007).
94. Julian, L. M. *et al.* Opposing regulation of Sox2 by cell-cycle effectors E2f3a and E2f3b in neural stem cells. *Cell Stem Cell* **12**, 440–452 (2013).
95. Julian, L. M. *et al.* Tissue-specific targeting of cell fate regulatory genes by E2f factors. *Cell Death Differ.* **23**, 565–575 (2016).
96. Caubit, X. *et al.* Mouse Dac, a novel nuclear factor with homology to Drosophila dachshund shows a dynamic expression in the neural crest, the eye, the neocortex, and the limb bud. *Dev. Dyn. an Off. Publ. Am. Assoc.*

Anat. **214**, 66–80 (1999).

97. Wu, K. *et al.* Cell fate determination factor Dachshund reprograms breast cancer stem cell function. *J. Biol. Chem.* **286**, 2132–2142 (2011).
98. Hu, X. *et al.* Organoid modelling identifies that DACH1 functions as a tumour promoter in colorectal cancer by modulating BMP signalling. *EBioMedicine* **56**, 102800 (2020).
99. Bond, A. M., Bhalala, O. G. & Kessler, J. A. The dynamic role of bone morphogenetic proteins in neural stem cell fate and maturation. *Dev. Neurobiol.* **72**, 1068–1084 (2012).
100. Polanco, J., Reyes-Vigil, F., Weisberg, S. D., Dhimitruka, I. & Brusés, J. L. Differential Spatiotemporal Expression of Type I and Type II Cadherins Associated With the Segmentation of the Central Nervous System and Formation of Brain Nuclei in the Developing Mouse. *Front. Mol. Neurosci.* **14**, 633719 (2021).
101. Machon, O., Masek, J., Machonova, O., Krauss, S. & Kozmik, Z. Meis2 is essential for cranial and cardiac neural crest development. *BMC Dev. Biol.* **15**, 40 (2015).
102. Fabik, J., Kovacova, K., Kozmik, Z. & Machon, O. Neural crest cells require Meis2 for patterning the mandibular arch via the Sonic hedgehog pathway. *Biol. Open* **9**, (2020).
103. Lui, N. C. *et al.* Lhx1/5 control dendritogenesis and spine morphogenesis of Purkinje cells via regulation of Espin. *Nat. Commun.* **8**, 15079 (2017).
104. Zhao, Y. *et al.* LIM-homeodomain proteins Lhx1 and Lhx5, and their cofactor Ldb1, control Purkinje cell differentiation in the developing cerebellum. *Proc.*

Natl. Acad. Sci. U. S. A. **104**, 13182–13186 (2007).

105. Heide, M. *et al.* Lhx5 controls mamillary differentiation in the developing hypothalamus of the mouse. *Front. Neuroanat.* **9**, 113 (2015).
106. Sheng, H. Z. *et al.* Expression of murine Lhx5 suggests a role in specifying the forebrain. *Dev. Dyn. an Off. Publ. Am. Assoc. Anat.* **208**, 266–277 (1997).
107. Miquelajáuregui, A. *et al.* LIM-homeobox gene Lhx5 is required for normal development of Cajal-Retzius cells. *J. Neurosci. Off. J. Soc. Neurosci.* **30**, 10551–10562 (2010).
108. Peng, G. & Westerfield, M. Lhx5 promotes forebrain development and activates transcription of secreted Wnt antagonists. *Development* **133**, 3191–3200 (2006).
109. Sedykh, I. *et al.* Zebrafish Rfx4 controls dorsal and ventral midline formation in the neural tube. *Dev. Dyn. an Off. Publ. Am. Assoc. Anat.* **247**, 650–659 (2018).
110. Wang, J. *et al.* Transcription factor induction of human oligodendrocyte progenitor fate and differentiation. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2885-94 (2014).
111. Ashique, A. M. *et al.* The Rfx4 transcription factor modulates Shh signaling by regional control of ciliogenesis. *Sci. Signal.* **2**, ra70 (2009).
112. Tiklová, K. *et al.* Single cell transcriptomics identifies stem cell-derived graft composition in a model of Parkinson's disease. *Nat. Commun.* **11**, 2434 (2020).
113. Kawase, S. *et al.* Regulatory factor X transcription factors control Musashi1

transcription in mouse neural stem/progenitor cells. *Stem Cells Dev.* **23**, 2250–2261 (2014).

114. Xu, P. *et al.* Conditional ablation of the RFX4 isoform 1 transcription factor: Allele dosage effects on brain phenotype. *PLoS One* **13**, e0190561 (2018).
115. Boudjadi, S., Chatterjee, B., Sun, W., Vemu, P. & Barr, F. G. The expression and function of PAX3 in development and disease. *Gene* **666**, 145–157 (2018).
116. Moore, S. *et al.* Distinct regulatory mechanisms act to establish and maintain Pax3 expression in the developing neural tube. *PLoS Genet.* **9**, e1003811 (2013).
117. Sudiwala, S. *et al.* Cellular mechanisms underlying Pax3-related neural tube defects and their prevention by folic acid. *Dis. Model. Mech.* **12**, (2019).
118. Lin, S. *et al.* Aberrant methylation of Pax3 gene and neural tube defects in association with exposure to polycyclic aromatic hydrocarbons. *Clin. Epigenetics* **11**, 13 (2019).
119. Kim, K., Orvis, J. & Stolfi, A. Pax3/7 regulates neural tube closure and patterning in a non-vertebrate chordate. *Front. cell Dev. Biol.* **10**, 999511 (2022).
120. Epstein, J. A. Pax3, Neural Crest and Cardiovascular Development. *Trends Cardiovasc. Med.* **6**, 255–260 (1996).
121. Sato, T., Sasai, N. & Sasai, Y. Neural crest determination by co-activation of Pax3 and Zic1 genes in Xenopus ectoderm. *Development* **132**, 2355–2363 (2005).
122. Parker, H. J., De Kumar, B., Pushel, I., Bronner, M. E. & Krumlauf, R. Analysis

of lamprey meis genes reveals that conserved inputs from Hox, Meis and Pbx proteins control their expression in the hindbrain and neural tube. *Dev. Biol.* **479**, 61–76 (2021).

123. Schulte, D. & Geerts, D. MEIS transcription factors in development and disease. *Development* **146**, (2019).
124. Tucker, E. S. *et al.* Proliferative and transcriptional identity of distinct classes of neural precursors in the mammalian olfactory epithelium. *Development* **137**, 2471–2481 (2010).
125. Ki, S. M., Jeong, H. S. & Lee, J. E. Primary Cilia in Glial Cells: An Oasis in the Journey to Overcoming Neurodegenerative Diseases. *Front. Neurosci.* **15**, 736888 (2021).
126. Bonnafe, E. *et al.* The transcription factor RFX3 directs nodal cilium development and left-right asymmetry specification. *Mol. Cell. Biol.* **24**, 4417–4427 (2004).
127. Chen, B. *et al.* Auto-fatty acylation of transcription factor RFX3 regulates ciliogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E8403–E8412 (2018).
128. Benadiba, C. *et al.* The ciliogenic transcription factor RFX3 regulates early midline distribution of guidepost neurons required for corpus callosum development. *PLoS Genet.* **8**, e1002606 (2012).
129. Lee, J. E. & Gleeson, J. G. Cilia in the nervous system: linking cilia function and neurodevelopmental disorders. *Curr. Opin. Neurol.* **24**, 98–105 (2011).
130. Lim, J. W. C. *et al.* EMX2 transcriptionally regulates Nfib expression in neural progenitor cells during early cortical development. *bioRxiv* 2021.12.26.474186 (2021)

doi:10.1101/2021.12.26.474186.

131. Galli, R. *et al.* Emx2 regulates the proliferation of stem cells of the adult mammalian central nervous system. *Development* **129**, 1633–1644 (2002).
132. Bishop, K. M., Garel, S., Nakagawa, Y., Rubenstein, J. L. R. & O'Leary, D. D. M. Emx1 and Emx2 cooperate to regulate cortical size, lamination, neuronal differentiation, development of cortical efferents, and thalamocortical pathfinding. *J. Comp. Neurol.* **457**, 345–360 (2003).
133. Muzio, L. & Mallamaci, A. Emx1, Emx2 and Pax6 in Specification, Regionalization and Arealization of the Cerebral Cortex. *Cereb. Cortex* **13**, 641–647 (2003).
134. Brancaccio, M., Pivetta, C., Granzotto, M., Filippis, C. & Mallamaci, A. Emx2 and Foxg1 inhibit gliogenesis and promote neuronogenesis. *Stem Cells* **28**, 1206–1218 (2010).
135. Mallamaci, A. *et al.* The Lack of Emx2 Causes Impairment of Reelin Signaling and Defects of Neuronal Migration in the Developing Cerebral Cortex. *J. Neurosci.* **20**, 1109 LP – 1118 (2000).
136. Bishop, K. M., Rubenstein, J. L. R. & O'Leary, D. D. M. Distinct Actions of Emx1, Emx2, and Pax6 in Regulating the Specification of Areas in the Developing Neocortex. *J. Neurosci.* **22**, 7627 LP – 7638 (2002).
137. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
138. Teif, V. B. *et al.* Genome-wide nucleosome positioning during embryonic stem cell development. *Nat. Struct. Mol. Biol.* **19**, 1185–1192 (2012).

139. Park, P. J. ChIP–seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* **10**, 669–680 (2009).

140. Pajoro, A., Muiño, J. M., Angenent, G. C. & Kaufmann, K. Profiling Nucleosome Occupancy by MNase-seq: Experimental Protocol and Computational Analysis. *Methods Mol. Biol.* **1675**, 167–181 (2018).

141. Huang, X. *et al.* Stable H3K4me3 is associated with transcription initiation during early embryo development. *Bioinformatics* **35**, 3931–3936 (2019).

142. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, e21856 (2017).

APPENDIX

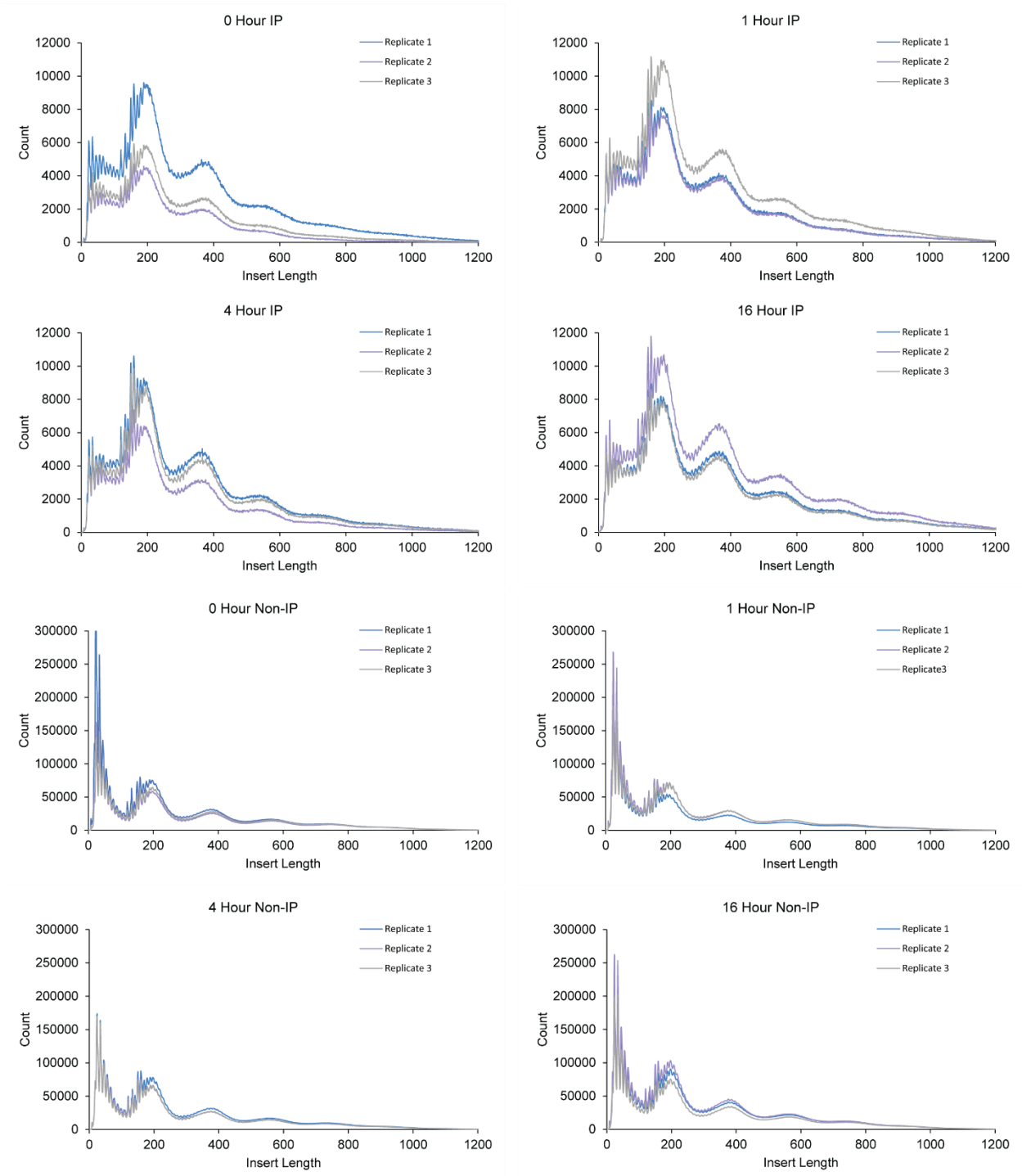


Figure A.1. Smooth curved scatter plots for the counts of fragments at each insert length in Repli-ATAC-seq replicates.

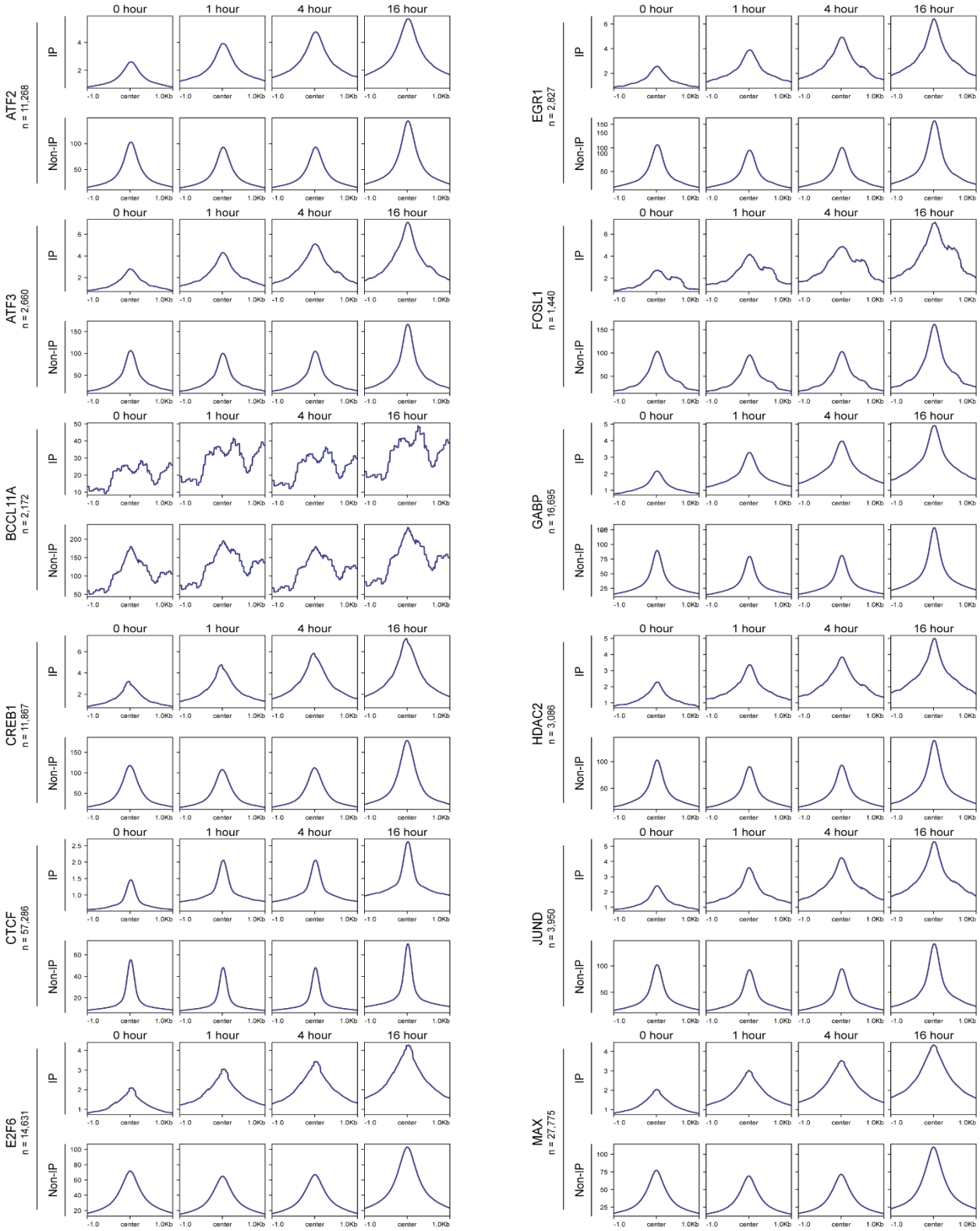


Figure A.2. Various TF binding site enrichment scores for Repli-ATAC-seq IP and Non-IP samples from each time point.

