Single Cell and Combinatorial Analyses of Chromatin Accessibility

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Roman Spektor

May 2019

Abstract

SINGLE CELL AND COMBINATORIAL ANALYSES OF CHROMATIN ACCESSIBILITY

Roman Spektor Ph. D.

Cornell University 2019

The study of chromatin accessibility and DNA methylation have been fundamentally important to the understanding of gene regulation and disease. I developed methyl-ATAC-seq to query the relationship between DNA methylation and accessibility at transposase-hypersensitive chromatin, and to characterize sites of methylation-dependent accessibility in human colorectal tumor cells. Furthermore, I have performed single-cell combinatorial indexing assay for transposase accessible chromatin using sequencing (sci-ATAC-seq) to characterize the changes in cellular heterogeneity and chromatin accessibility in the cortices of Ts65Dn Down syndrome model mice. Using sci-ATAC-seq I identified 26 distinct cell-types in the cortex; I found broad changes in cell-type distribution of varying severity, including a substantial increase in abundance of several classes of interneurons corresponding to a decrease in excitatory neuron abundance. These efforts provide novel tools for analyzing chromatin states and a high-resolution assay of cellular changes that accompany Down syndrome.

Biographical sketch

Roman Spektor was born on November 18th, 1986 in the city of Vinnitsa, in the Ukraine. In 1991, his family immigrated to the United States, landing in Brooklyn, NY where they settled in Bensonhurst. He then spent approximately 13 years aging.

He attended Stony Brook University, earning a B.S. in Biochemistry in May 2008, having done research in the lab of Dr. Vitaly Citovsky under Dr. Shoko Ueki studying Tobacco mosaic virus movement protein. After graduating he continued research at Stony Brook in the lab of Dr. Nurit Ballas, the most competent scientist he's ever met, studying Rett syndrome. He later moved back to New York working for Dr. Roger Pearse in Weill Cornell Medical College studying multiple myeloma, and subsequently with Dr. Joseph Scandura, studying acute myeloid leukemia. At Memorial Sloan Kettering Cancer Center, he worked in the GCL headed by Dr. Agnes Viale, where he learned the value of genomics.

When it came time to get a PhD, he joined the lab of Dr. Paul Soloway at Cornell University, specifically, to study imprinting at Rasgrf1. There he devoted several years of his life to the study of the piRNA targeted long non-coding RNA (lncRNA), the pit-RNA, a ~3kb-long transcript ~30kb upstream of Rasgrf1 shown by his predecessors to be transcribed from the Rasgrf1 imprinting control region (ICR) and direct DNA methylation at the Rasgrf1 ICR in the mouse germline, investigating how the pit-RNA recruits DNA methyltransferases to the Rasgrf1 ICR. Completing this work, he chose to no longer study lncRNAs. For his thesis he assayed DNA methylation at accessible chromatin using mATAC-seq and characterized disease using single-cell ATAC-seq.

“As soon as you’re sure you’re right, there’s no point in your being here.”

- Neal Stephenson, *Anathem*

Acknowledgements

First and foremost, I thank the friends that I've made along the way. So that nobody feels special, I'll name none of you.

Paul Soloway; Thanks for tolerating me and giving me the freedom to eventually do what I think was pretty good work.

Andrew Grimson; Without *my admittedly warped interpretation of* your advice over the years I am positive that I would not have graduated.

John Lis; You have trained all the best graduate students I met along the way. Thanks for letting me into your house.

John Schimenti; You refused to be on my committee, agreed to be on my committee, and then refused to be on my committee.

Eric Alani and Kelly Liu; You'll always be the Chair and DGS of my heart.

Erin Chu; You're the best lab mate to have suffered a few small victories and many major defeats with.

Claudia Mimoso; Best undergrad. Now best grad student. Who knows what you'll be the best at next?

My predecessors in the Soloway lab; You helped me learn the value of rigorous, reproducible, and responsible research and to never compromise my values.

Genomics Facility staff; Nothing I've done would be possible without you.

MBG administrative staff; similarly, nothing I've done would be possible without your help.

Weill Hall CARE staff; So long and thanks for all the mice.

Contents:

List of figures

Chapter 2:

Chapter 3:

List of tables

List of abbreviations

5caC - 5-carboxylcytosine
5fC - 5-formylcytosine
5hmC - 5-hydroxymethylcytosine
5mC - 5-methyl-cytosine
5-mdCTP - 5-methyldeoxycytosine triphosphate
AC - Astrocytes
ACE-seq - APOBEC-coupled epigenetic sequencing
AID - Activation-induced cytidine deaminase
APOBEC - apolipoprotein B editing complex
ATAC-seq - assay for transposase accessible chromatin using sequencing
ATP - Adenosine triphosphate
BER - base excision repair
bHLH - Basic helix loop helix
BORIS - Brother of the regulator of imprinted sites
BS-Seq - bisulfite sequencing
bZIP - Basic leucine zipper
ChIP-Exo - Chromatin immunoprecipitation exonuclease sequencing
ChIP-NEXUS - Chromatin immunoprecipitation experiments with nucleotide resolution through exonuclease, unique barcode and single ligation
ChIP-seq - Chromatin immunoprecipitation Sequencing
CNS - Central nervous system
CpG - Cytosine followed directly by Guanine
CTCF - CCCTC-binding factor
CUT&RUN - cleavage under targets and release using nuclease
DKO - double knock-out cells
DNA - Deoxyribonucleic acid
DNase-seq - DNase I hypersensitive sites sequencing
DNMT - DNA methyltransferase
DS - Down syndrome
EC - Endothelial cells
ENCODE - Encyclopedia of DNA Elements
EX - Excitatory neurons
FACS - fluorescence activated cell sorting
FAIRE-seq - Formaldehyde-Assisted Isolation of Regulatory Elements
FANS - Fluorescence activated nuclei sorting
FRiP - Fraction of reads in peaks
GpC - Guanine followed directly by Cytosine
ICR - imprinting control region
IN - Interneurons
lncRNA - long non-coding RNA
mATAC-seq - methylation assay for transposase accessible chromatin using

sequencing
ME - mosaic-end
Methyl-seq - Methylation sequencing
MG - Microglia
MNase-seq - Micrococcal nuclease digestion sequencing
mSTARR-seq - methyl self-transcribing active regulatory region sequencing
NDR - nucleosome depleted region
NOMe-Seq - Nucleosome Occupancy and Methylome Sequencing
OG - Oligodendrocytes
OxBS-seq - Oxidative bisulfite sequencing
pA-MN - Protein A and MNase
PBS - Phosphate buffered saline
PCR - Polymerase chain reaction
Pro-seq - Precision nuclear run-on sequencing
PWM - position weight matrix
RNA - Ribonucleic acid
RNA-seq - RNA sequencing
RRBS - Reduced Representation Bisulfite Sequencing
sci-ATAC-seq - single-cell combinatorial indexing assay for transposase accessible chromatin using sequencing
sci-seq - Single-cell combinatorial indexing sequencing
SELEX - Systematic evolution of ligands by exponential enrichment
TAB-seq - TET-assisted bisulfite sequencing
TDG - thymine–DNA–glycosylase
TET - Ten-Eleven-Translocation
TF - Transcription factor
TFBS - transcription factor binding site
TF-IDF - Term frequency–inverse document frequency
THS - Transposase hypersensitive sites
Ts - Trisomy
Ts65Dn - trisomy, Chromosome 16 translocation to Chromosome 17, Davisson 65
WGBS - Whole-genome bisulfite sequencing
WT - Wild type

Chapter 1: Introduction

1. Regulation of chromatin accessibility

Despite containing identical DNA, each cell in an organism carries out distinct gene expression patterns during development and in response to environmental cues. The organization of nucleosomes, binding of transcription factors, and DNA modifications are reflective of transcription in a cell. Our understanding of these interactions potentially provides a wealth of information to infer activity within cells, how that signaling is directed, and how it can go awry in disease. Because of the predictive power of these inferences, projects such as ENCODE are a large investment to characterize the chromatin landscape of a wide array of tissues and cell types as a regulatory reference (Kundaje et al., 2015).

1.1 An overview of chromatin organization

In eukaryotes, the process of regulating chromatin organization at genes is tightly regulated. In the nucleus, DNA is wrapped around nucleosomes to form a DNA-protein complex, chromatin. Chromatin is formed of units of nucleosomes, 4-subunit complexes composed of a pair of each histone: H2A, H2B, H3, and H4 (Richmond et al., 1984)**.** Approximately 146bp of DNA wraps tightly around each nucleosome (Luger et al., 1997); DNA in between each nucleosome is bound by histone H1 (Thoma et al., 1979).

The assembly of histones occurs during DNA replication. During S-phase, CAF-1 is

recruited to the replication fork by PCNA and deposits ASF-1-bound H3-H4 complexes onto newly synthesized DNA (Sauer et al., 2018). H2A and H2B are then subsequently deposited by the NAP-1 complex (Aguilar-Gurrieri et al., 2016). DNA sequence encodes nucleosome deposition; *in vitro*, salt gradient dialysis assays reveal that, in yeast, nucleosomes bind to DNA in a pattern roughly resembling *in vivo* placement at promoters, which remain largely nucleosome free (Segal et al., 2006; Zhang et al., 2011). Spacing downstream and upstream of promoters appear disorganized; to recapitulate *in vivo* nucleosome positioning requires the addition of nuclear extract in the presence of ATP, actively regulating nucleosome organization.

The arrangement of nucleosomes in each cell-type is mediated by ATP-dependent chromatin remodeling and accounts for processes such as sliding, ejection of nucleosomes, and exchanging of nucleosomes for variants (Clapier et al., 2017). ISWI and CHD subfamily factors evenly space deposited nucleosomes. SWI/SNF factors remodel and reposition chromatin, altering accessibility. Transcriptional activity at these organized regions can be assayed indirectly by querying the positioning and organization of nucleosomes, covalent modifications of histones, and DNA modifications.

When measuring the positioning of nucleosomes, areas depleted of nucleosomes, nucleosome depleted regions (NDRs) are referred to as “open chromatin”, while areas containing nucleosomes are referred to as “closed chromatin” or “heterochromatin”, in the context of large, chromatin-dense domains. Open chromatin is often capable of supporting transcription factor binding and is

permissive to transcription, whereas closed chromatin is refractive to these.

1.2 Histone modifications

Each histone subunit contains several amino acids that are amenable to modifications, including a N-terminal tail which protrudes out of the nucleosome; these modifications are frequently referred to as the "histone code" (Strahl & Allis, 2000). While dozens of histone modifications have been observed (Zhao & Garcia, 2015); modifications at H3 are currently the most studied in the context of transcriptional regulation. The N-terminal tail of histone H3, contains several lysine groups that interact with writers, readers, and erasers that act upon the histone code (Arrowsmith et al., 2012; Patel & Wang, 2013).

Thus far, we have a limited, though quickly expanding, interpretation of the histone code. The most characterized and studied modifications are at Histone H3. H3K4me1, H3K4me3, and H3K27Ac correlate with positive transcription at open chromatin and are present at promoters and enhancers, each carrying distinctive combinations of "activating marks". Opposite of these activating marks, H3K27me3 is present at inactive enhancers and H3K9me3 is present at closed chromatin, heterochromatin, and correlate inversely to transcription.

Further marks on histone H3 are reflective of processes such as transcriptional elongation (H3K36me3, present over gene bodies), mitosis (H3T3P, H3S10P, H3T11/S28P), and DNA repair (H3K14ac, H3K23ac,H4K20me) (Lawrence et al., 2016). Similarly, variant histones can mark cellular processes such as phosphorylation of

S139 at the variant histone, H2A.X, marking DNA damage (yH2A.X) (van Attikum & Gasser, 2009) and H2A.Z, which potentially marks nucleosome free regions (Jin et al., 2009).

Many proteins that act on these histone marks to have been characterized. Writers such as P300/CBP acetylate histones at H3K27, and SET1A/SET1B, MLL1/MLL2/MLL3/MLL4, and PRDM9 methylate histones at H3K4, marking active transcription at promoters and enhancers. Opposed to these are writers such a Suv39H1/Suv39H2, G9a, and SETDB1, which methylate histones at H3K9 at heterochromatin and EZH1/EZH2 which methylate H3K27 at inactive promoters and enhancers. Reader proteins feature several functional domains, for example, Bromo-domain containing proteins bind acetylated histones, while Chromo domains bind methylated histones. Erasers such as Histone Deacetylases (HDACs) act upon acetylated histones, LSD1/LSD2 and JARID1A/JARID1B/JARID1C deacetylate H3K4; JHDM2/JHDM3 and PHF8 demethylate H3K9; UTX, UTY, JMJD3, PHF8, and KIAA1718 demethylate H3K27 (Calo & Wysocka, 2013; Hyun et al., 2017).