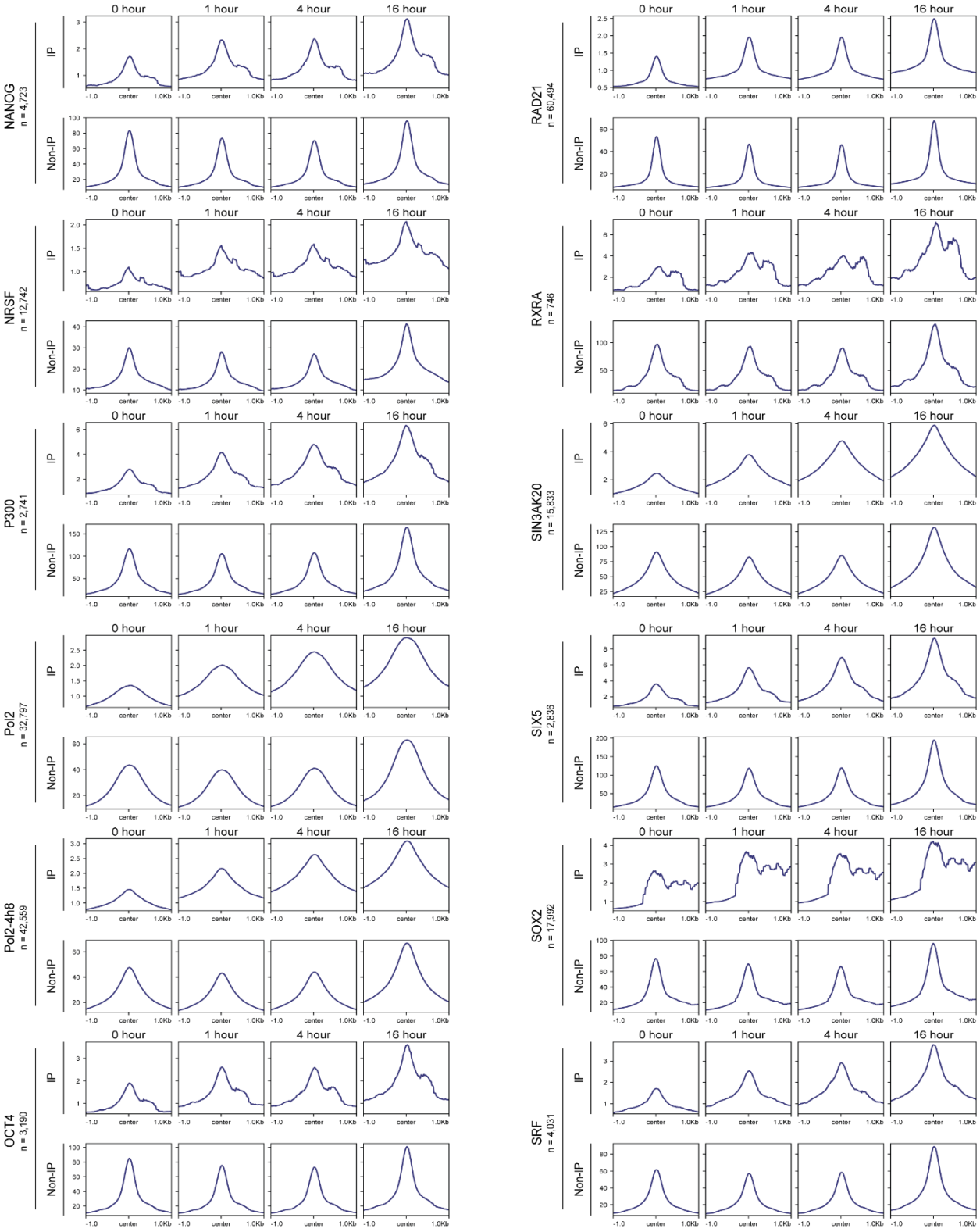


Figure A.3. Various TF binding site enrichment scores for Repli-ATAC-seq IP and Non-IP samples from each time point.

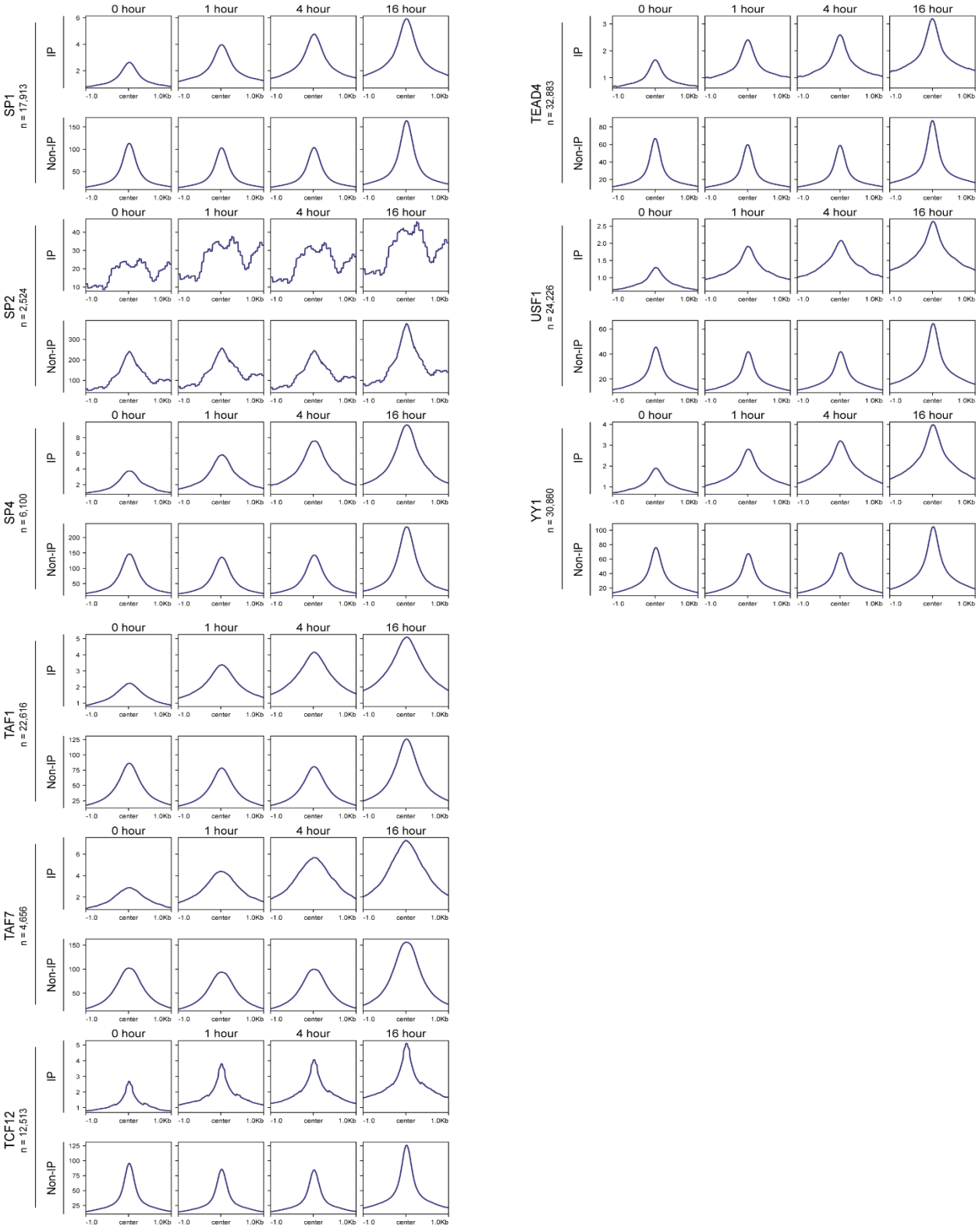


Figure A.4. Various TF binding site enrichment scores for Repli-ATAC-seq IP and Non-IP samples from each time point.

Table A.3.1. Multiome Replicates QC

Sample ID	Run1_S1	Run1_S2	Run1_S3	Run1_S4	Run2_S1	Run2_S2	Run2_S3	Run2_S4
Genome	GRCh38	GRCh38	GRCh38	GRCh38	GRCh38	GRCh38	GRCh38	GRCh38
Pipeline version	arc-2.0.1	arc-2.0.1	arc-2.0.1	arc-2.0.1	arc-2.0.2	arc-2.0.2	arc-2.0.2	arc-2.0.2
Estimated number of cells	1773	2069	2951	2907	7512	4689	2044	4124
Feature linkages detected	1162	1239	1120	1329	5643	11321	2907	14066
Linked genes	706	674	662	772	1573	1747	964	1902
Linked peaks	1408	1538	1355	1659	5349	7816	3330	10954
ATAC Confidently mapped read pairs	0.8835	0.8912	0.8855	0.8884	0.875	0.8745	0.8708	0.8779
ATAC Fraction of genome in peaks	0.0297	0.0278	0.0281	0.0315	0.0736	0.0603	0.0529	0.0672
ATAC Fraction of high-quality fragments in cells	0.9583	0.9622	0.9559	0.9674	0.9731	0.9496	0.9665	0.9699
ATAC Fraction of high-quality fragments overlapping TSS	0.1707	0.1736	0.1811	0.1895	0.2224	0.2307	0.2486	0.256
ATAC Fraction of high-quality fragments overlapping peaks	0.1882	0.183	0.1897	0.2099	0.3775	0.3517	0.3498	0.3895
ATAC Fraction of transposition events in peaks in cells	0.1728	0.1681	0.1734	0.1931	0.3377	0.3136	0.3188	0.35
ATAC Mean raw read pairs per cell	58792	59113	58927	51906	164601	312646	156975	250030
ATAC Median high-quality fragments per cell	29343	30262	34324	30102	77509	133174	87454	110793
ATAC Non-nuclear read pairs	0.0095	0.0097	0.0083	0.006	0.0226	0.0217	0.0218	0.0211
ATAC Number of peaks	105588	98643	101301	112546	276555	226386	196749	251915
ATAC Percent duplicates	0.2767	0.2789	0.2805	0.2801	0.3409	0.4692	0.3232	0.4346
ATAC Q30 bases in barcode	0.9185	0.9252	0.9211	0.9206	0.8933	0.92	0.9031	0.9138
ATAC Q30 bases in read 1	0.9414	0.9406	0.9246	0.9405	0.9367	0.9414	0.9315	0.94
ATAC Q30 bases in read 2	0.9425	0.9445	0.9458	0.9451	0.9181	0.9339	0.93	0.9314

ATAC Q30 bases in sample index i1	0.914	0.9284	0.9143	0.9324	0.9463	0.9191	0.9387	0.9268
ATAC Sequenced read pairs	1.04E+08	1.22E+08	1.74E+08	1.51E+08	1.24E+09	1.47E+09	3.21E+08	1.03E+09
ATAC TSS enrichment score	3.7071	3.6307	3.6791	4.0455	4.9495	4.6511	5.0435	5.3521
ATAC Unmapped read pairs	0.0064	0.0063	0.0066	0.0065	0.0101	0.0099	0.0117	0.0094
ATAC Valid barcodes	0.9877	0.9263	0.9864	0.9881	0.9752	0.9829	0.9789	0.9824
GEX Fraction of transcriptomic reads in cells	0.9184	0.9218	0.8954	0.9423	0.9501	0.9162	0.936	0.9529
GEX Mean raw reads per cell	132528	92260	66856	48130	26638	34932	50510	33090
GEX Median UMI counts per cell	1767	1763	1496	1371	1729	2329	2357.5	2231.5
GEX Median genes per cell	845	814	800	816	929	1164	1165	1204
GEX Percent duplicates	0.9674	0.9598	0.951	0.9278	0.8711	0.875	0.9131	0.8673
GEX Q30 bases in UMI	0.929	0.9288	0.9281	0.9279	0.9466	0.9426	0.9324	0.9426
GEX Q30 bases in barcode	0.9317	0.9313	0.9309	0.9302	0.954	0.9527	0.9466	0.9522
GEX Q30 bases in read 2	0.9207	0.9109	0.9191	0.9076	0.9386	0.9299	0.9348	0.9408
GEX Q30 bases in sample index i1	0.8692	0.9373	0.9199	0.9299	0.9528	0.9618	0.942	0.9375
GEX Q30 bases in sample index i2	0.9253	0.885	0.922	0.8566	0.9286	0.9218	0.9204	0.9583
GEX Reads mapped antisense to gene	0.107	0.0934	0.1434	0.2377	0.1025	0.1104	0.1233	0.1597
GEX Reads mapped confidently to exonic regions	0.5961	0.6218	0.5308	0.3228	0.5497	0.5408	0.5219	0.4314
GEX Reads mapped confidently to genome	0.9213	0.9161	0.9239	0.9188	0.9242	0.9205	0.9086	0.9299
GEX Reads mapped confidently to intergenic regions	0.136	0.1277	0.1264	0.0927	0.11	0.1076	0.114	0.0967
GEX Reads mapped confidently to intronic regions	0.1892	0.1666	0.2667	0.5033	0.2646	0.2721	0.2727	0.4018
GEX Reads mapped confidently to transcriptome	0.6744	0.6913	0.6495	0.5821	0.7062	0.6971	0.6661	0.6674

GEX Reads mapped to genome	0.9604	0.9526	0.9647	0.953	0.9696	0.9635	0.9586	0.9703
GEX Reads with TSO	0.0996	0.0742	0.1239	0.1385	0.1923	0.1333	0.1545	0.1407
GEX Sequenced read pairs	2.35E+08	1.91E+08	1.97E+08	1.40E+08	2.00E+08	1.64E+08	1.03E+08	1.36E+08
GEX Total genes detected	27139	26241	28397	28096	27931	27104	24792	26846
GEX Valid UMIs	0.9995	0.9995	0.9994	0.9994	0.9997	0.9996	0.9996	0.9996
GEX Valid barcodes	0.9568	0.9592	0.9465	0.9336	0.9698	0.9643	0.9562	0.9564

Table A.3.2. Replicate 2 top 10 Up- and Down-regulated DEGs between S-phase fractions

S1 vs S2						
	p-value	Avg Log2FC	pct.1	pct.2	p_val_adj	Expression
ROBO2	1.76E-30	-0.3962563	0.518	0.607	4.29E-26	Down-regulated
RPS10	4.13E-101	-0.3789427	0.36	0.57	1.01E-96	Down-regulated
DIAPH3	1.70E-85	-0.3498519	0.301	0.496	4.16E-81	Down-regulated
RPL39	2.09E-60	-0.3289903	0.588	0.732	5.10E-56	Down-regulated
RPL28	1.95E-36	-0.3280765	0.72	0.801	4.77E-32	Down-regulated
RPL36	5.30E-45	-0.3204986	0.414	0.551	1.29E-40	Down-regulated
RPS19	2.47E-37	-0.3171548	0.748	0.83	6.04E-33	Down-regulated
RPS28	1.18E-44	-0.3139522	0.567	0.691	2.88E-40	Down-regulated
CADM2	4.12E-30	-0.3119336	0.694	0.772	1.01E-25	Down-regulated
RPL41	5.28E-50	-0.3118343	0.896	0.937	1.29E-45	Down-regulated
S1 vs S3						
RPL6	4.82E-28	0.3166611	0.988	0.965	1.18E-23	Up-regulated
DLGAP1	3.96E-21	0.2826533	0.865	0.8	9.66E-17	Up-regulated
RPL4	9.24E-27	0.2816180	0.894	0.825	2.26E-22	Up-regulated
EEF1A1	4.77E-23	0.2534027	0.99	0.968	1.16E-18	Up-regulated
CADM21	1.45E-64	-0.6029503	0.694	0.832	3.55E-60	Down-regulated
WWOX	6.66E-76	-0.5324779	0.557	0.772	1.62E-71	Down-regulated
MTRNR2L12	3.70E-76	-0.4876604	0.958	0.963	9.03E-72	Down-regulated
DACH1	1.09E-08	-0.4630364	0.129	0.181	2.66E-04	Down-regulated
ROBO21	2.16E-18	-0.4397071	0.518	0.605	5.28E-14	Down-regulated
CDH18	3.99E-04	-0.3885473	0.1	0.129	1	Down-regulated
RPS101	6.41E-49	-0.3852980	0.36	0.55	1.57E-44	Down-regulated
HIST1H4C	2.07E-73	-0.3818950	0.25	0.48	5.06E-69	Down-regulated

CTNNA2	9.08E-37	-0.3777780	0.395	0.559	2.22E-32	Down-regulated
DIAPH31	2.83E-49	-0.3739065	0.301	0.492	6.92E-45	Down-regulated
S1 vs S4						
PRKDC	9.50E-252	0.6585263	0.896	0.71	2.32E-247	Up-regulated
RPL61	3.53E-167	0.6301253	0.988	0.888	8.63E-163	Up-regulated
EEF1A11	1.92E-193	0.6061974	0.99	0.906	4.69E-189	Up-regulated
HSP90AB1	2.35E-173	0.5612087	0.861	0.644	5.73E-169	Up-regulated
TPT1	3.01E-95	0.5587674	0.721	0.56	7.34E-91	Up-regulated
RPL42	3.77E-144	0.5275355	0.894	0.73	9.21E-140	Up-regulated
GAPDH	2.37E-128	0.4961101	0.924	0.794	5.78E-124	Up-regulated
ENO1	7.47E-131	0.4882790	0.889	0.73	1.82E-126	Up-regulated
RPL5	3.06E-108	0.4831676	0.957	0.826	7.48E-104	Up-regulated
FLNA	2.50E-121	0.4630679	0.694	0.485	6.10E-117	Up-regulated
ROBO22	4.50E-86	-1.0633361	0.518	0.643	1.10E-81	Down-regulated
WWOX1	6.79E-229	-0.8199320	0.557	0.817	1.66E-224	Down-regulated
NLGN1	2.06E-138	-0.8103524	0.485	0.692	5.03E-134	Down-regulated
GRID2	5.77E-33	-0.7037944	0.854	0.818	1.41E-28	Down-regulated
RBFOX1	2.06E-126	-0.6563775	0.433	0.644	5.03E-122	Down-regulated
IMMP2L	1.21E-145	-0.6472090	0.546	0.761	2.94E-141	Down-regulated
LINGO2	2.22E-184	-0.6164277	0.119	0.365	5.43E-180	Down-regulated
GPC6	4.91E-102	-0.5818877	0.576	0.754	1.20E-97	Down-regulated
CENPE	1.44E-229	-0.5625534	0.124	0.409	3.51E-225	Down-regulated
MYEF2	4.98E-226	-0.5372353	0.162	0.462	1.22E-221	Down-regulated
S2 vs S3						
TERF1	7.84E-15	0.2594579	0.711	0.62	1.91E-10	Up-regulated
RPL62	6.23E-16	0.2551810	0.98	0.965	1.52E-11	Up-regulated
ANK2	1.13E-24	-0.4222046	0.736	0.799	2.76E-20	Down-regulated
DACH11	3.04E-07	-0.3924392	0.13	0.181	7.41E-03	Down-regulated
MTRNR2L121	3.00E-38	-0.3575424	0.954	0.963	7.32E-34	Down-regulated

CDH181	1.56E-03	-0.3061523	0.1	0.129	1	Down-regulated
WWOX2	3.23E-25	-0.2959049	0.653	0.772	7.88E-21	Down-regulated
CADM22	7.43E-17	-0.2910167	0.772	0.832	1.81E-12	Down-regulated
GREB1L	3.07E-13	-0.2704935	0.197	0.283	7.49E-09	Down-regulated
NR6A1	1.73E-10	-0.2703911	0.581	0.627	4.22E-06	Down-regulated
MT-ATP8	1.13E-20	-0.2533685	0.989	0.985	2.76E-16	Down-regulated
S2 vs S4						
EEF1A12	2.69E-157	0.5821011	0.989	0.906	6.56E-153	Up-regulated
HSP90AB11	5.68E-158	0.5774101	0.878	0.644	1.39E-153	Up-regulated
RPL63	8.48E-112	0.5686452	0.98	0.888	2.07E-107	Up-regulated
PRKDC1	1.19E-156	0.5454300	0.891	0.71	2.91E-152	Up-regulated
TPT11	5.31E-86	0.5369015	0.747	0.56	1.30E-81	Up-regulated
RPS25	6.35E-115	0.5190049	0.952	0.846	1.55E-110	Up-regulated
RPS29	2.12E-78	0.5110423	0.833	0.692	5.16E-74	Up-regulated
RPS23	5.31E-95	0.4912102	0.9	0.771	1.30E-90	Up-regulated
GAPDH1	3.97E-101	0.4901820	0.915	0.794	9.69E-97	Up-regulated
FLNA1	1.28E-126	0.4893833	0.736	0.485	3.12E-122	Up-regulated
ANK21	2.17E-147	-0.7271264	0.736	0.882	5.30E-143	Down-regulated
ROBO23	1.51E-19	-0.6670798	0.607	0.643	3.68E-15	Down-regulated
MALAT1	2.13E-55	-0.6468422	0.944	0.958	5.19E-51	Down-regulated
ROR1	3.24E-84	-0.6020003	0.691	0.788	7.92E-80	Down-regulated
WWOX3	1.68E-100	-0.5833590	0.653	0.817	4.09E-96	Down-regulated
RBFOX11	5.18E-71	-0.5695772	0.496	0.644	1.26E-66	Down-regulated
NRXN3	2.38E-76	-0.5573043	0.416	0.597	5.80E-72	Down-regulated
PRKG1	5.15E-70	-0.5329453	0.718	0.821	1.26E-65	Down-regulated
NLGN11	3.76E-44	-0.5291723	0.591	0.692	9.18E-40	Down-regulated
PLEKHA5	9.13E-81	-0.5120595	0.555	0.701	2.23E-76	Down-regulated
S3 vs S4						
MTRNR2L122	1.88E-128	0.7050761	0.963	0.877	4.59E-124	Up-regulated

MT-ATP81	4.52E-68	0.4772670	0.985	0.965	1.10E-63	Up-regulated
ENO11	1.11E-43	0.4442064	0.861	0.73	2.70E-39	Up-regulated
PRKDC2	3.22E-62	0.4437250	0.855	0.71	7.85E-58	Up-regulated
HSP90AB12	4.64E-59	0.4365518	0.843	0.644	1.13E-54	Up-regulated
RPS291	5.96E-28	0.4234082	0.79	0.692	1.46E-23	Up-regulated
FLNA2	4.79E-54	0.4204665	0.693	0.485	1.17E-49	Up-regulated
MT-ND2	3.25E-61	0.4045469	0.992	0.981	7.93E-57	Up-regulated
MT-ND3	2.02E-63	0.4034390	0.995	0.991	4.94E-59	Up-regulated
MT-ND5	1.43E-50	0.3950482	0.974	0.938	3.48E-46	Up-regulated
GRID21	4.43E-19	-0.6324292	0.791	0.818	1.08E-14	Down-regulated
ROBO24	5.11E-11	-0.6236289	0.605	0.643	1.25E-06	Down-regulated
DLGAP11	2.32E-33	-0.5739120	0.8	0.851	5.67E-29	Down-regulated
NLGN12	1.82E-23	-0.5092912	0.597	0.692	4.45E-19	Down-regulated
PLEKHA51	2.68E-49	-0.4983000	0.529	0.701	6.54E-45	Down-regulated
RBFOX12	2.50E-23	-0.4306536	0.544	0.644	6.11E-19	Down-regulated
UNC5D	2.69E-21	-0.4173013	0.738	0.799	6.56E-17	Down-regulated
PTPRG	1.55E-24	-0.4157366	0.848	0.879	3.78E-20	Down-regulated
GULP1	7.96E-35	-0.4081698	0.626	0.749	1.94E-30	Down-regulated
MALAT11	3.06E-13	-0.4076441	0.952	0.958	7.48E-09	Down-regulated

Table A.3.3. Replicate 2 cluster gene markers

Gene	p-value	Avg Log2FC	Pct.1	Pct.2	Adj. p-value	Cluster
RPL6	3.47E-143	0.5144481	0.994	0.954	8.47E-139	0
TERF1	1.16E-128	0.4407794	0.848	0.612	2.83E-124	0
MT-CO2	6.87E-112	0.3138252	1	0.999	1.68E-107	0
TPT1	6.21E-107	0.5202396	0.824	0.66	1.51E-102	0
RPL4	7.37E-106	0.4302548	0.941	0.83	1.80E-101	0
RPL5	1.47E-104	0.4584256	0.978	0.909	3.58E-100	0
MT-ATP6	1.61E-103	0.2798833	1	0.999	3.93E-99	0
MT-CYB	5.18E-98	0.2858903	1	0.999	1.26E-93	0
RPS25	4.01E-96	0.4156213	0.971	0.914	9.78E-92	0
MT-ND3	5.90E-96	0.2803985	1	0.996	1.44E-91	0
MT-CO3	1.08E-94	0.2673095	1	1	2.64E-90	0
GRID2	5.82E-91	0.2631713	0.956	0.811	1.42E-86	0
EEF1A1	3.77E-89	0.4033690	0.993	0.964	9.19E-85	0
MT-ND4	4.60E-89	0.2631770	1	0.996	1.12E-84	0
RPS23	6.58E-87	0.4255324	0.934	0.848	1.61E-82	0
RPS24	4.94E-84	0.3710826	0.931	0.863	1.21E-79	0
AL354821.1	6.73E-84	0.3172822	0.619	0.401	1.64E-79	0
MT-ATP8	5.48E-82	0.2892678	0.998	0.983	1.34E-77	0
RPS8	4.97E-81	0.3218365	0.956	0.872	1.21E-76	0
RPS27A	1.55E-80	0.3815273	0.941	0.863	3.77E-76	0
AC022140.1	7.54E-77	0.3335513	0.837	0.67	1.84E-72	0
MT-ND5	3.38E-76	0.2778752	0.994	0.97	8.25E-72	0
USP9X	3.63E-76	0.3233588	0.758	0.574	8.87E-72	0
RPL17	7.10E-76	0.3546416	0.877	0.779	1.73E-71	0
SNHG14	3.28E-75	0.2652765	0.992	0.969	8.02E-71	0
APOE	7.07E-73	0.2879314	0.98	0.931	1.72E-68	0
RPL26	1.21E-71	0.3999374	0.863	0.76	2.95E-67	0
UGP2	1.83E-70	0.3428999	0.686	0.504	4.47E-66	0
RPL21	2.24E-66	0.3783998	0.777	0.629	5.46E-62	0
RPL32	2.37E-62	0.3281345	0.812	0.688	5.79E-58	0
SLC16A1	9.79E-60	0.2880056	0.617	0.451	2.39E-55	0
VCAN	4.24E-58	0.2843222	0.612	0.441	1.04E-53	0
PABPC1	1.18E-57	0.3103723	0.955	0.918	2.87E-53	0
IGFBP2	2.54E-57	0.3090643	0.947	0.887	6.20E-53	0
ZNF208	2.36E-55	0.2622785	0.597	0.429	5.76E-51	0
RPS3A	3.65E-55	0.3224417	0.823	0.716	8.92E-51	0
RPL14	8.45E-54	0.2805013	0.768	0.63	2.06E-49	0
RPS6	1.68E-52	0.3085801	0.826	0.726	4.09E-48	0

MFGE8	5.45E-52	0.2727032	0.571	0.422	1.33E-47	0
CD24	4.81E-51	0.2539875	0.52	0.363	1.17E-46	0
NAP1L1	8.66E-48	0.2511697	0.971	0.926	2.11E-43	0
RPL7	1.81E-46	0.2794509	0.709	0.568	4.43E-42	0
AASS	1.23E-44	0.2502317	0.766	0.63	3.01E-40	0
RPL31	1.03E-43	0.2511039	0.802	0.686	2.50E-39	0
RPS20	2.71E-43	0.2686606	0.804	0.708	6.62E-39	0
SNHG5	3.00E-39	0.2563900	0.581	0.446	7.33E-35	0
TERF11	0.00E+00	0.8411328	0.913	0.604	0.00E+00	1
GRID21	6.55E-224	0.5938581	0.985	0.808	1.60E-219	1
KIAA0825	1.83E-164	0.5609443	0.713	0.403	4.47E-160	1
MT-ND41	1.09E-157	0.3617643	0.999	0.996	2.66E-153	1
MT-CO21	1.38E-153	0.3719528	1	0.999	3.38E-149	1
MT-ATP61	1.87E-146	0.3353499	1	0.999	4.55E-142	1
KCND2	7.26E-145	0.5298703	0.753	0.47	1.77E-140	1
AL354821.11	1.22E-132	0.4171740	0.675	0.394	2.99E-128	1
MT-CO31	1.71E-130	0.3130924	1	1	4.17E-126	1
MT-ND1	4.08E-128	0.3169496	0.998	0.996	9.97E-124	1
ZNF2081	1.69E-113	0.3948666	0.659	0.42	4.12E-109	1
RIMS2	4.89E-111	0.4285142	0.784	0.556	1.19E-106	1
MT-CO1	8.14E-111	0.2649427	1	1	1.99E-106	1
MT-CYB1	5.87E-109	0.3052783	1	0.999	1.43E-104	1
MT-ND2	8.34E-108	0.3077802	0.997	0.992	2.04E-103	1
CCNG1	3.46E-102	0.3952331	0.626	0.392	8.44E-98	1
MT-ND51	7.69E-97	0.3221462	0.994	0.97	1.88E-92	1
AASS1	2.30E-95	0.3598037	0.833	0.62	5.62E-91	1
ROBO2	2.04E-83	0.3952809	0.74	0.548	4.97E-79	1
MT-ND31	4.70E-78	0.2604337	0.999	0.996	1.15E-73	1
MT-ATP81	7.90E-78	0.2843167	0.994	0.984	1.93E-73	1
NOP56	6.37E-70	0.3244345	0.692	0.498	1.55E-65	1
RPS21	1.75E-66	0.2902659	0.961	0.907	4.27E-62	1
L1TD1	4.23E-65	0.2818332	0.608	0.415	1.03E-60	1
VCAN1	3.82E-63	0.2937731	0.626	0.441	9.32E-59	1
FAM155A	2.21E-61	0.3156499	0.79	0.609	5.40E-57	1
TPT11	6.52E-61	0.3771107	0.8	0.667	1.59E-56	1
AC009446.1	3.85E-60	0.2501700	0.549	0.358	9.39E-56	1
AC022140.11	1.35E-59	0.3159162	0.82	0.676	3.30E-55	1
MAGI2	2.12E-59	0.2896943	0.598	0.413	5.18E-55	1
MTRNR2L12	2.41E-49	0.2514178	0.969	0.934	5.89E-45	1
FOXN3	1.21E-47	0.2608559	0.711	0.539	2.95E-43	1
AL157778.1	7.05E-39	0.2996642	0.707	0.583	1.72E-34	1