1.3 DNA methylation

DNA is further subject to modification. The most common DNA modification in eukaryotic organisms is 5-methyl-cytosine (5mC). Most studies of 5mC have been focused at sequences of cytosines followed by guanines (CpGs). In animals that have evolved DNA methylation as a regulatory mechanism, this covalent modification to cytosine is indicative of inactive regulatory elements. In mammals, CpGs are

frequently found in clusters, “CpG Islands” (Gardiner-Garden & Frommer, 1987).

Cytosines are modified by the addition of a methyl group to the 5th carbon of cytosine. In mammalian somatic cells, methylation is mediated by three DNA methyltransferases (DNMTs), DNMT1, DNMT3A, and DNMT3B (Reviewed in Smith & Meissner, 2013)**.** DNMT1 is a maintenance methyltransferase, trailing the DNA replication complex methylating cytosines complementary to CpGs that are methylated on a single strand (hemi-methylated). DNMT3A and DNMT3B are de novo methyltransferases, methylating previously unmethylated cytosines. Each of these enzymes is essential for viability but are amenable to knockouts in cell-culture models (Liao et al., 2015; Sakaue et al., 2010).

DNA methylation is erased by two known mechanisms; passive and active demethylation. Passive demethylation occurs when DNA methylation is not maintained across DNA replication. It can be regulated by factors that are antagonistic DNMT recruitment and by overall expression of DNMTs (Reviewed in Piccolo & Fisher, 2014). Active demethylation is driven by two known processes, by TET enzymes and by AID/APOBEC enzymes. Ten-Eleven-Translocation (TET) enzymes, which mediate the oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), which is processed into 5-formylcytosine (5fC), and ultimately to 5-carboxylcytosine (5caC) (Ito et al., 2011), which are recognized and excised by thymine–DNA–glycosylase (TDG). The remaining abasic site is recognized base excision repair (BER) machinery and repaired, leaving behind an unmethylated cytosine (Rasmussen & Helin, 2016). The Activation-induced cytidine deaminase/apolipoprotein B editing complex family

of enzymes, AID/APOBECs, directly catalyze the deamination of 5mC to Uracil (Wijesinghe & Bhagwat, 2012); the resulting mismatch between U and the G on its complementary strand is then recognized and repaired by BER machinery.

Methylation states at regulatory regions are reflective of chromatin state, further discussed in Chapter 2. CpGs at open chromatin are predominantly "hypo-methylated", being depleted of DNA methylation; similarly, closed, heterochromatin, tends to be hyper-methylated and enriched for H3K9me3. Histones at gene bodies are methylated at H3K36me3, contain hyper-methylated DNA, and interacts with nucleosome placement to prevent "leaky" transcription (Teissandier & Bourc'his, 2017).

Though DNMTs have been shown to be catalytically active on naked DNA *in vitro* (Felle et al., 2011), their activities appear to be highly regulated *in vivo*. DNMTs are recruited as part of several repressive complexes; DNMT3B has been demonstrated to bind proteins such as SUV39H1, a writer of H3K9me3 and HP1, a protein that binds these marks (Fuks et al., 2003)**,** while DNMT3A has been demonstrated to interact with SETDB1, another writer of H3K9me3 (Li et al., 2006).

DNMT knockouts are viable in cell culture models. A major model is the human colorectal tumor cell line, HCT116, developed by the Vogelstein group (Rhee et al., 2002). The DNMT knockout phenotype maintains viability in the absence of DNMT3B and of DNMT1. This model has been used to characterize the effect of 5mC depletion on transcription factor binding, nucleosome placement (Lay et al., 2015), and the interactions between DNA methylation and the histone code (Suzuki et al., 2011).

1.4 Transcription Factors

In order to initiate transcription, transcription factors bind to short, specific sequences (motifs) at regulatory elements. Upon binding DNA, transcription factors recruit chromatin remodelers, readers and writers of the histone code, DNMTs and TET enzymes, and recruit RNA polymerase II. These enzymes arrange nucleosomes, modify surrounding nucleosomes, modify the underlying DNA, and transcribe RNA (Jonkers & Lis, 2015). Despite the frequent presence of motifs across the genome, most transcription factors that are bound to DNA are found at open chromatin. The surrounding chromatin context of a transcription factor binding site (TFBS) and underlying DNA modifications modulate transcription factor activity binding.

Because of their direct effect on transcription and cell fate, the discovery of transcription factor motifs and binding sites is important. By sequencing immunoprecipitated transcription factor DNA by ChIP-seq, described in section 2.2, and forming a position weight matrix (PWM) of represented sequences, motifs can be derived. Also, TFBSs have been identified *in vitro* using systematic evolution of ligands by exponential enrichment (SELEX) (Roulet et al., 2002), in which *in vitro* synthesized TFs are bound to known nucleotide sequences on either array chips or sequencers. These interactions at transcription factor binding sites are determined by their structural interactions with the major and minor grooves of DNA. Most transcription factors cannot bind closed chromatin, although class of transcription factors, Pioneer Transcription Factors, bind DNA at closed chromatin (Reviewed in

Zaret & Mango, 2016).

Many transcription factors are sensitive to DNA methylation. These interactions have been measured *in vitro*, using a methylation-informative variant of SELEX, methyl-SELEX (Yin et al., 2017). The weakness of such assays is apparent because of their lack of cellular context such as protein modifications, chromatin accessibility, and integration to nearby regulatory elements. Systems to study the cellular contexts of DNA methylation have relied on inhibiting DNA methylation using inhibitors such as 5-Aza-deoxycytosine and DNA methyltransferase knockouts (Christman, 2002). Similarly, approaches such as mSTARR-seq (Lea et al., 2018), quantify this change by transfecting methylated (or unmethylated) DNA to query reporter activity states.

2. High-throughput sequencing assays to measure chromatin states, transcription factor binding, and DNA modifications

Section 1 detailed the importance of chromatin organization and DNA modifications to our understanding of gene regulation. Several high-throughput sequencing assays have been developed in order to assay accessibility, histone marks, transcription factor binding, and DNA methylation. These assays rely on the same fundamental approach, to isolate DNA at protein or regions of interest, perform high-throughput sequencing, and to quantify the result. Inherent limitations of whole-tissue or multiple cell, “bulk analysis”, is the investigation of average activity of all tissues queried; these limitations and their solutions are discussed in section 3. Furthermore, the assays described in this section are limited to querying DNA, they

seek to indirectly infer transcriptional activity; assays which directly query transcription, such as PRO-seq (Mahat et al., 2016), which are specifically informative of actual, rather than inferred, transcription activity are outside the confines of this thesis.

2.1 Accessibility

Chromatin, in the study of accessibility, can be thought of in two contexts, accessible euchromatin and inaccessible heterochromatin. Within euchromatin, "open" accessible NDRs are indicative of potential transcriptional activation; these NDRs are often occupied by transcription factors, transcriptional machinery, and contain DNA that is frequently unmethylated.

Several methods exist to quantify the localization of nucleosomes: DNase-seq (Song & Crawford, 2010), MNase-seq (Schones et al., 2008), FAIRE-seq (Giresi, Kim, McDaniell, Iyer, & Lieb, 2007), and ATAC-seq (Buenrostro, Giresi, Zaba, Chang, & Greenleaf, 2013). DNase-seq and MNase-seq are similar enzymatic treatment to target nucleosome depleted regions. DNase-seq relies on sequencing digested DNA at NDRs while MNase-seq relies on sequencing DNA protected by nucleosomes during digestion. FAIRE-seq relies on sequencing crosslinking-resistant NDRs by removing protein-crosslinked-DNA by phenol extraction.

Finally, ATAC-seq is the current leading method to measure accessibility in the genome. ATAC-seq relies on hyperactive mutants of Tn5 Transposase, a bacterially derived enzyme. Tn5 Transposase binds mosaic-end (ME) sequences defined by a

19bp sequence, 5'-CTGTCTCTTATACACATCT, forming a dimer when bound to DNA. Once bound, and in the presence of Magnesium, Tn5 will insert mosaic-end containing sequences into DNA in a process referred to as tagmentation (Picelli et al., 2014). This process is exploited for library generation using PCR-handle containing ME sequences, dubbed ME-A and ME-B. Two tagmentation events, leading to fragments containing 5' ME-A sequence and 3' ME-B sequences must occur to yield amplifiable fragments. Tagmentation occurs preferentially at exposed DNA, thus at NDRs, yielding fragments enriched at NDRs. Upon PCR amplification, libraries can be sequenced in nearly a single step. ATAC-seq exploits this by directly treating nuclei with ME-A/ME-B carrying Tn5 transposase yielding directly amplifiable fragments in a single step.

Presently, ATAC-seq is the leading method to measure accessibility in the genome owing to its simplified protocol, low background, and low input requirements. Overall, it stands as a large improvement over previous protocols because of its range of sensitivity across cell/tissue types and quantities.

2.2 Histone modifications and transcription factors

Sequencing methods to localize histone modifications and transcription factor occupancy both rely on identical underlying basic techniques. The most commonly used method to isolate and sequence chromatin to localize histone modifications and transcription factor occupancy is dubbed, appropriately, Chromatin immunoprecipitation sequencing (ChIP-seq) (Johnson et al., 2007). First, DNA is

fragmented either via sonication or enzymatically. In order to pull out proteins of interest and their underlying DNA, ChIP-seq relies on crosslinking factors to their underlying DNA (Gilmour & Lis, 1985), though methods such as Native-ChIP-seq do so under physiological (native) conditions. These factors are then isolated, frequently by immunoprecipitation. DNA bound to immunoprecipitated crosslinked proteins of interest is then freed from surrounding/attached proteins. Finally, adapters are added via ligation and this DNA is then sequenced. This method has been employed to great success but requires large amounts of input and provides a low-resolution localization of TFs and histone modifications in the genome. Re-ChIP assays, by performing two consecutive immunoprecipitations of two different factors, query coincident factors of interest in the genome (Furlan-Magaril et al., 2009; J. V. Geisberg & Struhl, 2004; Joseph V Geisberg & Struhl, 2005).

Recent advancements have lowered the input requirements of ChIP-seq but have been largely superseded by new techniques such as ChIP-Exo (Rossi, Lai, & Pugh, 2018) and ChIP-NEXUS (He, Johnston, & Zeitlinger, 2015), which attempt to solve low-resolution issues by exonuclease digestion to yield strand-specific products terminating at bound proteins, which physically block exonucleases.

Most recently, cleavage under targets and release using nuclease (CUT&RUN) has been introduced, allowing for the lowest current input quantities (Skene & Henikoff, 2017). CUT&RUN relies on a fusion of Protein A and MNase (pA-MN), yielding a MNase that is targetable to antibodies. By incubating nuclei with antibodies against

proteins of interest, washing off excess, and treating with pA-MN to bind antibodies, and adding Ca^{2+} to activate MNase, DNA nearby antibody-bound proteins is selectively cleaved. By allowing this DNA to either diffuse out of nuclear pores or size-selecting for small fragments, highly targeted ChIP-seq libraries can be easily generated using conventional library preparation. Besides its speed and simplicity, CUT&RUN is advantageous because of its ability to query insoluble proteins which usually reside in the nucleus.

2.3 DNA methylation

Sites harboring 5mC are identified by bisulfite sequencing (BS-Seq). 5mC primarily measured using bisulfite conversion in which an unmethylated cytosine is deaminated to uracil, leaving 5mC unchanged (Frommer et al., 1992). Upon PCR, these deaminated U bases are written as T. Using this method, methylation at CpGs is measured as the proportion of aligned CpGs that either retain a C (methylated, read as C) or have been converted to a U (unmethylated, read as T) after bisulfite conversion. Whole-genome bisulfite sequencing (WGBS or Methyl-seq) (Lister et al., 2008, 2009) relies on this process across the genome to measure DNA methylation at all CpGs. This method is extremely costly because of two fundamental issues. First, WGBS is an untargeted approach requiring whole-genome sequencing. Second, because bisulfite conversion depletes Cs in sequencing data, it must be aligned to a limited 3-nucleotide genome, decreasing coverage, requiring even further sequencing depth. In order to assuage the costs of bisulfite-sequencing, targeted

approaches have been developed.

Reduced Representation Bisulfite Sequencing (RRBS) (Meissner et al., 2005) relies on the high CpG content at CpG islands as a method of enrichment. Enrichment is performed by enzymatically digesting DNA, usually by Mspi at CCGG sequences, enriching for CpG-dense regions within a size-distribution suitable for sequencing. The weakness of RRBS rests in its highly-targeted approach, sequencing only CCGG-flanked regions, which, though frequently at regulatory regions, have exceedingly low complexity and frequently do not reflect surrounding DNA (Yong, Hsu, & Chen, 2016).

The primary weakness of all BS-seq assays is in its inability to discern the presence of additional, less frequent, modifications, described in section 1.3, such as 5hmC, which are represented as 5mC in BS-seq data. Further modifications to existing BS-seq have been investigated to clarify this issue, such as 5hmC specific enzymatic modification via TET enzymes using TAB-seq (Yu et al., 2012), APOBECs using ACE-seq (Schutsky et al., 2018), and further chemical modifications to discern 5hmC from 5mC using OxBS-seq (Booth et al., 2012).