ENO1	3.27E-94	0.4614196	0.953	0.837	7.99E-90	2
DAB1	3.81E-83	0.5721934	0.756	0.606	9.31E-79	2
GAPDH	5.40E-83	0.4300594	0.965	0.88	1.32E-78	2
VIM	7.64E-80	0.4874520	0.558	0.356	1.86E-75	2
RPL3	5.72E-69	0.4097345	0.909	0.808	1.40E-64	2
TUBA1A	6.29E-58	0.4211327	0.631	0.472	1.53E-53	2
FZD7	2.54E-56	0.4044677	0.544	0.383	6.21E-52	2
HSP90AB1	7.90E-50	0.3529774	0.899	0.805	1.93E-45	2
TUBB	1.50E-43	0.3252053	0.838	0.744	3.67E-39	2
PRKDC	9.20E-42	0.3139408	0.902	0.842	2.25E-37	2
TUBA1B	1.13E-40	0.3238602	0.804	0.711	2.75E-36	2
ANK2	1.35E-40	0.4432694	0.864	0.784	3.29E-36	2
NR6A1	2.28E-35	0.3272182	0.714	0.615	5.56E-31	2
PRKG1	3.22E-34	0.3877189	0.811	0.738	7.86E-30	2
MAPK10	3.62E-34	0.3497261	0.513	0.384	8.84E-30	2
XACT	3.99E-34	0.3688046	0.805	0.713	9.74E-30	2
PGK1	1.30E-29	0.2852786	0.623	0.522	3.17E-25	2
PODXL	7.89E-28	0.2967872	0.652	0.582	1.92E-23	2
SFRP1	9.05E-28	0.3174246	0.619	0.548	2.21E-23	2
SFRP2	6.03E-22	0.2521508	0.62	0.547	1.47E-17	2
SON	2.97E-21	0.2603605	0.615	0.566	7.24E-17	2
DLGAP1	1.71E-18	0.2711788	0.899	0.842	4.17E-14	2
RACK1	1.59E-17	0.2534933	0.599	0.554	3.88E-13	2
WWOX	1.18E-127	0.6329658	0.881	0.639	2.89E-123	3
DAB11	3.09E-106	0.7018491	0.804	0.602	7.53E-102	3
FZD71	9.84E-75	0.4573230	0.591	0.379	2.40E-70	3
CADM2	9.25E-65	0.5834399	0.862	0.745	2.26E-60	3
XACT1	1.38E-62	0.5748826	0.834	0.711	3.37E-58	3
TUBA1B1	2.08E-46	0.3484796	0.83	0.71	5.08E-42	3
RPL33	5.15E-43	0.3334753	0.889	0.811	1.26E-38	3
ENO11	2.04E-37	0.2761936	0.926	0.841	4.98E-33	3
VIM1	4.71E-37	0.2937868	0.521	0.362	1.15E-32	3
COL1A1	2.98E-36	0.3276667	0.546	0.397	7.27E-32	3
NRG3	3.13E-35	0.3119891	0.793	0.68	7.65E-31	3
TUBB1	6.19E-35	0.2880120	0.833	0.745	1.51E-30	3
NRXN3	9.19E-35	0.3934581	0.581	0.442	2.24E-30	3
TUBA1A1	2.87E-34	0.2619149	0.625	0.474	7.00E-30	3
CTNNA2	5.23E-30	0.3160751	0.608	0.477	1.28E-25	3
PRKDC1	9.75E-29	0.2523128	0.909	0.842	2.38E-24	3
COX6A1	4.49E-28	0.2789362	0.78	0.699	1.10E-23	3
FASN	3.06E-25	0.2576985	0.585	0.464	7.47E-21	3

MIR924HG	5.36E-20	0.2745403	0.686	0.59	1.31E-15	3
SFRP21	1.40E-18	0.2521269	0.629	0.547	3.42E-14	3
ANK21	5.58E-18	0.2612413	0.838	0.788	1.36E-13	3
GRID22	5.31E-137	0.7211856	0.982	0.823	1.30E-132	4
ROBO21	6.42E-100	0.7292429	0.8	0.558	1.57E-95	4
TERF12	2.32E-91	0.5467738	0.877	0.632	5.66E-87	4
KIAA08251	7.13E-91	0.6455294	0.702	0.429	1.74E-86	4
MGAT4C	1.49E-81	0.5621873	0.555	0.299	3.63E-77	4
AL354821.12	2.25E-79	0.4696473	0.685	0.416	5.49E-75	4
ANK3	5.62E-59	0.3809128	0.64	0.404	1.37E-54	4
AASS2	1.08E-58	0.4195280	0.829	0.638	2.63E-54	4
KCND21	2.40E-52	0.4097447	0.731	0.495	5.86E-48	4
AC009446.11	1.14E-51	0.3803067	0.581	0.37	2.78E-47	4
CADM21	1.73E-51	0.4610185	0.891	0.744	4.23E-47	4
BNC2	4.41E-45	0.3771868	0.66	0.458	1.08E-40	4
SNHG141	6.11E-42	0.3182889	0.988	0.971	1.49E-37	4
TCF7L2	8.29E-41	0.3579318	0.611	0.429	2.02E-36	4
RIMS21	6.90E-40	0.3382061	0.759	0.577	1.68E-35	4
IMMP2L	2.79E-37	0.3794340	0.79	0.624	6.82E-33	4
PCDH11X	9.45E-36	0.4077525	0.691	0.527	2.31E-31	4
FOXN31	4.70E-35	0.3424531	0.723	0.552	1.15E-30	4
FAM155A1	5.82E-35	0.3397013	0.798	0.623	1.42E-30	4
L1TD11	1.13E-30	0.2582031	0.598	0.432	2.77E-26	4
MTRNR2L121	1.42E-28	0.2565474	0.958	0.938	3.47E-24	4
FGD4	1.99E-28	0.3055086	0.625	0.472	4.85E-24	4
BBS9	2.02E-28	0.3586303	0.521	0.369	4.93E-24	4
JARID2	7.57E-28	0.3269987	0.68	0.533	1.85E-23	4
NRXN1	3.50E-25	0.2868771	0.83	0.7	8.54E-21	4
YBX1	1.89E-14	0.2867157	0.706	0.619	4.60E-10	4
CENPE	6.75E-183	0.6720945	0.504	0.179	1.65E-178	5
MYEF2	5.33E-169	0.6374000	0.551	0.219	1.30E-164	5
WWOX1	1.21E-132	0.8994921	0.843	0.646	2.95E-128	5
NRXN31	1.24E-126	0.9139747	0.701	0.433	3.02E-122	5
ANK22	1.99E-113	0.8615515	0.904	0.782	4.85E-109	5
EFNA5	9.41E-113	0.7746007	0.796	0.569	2.30E-108	5
LSAMP	5.11E-112	0.7655711	0.713	0.459	1.25E-107	5
ST6GALNAC5	1.11E-106	0.7178929	0.513	0.249	2.72E-102	5
BCKDHB	1.83E-97	0.6215219	0.737	0.501	4.46E-93	5
SSBP2	1.12E-93	0.6882231	0.604	0.356	2.74E-89	5
IL1RAPL1	8.95E-92	0.8715030	0.521	0.278	2.19E-87	5
TOP2A	8.87E-87	0.5922630	0.624	0.391	2.17E-82	5

PRKG11	1.54E-85	0.7249174	0.855	0.736	3.75E-81	5
RMST	2.46E-81	0.7285198	0.754	0.563	6.01E-77	5
ASPM	8.04E-81	0.4641158	0.5	0.268	1.96E-76	5
KPNA2	1.03E-77	0.6100623	0.623	0.406	2.51E-73	5
UBE2E2	8.46E-72	0.5451422	0.751	0.575	2.07E-67	5
ROR1	3.95E-71	0.6484584	0.82	0.722	9.64E-67	5
PAM	5.19E-71	0.5774822	0.579	0.359	1.27E-66	5
SMC4	1.17E-70	0.4728784	0.598	0.382	2.86E-66	5
PTPRD	6.25E-69	0.6604755	0.697	0.513	1.53E-64	5
PLEKHA5	1.80E-67	0.5612669	0.732	0.56	4.39E-63	5
ROBO1	7.39E-66	0.6336786	0.724	0.558	1.80E-61	5
ADGRL2	4.12E-59	0.5197518	0.761	0.595	1.01E-54	5
FBXL7	2.53E-58	0.5003600	0.698	0.511	6.16E-54	5
LINC01473	5.27E-58	0.4834196	0.689	0.511	1.29E-53	5
TBL1XR1	5.57E-55	0.4751983	0.623	0.441	1.36E-50	5
FTX	6.26E-53	0.4969834	0.798	0.687	1.53E-48	5
DIAPH3	2.26E-52	0.4037861	0.611	0.414	5.52E-48	5
RBFOX1	4.97E-52	0.5947733	0.65	0.495	1.21E-47	5
SYT1	1.80E-50	0.4960949	0.757	0.62	4.40E-46	5
TENM3	2.83E-50	0.4796623	0.857	0.773	6.92E-46	5
EXT1	2.99E-50	0.4816275	0.657	0.482	7.29E-46	5
NRG31	5.64E-50	0.5771840	0.801	0.681	1.38E-45	5
PPP1R9A	3.81E-49	0.4375295	0.607	0.435	9.30E-45	5
MAPK101	4.72E-49	0.4642533	0.572	0.382	1.15E-44	5
PTBP2	1.75E-47	0.3955043	0.653	0.482	4.27E-43	5
LGR4	8.18E-46	0.4718271	0.52	0.351	2.00E-41	5
NR6A11	8.35E-46	0.5360099	0.726	0.616	2.04E-41	5
GRIP1	2.19E-45	0.4957012	0.587	0.433	5.34E-41	5
RAD51B	2.36E-45	0.5514805	0.73	0.624	5.77E-41	5
SPATA5	2.77E-45	0.3885873	0.569	0.387	6.76E-41	5
ARL15	4.78E-45	0.4648775	0.618	0.469	1.17E-40	5
GPC6	9.78E-44	0.5008421	0.762	0.647	2.39E-39	5
PUM2	1.22E-43	0.3624946	0.54	0.361	2.97E-39	5
FMNL2	3.24E-43	0.4396762	0.567	0.409	7.90E-39	5
DLG2	7.92E-43	0.4926945	0.631	0.486	1.93E-38	5
R3HDM1	9.19E-43	0.3872594	0.558	0.391	2.24E-38	5
POU2F1	4.19E-40	0.3925697	0.546	0.389	1.02E-35	5
VPS13B	3.16E-38	0.3668114	0.563	0.406	7.72E-34	5
RBM26	3.89E-38	0.3486263	0.589	0.432	9.50E-34	5
ANKRD17	7.86E-38	0.3783716	0.664	0.527	1.92E-33	5
ARID1B	1.96E-37	0.4245395	0.67	0.539	4.78E-33	5

UST	2.90E-35	0.4150158	0.572	0.425	7.07E-31	5
DAB12	7.11E-35	0.5689573	0.723	0.613	1.73E-30	5
CTNNA21	2.31E-34	0.3795762	0.625	0.477	5.64E-30	5
SUMF1	2.41E-34	0.3415697	0.566	0.415	5.87E-30	5
MACROD2	8.36E-34	0.4041705	0.535	0.388	2.04E-29	5
CENPF	9.33E-34	0.3494912	0.612	0.47	2.28E-29	5
JMJD1C	1.30E-33	0.4054498	0.726	0.631	3.18E-29	5
DLGAP11	1.31E-33	0.6179170	0.867	0.847	3.19E-29	5
IMMP2L1	3.28E-33	0.4338259	0.74	0.628	8.00E-29	5
NIPBL	7.03E-33	0.3487817	0.517	0.373	1.72E-28	5
MAGI1	8.73E-33	0.4340751	0.724	0.635	2.13E-28	5
BBX	1.08E-32	0.3578379	0.518	0.379	2.64E-28	5
ADGRA3	2.94E-32	0.3329227	0.525	0.38	7.19E-28	5
RBPMS	8.02E-32	0.4350500	0.649	0.536	1.96E-27	5
SMYD3	9.36E-32	0.4184402	0.713	0.605	2.28E-27	5
PKN2	2.45E-31	0.3069206	0.521	0.372	5.99E-27	5
CHD7	4.31E-31	0.3010479	0.585	0.432	1.05E-26	5
LINC01876	6.17E-31	0.4401801	0.639	0.529	1.51E-26	5
TTC28	1.09E-30	0.4116447	0.581	0.45	2.65E-26	5
JAKMIP2-AS1	2.04E-30	0.4776282	0.621	0.514	4.98E-26	5
AUTS2	2.66E-30	0.4278879	0.604	0.489	6.50E-26	5
FNDC3B	4.58E-30	0.3203819	0.516	0.372	1.12E-25	5
BCAS3	6.13E-30	0.3789258	0.52	0.392	1.50E-25	5
TCF4	3.17E-29	0.4461013	0.721	0.652	7.75E-25	5
CASC15	4.28E-29	0.3569800	0.748	0.67	1.04E-24	5
JAKMIP2	1.04E-28	0.3316336	0.536	0.401	2.54E-24	5
LIN28B	7.31E-28	0.3385471	0.576	0.453	1.78E-23	5
TCF12	1.13E-26	0.4216097	0.598	0.497	2.77E-22	5
SRSF11	1.35E-26	0.2790855	0.538	0.407	3.30E-22	5
PHIP	3.11E-26	0.2945849	0.599	0.474	7.59E-22	5
PCCA	5.31E-26	0.3319692	0.573	0.456	1.30E-21	5
ST6GALNAC3	1.94E-25	0.3110246	0.642	0.519	4.73E-21	5
TBC1D5	2.51E-25	0.3278175	0.657	0.541	6.12E-21	5
EXOC4	2.94E-25	0.3763971	0.602	0.506	7.18E-21	5
NLGN1	3.66E-25	0.3907511	0.663	0.562	8.93E-21	5
PKIB	6.62E-25	0.3157387	0.636	0.522	1.61E-20	5
GULP1	6.20E-24	0.3128914	0.747	0.65	1.51E-19	5
IGF1R	1.64E-23	0.3225883	0.668	0.578	4.00E-19	5
SUPT3H	2.00E-23	0.3303204	0.624	0.524	4.89E-19	5
MIR924HG1	2.47E-23	0.4541568	0.662	0.594	6.02E-19	5
LINC02315	2.75E-22	0.4029394	0.613	0.54	6.72E-18	5

CADM22	5.84E-22	0.3490272	0.807	0.752	1.43E-17	5
PBX3	6.99E-22	0.4352872	0.59	0.5	1.71E-17	5
ZMYM2	9.29E-22	0.3018153	0.554	0.443	2.27E-17	5
ZNF292	1.45E-21	0.2778500	0.585	0.47	3.53E-17	5
TRIO	3.46E-21	0.2544074	0.517	0.393	8.45E-17	5
CDKAL1	3.86E-21	0.2647136	0.505	0.385	9.42E-17	5
ADGRL3	7.95E-21	0.3497609	0.572	0.468	1.94E-16	5
NBEA	4.05E-19	0.3175076	0.619	0.544	9.89E-15	5
PARD3	3.46E-18	0.3412186	0.598	0.52	8.45E-14	5
XACT2	4.95E-16	0.4908042	0.756	0.72	1.21E-11	5
IGF2BP2	8.51E-16	0.2662646	0.524	0.444	2.08E-11	5
ADGRV1	1.11E-15	0.2520109	0.742	0.679	2.72E-11	5
PRELID2	2.65E-15	0.2672052	0.503	0.407	6.46E-11	5
PTK2	1.99E-13	0.2526544	0.541	0.459	4.85E-09	5
MALAT1	2.30E-13	0.4294557	0.956	0.963	5.62E-09	5
NRXN11	3.51E-12	0.3019982	0.734	0.709	8.56E-08	5
PBX1	2.62E-11	0.2697020	0.696	0.653	6.39E-07	5
PTPRG	2.78E-10	0.3025724	0.851	0.859	6.78E-06	5
LRRTM4	4.91E-04	0.2943739	0.564	0.569	1.00E+00	5
CDH18	0.00E+00	2.3238132	0.627	0.079	0.00E+00	6
PRTG	0.00E+00	1.6135636	0.865	0.273	0.00E+00	6
DACH1	0.00E+00	1.5935690	0.653	0.116	0.00E+00	6
ZFHX4	1.80E-233	0.9377319	0.553	0.146	4.40E-229	6
EPHA7	3.38E-218	1.1046961	0.572	0.166	8.26E-214	6
PCDH7	2.13E-164	1.1815392	0.572	0.214	5.21E-160	6
SOX5	6.02E-133	0.9116610	0.535	0.211	1.47E-128	6
TUBA1A2	8.40E-130	0.9049213	0.755	0.472	2.05E-125	6
CDH2	5.24E-128	0.7295518	0.613	0.268	1.28E-123	6
ENO12	1.18E-111	0.7684030	0.963	0.842	2.89E-107	6
GREB1L	5.09E-108	0.6982405	0.568	0.247	1.24E-103	6
EFNA51	1.47E-98	0.7665740	0.815	0.573	3.58E-94	6
MAPK102	4.18E-98	0.8608348	0.658	0.381	1.02E-93	6
CDK14	1.02E-91	0.8019917	0.557	0.292	2.48E-87	6
PTPRD1	1.06E-84	0.7302276	0.738	0.515	2.58E-80	6
TLE4	4.80E-82	0.8497774	0.611	0.368	1.17E-77	6
PKM	1.46E-67	0.6378992	0.704	0.531	3.56E-63	6
IGFBP21	2.69E-61	0.5792989	0.962	0.893	6.58E-57	6
PGK11	1.05E-55	0.6432523	0.685	0.523	2.56E-51	6
LSAMP1	3.68E-54	0.5996824	0.664	0.468	8.97E-50	6
ADGRL31	3.35E-51	0.6314197	0.641	0.467	8.18E-47	6
SSBP21	3.86E-51	0.5703422	0.569	0.364	9.42E-47	6

VIM2	3.70E-50	0.4954521	0.58	0.365	9.04E-46	6
IL1RAPL11	1.67E-48	0.4857916	0.507	0.286	4.06E-44	6
GPC61	9.96E-48	0.5287155	0.792	0.648	2.43E-43	6
TTC281	2.65E-46	0.5114885	0.638	0.45	6.47E-42	6
GAPDH1	2.77E-43	0.4312092	0.958	0.885	6.76E-39	6
ROBO11	2.67E-40	0.6431708	0.685	0.564	6.53E-36	6
TENM31	8.33E-39	0.4366497	0.861	0.775	2.03E-34	6
MDK	9.55E-38	0.5028427	0.617	0.483	2.33E-33	6
PAM1	5.11E-37	0.4571148	0.533	0.368	1.25E-32	6
SLC2A1	4.91E-35	0.4529050	0.552	0.406	1.20E-30	6
RPL34	3.33E-34	0.3572818	0.9	0.814	8.12E-30	6
TUBB2	8.07E-30	0.3456878	0.843	0.748	1.97E-25	6
PTPRS	1.81E-27	0.3809727	0.559	0.439	4.42E-23	6
TCF121	3.24E-27	0.4350993	0.618	0.499	7.90E-23	6
NR6A12	3.33E-26	0.3431944	0.73	0.619	8.12E-22	6
H3F3B	1.75E-25	0.3735296	0.501	0.372	4.28E-21	6
TUBA1B2	4.30E-24	0.3399857	0.799	0.716	1.05E-19	6
RMST1	5.90E-23	0.6186686	0.641	0.575	1.44E-18	6
PRKG12	1.13E-21	0.3605334	0.815	0.742	2.75E-17	6
ADGRL21	4.11E-21	0.3745195	0.681	0.605	1.00E-16	6
SYNE2	2.08E-18	0.3595059	0.52	0.425	5.08E-14	6
AUTS21	3.21E-17	0.3363974	0.588	0.494	7.84E-13	6
FASN1	3.93E-15	0.2971550	0.549	0.471	9.60E-11	6
MACF1	2.15E-12	0.2864444	0.604	0.56	5.24E-08	6
CTNNA22	7.76E-12	0.3360610	0.55	0.486	1.89E-07	6
DACH11	0.00E+00	1.8814215	0.769	0.113	0.00E+00	7
DLK1	0.00E+00	1.5493285	0.565	0.078	0.00E+00	7
MECOM	0.00E+00	1.2827160	0.518	0.063	0.00E+00	7
LRP2	3.63E-265	0.8715423	0.501	0.099	8.86E-261	7
PRTG1	4.47E-249	1.2010088	0.792	0.281	1.09E-244	7
TPBG	2.30E-240	0.8827429	0.518	0.117	5.61E-236	7
CDH21	4.49E-238	1.1501204	0.737	0.263	1.10E-233	7
ZFHX41	2.15E-233	0.9926392	0.575	0.147	5.25E-229	7
NAV3	8.70E-226	1.4719901	0.647	0.207	2.12E-221	7
GPC3	3.65E-203	1.2754733	0.793	0.381	8.92E-199	7
EPHA71	1.21E-177	1.2481769	0.547	0.17	2.95E-173	7
ZNF521	6.16E-162	1.1210074	0.658	0.283	1.50E-157	7
EFNA52	6.75E-150	1.2193782	0.869	0.572	1.65E-145	7
SOX51	2.11E-147	0.9688597	0.57	0.211	5.15E-143	7
AC233296.1	1.44E-124	0.7718601	0.534	0.206	3.52E-120	7
GREB1L1	4.08E-120	0.8726823	0.608	0.247	9.95E-116	7

ADGRL32	3.03E-113	0.9787462	0.741	0.462	7.39E-109	7
IL1RAPL12	2.31E-110	0.9104000	0.62	0.281	5.64E-106	7
TLE41	1.36E-102	0.9250328	0.666	0.366	3.31E-98	7
ST6GALNAC51	4.08E-102	0.9328787	0.564	0.255	9.96E-98	7
FBN2	1.27E-101	0.7915692	0.594	0.295	3.11E-97	7
RMST2	8.56E-98	0.9761357	0.795	0.567	2.09E-93	7
GLI3	7.63E-94	0.7454280	0.511	0.223	1.86E-89	7
FZD3	3.65E-88	0.6661177	0.55	0.266	8.90E-84	7
SHROOM3	7.82E-76	0.6594371	0.513	0.25	1.91E-71	7
GPC62	5.25E-74	0.7143291	0.842	0.646	1.28E-69	7
MAPK103	1.49E-70	0.6978210	0.638	0.384	3.63E-66	7
TTC282	2.22E-63	0.6298441	0.68	0.448	5.42E-59	7
TENM32	6.76E-54	0.6194817	0.907	0.773	1.65E-49	7
ROBO12	1.41E-50	0.6207185	0.726	0.563	3.45E-46	7
AUTS22	1.04E-49	0.6054673	0.67	0.489	2.54E-45	7
PTPRS1	6.46E-46	0.4800850	0.628	0.435	1.58E-41	7
TUBA1A3	8.65E-44	0.5533863	0.659	0.479	2.11E-39	7
SSBP22	5.56E-42	0.5297619	0.56	0.366	1.36E-37	7
PTPRD2	1.28E-40	0.4940451	0.694	0.519	3.12E-36	7
LSAMP2	4.17E-39	0.5191421	0.647	0.471	1.02E-34	7
SYNE21	3.81E-37	0.5137909	0.582	0.422	9.30E-33	7
TOP2A1	8.96E-29	0.5662037	0.54	0.403	2.19E-24	7
CENPF1	4.22E-27	0.4730863	0.594	0.475	1.03E-22	7
SOX4	4.16E-21	0.3633738	0.515	0.385	1.01E-16	7
MACF11	4.79E-21	0.3566640	0.659	0.557	1.17E-16	7
PKM1	5.18E-20	0.3612661	0.651	0.535	1.26E-15	7
DST	7.21E-20	0.3708077	0.652	0.563	1.76E-15	7
CHD71	8.61E-16	0.2982978	0.54	0.44	2.10E-11	7
GRIP11	2.36E-15	0.3191093	0.548	0.44	5.76E-11	7
NLGN11	2.84E-15	0.3150417	0.672	0.565	6.92E-11	7
TUBB3	4.74E-15	0.2767454	0.821	0.75	1.16E-10	7
TUBA1B3	1.52E-11	0.2722196	0.765	0.719	3.72E-07	7
FLNA	1.63E-11	0.2667805	0.71	0.655	3.98E-07	7
UBE2E21	1.71E-11	0.2856816	0.651	0.587	4.17E-07	7
MSI2	3.94E-10	0.2923453	0.531	0.469	9.62E-06	7
PHIP1	4.78E-10	0.2596330	0.551	0.481	1.17E-05	7
TCF122	1.43E-08	0.2717330	0.56	0.503	3.49E-04	7
RIMS22	1.64E-24	0.3521812	0.745	0.584	4.01E-20	8
MT-ATP82	1.97E-23	0.2700161	0.999	0.985	4.80E-19	8
TERF13	8.38E-22	0.3675161	0.782	0.646	2.04E-17	8
MT-ND52	1.35E-20	0.2645910	0.992	0.973	3.30E-16	8

GRID23	4.25E-20	0.3212875	0.91	0.832	1.04E-15	8
KCND22	3.15E-17	0.2568825	0.678	0.506	7.70E-13	8
KIAA08252	1.02E-15	0.2993566	0.604	0.444	2.48E-11	8
FAM155A2	1.82E-15	0.2976462	0.759	0.631	4.44E-11	8
GRID24	1.04E-243	1.5924651	0.997	0.828	2.54E-239	9
ROBO22	1.42E-215	1.8474649	0.914	0.561	3.45E-211	9
ITPR2	7.83E-168	0.8869902	0.505	0.144	1.91E-163	9
NLGN12	1.73E-155	1.3421486	0.867	0.555	4.23E-151	9
LINGO2	5.54E-149	0.8672169	0.583	0.2	1.35E-144	9
ADCY2	1.03E-145	1.1325731	0.574	0.205	2.51E-141	9
ANK31	2.26E-133	0.9247684	0.786	0.405	5.51E-129	9
KCND23	8.46E-122	0.9327005	0.862	0.496	2.07E-117	9
UNC5D	7.75E-119	1.0978585	0.929	0.743	1.89E-114	9
RIMS23	1.28E-115	0.8977941	0.865	0.578	3.11E-111	9
PHC1	7.89E-98	0.4980000	0.529	0.208	1.93E-93	9
GULP11	1.34E-97	0.8047247	0.873	0.647	3.26E-93	9
FAT3	2.05E-94	0.6695132	0.62	0.293	5.01E-90	9
LRBA	8.65E-94	0.8011891	0.847	0.616	2.11E-89	9
KCTD8	5.55E-93	0.6951312	0.602	0.276	1.35E-88	9
AC073050.1	8.68E-92	0.7069032	0.739	0.404	2.12E-87	9
BNC21	1.02E-88	0.7694555	0.786	0.459	2.49E-84	9
MYEF21	6.50E-88	0.4660302	0.553	0.231	1.59E-83	9
PITPNC1	7.94E-87	0.5728104	0.589	0.273	1.94E-82	9
SYT11	9.12E-87	0.7608980	0.858	0.62	2.22E-82	9
IMMP2L2	9.98E-87	0.7424915	0.876	0.625	2.44E-82	9
OSBPL10	6.90E-86	0.5343952	0.508	0.214	1.68E-81	9
LDB2	4.15E-84	0.7299758	0.686	0.382	1.01E-79	9
ARL151	8.76E-83	0.6306346	0.776	0.466	2.14E-78	9
SH3D19	1.95E-82	0.5865879	0.553	0.254	4.76E-78	9
CACNA2D1	2.26E-82	0.6512868	0.704	0.396	5.53E-78	9
NRXN12	4.95E-80	0.7316702	0.906	0.701	1.21E-75	9
TMEM132B	5.64E-80	0.6055384	0.619	0.313	1.38E-75	9
KIAA08253	7.03E-79	0.9518946	0.743	0.437	1.71E-74	9
AC003975.1	1.43E-78	0.7069741	0.53	0.236	3.48E-74	9
RBFOX11	4.02E-74	0.7572585	0.754	0.495	9.81E-70	9
JARID21	6.14E-74	0.8581541	0.764	0.534	1.50E-69	9
SPATA51	1.80E-73	0.5780324	0.689	0.387	4.40E-69	9
CASC151	2.06E-72	0.6704739	0.855	0.668	5.02E-68	9
JADE1	5.93E-72	0.4690898	0.51	0.229	1.45E-67	9
PCNX2	1.38E-70	0.5974940	0.561	0.289	3.37E-66	9
PTBP21	1.12E-69	0.5637097	0.766	0.482	2.73E-65	9

NLGN4X	5.43E-69	0.8809216	0.619	0.38	1.33E-64	9
UBE2E22	1.21E-68	0.6283639	0.825	0.578	2.96E-64	9
PKP4	1.56E-68	0.5126336	0.623	0.331	3.81E-64	9
PHF21B	1.96E-67	0.4524133	0.511	0.238	4.80E-63	9
MAGI21	3.60E-67	0.5685350	0.734	0.427	8.79E-63	9
TMEM132D	4.78E-67	0.5551805	0.557	0.281	1.17E-62	9
FAM155A3	1.86E-64	0.6738046	0.855	0.626	4.55E-60	9
PTPRG1	7.49E-64	0.7901477	0.917	0.855	1.83E-59	9
AC008014.1	3.29E-62	0.5824445	0.598	0.324	8.02E-58	9
ROR11	1.00E-61	0.6664076	0.866	0.723	2.44E-57	9
UST1	1.20E-61	0.6405156	0.672	0.425	2.94E-57	9
AASS3	3.19E-61	0.6037359	0.851	0.643	7.78E-57	9
PLEKHA51	1.02E-60	0.6244008	0.781	0.564	2.49E-56	9
GPR176	1.93E-60	0.5597266	0.53	0.271	4.70E-56	9
RABGAP1L	7.21E-60	0.5945430	0.727	0.479	1.76E-55	9
GMDS	2.81E-59	0.5514799	0.526	0.278	6.85E-55	9
RAPGEF2	4.68E-58	0.4618975	0.504	0.251	1.14E-53	9
SLC4A7	1.37E-57	0.5359082	0.648	0.395	3.34E-53	9
RASAL2	7.10E-57	0.5268344	0.619	0.361	1.73E-52	9
ST6GALNAC31	9.43E-57	0.6279874	0.75	0.518	2.30E-52	9
PHLPP1	1.16E-56	0.5047551	0.607	0.35	2.83E-52	9
R3HDM11	1.96E-56	0.5052946	0.645	0.393	4.78E-52	9
TRIO1	1.10E-55	0.4981375	0.649	0.39	2.67E-51	9
AC009446.12	4.14E-55	0.5555256	0.637	0.375	1.01E-50	9
SBF2	5.33E-55	0.5350573	0.6	0.348	1.30E-50	9
FBXW7	1.13E-54	0.4687324	0.581	0.323	2.75E-50	9
FOXN32	6.98E-54	0.5383713	0.777	0.555	1.70E-49	9
WWC2	4.91E-53	0.4949557	0.604	0.352	1.20E-48	9
ETNK1	5.28E-53	0.4410579	0.574	0.322	1.29E-48	9
XKR6	1.10E-52	0.4587434	0.687	0.435	2.68E-48	9
EXT11	1.32E-52	0.5169578	0.728	0.484	3.23E-48	9
PKIB1	3.72E-52	0.5411231	0.746	0.52	9.08E-48	9
LPP	4.76E-52	0.7206768	0.676	0.448	1.16E-47	9
PAWR	3.49E-51	0.5268326	0.63	0.386	8.51E-47	9
KCNT2	5.95E-51	0.5315984	0.803	0.605	1.45E-46	9
MLLT10	2.52E-50	0.4553695	0.583	0.331	6.16E-46	9
CAMKMT	1.96E-49	0.5410720	0.676	0.445	4.77E-45	9
FRAS1	2.54E-49	0.5731261	0.775	0.569	6.20E-45	9
ANKRD171	2.87E-49	0.5010580	0.747	0.528	7.01E-45	9
PTK21	3.37E-49	0.5942268	0.667	0.456	8.22E-45	9
PAN3	9.26E-49	0.5161780	0.611	0.368	2.26E-44	9