2.4 Combinatorial assays

Several assays have been developed to assay the interplay between DNA modifications, histone modifications, and accessibility. These assays have been developed to measure several types of data simultaneously to lower cost and provide greater information content per read.

NOMe-Seq (Kelly et al., 2012) identifies NDRs and measures DNA methylation genome wide; this method exploits the pattern of DNA methylation observed in mammals at CpG dinucleotides and depletion of DNA methylation at GpC dinucleotides. NOMe-seq consists of a GpC methyltransferase treatment of isolated nuclei to place 5mC at accessible GpCs that are not occluded by nucleosomes. Following this, standard bisulfite treatment, library preparation, and sequencing is performed. 5mC measurements at naturally occurring 5mC at CpGs remains identical to its Methyl-seq counterpart, while NDRs are identified by the presence of unnaturally occurring 5mC at GpCs; accordingly, nucleosome free regions are enriched in GpC methylation while nucleosome-containing regions are depleted in GpC methylation. This method allows for the identification of simultaneous placement of nucleosome and DNA methylation. The weakness of this method is identical to that of Methyl-seq in its high expense because of its read-coverage requirements.

Like NOMe-seq, methyl-ATAC-seq, described in Chapter 2, identifies and quantifies NDRs via transposase accessibility and queries the underlying DNA methylation at NDRs. In order to do so, standard ATAC-seq library preparation is performed using modified adapters containing methylated Cytosines, followed by a methylated end-repair to preserve complexity during bisulfite treatment. Unlike NOMe-seq, it has little coverage of closed chromatin, providing limited information about its underlying 5mC. The main advantage of this method is comparably low read requirements comparable to ATAC-seq to ascertain the underlying 5mC state at

accessible chromatin. The primary weakness of this method is in the limited characterization of DNA methylation at closed chromatin.

Methyl-ChIP is a modification of ChIP-seq in which the final library is bisulfite-converted in order to assay DNA methylation at immunoprecipitated DNA-bound proteins.

3. Single-cell DNA analysis

Section 2 of this introduction describes bulk measurements of chromatin state that measure the average state of tissues and hundreds-to-thousands of cells. While this is highly informative when measuring large phenotypic changes in a homogeneous population of cells, subtle changes are often overshadowed. For example, in a disease caused by a small population of cells undergoing a dramatic change, where a 5-fold change in population occurs, from 1% of total cells to 5% of total cells, can be missed by bulk analysis, accounting for a minority of change in abundance of sequenced data of the whole tissue. Efforts to enrich for specific cell populations in a tissue by cell-sorting requires prior knowledge of cell populations, specific antibodies present, and means of purifying these populations. Any method other than unbiased sampling of cells from a tissue necessarily imposes a sampling bias. Furthermore, all tissue and cell culture are highly heterogeneous, containing many independent cell-types.

Single-cell approaches are currently being developed order to solve issues in dealing with tissue heterogeneity. The three common strategies that have emerged

for single-cell sequencing consisting of: low-throughput single-cell library preparation in wells, high-throughput encapsulation of cells using microfluidic devices, and high-throughput barcoding of cells using combinatorial indexing. For the purposes of this thesis, the methods described in the following sections will be focused on the use of single cell ATAC-seq, specifically, the use of combinatorial indexing in section 3.4, which has been used in our lab to query cellular heterogeneity in the brains of Down syndrome model mice. The primary advantage of this method is its focus on regulatory regions, requiring less sequencing coverage compared to whole genome sequencing methods while providing regulatory information at queried regions.

3.1 Applications

Though there are several relevant applications to these methods, they are fundamentally identical to those in Chapter 2. In studies measuring chromatin states, transcription factor binding, and DNA modifications, single-cell sequencing has been predominantly used to query cellular heterogeneity in normal and perturbed tissues. A frequent application of single-cell sequencing is the query of cellular heterogeneity and lineage in cancer populations; a simple strategy is the use of methods such as single-cell whole-genome sequencing to identify abnormal chromosomal duplications, deletions, inversion, and translocations (Reviewed in Gawad, Koh, & Quake, 2016). This is limited by high coverage required in whole genome sequencing, high error rates, and amplification bias.

The application of single cell methylation has shown promise assaying cellular

heterogeneity in tissue, though still requiring higher coverage, regulatory information can be inferred from the methylation state of known regulatory regions. Single cell bisulfite sequencing has been successfully used to identify cell-types in the brain and tumors (Cancer Genome Atlas Research Network et al., 2017; Luo et al., 2017; Mulqueen et al., 2018; Smallwood et al., 2014).

3.2 Well-based methods

The earliest single-cell library preparations consisted of manually pipetting single cells into multi-well plates and assembling their libraries using conventional bulk protocols. This strategy has been applied to most library-types (Chen et al., 2018; Kalisky, Blainey, & Quake, 2011; Navin et al., 2011; Stevens et al., 2017). Frequently, single cells are isolated using either fluorescence activated cell sorting (FACS), limiting dilutions, or single-cell pipetting is used to isolate cells. Well-based strategies exclusively use PCR-based barcoding to identify cells. This strategy has limited application because of large reaction volumes, considerable pipetting loss, and unoptimized reactions. As a strategy, this methodology is frequently no longer considered ideal due to its comparably high cost-per-cell, low-throughput, and overall tediousness.

3.3 Microfluidics

Microfluidics have emerged as a practical strategy to prepare single-cell RNA-seq

libraries (Reviewed in Hwang, Lee, & Bang, 2018). Primarily, microfluidics strategies rely on forming reaction-compartments in aqueous droplets in an oil medium containing a single cell and barcode (usually on a bead). Because of scaled-down and optimized reactions, read depth from these libraries, when successful, shows promise.

Much of this approach relies on costly, frequently proprietary machinery and bead-based barcoding. Because of this high cost, this method has a wide commercial adoption, with multiple devices available on the market, (Reviewed in Haque et al., 2017).

3.4 Combinatorial indexing/split-pool

Single-cell combinatorial indexing (sci-seq) is a more recent strategy for generating large-scale single-cell library preparation. Initially used to sequence whole-genome DNA (Adey et al., 2014), recently sci-seq strategies have been applied to RNA-seq (Cao et al., 2017), Bisulfite-seq (Mulqueen et al., 2018), and ATAC-seq (Darren A. Cusanovich, Reddington, et al., 2018). Sci-seq relies on a split-pool strategy described in a paper aptly titled “DNA Sudoku” (Erlich et al., 2009). In this strategy, cells themselves act as virtually partitioned barcodes; pools of cells are individually barcoded, pooled, and subjected to further barcoding, increasing the overall complexity of each barcode. This process can be repeated as many times as necessary, leading to large pools of uniquely barcoded cells. Combinatorial indexing has been implemented in assays such as ATAC-seq, BS-seq, and RNA-seq. In chapter

3, single-cell combinatorial indexing assay for transposase accessible chromatin using sequencing (sci-ATAC-seq) (Cusanovich et al., 2015) is leveraged to assay changes in the cortex of Down syndrome model mice compared to their control counterparts.

Sci-ATAC-seq uses combinatorial indexing to perform ATAC-seq inexpensively on a large scale. First, tagmentation is performed using Tn5 transposase carrying barcoded oligonucleotides to assign unique barcode combinations to 96 initial wells containing thousands of nuclei per well. These wells are pooled and FACS sorted at 20-25 nuclei/well. An additional set of barcodes is added to each well using PCR, yielding 9,216 barcode combinations per 96-well plate containing 1,920-2400 nuclei. Using this strategy, 87.5%-90% of sequenced cells carry a unique final barcode combination. This strategy is limited by its reagent costs which add considerable expense to the process, though becomes viable when producing Tn5 transposase in-house (Hennig et al., 2018; Picelli et al., 2014). This strategy has allowed for large-scale library preparations of organs and cells and have resulted in several publications showing tissue-wide (Cusanovich, Hill, et al., 2018) and organism-wide (Cusanovich, Reddington, et al., 2018) atlases of accessibility.

Within a common disease model, Down syndrome, most analysis studying changes in cell-type distribution in the central nervous system (CNS) of mouse-model and post-mortem patient tissue has focused on the use of immunofluorescence (Reviewed in Contestabile, Magara, & Cancedda, 2017) and bulk sequencing methods (Guedj et al., 2016); this has led to insight into broad dysregulation that occurs in Down syndrome but fails to accurately quantify specific changes. In chapter 3, I have

characterized the change in cell-type distribution and regulatory changes in the brains of Ts65Dn Down syndrome model mice using sci-ATAC-seq.

Chapter 2: methyl-ATAC-seq measures DNA methylation at accessible chromatin

methyl-ATAC-seq measures DNA methylation at accessible chromatin

Spektor R[1], Tippens ND[2], Mimoso CA[3], Soloway PD[4,5]

1 Department of Molecular Biology and Genetics, Field of Genetics, Genomics, and Development, Cornell University, Ithaca, New York 14853, USA.
2 Tri-Institutional Training Program in Computational Biology and Medicine, Cornell University, Ithaca, New York 14853, USA.
3 College of Agricultural and Life Sciences, Cornell University, Ithaca, New York 14853, USA
4 College of Veterinary Medicine, Department of Biomedical Sciences, Cornell University, Ithaca, New York 14853, USA.
5 College of Agriculture and Life Sciences, Division of Nutritional Sciences, Cornell University, Ithaca, New York 14853, USA

Author Contributions:

RS, NT, and PS conceived this project. RS performed all analysis and generation of libraries. CM performed genotyping and validation experiments.

Abstract:

Chromatin features are characterized by genome-wide assays for nucleosome location, protein binding sites, 3-dimensional interactions, and modifications to histones and DNA. For example, Assay for Transposase Accessible Chromatin sequencing (ATAC-seq) identifies nucleosome-depleted (open) chromatin, which harbors potentially active gene regulatory sequences; and bisulfite sequencing (BS-seq) quantifies DNA methylation. When two distinct chromatin features like these are assayed separately in populations of cells, it is impossible to determine, with certainty, where the features are coincident in the genome by simply overlaying datasets. Here we describe methyl-ATAC-seq (mATAC-seq), which implements modifications to ATAC-seq, including subjecting the output to BS-seq. Merging these

assays into a single protocol identifies the locations of open chromatin, and reveals, unambiguously, the DNA methylation state of the underlying DNA. Such combinatorial methods eliminate the need to perform assays independently and infer where features are coincident.

Introduction:

Active promoters, enhancers, and other gene regulatory sequences are typically bound by sequence-specific transcription factors (TFs), free of nucleosomes, and these facilitate transcription. Such regulatory sequences can be identified by methods that detect nucleosome-depleted regions (NDRs), including DNase-seq, which identifies NDRs by their hypersensitivity to DNase I (Thurman et al. 2012); FAIRE-seq, which identifies NDRs according to their reduced protein content (Gaulton et al. 2010); and ATAC-seq, which identifies NDRs based on their increased accessibility to Tn5 transposase integration, and accordingly are called Transposase hypersensitive sites (THS) (Buenrostro et al. 2013). There is considerable agreement among the regions identified by each assay. ATAC-seq has received further use recently owing to its simplified workflow, reduced material requirements and lower background signals. Additional advancements such as Omni-ATAC (Corces et al. 2017) and Fast-ATAC (Corces et al. 2016) have further improved the utility of ATAC-seq.

DNA within NDRs may have different modification states, including methylation at the fifth carbon of Cytosine (5mC), and oxidized derivatives. In the mammalian

genome, most 5mC is found at CpG dinucleotides, and is generally associated with transcriptionally inactive regions. Bisulfite sequencing (BS-seq) uses selective chemical deamination of unmodified cytosines to uracil, leaving 5mC unchanged. The extent of methylation at a given CpG in a sample is detected after amplification, sequencing, aligning reads to the genome, and then assessing the proportion of aligned reads that retained a C at a CpG, diagnostic of methylation, vs. a T, which reports an unmethylated residue.

Two features of BS-seq dramatically increase costs compared to routine sequencing assays. First, bisulfite treatment reduces the yield and complexity of DNA libraries, resulting in fewer reads uniquely aligning to the genome. Second, to reliably quantify the extent of methylation of a given CpG requires high read coverage. For these reasons, Reduced Representation Bisulfite Sequencing (RRBS) (Meissner et al. 2005) and derivatives (Boyle et al. 2012; Chatterjee et al. 2012; Garrett-Bakelman et al. 2015) have been used to focus analysis on CpG dense regions. However, not all gene regulatory sequences are detected by RRBS, and many regions that are detected are not regulatory.