ADGRV11	1.14E-48	0.5185236	0.858	0.675	2.77E-44	9
CDKAL11	2.19E-48	0.4874534	0.623	0.383	5.34E-44	9
GNAQ	5.12E-48	0.4538615	0.613	0.365	1.25E-43	9
EXOC41	9.41E-48	0.5811123	0.712	0.504	2.30E-43	9
PTPRZ1	1.14E-47	0.5455891	0.807	0.648	2.79E-43	9
AC068587.4	4.36E-47	0.3684531	0.602	0.349	1.06E-42	9
CUX1	3.18E-46	0.4458063	0.549	0.322	7.75E-42	9
MGAT4C1	6.03E-46	0.5204375	0.552	0.309	1.47E-41	9
FANCL	6.67E-46	0.4025396	0.572	0.334	1.63E-41	9
AL157778.11	1.30E-45	0.7199415	0.755	0.595	3.17E-41	9
NBEA1	1.64E-45	0.5577238	0.74	0.54	3.99E-41	9
ERC2	2.98E-45	0.4906562	0.544	0.321	7.27E-41	9
FNDC3B1	3.74E-45	0.4768112	0.6	0.372	9.14E-41	9
SAMD12	4.19E-45	0.3814426	0.508	0.274	1.02E-40	9
GABRB3	4.99E-45	0.4994849	0.571	0.347	1.22E-40	9
GPC63	9.29E-45	0.6284064	0.84	0.647	2.27E-40	9
CMSS1	2.36E-44	0.4435818	0.622	0.381	5.75E-40	9
MED12L	4.89E-44	0.3809366	0.549	0.313	1.19E-39	9
PLXDC2	4.30E-43	0.5301190	0.637	0.416	1.05E-38	9
MAGI11	6.36E-43	0.5531536	0.795	0.634	1.55E-38	9
JMJD1C1	1.13E-42	0.5270255	0.777	0.631	2.77E-38	9
ADAMTS19	1.57E-42	0.3976251	0.6	0.364	3.83E-38	9
CTNNA23	3.08E-42	0.4524157	0.705	0.478	7.52E-38	9
ADGRA31	4.18E-42	0.4028847	0.615	0.38	1.02E-37	9
VTI1A	4.62E-42	0.4080126	0.525	0.3	1.13E-37	9
SUPT3H1	1.21E-41	0.5583954	0.709	0.524	2.95E-37	9
DLG21	3.34E-41	0.4617560	0.71	0.487	8.16E-37	9
PVT1	5.76E-41	0.5250594	0.705	0.512	1.41E-36	9
STK33	1.00E-40	0.3869635	0.577	0.342	2.44E-36	9
STRN3	4.95E-40	0.3840651	0.541	0.319	1.21E-35	9
WWOX2	5.32E-40	0.5339967	0.817	0.655	1.30E-35	9
COMMD10	2.15E-39	0.3260428	0.514	0.291	5.25E-35	9
SRSF111	6.02E-39	0.3559482	0.63	0.407	1.47E-34	9
PUM21	8.29E-39	0.3932023	0.585	0.366	2.02E-34	9
FOXO1	1.46E-38	0.4144562	0.553	0.333	3.56E-34	9
MMP16	3.04E-38	0.4853707	0.604	0.393	7.43E-34	9
TCF123	5.42E-38	0.4891371	0.676	0.497	1.32E-33	9
COG5	2.01E-37	0.3571726	0.507	0.295	4.89E-33	9
FBXL71	2.36E-37	0.4129481	0.74	0.516	5.76E-33	9
WDFY3	2.64E-37	0.3816172	0.519	0.31	6.43E-33	9
POU2F11	3.74E-37	0.3917284	0.608	0.392	9.13E-33	9

DPP10	4.28E-37	0.4062902	0.594	0.374	1.04E-32	9
BMPR1A	5.19E-37	0.4103787	0.534	0.325	1.27E-32	9
UBE2E3	1.19E-36	0.3352668	0.527	0.31	2.91E-32	9
EXOC6B	1.98E-36	0.4567702	0.57	0.367	4.84E-32	9
ARID1B1	2.64E-36	0.4716275	0.727	0.541	6.44E-32	9
AC022140.12	3.83E-36	0.5419351	0.829	0.692	9.34E-32	9
EML4	6.13E-36	0.3189884	0.503	0.289	1.50E-31	9
FTX1	6.76E-36	0.4057730	0.848	0.688	1.65E-31	9
RORA	6.80E-36	0.4930620	0.641	0.444	1.66E-31	9
GALNT13	7.24E-36	0.3957619	0.563	0.35	1.77E-31	9
DLGAP12	7.42E-36	0.5102067	0.919	0.845	1.81E-31	9
TBL1XR11	8.70E-36	0.3708198	0.663	0.446	2.12E-31	9
SND1	9.83E-36	0.3795807	0.525	0.324	2.40E-31	9
FGD41	1.46E-35	0.4502311	0.665	0.475	3.57E-31	9
TCF41	3.59E-35	0.4297835	0.809	0.65	8.75E-31	9
AL354821.13	5.64E-35	0.3968164	0.653	0.427	1.38E-30	9
BBS91	6.39E-35	0.5738507	0.578	0.372	1.56E-30	9
CLASP2	1.49E-34	0.3507765	0.559	0.346	3.63E-30	9
ATF7IP	1.88E-34	0.3797190	0.561	0.35	4.58E-30	9
FMNL21	4.25E-34	0.4043299	0.616	0.412	1.04E-29	9
PTPRK	4.74E-34	0.3948051	0.746	0.532	1.16E-29	9
GALNT7	4.93E-34	0.3111401	0.522	0.307	1.20E-29	9
SRPK2	6.96E-34	0.3479861	0.579	0.371	1.70E-29	9
DIAPH31	9.60E-34	0.3791566	0.628	0.42	2.34E-29	9
TBC1D51	6.69E-33	0.4240594	0.728	0.541	1.63E-28	9
SNTG2	8.28E-33	0.3158006	0.5	0.294	2.02E-28	9
KHDRBS2	1.19E-32	0.4277245	0.712	0.527	2.92E-28	9
BCAS31	1.50E-32	0.3603164	0.604	0.393	3.66E-28	9
RBPMS1	2.63E-32	0.4854491	0.69	0.538	6.41E-28	9
COL4A5	3.67E-32	0.2972889	0.538	0.326	8.96E-28	9
TCF7L21	8.44E-32	0.4287130	0.624	0.435	2.06E-27	9
HS2ST1	8.81E-32	0.3322231	0.556	0.351	2.15E-27	9
FRMD5	9.77E-32	0.3936989	0.563	0.368	2.38E-27	9
SENP6	1.10E-31	0.3333613	0.531	0.334	2.69E-27	9
GALNT17	1.81E-31	0.4842927	0.63	0.459	4.42E-27	9
COP1	1.88E-31	0.3334880	0.522	0.323	4.58E-27	9
IGF2BP21	2.97E-31	0.4005904	0.628	0.441	7.26E-27	9
RBM261	3.13E-31	0.3300600	0.646	0.434	7.63E-27	9
PPM1E	3.68E-31	0.3709596	0.523	0.327	8.98E-27	9
ADK	7.04E-31	0.4237478	0.567	0.388	1.72E-26	9
TCF7L1	1.99E-30	0.3518891	0.527	0.336	4.85E-26	9

DIAPH2	2.40E-30	0.3860381	0.605	0.41	5.86E-26	9
UBR5	3.62E-30	0.3677000	0.609	0.42	8.84E-26	9
PARD31	3.74E-30	0.4528823	0.676	0.519	9.12E-26	9
IGF1R1	5.28E-30	0.4008876	0.742	0.578	1.29E-25	9
RANBP17	8.19E-30	0.3747329	0.54	0.353	2.00E-25	9
UBE2K	1.04E-29	0.3127828	0.508	0.317	2.53E-25	9
LINC01572	1.12E-29	0.3211493	0.522	0.322	2.74E-25	9
ROBO13	1.96E-29	0.3459632	0.753	0.562	4.77E-25	9
PPP3CA	2.25E-29	0.3384901	0.503	0.312	5.50E-25	9
SIPA1L1	3.18E-29	0.4110347	0.54	0.359	7.77E-25	9
MACROD21	3.51E-29	0.3783671	0.582	0.391	8.58E-25	9
METTL15	3.84E-29	0.3192016	0.582	0.381	9.36E-25	9
ZMYM21	5.59E-29	0.3876452	0.626	0.443	1.36E-24	9
PCCA1	8.86E-29	0.3837548	0.649	0.457	2.16E-24	9
PPP1R9A1	1.11E-28	0.3441466	0.634	0.439	2.71E-24	9
CPSF6	1.35E-28	0.2862521	0.553	0.353	3.30E-24	9
BIRC6	3.62E-28	0.3235039	0.638	0.438	8.83E-24	9
JAKMIP21	3.65E-28	0.3278651	0.594	0.403	8.92E-24	9
BCKDHB1	5.38E-28	0.3246688	0.708	0.511	1.31E-23	9
ZNF827	7.71E-28	0.2856510	0.501	0.313	1.88E-23	9
PHF14	1.11E-27	0.3413280	0.594	0.408	2.70E-23	9
KIAA1958	2.09E-27	0.3282233	0.538	0.35	5.11E-23	9
GRIP12	3.45E-27	0.3656016	0.616	0.437	8.42E-23	9
TENM33	4.38E-27	0.3712885	0.87	0.776	1.07E-22	9
IGF2BP3	5.92E-27	0.3715202	0.657	0.475	1.45E-22	9
SLC25A21	7.12E-27	0.3322759	0.508	0.325	1.74E-22	9
APP	1.01E-26	0.3859564	0.634	0.467	2.47E-22	9
STAU2	1.47E-26	0.2889385	0.516	0.334	3.58E-22	9
BAZ2B	1.64E-26	0.2916510	0.559	0.369	4.01E-22	9
LINC02335	6.05E-26	0.4616997	0.544	0.373	1.48E-21	9
MAN2A1	9.32E-26	0.3095382	0.592	0.412	2.27E-21	9
FOXP1	1.49E-25	0.3149157	0.51	0.33	3.63E-21	9
SMYD31	2.78E-25	0.3908066	0.73	0.608	6.79E-21	9
CDYL	2.96E-25	0.3722294	0.54	0.37	7.23E-21	9
CTBP2	3.17E-25	0.3817224	0.516	0.358	7.73E-21	9
AKT3	6.38E-25	0.3164035	0.511	0.336	1.56E-20	9
USP34	1.07E-24	0.3193698	0.597	0.422	2.60E-20	9
LINC01090	1.72E-24	0.2604705	0.531	0.34	4.19E-20	9
ANK23	2.22E-24	0.2873439	0.893	0.787	5.42E-20	9
ZNF644	2.72E-24	0.3004051	0.522	0.347	6.63E-20	9
KANSL1	5.27E-24	0.3520641	0.533	0.373	1.29E-19	9

CENPP	5.66E-24	0.3016479	0.504	0.332	1.38E-19	9
LIN28B1	6.38E-24	0.3394868	0.63	0.455	1.56E-19	9
LARP7	1.63E-23	0.2611588	0.551	0.372	3.97E-19	9
QSER1	1.64E-23	0.2702473	0.533	0.355	4.01E-19	9
CCDC91	3.86E-23	0.3332456	0.56	0.386	9.42E-19	9
WDPCP	4.00E-23	0.3089222	0.619	0.443	9.75E-19	9
SETD5	6.11E-23	0.2769889	0.557	0.383	1.49E-18	9
FOCAD	7.19E-23	0.2861124	0.53	0.357	1.75E-18	9
NLGN4Y	1.01E-22	0.4143648	0.582	0.434	2.47E-18	9
SEMA6A	1.35E-22	0.3385785	0.608	0.439	3.31E-18	9
ARID2	1.28E-21	0.2942128	0.511	0.348	3.12E-17	9
ENAH	2.22E-21	0.3408549	0.585	0.431	5.43E-17	9
SDK1	2.37E-21	0.3644448	0.577	0.428	5.79E-17	9
RAD51B1	2.38E-21	0.3918829	0.74	0.627	5.81E-17	9
PHIP2	2.40E-21	0.2554120	0.657	0.476	5.85E-17	9
CADM23	2.03E-20	0.3390425	0.839	0.752	4.95E-16	9
VPS13B1	1.17E-19	0.2587700	0.57	0.412	2.86E-15	9
PBX11	1.37E-19	0.3431958	0.769	0.651	3.34E-15	9
LINC023151	5.13E-19	0.3655509	0.671	0.54	1.25E-14	9
EIF4G3	7.87E-19	0.3740260	0.574	0.452	1.92E-14	9
TRIM24	2.70E-18	0.2760493	0.568	0.41	6.59E-14	9
MALAT11	5.49E-18	0.7481753	0.954	0.963	1.34E-13	9
GPHN	5.71E-18	0.2605137	0.5	0.344	1.39E-13	9
MDN1	1.92E-17	0.2827759	0.507	0.367	4.67E-13	9
LINC014731	3.25E-17	0.2506601	0.668	0.518	7.94E-13	9
XACT3	7.22E-17	0.2925270	0.833	0.717	1.76E-12	9
LARGE1	7.26E-17	0.3844652	0.642	0.525	1.77E-12	9
PRKG13	6.32E-16	0.2714095	0.839	0.741	1.54E-11	9
HNRNPC	2.77E-15	0.2689552	0.575	0.444	6.76E-11	9
CECR2	5.59E-15	0.2732979	0.5	0.37	1.36E-10	9
NRG32	1.01E-14	0.2824166	0.786	0.686	2.46E-10	9
MIR924HG2	7.81E-14	0.3594342	0.675	0.596	1.91E-09	9
LINC018761	2.08E-13	0.2706018	0.663	0.532	5.07E-09	9
AGAP1	7.57E-13	0.2596501	0.536	0.421	1.85E-08	9
PBX31	3.59E-12	0.2683320	0.607	0.502	8.77E-08	9
LRRTM41	9.22E-11	0.2646811	0.673	0.563	2.25E-06	9
RERE	2.23E-09	0.4145637	0.654	0.641	5.44E-05	9
UGP21	2.38E-39	0.5106097	0.744	0.526	5.80E-35	10
USP9X1	9.68E-25	0.3987876	0.757	0.599	2.36E-20	10
BNC22	4.75E-20	0.3458769	0.646	0.468	1.16E-15	10
EEF1A11	9.75E-16	0.2753484	0.995	0.968	2.38E-11	10

RPL41	4.68E-15	0.2759150	0.934	0.846	1.14E-10	10
VCAN2	1.14E-12	0.3032193	0.591	0.465	2.77E-08	10
RPL61	3.34E-12	0.2682272	0.992	0.96	8.16E-08	10
TPT12	1.72E-11	0.3313666	0.792	0.684	4.20E-07	10
CTBP21	9.36E-16	0.3295193	0.52	0.361	2.29E-11	11
TRIM241	1.09E-09	0.3326746	0.531	0.414	2.66E-05	11
HNRNPC1	1.22E-09	0.3205514	0.558	0.448	2.97E-05	11
UGP22	9.72E-09	0.2964815	0.636	0.532	2.37E-04	11
JARID22	2.63E-08	0.3325694	0.647	0.542	6.43E-04	11
MALAT12	4.64E-06	0.3237877	0.973	0.962	1.13E-01	11
RPL36	7.84E-119	2.0214516	0.994	0.449	1.91E-114	12
RPS12	6.21E-110	2.2041551	1	0.586	1.52E-105	12
RPL28	8.44E-110	2.2818723	1	0.729	2.06E-105	12
RPS19	1.06E-108	2.3067127	1	0.761	2.60E-104	12
RPS27	3.09E-107	2.1530851	1	0.842	7.55E-103	12
RPL13	2.29E-105	1.9549055	1	0.658	5.58E-101	12
RPS15	5.87E-105	2.1783572	0.989	0.6	1.43E-100	12
RPL35	1.03E-104	2.1007643	0.989	0.598	2.51E-100	12
RPL37	1.47E-103	1.8924461	1	0.718	3.58E-99	12
RPL411	3.88E-103	1.8015018	1	0.897	9.48E-99	12
RPL37A	3.40E-101	1.9734341	1	0.864	8.29E-97	12
RPS18	1.01E-100	1.8523150	0.994	0.72	2.46E-96	12
RPLP2	1.86E-99	1.8703821	1	0.625	4.55E-95	12
RPL27A	1.12E-98	1.8724357	0.994	0.65	2.73E-94	12
RPS29	6.79E-98	1.8597286	1	0.779	1.66E-93	12
RPS17	6.42E-97	1.8456845	0.949	0.428	1.57E-92	12
RPS28	6.94E-97	1.8280804	0.994	0.6	1.69E-92	12
RPL11	8.02E-97	1.8684837	0.944	0.447	1.96E-92	12
TMSB10	3.67E-94	1.5851295	0.932	0.376	8.97E-90	12
RPL38	5.45E-94	1.8470400	0.994	0.725	1.33E-89	12
RPL13A	1.58E-92	1.6067989	1	0.919	3.86E-88	12
ELOB	7.38E-90	1.2295015	0.729	0.197	1.80E-85	12
RPS26	1.75E-87	1.7201889	0.977	0.627	4.26E-83	12
RPS16	9.12E-87	1.5608923	0.966	0.578	2.23E-82	12
RPL18A	1.75E-86	1.6836058	0.994	0.704	4.28E-82	12
RPL23	2.08E-85	1.5315292	0.966	0.572	5.09E-81	12
RPS2	3.30E-85	1.7411432	0.994	0.733	8.06E-81	12
SNRPF	1.07E-84	0.9060817	0.672	0.156	2.61E-80	12
RPL311	3.37E-84	1.6612524	0.989	0.703	8.23E-80	12
ATP5ME	4.92E-84	1.2352988	0.808	0.256	1.20E-79	12
UBL5	5.45E-84	1.0424338	0.74	0.198	1.33E-79	12

RPL36A	7.12E-84	1.5370807	0.938	0.481	1.74E-79	12
RPS11	2.50E-81	1.6521157	0.983	0.723	6.10E-77	12
RPL39	9.26E-80	1.5019078	0.972	0.623	2.26E-75	12
RPS14	1.88E-78	1.4581804	0.944	0.521	4.60E-74	12
SNRPD2	3.57E-78	1.0694978	0.791	0.246	8.72E-74	12
POLR2L	4.16E-77	0.8741004	0.65	0.161	1.02E-72	12
RPS15A	5.44E-72	1.4310178	0.989	0.591	1.33E-67	12
RPL30	1.47E-71	1.3665182	0.955	0.563	3.59E-67	12
COX7C	3.87E-71	1.3666596	0.91	0.468	9.44E-67	12
RPS10	5.15E-71	1.3719333	0.915	0.447	1.26E-66	12
ATP5IF1	1.05E-69	0.9205431	0.627	0.163	2.57E-65	12
UBA52	1.07E-69	1.2445949	0.876	0.394	2.62E-65	12
ATP5F1E	2.93E-69	1.0970240	0.78	0.275	7.15E-65	12
RPL12	8.02E-69	1.3024543	0.949	0.531	1.96E-64	12
RPS9	8.16E-69	1.3341049	0.938	0.591	1.99E-64	12
SELENOH	1.18E-67	0.8782947	0.678	0.195	2.88E-63	12
GPX4	8.69E-67	0.8864206	0.678	0.2	2.12E-62	12
RPL29	1.43E-66	1.2636237	0.881	0.436	3.50E-62	12
RPL22	1.55E-66	1.1749115	0.927	0.493	3.78E-62	12
RPS201	1.05E-65	1.4440556	0.977	0.721	2.55E-61	12
RPS211	1.13E-65	1.0765901	1	0.914	2.76E-61	12
RPL18	1.62E-65	1.1822192	0.955	0.55	3.96E-61	12
RPS81	2.27E-65	1.1200752	0.989	0.885	5.54E-61	12
ATP5MD	3.95E-65	0.6029867	0.525	0.115	9.64E-61	12
NDUFA3	2.85E-64	0.6655484	0.542	0.126	6.96E-60	12
RPL10A	3.16E-64	1.2062864	0.927	0.524	7.71E-60	12
RPL8	1.50E-63	1.2131814	0.972	0.612	3.66E-59	12
MIF	8.45E-63	1.3853898	0.898	0.505	2.06E-58	12
TOMM7	9.31E-63	0.6987784	0.582	0.147	2.27E-58	12
SERF2	3.28E-62	1.1978538	0.87	0.422	7.99E-58	12
ATP5MPL	7.55E-62	0.6953539	0.531	0.124	1.84E-57	12
RPS3	1.09E-61	1.1899179	0.966	0.64	2.66E-57	12
NDUFS6	3.08E-61	0.7631679	0.633	0.178	7.51E-57	12
RPL7A	2.68E-60	1.2042179	0.994	0.677	6.55E-56	12
RPL171	3.68E-60	1.1185826	0.994	0.794	8.99E-56	12
RPLP1	3.76E-60	0.9592755	1	0.969	9.17E-56	12
RPL35A	3.76E-60	1.1395719	0.977	0.672	9.19E-56	12
PFDN5	9.21E-60	0.6658534	0.537	0.131	2.25E-55	12
RPS13	1.73E-59	1.1951193	0.921	0.498	4.21E-55	12
RPS231	3.88E-59	1.1488312	0.994	0.861	9.48E-55	12
RPL36AL	2.78E-58	0.6211184	0.514	0.122	6.79E-54	12

RPL321	7.98E-58	1.2196760	0.949	0.706	1.95E-53	12
RPL23A	1.40E-57	1.3066971	0.977	0.809	3.42E-53	12
TMSB4X	6.04E-57	0.8890069	0.751	0.263	1.47E-52	12
RPL19	2.09E-56	1.1047780	0.983	0.836	5.11E-52	12
ROMO1	2.97E-56	0.7029502	0.542	0.142	7.25E-52	12
RPS241	2.55E-55	1.0715030	1	0.873	6.23E-51	12
RPS5	3.86E-55	1.1647262	0.927	0.592	9.42E-51	12
RPL24	6.51E-55	1.2311000	0.955	0.649	1.59E-50	12
NDUFA13	3.32E-53	1.0583614	0.831	0.384	8.11E-49	12
NDUFS5	6.74E-53	0.9326868	0.678	0.247	1.64E-48	12
RPL10	1.61E-52	1.0590284	0.972	0.71	3.93E-48	12
RPL15	2.92E-52	1.0806151	0.91	0.537	7.12E-48	12
HSPE1	2.14E-49	0.8182222	0.684	0.248	5.22E-45	12
RPS27A1	2.82E-49	0.9722546	0.989	0.875	6.89E-45	12
RPL9	4.12E-47	0.9601344	0.893	0.503	1.01E-42	12
UQCR11	5.94E-47	0.6298878	0.588	0.186	1.45E-42	12
PFN1	1.96E-44	1.0307703	0.819	0.441	4.79E-40	12
SNRPG	2.77E-44	0.6287423	0.588	0.189	6.77E-40	12
FAU	1.65E-43	0.8893792	0.898	0.542	4.03E-39	12
RPL211	9.31E-43	0.9961587	0.955	0.651	2.27E-38	12
SLIRP	2.40E-42	0.6808359	0.633	0.227	5.86E-38	12
EEF1D	3.42E-42	0.6482733	0.61	0.213	8.36E-38	12
COX6B1	3.91E-42	0.6574584	0.644	0.237	9.53E-38	12
GADD45GIP1	6.94E-42	0.5925069	0.571	0.185	1.69E-37	12
RPS7	7.74E-42	0.9656918	0.904	0.6	1.89E-37	12
EDF1	3.35E-41	0.5268098	0.508	0.15	8.18E-37	12
RPL341	1.21E-40	1.0327575	0.904	0.607	2.95E-36	12
UQCRH	2.53E-40	0.7376930	0.593	0.218	6.17E-36	12
RPL27	2.75E-39	0.7985387	0.887	0.53	6.71E-35	12
TMA7	3.04E-38	0.8499516	0.723	0.34	7.42E-34	12
MYL6	5.74E-38	0.7758308	0.729	0.335	1.40E-33	12
COX5B	8.42E-38	0.6537408	0.678	0.276	2.05E-33	12
NME2	1.31E-37	0.8169850	0.751	0.373	3.19E-33	12
RPS4X	1.64E-37	0.7753948	0.757	0.356	4.01E-33	12
COX6C	2.47E-37	0.7042538	0.616	0.234	6.02E-33	12
HMGA1	4.69E-37	0.7500286	0.718	0.334	1.15E-32	12
ATP5MF	5.40E-37	0.6033065	0.565	0.2	1.32E-32	12
NDUFB7	4.13E-36	0.5885624	0.525	0.178	1.01E-31	12
RPLP0	1.55E-35	0.8713719	0.898	0.645	3.77E-31	12
SNHG51	1.73E-35	0.8088035	0.825	0.465	4.22E-31	12
GAPDH2	1.82E-35	0.8140976	0.989	0.888	4.45E-31	12

RPS3A1	3.20E-35	0.8541611	0.966	0.731	7.81E-31	12
RPL141	4.29E-35	0.8561171	0.91	0.651	1.05E-30	12
SNRPE	7.41E-34	0.5758781	0.559	0.206	1.81E-29	12
NDUFB4	2.04E-33	0.5100051	0.559	0.203	4.98E-29	12
RPS251	3.39E-33	0.6894475	1	0.923	8.27E-29	12
RARRES2	3.12E-32	0.6538711	0.633	0.272	7.63E-28	12
SNHG6	1.08E-31	0.6117887	0.576	0.229	2.65E-27	12
ZFAS1	2.60E-31	0.5269020	0.537	0.199	6.34E-27	12
H3F3A	4.51E-31	0.7873104	0.859	0.559	1.10E-26	12
RPSA	2.39E-30	0.7589920	0.944	0.758	5.82E-26	12
GSTP1	9.02E-30	0.7350179	0.78	0.434	2.20E-25	12
TMEM258	1.82E-29	0.5359472	0.565	0.22	4.45E-25	12
ATP5MG	2.00E-28	0.5622610	0.508	0.195	4.89E-24	12
RPL261	2.10E-27	0.6679683	0.966	0.776	5.14E-23	12
PPIA	3.11E-27	0.6746210	0.814	0.507	7.58E-23	12
RACK11	2.71E-26	0.6894331	0.836	0.555	6.60E-22	12
NDUFA4	1.09E-25	0.5706649	0.706	0.36	2.66E-21	12
APOE1	2.90E-23	0.5995230	0.989	0.939	7.08E-19	12
UQCRB	3.94E-22	0.5425946	0.678	0.367	9.61E-18	12
RPS61	1.68E-21	0.6426380	0.904	0.741	4.11E-17	12
EEF1B2	3.25E-21	0.5284442	0.616	0.313	7.92E-17	12
ANAPC11	5.56E-20	0.4596042	0.52	0.241	1.36E-15	12
FTL	2.28E-19	0.6014515	0.718	0.441	5.57E-15	12
COX7A2	6.02E-18	0.4055822	0.588	0.305	1.47E-13	12
NACA	9.66E-18	0.5083853	0.672	0.407	2.36E-13	12
RPL71	2.06E-17	0.5618695	0.831	0.589	5.03E-13	12
FTH1	7.04E-17	0.5241979	0.734	0.473	1.72E-12	12
TPI1	4.59E-16	0.4307030	0.633	0.364	1.12E-11	12
RPL310	4.97E-16	0.4900442	0.944	0.817	1.21E-11	12
SLC25A6	6.56E-16	0.3405493	0.548	0.273	1.60E-11	12
EEF1G	9.70E-16	0.4589211	0.661	0.391	2.37E-11	12
ATP5MC2	1.96E-15	0.4392518	0.616	0.345	4.77E-11	12
COX4I1	2.46E-15	0.5123704	0.599	0.364	5.99E-11	12
BTF3	2.79E-14	0.3540291	0.537	0.284	6.80E-10	12
SUMO2	1.66E-13	0.3544733	0.514	0.275	4.06E-09	12
PTMA	7.27E-13	0.4380740	0.921	0.763	1.77E-08	12
PRDX2	2.00E-12	0.3641521	0.508	0.28	4.88E-08	12
HIST1H4C	1.46E-11	0.3950492	0.565	0.336	3.58E-07	12
PRDX1	9.64E-11	0.3236089	0.751	0.493	2.35E-06	12
H2AFZ	1.32E-10	0.3292910	0.554	0.334	3.23E-06	12
MDK1	7.76E-10	0.3485400	0.712	0.488	1.89E-05	12

CFL1	2.07E-09	0.2944670	0.599	0.385	5.05E-05	12
CD241	3.39E-09	0.3391344	0.582	0.388	8.28E-05	12
H3F3B1	6.30E-09	0.2926507	0.582	0.377	1.54E-04	12
HMGB1	1.99E-08	0.3371788	0.638	0.453	4.85E-04	12
MIR302CHG	1.19E-06	0.3524208	0.672	0.533	2.90E-02	12
COX6A11	1.21E-06	0.2695773	0.847	0.705	2.96E-02	12
CLDN6	1.44E-06	0.2560295	0.655	0.47	3.51E-02	12
BSG	2.46E-06	0.2531286	0.633	0.474	5.99E-02	12
NDUFB9	4.36E-06	0.2538259	0.542	0.384	1.06E-01	12
ACTB	1.39E-04	0.2751062	0.802	0.727	1.00E+00	12

Table A.3.4. Replicate 2 chromVAR top motifs in S-phase fractions

	Gene	RNA AUC	RNA p-val	Motif Feature	Motif AUC	Motif p-val	Avg AUC
s1	SOX4	0.51547	2.60E-04	MA0867.2	0.60919	2.95E-112	0.56233
s1	SOX2	0.51394	2.54E-04	MA0143.4	0.55924	2.53E-34	0.53659
s2	SREBF1	0.50829	6.90E-04	MA0595.1	0.67261	4.57E-221	0.59045
s2	MGA	0.51257	1.69E-04	MA0801.1	0.65889	1.19E-187	0.58573
s2	NFKB2	0.50977	6.33E-07	MA0778.1	0.65114	5.66E-170	0.58045
s2	ZNF682	0.50925	2.45E-04	MA1599.1	0.64381	4.30E-154	0.57653
s2	ZBTB6	0.50263	4.48E-04	MA1581.1	0.64079	9.25E-148	0.57171
s2	MYCN	0.51706	4.52E-12	MA0104.4	0.56268	9.80E-31	0.53987
s2	POU5F1	0.52524	1.67E-12	MA1115.1	0.53729	7.06E-12	0.53127
s2	FOSL1	0.50333	1.16E-04	MA0477.2	0.55174	1.84E-21	0.52753
s2	ZIC3	0.50892	4.96E-05	MA0697.1	0.53717	8.19E-12	0.52305
s2	MAZ	0.50872	7.10E-04	MA1522.1	0.53386	4.80E-10	0.52129
s2	JUND	0.50893	8.63E-05	MA0491.2	0.53114	1.03E-08	0.52003
s2	XBP1	0.51421	3.69E-08	MA0844.1	0.52360	1.43E-05	0.51890
s2	ATF4	0.50769	7.42E-04	MA0833.2	0.53005	3.29E-08	0.51887
s2	ELF3	0.50524	1.49E-06	MA0640.2	0.51450	7.67E-03	0.50987
s4	FOXP1	0.55754	3.86E-33	MA0481.3	0.62945	3.57E-113	0.59349
s4	FOXN3	0.54632	1.66E-17	MA1489.1	0.62827	3.76E-111	0.58730
s4	FOXO3	0.51753	5.31E-10	MA0157.2	0.63264	9.97E-119	0.57509
s4	ELF2	0.53537	3.33E-20	MA1483.1	0.60689	9.12E-78	0.57113
s4	FOXK1	0.50638	2.24E-03	MA0852.2	0.63102	6.81E-116	0.56870
s4	FOXG1	0.50227	1.63E-03	MA0613.1	0.61803	2.08E-94	0.56015
s4	E2F3	0.54282	8.97E-25	MA0469.3	0.56810	1.28E-32	0.55546
s4	NFYC	0.51370	1.18E-09	MA1644.1	0.57770	6.02E-42	0.54570
s4	NRF1	0.53126	9.32E-24	MA0506.1	0.55556	2.89E-22	0.54341
s4	ETV6	0.53383	1.78E-29	MA0645.1	0.54949	5.50E-18	0.54166
s4	OVOL2	0.50524	1.78E-03	MA1545.1	0.57343	1.19E-37	0.53933
s4	HMBOX1	0.53160	5.32E-24	MA0895.1	0.54618	7.36E-16	0.53889
s4	SP3	0.51208	1.20E-03	MA0746.2	0.55792	4.68E-24	0.53500
s4	RORB	0.51045	7.14E-06	MA1150.1	0.55861	1.36E-24	0.53453
s4	DMRTA2	0.50118	7.18E-04	MA1478.1	0.55795	4.49E-24	0.52956
s4	NFATC3	0.52060	7.53E-11	MA0625.1	0.53805	3.03E-11	0.52933
s4	RUNX2	0.52809	5.79E-25	MA0511.2	0.52635	4.17E-06	0.52722
s4	EGR3	0.50165	2.12E-03	MA0732.1	0.55236	6.03E-20	0.52700
s4	ETV5	0.50741	2.78E-06	MA0765.2	0.54255	1.07E-13	0.52498
s4	HINFP	0.50421	1.63E-03	MA0131.2	0.54168	3.34E-13	0.52295
s4	CTCF	0.51187	2.68E-05	MA0139.1	0.53269	1.14E-08	0.52228
s4	ERG	0.50708	3.00E-09	MA0474.2	0.53448	1.72E-09	0.52078