Integrating results from assays for distinct chromatin features have defined novel categories of regulatory elements. These include bivalent promoters (Bernstein et al. 2006), enhancers (Heintzman et al. 2009), and widely observed chromatin states likely to harbor shared regulatory functions (Roadmap Epigenomics et al. 2015). In

most of these studies, results from assays for single features are superimposed, and when a given locus has signals for multiple features, the features are inferred to be coincident on the same molecule. Though many inferences might be accurate, there is uncertainty inherent in such approaches, owing to the fact that samples commonly contain multiple sub-populations of cells, each with a characteristic chromatin state. Accordingly, the population-averaged results might report chromatin states found in no individual subpopulation of cells. Methods that combine assays for multiple chromatin features in a single protocol can eliminate this ambiguity for the features assayed. Here, we describe methyl-ATAC-seq (mATAC-Seq), a modification of ATAC-seq that combines ATAC-seq with BS-seq, identifying the locations of open chromatin, and the methylation state of the underlying DNA. In addition to providing more reliable assignments of chromatin states, mATAC-seq can focus DNA methylation analyses to accessible regulatory regions of the genome.

Results:

Fig. 1 shows the workflow and sample results for mATAC-seq. It includes two primary modifications during the transposition step of the Omni-ATAC-seq protocol: (1), methylated oligonucleotides are loaded onto Tn5 to generate the transposome (Fig. 1A) which is then used to perform ATAC-seq (Fig. 1B); and (2), 5-methyldeoxycytosine triphosphate (5-mdCTP) is substituted for dCTP during the subsequent end repair step (Fig. 1C). These modifications protect the Nextera adapter sequences during the final step of mATAC-seq library preparation, bisulfite treatment of the tagmented

DNA (Fig. 1D). Use of methylated oligonucleotides, and 5-mdCTP during end repair protects cytosines in the adaptors from deamination caused by bisulfite treatment, which is necessary for successful PCR amplification and sequencing of the resulting libraries. Sequenced libraries provide information on both DNA methylation and Transposase hypersensitivity (Fig. 1E).

We applied mATAC-seq to nuclei prepared from HCT116 colorectal carcinoma cells. mATAC-seq reads in peaks were highly reproducible in biological replicates (r2=0.90, Fig. S1A). To validate that mATAC-seq captured open chromatin domains as well as conventional methods, we compared Transposase Hypersensitive (THS) sites found by mATAC-seq with those we identified using the standard Omni-ATAC-seq protocol (Fig. 2A-D) (Corces et al. 2017). Approximately 92% of called peaks found by Omni-ATAC-seq were found by mATAC-seq (Fig. 2A). There was also strong concordance between mATAC-seq and Omni-ATAC-seq with respect to gene features detected by both assays, with promoter regions being the most commonly identified features (Fig. 2B). In addition, reads in peaks identified by Omni-ATAC-seq and mATAC-seq were well correlated (Fig. S1A, B). Regions of greatest divergence include difficult to map regions such as repetitive elements, low complexity sequences, and simple repeat annotations (Fig. S1C). These analyses demonstrate that mATAC-seq detects open chromatin comparably to traditional Omni-ATAC-seq, and that protocol modifications that enable subsequent bisulfite sequencing do not compromise detection of open chromatin.

To validate that mATAC-seq identified DNA methylation patterns as reliably as conventional methods, we next compared the mATAC-seq methylation data with whole genome bisulfite sequencing (WGBS) data reported for HCT116 cells at THS sites and CpG islands (Blattler et al. 2014). DNA methylation detected by mATAC-seq replicates was highly reproducible at peaks (r2=0.83) and CpG Islands (r2=0.95) (Fig. S2A, B); and methylation levels reported by mATAC-seq correlated well with levels reported by WGBS at peaks (r2=0.68) and CpG Islands (r2=0.85) (Fig S2A, B). THS peaks identified by mATAC-Seq in HCT116 were predominantly unmethylated, and this agrees with existing WGBS data (Fig S2C, D). Fig. 2E and Fig. 2F report DNA methylation patterns assayed respectively by mATAC-seq and WGBS across gene bodies spanning from 2kb 5' of transcriptional start sites (TSS) to 2kb 3' of transcriptional end sites (TES). These patterns are consistent with the high correlations described above. We find these high correlations despite the fact that the assays were performed by different labs; also, WGBS and mATAC-seq assays are different in that mATAC-seq queries DNA methylation at open chromatin, whereas WGBS assays the entire genome, regardless of chromatin state. Our mATAC-seq data showed a reciprocal relationship between accessibility and 5mC density. These are in agreement with previous results from NOMe-seq (Kelly et al. 2012), which can also report sites of accessible chromatin and DNA methylation states but requires much greater sequencing depth. Both assays reveal that highly accessible chromatin is depleted of 5mC, and that there is an abundance of methylation in less accessible

chromatin over gene bodies (Fig. 2E, F). Having shown that sites of open chromatin and DNA methylation states reported by mATAC-seq, Omni-ATAC-seq, and WGBS are in agreement, we concluded that mATAC-seq can be used to simultaneously identify the locations of the genome with accessible chromatin, and the methylation state of the underlying DNA. Because mATAC-seq measures accessibility and methylation in a single assay, it eliminates the inherent uncertainty about coincidence of chromatin features that can arise when ATAC and bisulfite assays are performed independently, and inferences are made after overlaying the two datasets, and at lower costs.

We extended our analyses of HCT116 cells, performing mATAC-seq on HCT116-derived *DNMT1* and *DNMT3B* double knock-out cells (DKO) (Rhee et al. 2002) to assess the functional significance of these methyltransferases on chromatin accessibility and methylation states in parental HCT116 cells. In DKO cells, there were 23,301 hyper-accessible sites, and 3,166 hypo-accessible sites, compared to parental HCT116 cells (Fig. 3A; $|\log2 \text{ fold change}| > 1$, $q < 0.01$); 16,170 THS sites observed in HCT116 cells were unchanged in DKO cells ($|\log2 \text{ fold change}| < 1$, $q > 0.8$). Compared to unchanged sites, hyper-accessible sites in DKO cells were depleted of DNA methylation (Fig. 3B); these sites were enriched for ATF3, FOSL1, FOSL2, BATF, AP1 and JUNB binding motifs (Fig. 3C). These TFs were previously shown to interact more strongly to their binding motifs when unmethylated (methyl-minus TFs (Yin et al. 2017)). We infer that chromatin hyper-accessibility at these sites in DKO cells was due to enhanced binding of the methyl-minus TFs when methylation was diminished;

this had the effect of limiting nucleosome deposition, thus enabling increased chromatin accessibility. Conversely, hypo-accessible sites in DKO cells were modestly depleted of DNA methylation (Fig. 3B), and enriched for SP1, NFYA, SP5, KLF9, KLF14, and KLF3 binding motifs (Fig. 3D). These TFs were previously shown to exhibit less binding when their sites were unmethylated (methyl-plus TFs (Yin et al. 2017)). We infer that chromatin hypo-accessibility at these sites in DKO cells was due to reduced binding of the methyl-plus TFs when methylation was diminished, and that this led to increased nucleosome deposition, and reduced chromatin accessibility. In support of this is the observation that promoters showing the greatest increases in chromatin accessibility in DKO cells were also the promoter that were most extensively hypomethylated (Fig. 3E). These findings and conclusions are consistent with previously described mechanisms whereby TF binding can regulate nucleosome density (Zaret and Carroll 2011). These conclusions may be tempered by the fact that we are assaying methylation at accessible sequences, the same loci, when in an inaccessible state, are underrepresented in our methylation analyses.

To assess how promoter accessibility states detected by mATAC-seq relate to gene expression, we queried existing RNA-seq data from HCT116 and DKO cells (Blattler et al. 2014). Promoters that were hypo-accessible in DKO cells exhibited no significant gene expression changes relative to the corresponding promoters in parental HCT116 cells. At promoters that exhibited no differences in accessibility in the two cell types, there were significant, but very modest differences in mean expression levels. At

promoters that were hyper-accessible in DKO cells, we observed substantial and significantly higher levels of expression in DKO cells relative to HCT116, with expression differences increasing as accessibility increased (Fig. 3F). These are in accordance with previous findings (Kelly et al. 2012), further validating the utility of mATAC-seq, and demonstrating the concordance between the extent of chromatin accessibility at promoters, and promoter activity.

Our analyses so far have separately examined methylation and chromatin accessibility results from mATAC-seq. We next combined methylation and accessibility data to take advantage of added value of the combined results afforded by mATAC-seq. We first performed k-means clustering of DNA methylation levels at THS sites in HCT116 and DKO cells. DNA methylation at mATAC-seq peaks in HCT116 cells formed five distinct clusters (Fig. 4A). In Cluster 1, accessible peaks, and the 1kb intervals flanking the peaks, were hypermethylated in HCT116 relative to DKO cells, with the flanks exhibiting more hypermethylation. Clusters 2 and 3 were hypomethylated at peak centers in both cells; the clusters were respectively hypermethylated in HCT116 cells in one or the other of the two intervals flanking the peaks. Cluster 4 was hypermethylated over the peaks only in HCT116 cells, and hypomethylated in the peak and flanks in DKO cells. Cluster 5 was hypomethylated in the peaks and flanks of both cell types (Fig. 4A, C, D).

When we assessed mRNA expression from promoters within the five clusters,

differences between DKO and HCT116 emerged that varied according to cluster. Promoters in DKO cells from Clusters 1 and 4 were significantly more active than the corresponding promoters from the same clusters in HCT116 cells, with respective increases in mRNA of 2.5-, and 3.2- fold (Fig. 4B). Clusters 2 and 3 exhibited a modest change of 1.3-fold between the cell types. Cluster 5, which was both hypomethylated and hyper-accessible in both cell types, showed no difference in expression.

Besides the differences in DNA methylation and expression, the clusters have additional distinguishing features. There are more promoters, CpG islands, and exons in Cluster 1 compared to Cluster 4; and more intronic and distal intergenic elements in Cluster 4 compared to Cluster 1 (Fig. S3A, B). One feature is the broad domain of H2A.Z in Cluster 1 that accompanied the loss of DNA methylation in DKO cells (Fig. 4E). This finding is consistent with reports that DNA modification and H2A.Z are mutually antagonistic (Zilberman et al. 2008). In Cluster 4, where hypermethylation in HCT116 cells is largely confined to the mATAC-seq peak, there was also an increase in H2A.Z in DKO cells, with the increase being more modest and confined to a narrower portion of the 2kb window displayed. Additional histone modifications associated with active chromatin (H3K4me1, H3K4me3, H3K27ac) were elevated in DKO cells near Cluster 1 mATAC-seq peaks, but these effects were limited or absent in Cluster 4 (Fig. 4E). Like H2A.Z, H3K27me3 was increased in DKO cells at Cluster 1, with the effects also being more modest at Cluster 4 (Fig. 4F). This is also consistent with antagonism reported between H3K27me3 and DNA methylation (Lindroth et al.

2008). In contrast to these histone modifications and variants, H3K9me3 at mATAC peaks was largely unaffected by DNMT-loss. Cluster 5 shows no DNA methylation changes between the two conditions, and there were little to no changes in deposition of histone modifications and variants.

Motifs for TFs, and CTCF binding also varied by cluster. Cluster 1 is enriched for motifs recognized by DNA methyl-plus TFs such as CTCFL, MYC, and BHLHE40; ZFX and ZNF711 contain similar motifs to ZNF704, a methyl-minus transcription factor (Fig 4H). Of the top five TFs enriched in Cluster 4, three are MEF-family TFs, followed by ARNT, which was previously suggested to be methyl-sensitive (Lay et al. 2015) (Fig. 4I). ARNT motifs share substantial sequence identity with BHLHE40, a methyl-minus TF.

Discussion:

ATAC-seq identifies nucleosome-depleted regions of the genome, which are arguably the most relevant for gene regulation within cells. By including bisulfite treatment in the workflow, mATAC-seq targets DNA methylation profiling to open chromatin sites that are enriched for active regulatory regions of the genome. Accordingly, mATAC-seq queries the functional methylome of cells, using relatively few reads compared to WGBS. This is in contrast to other assays for DNA methylation that query the entire genome, or other domains that may not be regulatory.

By applying mATAC-seq to the well-characterized HCT116 cell line, and its DNA methylation-deficient DKO derivative, we demonstrated that mATAC-seq detects DNA methylation patterns that agree with both previously described WGBS results, and with our Omni-ATAC-seq results. These tests validated the fidelity, and compatibility of combining tagmentation and bisulfite treatment steps in the mATAC-seq workflow. DKO cells had many hyper-accessible sites relative to parental HCT116 cells, and these sites exhibited loss of methylation. These same regions were also enriched for methyl-minus TF binding sites, which interact more strongly with DNA when the sites are in an unmethylated context. This highlights the instructive role of TF binding for nucleosome occupancy in the genome. Specifically, our data indicate that when DNA is unmethylated, it facilitates the recruitment of methyl-minus TFs, and that these in turn enable chromatin to assume an open state. Our data also revealed that hyper-accessible and hypomethylated domains in DKO cells were enriched for the histone variant H2A.Z, implicating this factor in limiting DNA methylation, and nucleosome density at sequences where it is recruited. In contrast, regions that displayed no change in methylation showed little change in accessibility. We did not observe a depletion of H3K9me3 at sites with increased accessibility, and decreased DNA methylation, confirming statements in previous studies (Blattler et al. 2014). Such findings were made possible by the combination of DNA methylation, and open chromatin status provided by mATAC-seq.