Table A.3.5. Replicate 2 chromVAR top motifs in clusters

c0						
Gene	RNA AUC	RNA p-val	Motif Feature	Motif AUC	Motif p-val	Avg AUC
POU5F1	0.53802	6.65E-20	MA1115.1	0.73159	6.01E-293	0.63481
FOXN3	0.54926	2.62E-16	MA1489.1	0.68019	3.64E-178	0.61472
ZBTB33	0.50475	5.74E-04	MA0527.1	0.63612	1.55E-102	0.57043
TCF7L2	0.52162	1.56E-04	MA0523.1	0.55357	2.64E-17	0.53759
SOX4	0.52806	3.89E-07	MA0867.2	0.52766	1.25E-05	0.52786
c1						
Gene	RNA AUC	RNA p-val	Motif Feature	Motif AUC	Motif p-val	Avg AUC
NFKB2	0.50989	2.73E-05	MA0778.1	0.83759	0.00E+00	0.67374
MGA	0.51571	9.23E-05	MA0801.1	0.82701	0.00E+00	0.67136
POU2F2	0.51352	8.00E-07	MA0507.1	0.69966	4.96E-205	0.60659
NFIB	0.52545	1.74E-13	MA1643.1	0.68665	1.91E-179	0.60605
POU5F1	0.52711	2.81E-10	MA1115.1	0.68362	9.75E-174	0.60536
CEBPG	0.50554	8.67E-04	MA0838.1	0.66322	1.05E-137	0.58438
TP53	0.50974	2.07E-05	MA0106.3	0.65757	1.81E-128	0.58365
GRHL2	0.53059	1.96E-21	MA1105.2	0.63476	1.69E-94	0.58268
ZNF682	0.51170	1.14E-04	MA1599.1	0.63501	7.72E-95	0.57336
THRB	0.51626	1.80E-10	MA1574.1	0.62739	1.21E-84	0.57183
RORA	0.52145	2.99E-04	MA0071.1	0.61444	1.13E-68	0.56795
XBP1	0.51325	1.94E-05	MA0844.1	0.53200	9.70E-07	0.52263
ZBTB7A	0.51059	7.84E-06	MA0750.2	0.52372	2.84E-04	0.51715
BATF3	0.50175	2.81E-05	MA0835.2	0.52752	2.53E-05	0.51464
c2						
Gene	RNA AUC	RNA p-val	Motif Feature	Motif AUC	Motif p-val	Avg AUC
SOX2	0.51951	1.29E-03	MA0143.4	0.64891	5.87E-83	0.58421
LHX5	0.51138	3.69E-09	MA1519.1	0.60083	5.18E-39	0.55611
PHOX2A	0.50356	2.42E-06	MA0713.1	0.58083	1.13E-25	0.54220
LMX1A	0.50547	1.64E-04	MA0702.2	0.56463	5.54E-17	0.53505
PKNOX2	0.50855	7.46E-04	MA0783.1	0.55271	8.45E-12	0.53063
c3						
Gene	RNA AUC	RNA p-val	Motif Feature	Motif AUC	Motif p-val	Avg AUC
OTX2	0.51592	9.02E-08	MA0712.2	0.81096	0.00E+00	0.66344
PBX3	0.53317	9.56E-06	MA1114.1	0.70998	6.34E-150	0.62157
GLI3	0.52493	3.22E-05	MA1491.1	0.64880	2.93E-76	0.58687
LHX5	0.51420	1.76E-12	MA1519.1	0.62691	5.71E-56	0.57055
TEAD1	0.54419	1.95E-11	MA0090.3	0.57714	9.62E-22	0.56067

LMX1B	0.50268	3.84E-05	MA0703.2	0.60747	1.23E-40	0.55507
LMX1A	0.50620	4.26E-05	MA0702.2	0.59888	1.15E-34	0.55254
LEF1	0.51053	1.92E-04	MA0768.1	0.58696	3.45E-27	0.54874
NR3C1	0.51021	2.95E-04	MA0113.3	0.53761	3.00E-06	0.52391
c4						
Gene	RNA AUC	RNA p-val	Motif Feature	Motif AUC	Motif p-val	Avg AUC
MEF2B	0.50565	2.38E-05	MA0660.1	0.73791	1.16E-169	0.62178
MAFK	0.50421	7.83E-05	MA0496.3	0.71895	5.25E-144	0.61158
POU2F1	0.54469	3.06E-09	MA0785.1	0.67588	1.28E-93	0.61028
POU2F2	0.51933	7.30E-08	MA0507.1	0.66367	2.50E-81	0.59150
GRHL2	0.52967	2.03E-12	MA1105.2	0.62845	8.49E-51	0.57906
MLXIPL	0.50971	1.03E-05	MA0664.1	0.61344	5.24E-40	0.56158
RORA	0.56712	6.14E-18	MA0071.1	0.55342	4.55E-10	0.56027
ELF2	0.52134	2.06E-04	MA1483.1	0.59417	4.30E-28	0.55775
ZBTB7A	0.51231	7.41E-05	MA0750.2	0.59924	5.12E-31	0.55577
FOS	0.50851	2.83E-07	MA0476.1	0.59578	5.27E-29	0.55214
ATF2	0.52102	1.16E-04	MA1632.1	0.56560	1.93E-14	0.54331
PKNOX1	0.51172	1.38E-05	MA0782.2	0.57265	2.28E-17	0.54218
NRF1	0.51934	3.27E-05	MA0506.1	0.55818	1.12E-11	0.53876
CLOCK	0.52855	4.55E-08	MA0819.1	0.53942	4.22E-06	0.53398
RREB1	0.52206	8.16E-09	MA0073.1	0.53745	1.24E-05	0.52976
CREB1	0.52264	1.81E-04	MA0018.4	0.52993	4.77E-04	0.52629
SP4	0.51680	3.14E-04	MA0685.1	0.52932	6.22E-04	0.52306
EGR1	0.51507	1.22E-08	MA0162.4	0.52963	5.44E-04	0.52235
c5						
Gene	RNA AUC	RNA p-val	Motif Feature	Motif AUC	Motif p-val	Avg AUC
FOXP1	0.53169	1.05E-05	MA0481.3	0.65165	7.82E-70	0.59167
SOX13	0.51563	5.27E-05	MA1120.1	0.62164	1.42E-45	0.56863
E2F3	0.54112	4.59E-11	MA0469.3	0.57232	3.62E-17	0.55672
LEF1	0.53334	1.69E-28	MA0768.1	0.56786	2.69E-15	0.55060
RORB	0.51602	4.44E-06	MA1150.1	0.58318	3.36E-22	0.54960
HMBOX1	0.52603	2.84E-08	MA0895.1	0.54912	1.06E-08	0.53757
ELF2	0.53298	1.03E-08	MA1483.1	0.53876	6.34E-06	0.53587
NFATC3	0.51546	1.12E-03	MA0625.1	0.54893	1.20E-08	0.53220
RBPJ	0.53355	2.35E-06	MA1116.1	0.52642	2.08E-03	0.52999
NFATC4	0.52875	1.93E-16	MA1525.1	0.52992	4.92E-04	0.52933
TCF7L1	0.52407	8.61E-04	MA1421.1	0.53357	9.20E-05	0.52882
PRRX1	0.50924	3.73E-13	MA0716.1	0.53727	1.41E-05	0.52326
c6						

Gene	RNA AUC	RNA p-val	Motif Feature	Motif AUC	Motif p-val	Avg AUC
RFX4	0.52036	1.70E-19	MA0799.1	0.85206	6.26E-267	0.68621
PAX3	0.65357	4.62E-283	MA0780.1	0.63319	8.24E-40	0.64338
SOX4	0.56748	1.87E-14	MA0867.2	0.66949	2.29E-63	0.61848
MEIS2	0.53501	4.52E-25	MA0774.1	0.69180	1.27E-80	0.61340
LHX5	0.50948	1.72E-04	MA1519.1	0.71255	1.40E-98	0.61101
TEAD1	0.57614	2.78E-20	MA0090.3	0.63918	2.58E-43	0.60766
GBX2	0.52588	2.80E-50	MA0890.1	0.67346	2.78E-66	0.59967
MEIS3	0.52948	3.06E-18	MA0775.1	0.66357	3.79E-59	0.59653
MEOX2	0.50642	5.70E-19	MA0706.1	0.68625	3.89E-76	0.59633
MEIS1	0.50811	1.13E-07	MA0498.2	0.66745	6.77E-62	0.58778
MSX1	0.50982	8.94E-12	MA0666.1	0.65968	1.90E-56	0.58475
EN1	0.50778	2.09E-18	MA0027.2	0.64575	2.51E-47	0.57676
LEF1	0.53040	8.32E-18	MA0768.1	0.59202	7.31E-20	0.56121
SOX9	0.51070	3.64E-06	MA0077.1	0.61098	3.69E-28	0.56084
EMX2	0.50882	4.09E-06	MA0886.1	0.61154	2.00E-28	0.56018
PKNOX2	0.51880	1.38E-08	MA0783.1	0.58962	6.35E-19	0.55421
EN2	0.50323	1.08E-10	MA0642.1	0.60029	2.71E-23	0.55176
TFAP2A	0.50355	3.84E-08	MA0003.4	0.56395	2.30E-10	0.53375
DMRT3	0.50265	3.00E-08	MA0610.1	0.55438	6.97E-08	0.52852
c7						
Gene	RNA AUC	RNA p-val	Motif Feature	Motif AUC	Motif p-val	Avg AUC
RFX3	0.58135	1.15E-18	MA0798.2	0.82378	5.43E-205	0.70256
EMX2	0.55294	1.10E-152	MA0886.1	0.84457	6.97E-232	0.69875
GSC	0.50184	5.67E-08	MA0648.1	0.87628	4.09E-276	0.68906
SOX4	0.58728	4.16E-21	MA0867.2	0.78500	2.70E-159	0.68614
OTX2	0.52280	5.96E-09	MA0712.2	0.83968	2.06E-225	0.68124
LHX5	0.51233	3.31E-06	MA1519.1	0.84822	8.76E-237	0.68027
LHX2	0.51838	8.72E-65	MA0700.2	0.82829	1.08E-210	0.67334
RAX	0.52847	1.29E-88	MA0718.1	0.80086	2.76E-177	0.66467
RFX4	0.50926	9.19E-05	MA0799.1	0.81664	3.84E-196	0.66295
OTX1	0.50633	2.40E-14	MA0711.1	0.79287	4.26E-168	0.64960
LMX1A	0.51392	3.02E-12	MA0702.2	0.76687	6.35E-140	0.64039
SOX2	0.56090	2.60E-13	MA0143.4	0.67153	6.34E-59	0.61622
MEIS2	0.54002	2.31E-29	MA0774.1	0.69168	4.02E-73	0.61585
PAX6	0.52457	1.69E-69	MA0069.1	0.70672	9.80E-85	0.61564
SOX9	0.51693	3.16E-12	MA0077.1	0.69938	5.87E-79	0.60815
MEIS1	0.52896	1.07E-72	MA0498.2	0.67541	1.57E-61	0.60219
PAX3	0.53605	9.49E-16	MA0780.1	0.66123	2.89E-52	0.59864
LEF1	0.52461	3.53E-11	MA0768.1	0.67186	3.87E-59	0.59824

ZNF148	0.53177	1.35E-05	MA1653.1	0.65593	5.27E-49	0.59385
MEIS3	0.52062	6.72E-09	MA0775.1	0.66687	7.30E-56	0.59375
SOX21	0.51576	5.76E-11	MA0866.1	0.58174	1.23E-14	0.54875
ETV5	0.50986	7.50E-04	MA0765.2	0.57696	3.82E-13	0.54341
ATF6	0.51749	8.08E-04	MA1466.1	0.55864	3.15E-08	0.53806
DMRTA2	0.50811	1.83E-36	MA1478.1	0.56554	6.25E-10	0.53682
c9						
Gene	RNA AUC	RNA p-val	Motif Feature	Motif AUC	Motif p-val	Avg AUC
FOXN3	0.66074	6.98E-54	MA1489.1	0.66787	4.48E-53	0.66431
FOXP1	0.59584	1.49E-25	MA0481.3	0.67538	9.13E-58	0.63561
ELF2	0.54301	4.73E-09	MA1483.1	0.70877	4.46E-81	0.62589
TCF4	0.63145	3.59E-35	MA0830.2	0.61872	2.11E-27	0.62509
NR6A1	0.57283	5.52E-12	MA1541.1	0.67175	1.80E-55	0.62229
E2F3	0.58765	3.65E-28	MA0469.3	0.64378	2.11E-39	0.61572
POU2F1	0.62265	3.74E-37	MA0785.1	0.60317	4.34E-21	0.61291
FOXK2	0.54143	5.26E-11	MA1103.2	0.67891	4.88E-60	0.61017
FOXO3	0.53259	1.57E-09	MA0157.2	0.68085	2.62E-61	0.60672
ETV6	0.55771	8.39E-24	MA0645.1	0.65095	2.96E-43	0.60433
CTCF	0.51675	1.94E-03	MA0139.1	0.67130	3.42E-55	0.59403
RORA	0.62428	6.80E-36	MA0071.1	0.56105	2.45E-08	0.59267
ETS1	0.51005	1.51E-06	MA0098.3	0.66346	2.06E-50	0.58675
ETV5	0.51859	7.69E-10	MA0765.2	0.65386	7.22E-45	0.58622
ERG	0.52536	9.95E-29	MA0474.2	0.63967	2.78E-37	0.58252
ETV1	0.55335	1.12E-11	MA0761.2	0.60448	1.38E-21	0.57891
NFYC	0.53380	4.08E-15	MA1644.1	0.60764	8.14E-23	0.57072
MEF2A	0.56187	3.70E-14	MA0052.4	0.57797	1.06E-12	0.56992
SP4	0.54499	4.21E-14	MA0685.1	0.58606	3.81E-15	0.56552
TBX15	0.51640	4.31E-09	MA0803.1	0.61186	1.65E-24	0.56413
GRHL2	0.56493	2.05E-33	MA1105.2	0.56234	1.24E-08	0.56364
NRF1	0.54951	8.59E-17	MA0506.1	0.57376	1.60E-11	0.56164
VDR	0.50248	1.80E-06	MA0693.2	0.61370	2.88E-25	0.55809
BACH2	0.52111	4.20E-05	MA1101.2	0.56135	2.10E-08	0.54123
MEF2C	0.50829	7.25E-04	MA0497.1	0.57119	7.88E-11	0.53974
ESRRA	0.53367	4.22E-18	MA0592.3	0.53619	9.48E-04	0.53493
OVOL2	0.52617	3.16E-16	MA1545.1	0.53736	6.42E-04	0.53177
IRF2	0.50830	1.65E-03	MA0051.1	0.55220	1.86E-06	0.53025
HES6	0.52471	5.17E-14	MA1493.1	0.53038	5.52E-03	0.52755
ZNF136	0.50697	1.80E-03	MA1588.1	0.54734	1.53E-05	0.52716
MLXIPL	0.51114	7.54E-05	MA0664.1	0.53554	1.17E-03	0.52334
c10						

Gene	RNA AUC	RNA p-val	Motif Feature	Motif AUC	Motif p-val	Avg AUC
FOXN3	0.55822	3.40E-07	MA1489.1	0.58667	5.45E-13	0.57244
c11						
Gene	RNA AUC	RNA p-val	Motif Feature	Motif AUC	Motif p-val	Avg AUC
HES1	0.57603	2.85E-18	MA1099.2	0.59075	5.18E-11	0.58339
SP3	0.55493	1.08E-09	MA0746.2	0.60394	5.51E-14	0.57943
HIF1A	0.54772	6.98E-06	MA1106.1	0.58072	5.22E-09	0.56422
ETV6	0.53147	1.40E-05	MA0645.1	0.58230	2.62E-09	0.55689
c12						
Gene	RNA AUC	RNA p-val	Motif Feature	Motif AUC	Motif p-val	Avg AUC
POU5F1	0.65943	1.17E-28	MA1115.1	0.63720	3.29E-10	0.64832