We envisage that mATAC-seq could be applied to many other systems. For example,

HCT116 derivatives carrying single DNMT knockouts (Rhee et al. 2000; Rhee et al. 2002) would enable us to identify regulatory elements the different DNMTs individually target for methylation, and their respective influences on nucleosome placement. The various DNMTs have been shown to regulate DNA methylation states by independent, as well as cooperative mechanisms (Liang et al. 2002). Repetitive elements are common targets of the DNMTs, and the resulting DNA methylation contributes to their silencing. However, silencing can occur when DNMT activities are impaired, indicating that compensating mechanisms can silence transposons, likely involving H3K9 methylation, and possibly other chromatin modifications (Horard et al. 2009; Karimi et al. 2011; Walter et al. 2016; Jorda et al. 2017). Querying the specificities of the DNMTs, and their influences on chromatin accessibility at repetitive elements using mATAC-seq can elaborate mechanisms underlying repeat regulation.

Additionally, by using HCT116 DKO cells, we studied the effects of DNA methylation depletion that arose by passive mechanisms due to a lack of DNA methylation maintenance. Active demethylation by TET dioxygenases, and AID/APOBEC deaminases, occurring during differentiation and response to stimuli, is a distinct process. Applying mATAC-seq to stem cell differentiation – including under conditions where these active demethylation mechanisms are altered – can reveal both the combinatorial changes in accessibility and DNA methylation, and the effects active DNA methylation mechanisms have on chromatin state during differentiation.

Pioneer transcription factors have the unique property of binding chromatin that is generally inaccessible to other transcription factors (Zaret and Mango 2016). Application of mATAC-seq to systems where pioneer factor functions are altered can reveal the influences these factors have on both chromatin accessibility, and methylation state of the underlying DNA.

Our protocol for mATAC-seq can potentially be integrated with existing methods for combinatorial detection of other DNA modifications including 5-hydroxymethylcytosine (Yu et al. 2012; Booth et al. 2013), 5-formylcytosine (Song et al. 2013), and 5-carboxycytosine (Wu et al. 2016). ChIPmentation uses Tn5 tagmentation in a chromatin immunoprecipitation workflow (Schmidl et al. 2015). This too could be implemented, using steps we developed for mATAC-seq, to identify locations of DNA-bound proteins, and the underlying DNA modification states in a combinatorial detection strategy similar to other methyl-ChIP strategies (Brinkman et al. 2012; Statham et al. 2012). Combinatorial indexing as a low-cost strategy to query single cells can be used to enable the extension of mATAC-seq to a single cell format; specifically, methods such as single-cell combinatorial indexing assay for transposase accessible chromatin using sequencing (sci-ATAC-seq) (Cusanovich et al. 2015), and for methylation analysis (sci-MET) (Mulqueen et al. 2018). Some alterations are necessary to adapt our mATAC-seq protocol for single cell sequencing, including, extending the indexed adapter set to use methylated sci-ATAC-seq adapters during

tagmentation, followed by split-pooling, methylated end-repair, bisulfite conversion in a 96-well format, and PCR. The challenge to this approach is the depletion of reads due to the destructiveness of bisulfite conversion, and the limited sequence complexity in bisulfite converted reads.

Materials and Methods:

Cell Culture:

Cultured cells (#28 HCT116 Parental and #343 DKO) were procured from the Genetic Resources Core Facility at Johns Hopkins School of Medicine and cultured in McCoy's Modified 5A Medium containing 10% heat-inactivated FBS and 1× Penn/Strep (Gibco #15140122). Cells for each experiment were grown apart for at least 2 passages before library preparation.

Genotyping:

DNA from each cell line was extracted using EZ-10 Spin Columns (Bio Basic #BS427) following the manufacturer's protocol. Genotyping PCR was performed on 50ng genomic DNA using oligos from Table S2 from (Das and Chadwick 2016) for 40 cycles using GoTaq (Promega #M3001) (94°C 2 minutes, 40 cycles of: [94°C 30 seconds, 60°C 30 seconds, 72°C 30 seconds], 72°C 5 minutes) and run on a 2% agarose gel. Cells were confirmed to be Mycoplasma-free and HeLa-free via PCR (Rahbari et al. 2009; Young et al. 2010) on 50ng genomic DNA and cell-culture media (Figure S4).

Omni-ATAC-seq:

Cells were trypsinized and subsequently inactivated in cell culture media. Following inactivation, cells were pelleted and resuspended in cold PBS (without Ca++ and Mg++). Cells were stained with Trypan Blue and counted on a hemocytometer. Lysis and tagmentation were performed exactly as described (Corces et al. 2017) with modifications to inactivation and size selection. Briefly, 100,000 HCT116 Parental and DKO cells were lysed on ice for 3 minutes in 50 µL ice-cold Lysis Buffer (10mM Tris pH 7.4, 10mM NaCl, 3mM MgCl2, 0.1% NP-40, 0.1% TWEEN 20, 0.01% Digitonin in DEPC H2O), resuspended in 1mL ice-cold RBS-Wash (10mM Tris pH 7.4, 10mM NaCl, 3mM MgCl2, 0.1% TWEEN 20) and pelleted at 4□C at 500 × g for 10 minutes. Tagmentation was performed in 1 × Tagmentation Buffer (10mM Tris pH 7.4, 5mM MgCl2, 10% DMF, 33% PBS, 0.1% TWEEN 20, 0.01% Digitonin) using 100nM Tn5 Transposase for 30 minutes at 37°C. Tagmentation was inactivated with the addition of 5 volumes SDS Lysis Buffer (100mM Tris pH 7.4, 50mM NaCl, 10mM EDTA, 0.5% SDS in H2O) and 100µg Proteinase K (Invitrogen #25530049) for 30 minutes at 55°C followed by Isopropanol precipitation using GlycoBlue (Invitrogen #AM9516) as a carrier. DNA was size selected using Ampure XP beads (Beckman Coulter # A63880) using a 0.5 × volumes to remove large fragments followed by a 1.8 × final volume according to the manufacturer's instructions. PCR was performed for using Q5 DNA polymerase (NEB #M0491S) with 1 × GC buffer (72°C 5 minutes, 98C 30 seconds, 11 cycles of: [98°C 10 seconds, 65°C 30 seconds, 72°C 30 seconds], 72°C 5 minutes) followed by a final cleanup using a 1.8 × volumes of Ampure XP beads according to the manufacturer's

instructions.

methylATAC-seq:

Cell lysis was performed identically to Omni-ATAC-seq. Tagmentation was performed on 250,000 HCT116 Parental and DKO cells using 700nM Tn5 Transposase assembled using pre-annealed Tn5ME-A_mC and Tn5ME-B_mC (Table S3) for 30 minutes at 37°C following the addition of 0.01ng of unmethylated Lambda DNA (Promega #D1521). We recommend performing a titration of Tn5 transposase to nuclei input to assay minimum amounts required as in Figure S5. Inactivation and size-selection were performed identically to our modified Omni-ATAC-seq protocol. Tagmented DNA was End-Repaired for 30 minutes at 37°C (5U Klenow Exo- (NEB #M2012S), 1 × NEB Buffer 2, and 0.5 mM/each dATP, dGTP, dTTP, and 5-mdCTP (NEB #N0365S)) similar to T-WGBS (Lu et al. 2015) and X-WGBS (Suzuki et al. 2018). End repair was cleaned using a 1.8 × volumes of Ampure XP beads according to the manufacturer's instructions. 10% of the product was kept for quality control PCR (Fig. S6A). Bisulfite conversion was performed using EZ DNA Methylation-Lightning (Zymo #D5030T) following the manufacturer's protocol. PCR was immediately performed using PfuTurbo Cx (Agilent #600410) (94°C 2 minutes, 13 cycles of: [98°C 10 seconds, 6°5C 30 seconds, 72°C 30 seconds], 72°C 5 minutes) (Fig S6B) followed by a final cleanup using a 1.8 × volumes of Ampure XP beads according to the manufacturer's instructions.

Tn5 Transposase:

Tn5 was produced exactly as described (Picelli et al. 2014) with no modifications. For Omni-ATAC-seq, Tn5 transposase was assembled as described (Adey and Shendure 2012) using pre-Annealed Tn5MEDS-A and Tn5MEDS-B from Table S3. For methylATAC-seq, Tn5 transposase was assembled using pre-Annealed Tn5ME-A_5mC and Tn5MEB_5mC oligonucleotides from Table S3. Oligonucleotides were annealed by combining ME-A or ME-B oligos to Tn5MErev and incubating for 2 minutes at 94°C followed by a 0.1□C/s ramp to 25°C. Enzyme was stored at -80°C.

Data Analysis:

Libraries were quantified using the Qubit dsDNA HS Assay Kit (ThermoFisher #Q32854). High-throughput sequencing was performed by the Cornell University Genomics Facility on the Illumina NextSeq 500 with single-end 75bp reads. Trimming for mATAC-seq and Omni-ATAC-seq was performed using fastp (Chen et al. 2018) -q 20 -l 20 -a CTGTCTCTTATACACATCT. Trimming for ChIP-seq, RNA-seq, and WGBS data was performed using fastp -q 20 -l 20 -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC.

Alignment to hg19: In this study we used GRCh37 instead of GRCh38 to match previous studies using similar cells and methods. These results would not be affected by this change because we do not study centromeric sequences and predominantly discuss changes at promoters.

ChIP-seq hg19: Trimmed FASTQ files were aligned using BWA-MEM (Li and Durbin 2010) to hg19. Reads were deduplicated using Picard [http://broadinstitute.github.io/picard/] MarkDuplicates. HCT116 and DKO ChIP-seq data for H2A.Z, H3K4me3, H3K4me1, H3K27Ac, H3K27me3, H3K9me3, and H3K36me3 data (Lay et al. 2015) were downloaded from NCBI GEO database accession GSE58638. HCT116 and DKO ChIP-seq data for CTCF (Maurano et al. 2015) were downloaded from NCBI GEO database accession GSE50610.

RNA-seq hg19: Pair-end trimmed FASTQ files were aligned using HISAT2 (Kim et al. 2015) to hg19. HCT116 and DKO RNA-seq data (Blattler et al. 2014) were downloaded from NCBI GEO database accessions GSE52429 and GSE60106, respectively.

Omni-ATAC and mATAC-Seq: Trimmed FASTQ files were aligned using Bismark (Krueger and Andrews 2011) v0.19.0 to hg19 using the following settings: --score_min L,0,-0.6. Bisulfite reads to be used for MethylKit were filtered for non-conversion using Bismark's filter_non_conversion and deduplicated using deduplicate_bismark. Methylation was extracted using Bismark's methylation extractor --gzip --bedgraph --counts --ignore 9 --ignore_3prime 9. Reads used for peak calling and ATAC-seq visualization were deduplicated using deduplicate_bismark without filtering for non-conversion. Conversion rate (Table S1) was measured by aligning to the lambda genome (GenBank: J02459.1) and filtered as

above; percent conversion rate was calculated as (1-(Total methylated C's in all contexts)/(Total number of C's analyzed)) × 100.

WGBS hg19: Trimmed FASTQ files were aligned using Bismark v0.19.0 to hg19 using the following settings: --score_min L,0,-0.6. Bisulfite reads to be used for MethylKit were filtered for non-conversion using Bismark's filter_non_conversion and deduplicated using deduplicate_bismark. Methylation was extracted using Bismark's methylation extractor --gzip --bedgraph --counts. HCT116 and DKO WGBS data (Blattler et al. 2014) were downloaded from NCBI GEO database accession GSE60106.

Methylation: Differential methylation was quantified using MethylKit (Akalin et al. 2012) at merged HCT116 and DKO mATAC-seq peaks extended to 1kb tiles covering at least 3 CpGs. Promoters were defined as being within 1kb of a TSS using Genomation (Akalin et al. 2015).

Peak calling: ATAC-seq peaks were called using HOMER (Heinz et al. 2010) findPeaks localSize 50000 -size 150 -minDist 50 –fragLength 0 -style dnase. ChIP-seq peaks were called using HOMER findPeaks -style histone. CTCF ChIP-seq peaks were called using HOMER findPeaks -style factor. Reads were assigned to peaks merged from HCT116 and DKO cells using featurecounts (Liao et al. 2013) on reads filtered for a minimum log2 CPM of 0.5 in at least 2 samples. Differential accessibility was called using DESeq2 (Love et al. 2014) lfcShrink. Hyper- and hypo-accessible peaks were defined

as having a | log2 FC | > 1 with an adjusted p value < 0.01 in DKO compared to HCT116 parental cells. Promoters were defined as being within 1kb of a TSS using Genomation. FRiP scores in Table S1 and sample correlation in Fig. S2 were quantified using DiffBind (Stark and Brown 2018) on libraries downsampled to 5M reads using Picard DownsampleSam using peaks called by HOMER. Peak overlaps for Fig. 1A and Fig. S2E were generated using ChIPpeakAnno (Zhu et al. 2010). Feature overlaps for Fig. 1B and S3A were generated using ChIPseeker (Yu et al. 2015). Motif enriched in changed peaks were called using HOMER findMotifsGenome to the hg19 genome using unchanged peaks as background.

RNA-seq quantification: Unstranded hg19-aligned reads were assigned to hg19 genes using featurecounts inbuilt reference using default settings. Differential expression was quantified using DESeq2 lfcShrink on reads filtered for a minimum CPM of 0.5 in at least 2 samples.

Genome browser visualizations: ATAC-seq and mATAC-seq bigWig files were made using eepTools (Ramirez et al. 2016) bamCoverage --binSize 1 --normalizeUsing RPKM --ignoreForNormalization chrM --scaleFactor N and viewed on UCSC genome browser. ChIP-Seq bigWigs were made using deepTools bamCoverage --binSize 10 --normalizeUsing RPKM --ignoreForNormalization chrM --scaleFactor N. Scale factor was determined by coverage of peaks called by HOMER shared between HCT116 and DKO via bedops --intersect where N = (% reads in shared peaks in HCT116)/(% reads

in shared peaks in DKO) when N > 1.1. Scaling was applied to the following samples: DKO_mATAC 1.877, DKO_H3K27ac_R2 1.48, H3K4me3_R2 = 1.47.

Gene body heatmaps were produced using deepTools plotheatmap --beforeRegionStartLength 2000 --regionBodyLength 3000 --afterRegionStartLength 2000 to Ensembl hg19 APPRIS PRINCIPAL:1 flagged transcripts (Rodriguez et al. 2013). Heatmaps for differential peaks were centered on peaks called by HOMER. Peaks from HCT116 and DKO were combined using BEDOPS for clustering of DNA methylation at THS sites. Clustering was performed using deepTools plotheatmap --kmeans 5, the output and order of which was used for all subsequent heatmaps.

Data Access:

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE126215.

ACKNOWLEDGEMENTS:

Jennifer D. Mosher, Ann E. Tate, Jeff C. Mattison, and Peter A. Schweitzer from the Cornell University Biotechnology Resource Center (BRC) for genomic sequencing; Bert Vogelstein for HCT116 cells and derivatives; Önder Kartal, and Erin Chu for proofreading and helpful comments; the Cornell College of Veterinary Medicine and the National Institutes of Health for funding (R01HG006850 and R01GM105243).

Figures:

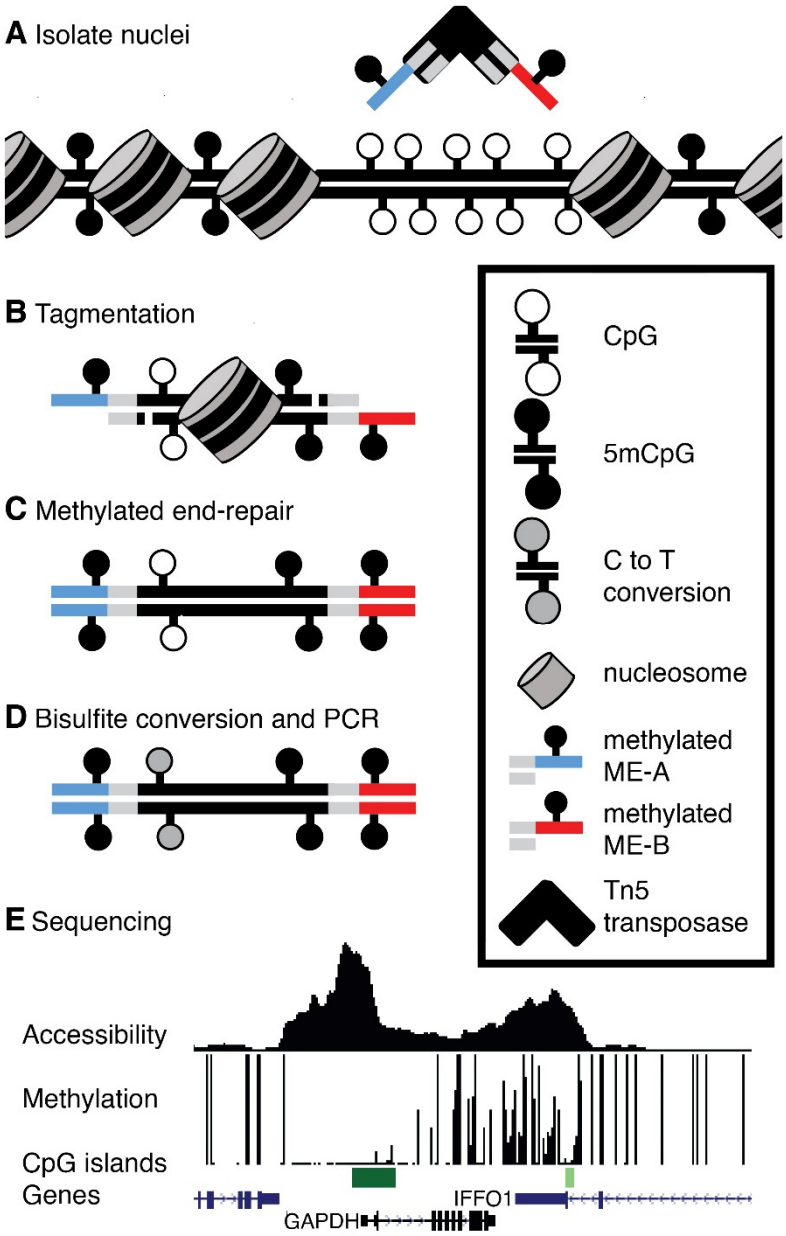


Figure 1:
Overview of mATAC-seq; (A) Tn5 carrying methylated oligonucleotides (red and blue segments) is used to (B) perform tagmentation on nuclei at THS sites. (C) Tagmented DNA is end-repaired using 5mdCTP + dDTPs, purified, (D) Bisulfite converted,

amplified, and (E) sequenced to measure DNA methylation and accessibility simultaneously; sample data are shown for one region in HCT116 cells. Peak height in accessibility track is proportional to read abundance; bar height in methylation track is proportional to extent of methylation at CpGs.

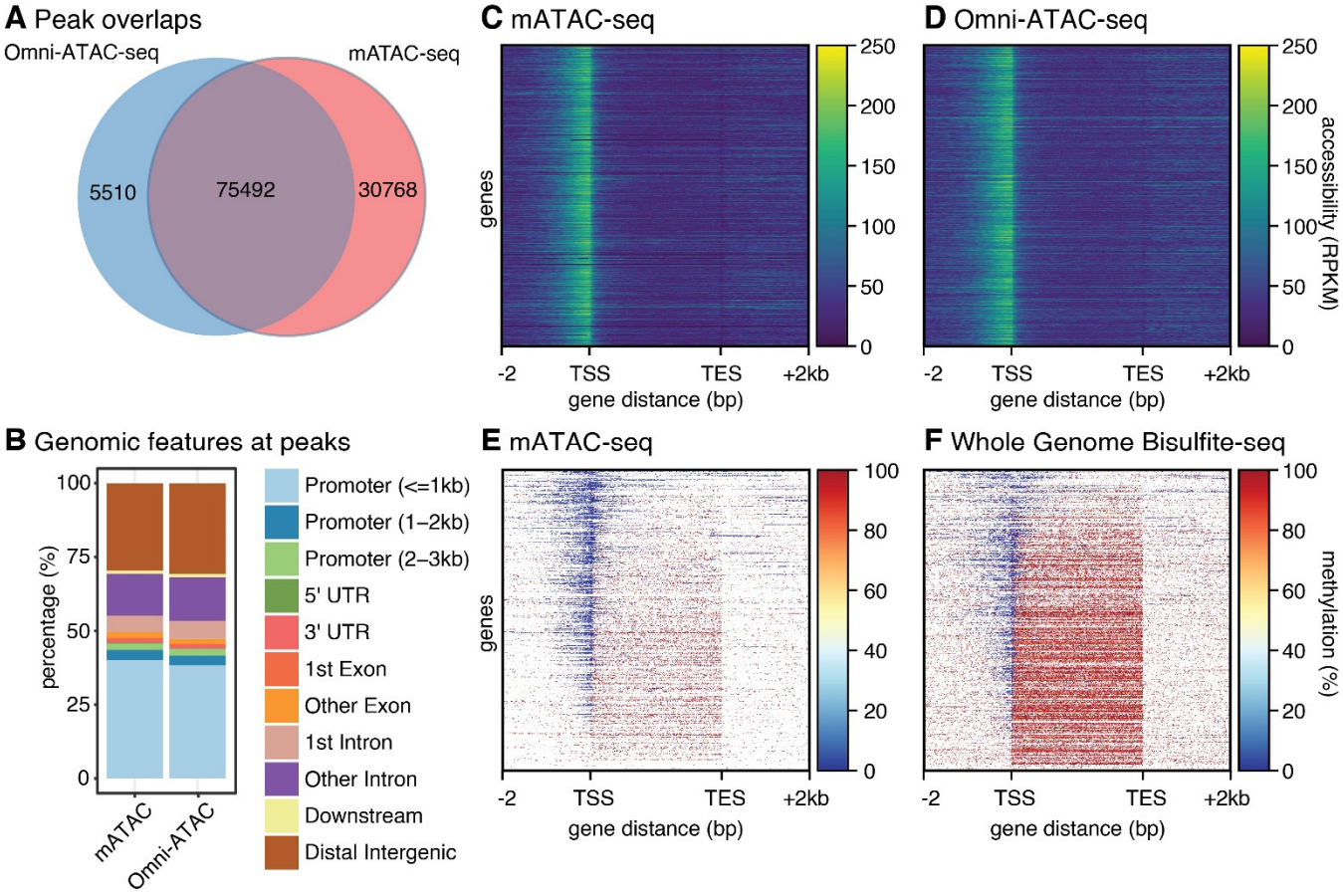


Figure 2:
Comparison of methods; (A) Omni-ATAC and mATAC share a majority of peaks (B) Features at peaks are similar for mATAC and Omni-ATAC. (C) Accessibility in mATAC-seq is comparable to (D) Omni-ATAC-seq at gene bodies +/- 2kb, n=21,305. (E) methylation reported by mATAC is comparable to (F) WGBS at gene bodies +/- 2kb, n=21,305, though WGBS includes data absent from mATAC.

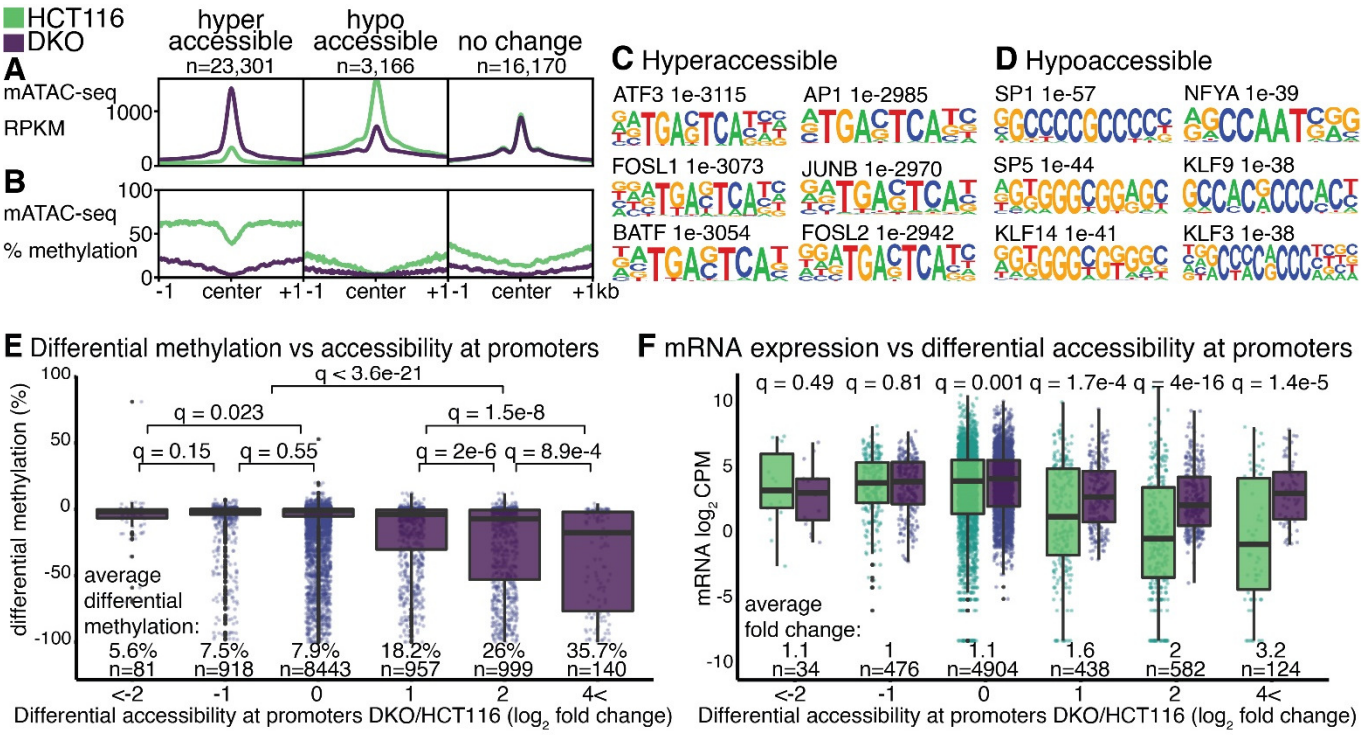


Figure 3:

Accessibility and methylation at peaks; (A, B) significantly changed mATAC-seq hyper-accessible (log2 fold change > 1, q < 0.01, n = 23,310 peaks), hypo-accessible (log2 fold change < -1, q < 0.01, n = 3,166 peaks), and unchanged peaks (| log2 fold change | < 1, q > 0.8 n = 16,170). Motifs enriched in (C) hyper- and (D) hypo-accessible sites compared to unchanged sites. (E) DNA methylation changes at promoters binned by accessibility, reported as the change in methylation ratio of DKO cells relative to HCT116 (DKO/HCT116). (F) mRNA expression changes in DKO cells relative to HCT116, reported as log2 CPM, at genes binned by differential accessibility of their promoters as in (E). q values are for Wilcoxon tests with Benjamini-Hochberg correction.

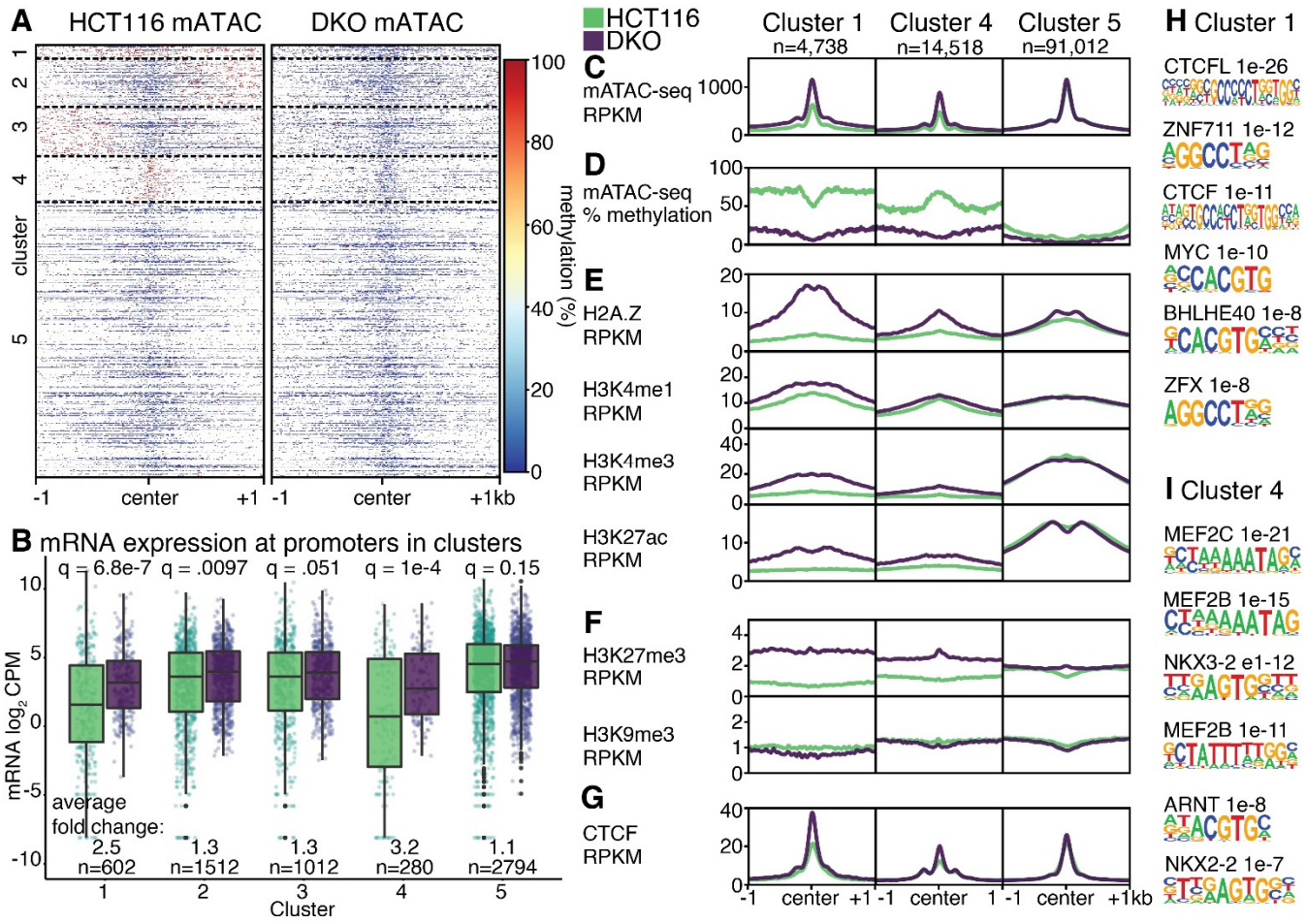
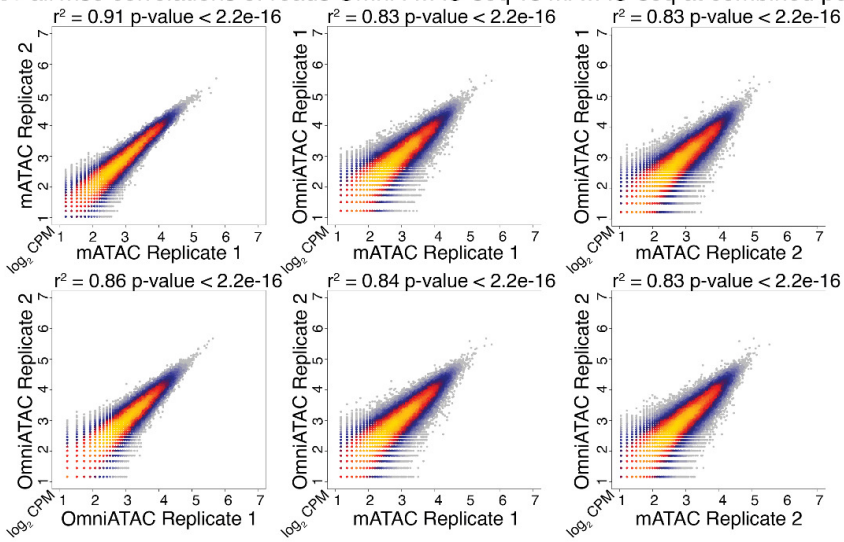


Figure 4:

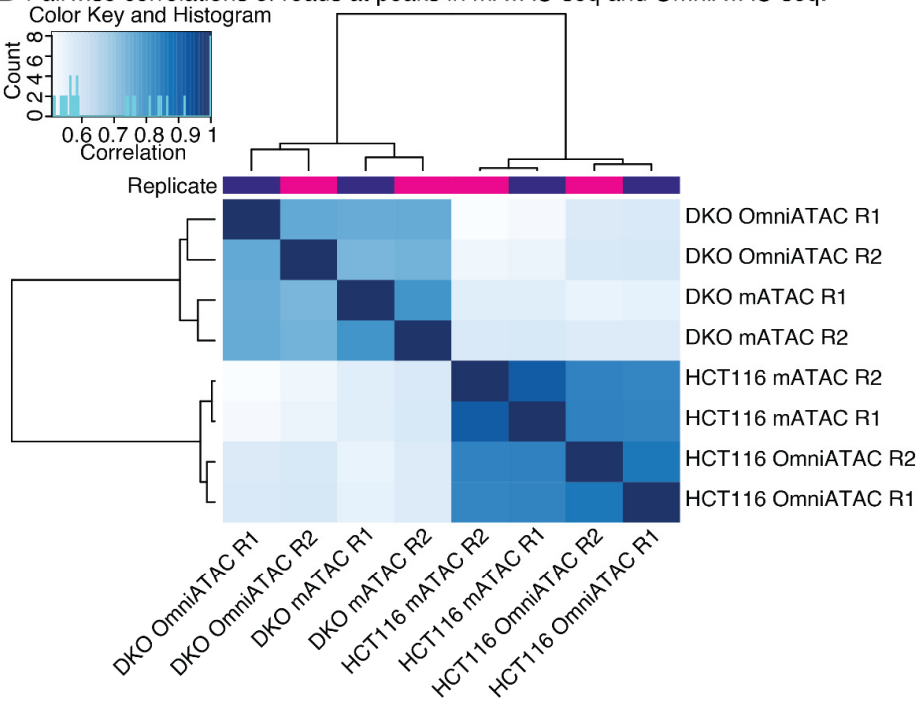
Combined Accessibility and methylation analysis. (A) DNA methylation at mATAC-seq peaks from HCT116 and DKO cells form 5 distinct clusters by DNA methylation. (B) mRNA expression in log2 CPM at identified clusters. Features in clusters 1, 4 and 5 are depicted according to (C) accessibility, (D) DNA methylation, (E) activating histone modifications, (F) silencing histone modifications, and (G) CTCF. Motifs enriched in cluster 1 (H), and cluster 4 (I), compared to cluster 5. q values are for Wilcoxon tests with Benjamini-Hochberg correction.

Supplemental figures:

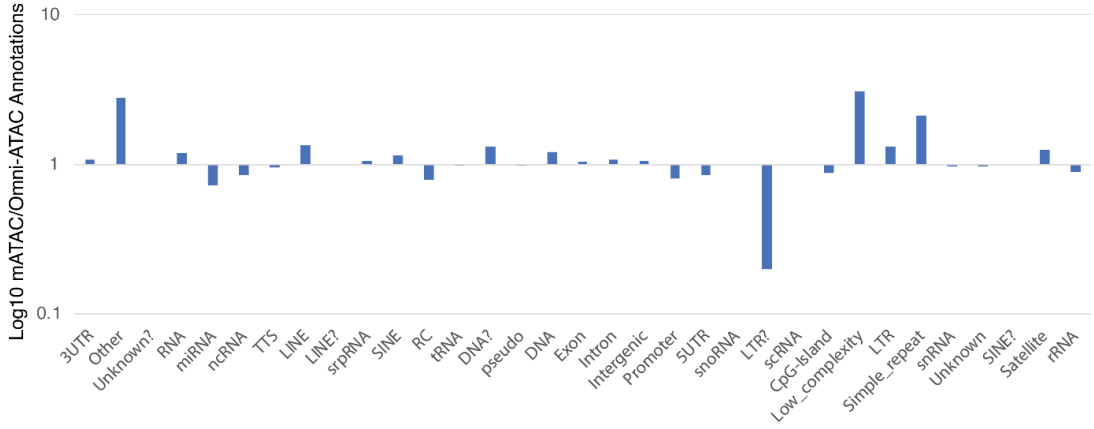
A Pairwise correlations of reads Omni-ATAC-seq vs mATAC-seq at combined peaks.



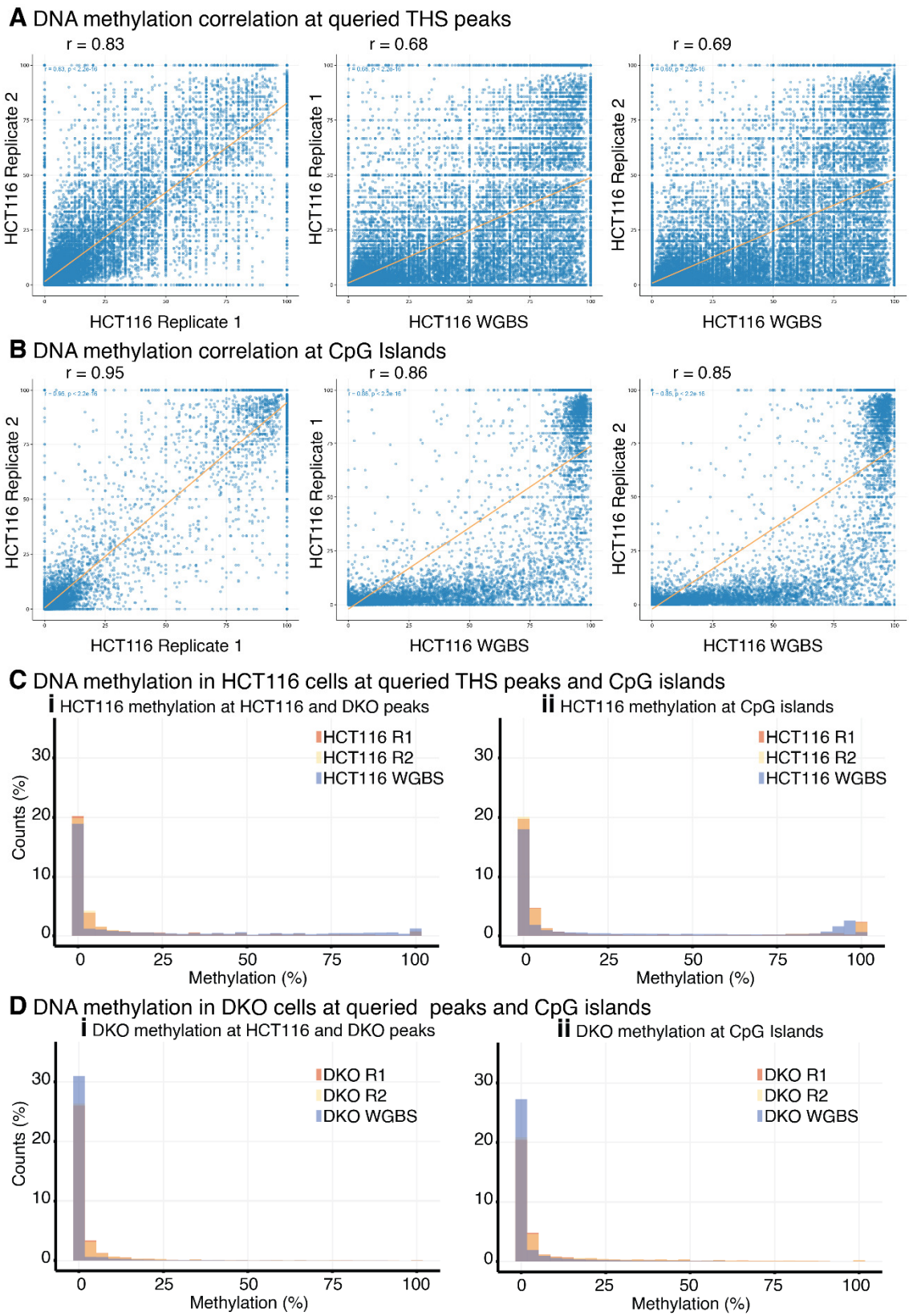
B Pairwise correlations of reads at peaks in mATAC-seq and OmniATAC-seq.



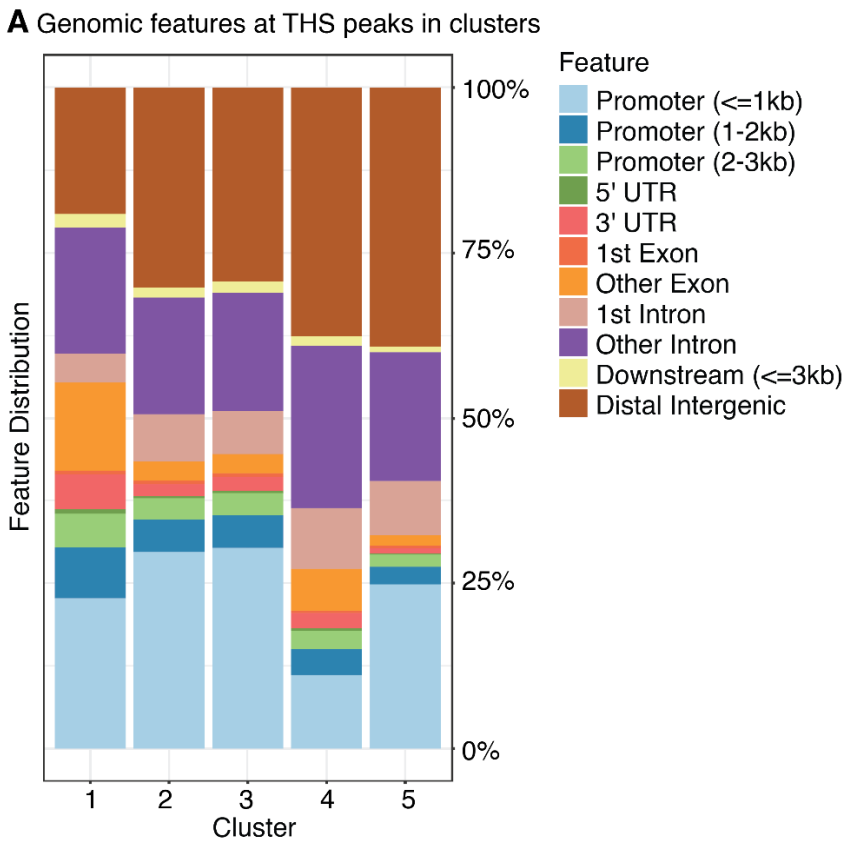
C HOMER peak annotations (% total) in mATAC-seq / Omni-ATAC-seq from Fig.2a

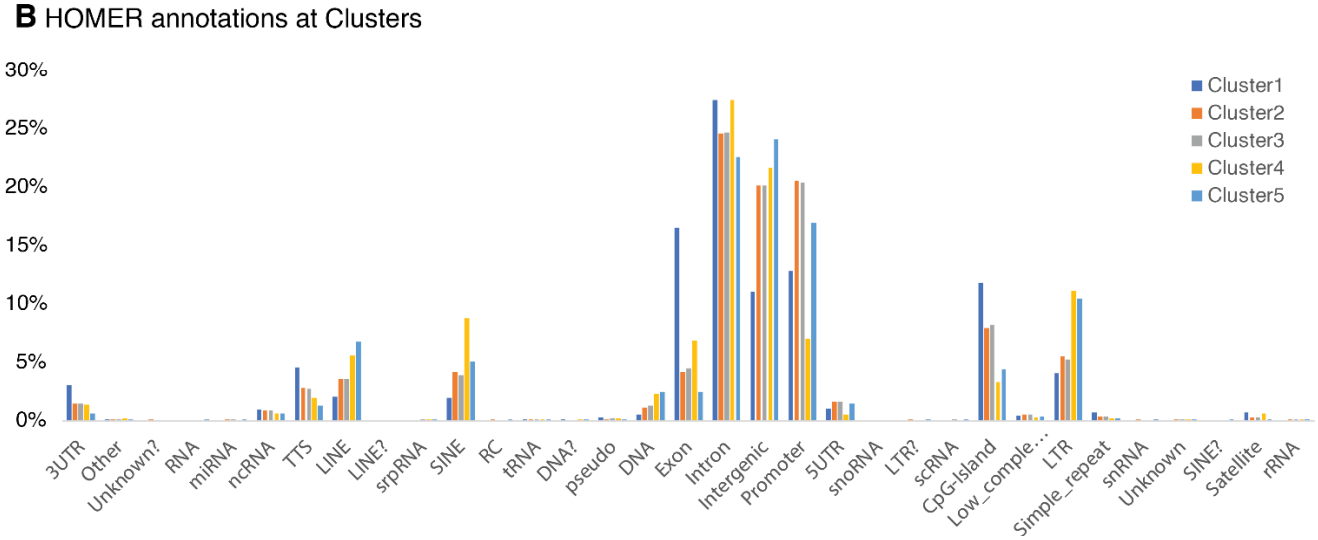


Supplemental Figure 1:
(A) Scatterplots of HCT116 mATAC-seq reads at peaks in libraries downsampled to 5M reads. (B) Pairwise Pearson correlations between mATAC-seq and Omni-ATAC-seq of HCT116 and *DNMT1*/*DNMT3B* Double Knockout (DKO) cells in downsampled libraries. (C) Annotations from peaks (% total) in Omni-ATAC-seq/mATAC-seq in Fig 2A.

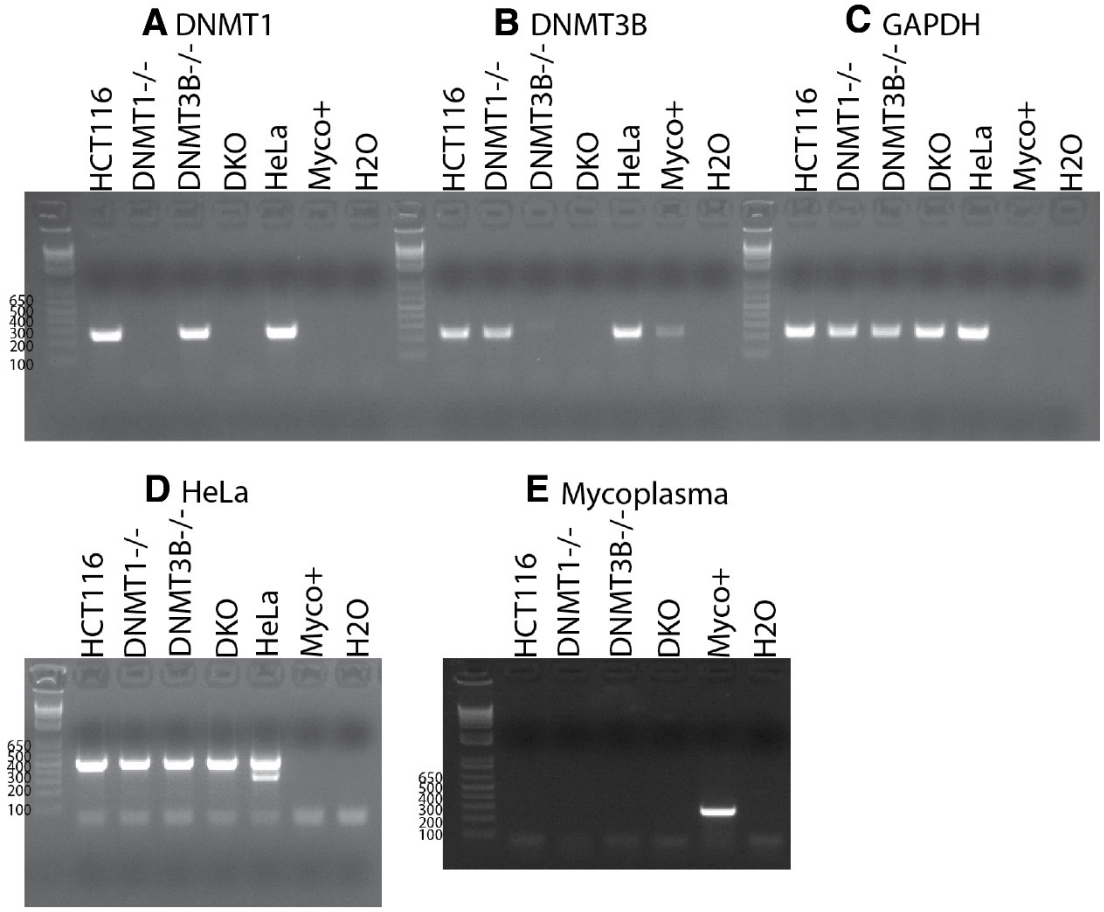

A DNA methylation correlation at queried THS peaks
r = 0.83
r = 0.68
r = 0.69
HCT116 Replicate 2
HCT116 Replicate 1
HCT116 WGBS
HCT116 Replicate 1
HCT116 WGBS
HCT116 Replicate 2
HCT116 WGBS
B DNA methylation correlation at CpG Islands
r = 0.95
r = 0.86
r = 0.85
HCT116 Replicate 2
HCT116 Replicate 1
HCT116 WGBS
HCT116 Replicate 1
HCT116 WGBS
HCT116 Replicate 2
HCT116 WGBS
C DNA methylation in HCT116 cells at queried THS peaks and CpG islands
i HCT116 methylation at HCT116 and DKO peaks
ii HCT116 methylation at CpG islands
HCT116 R1
HCT116 R2
HCT116 WGBS
Counts (%)
Methylation (%)
D DNA methylation in DKO cells at queried peaks and CpG islands
i DKO methylation at HCT116 and DKO peaks
ii DKO methylation at CpG Islands
DKO R1
DKO R2
DKO WGBS
Counts (%)
Methylation (%)

Supplemental Figure 2:
Scatterplots of HCT116 mATAC-seq libraries vs WGBS at (A) THS sites and (B) CpG Islands. DNA methylation in mATAC-seq and WGBS in (C) HCT116 at i. merged HCT116 and DKO peaks and ii. CpG islands and (D) DKO at i. merged HCT116 and DKO peaks and ii. CpG islands. Grey bars denote an overlap of all labeled samples, orange bars denote an overlap of mATAC-seq Replicate1 and Replicate2 samples.

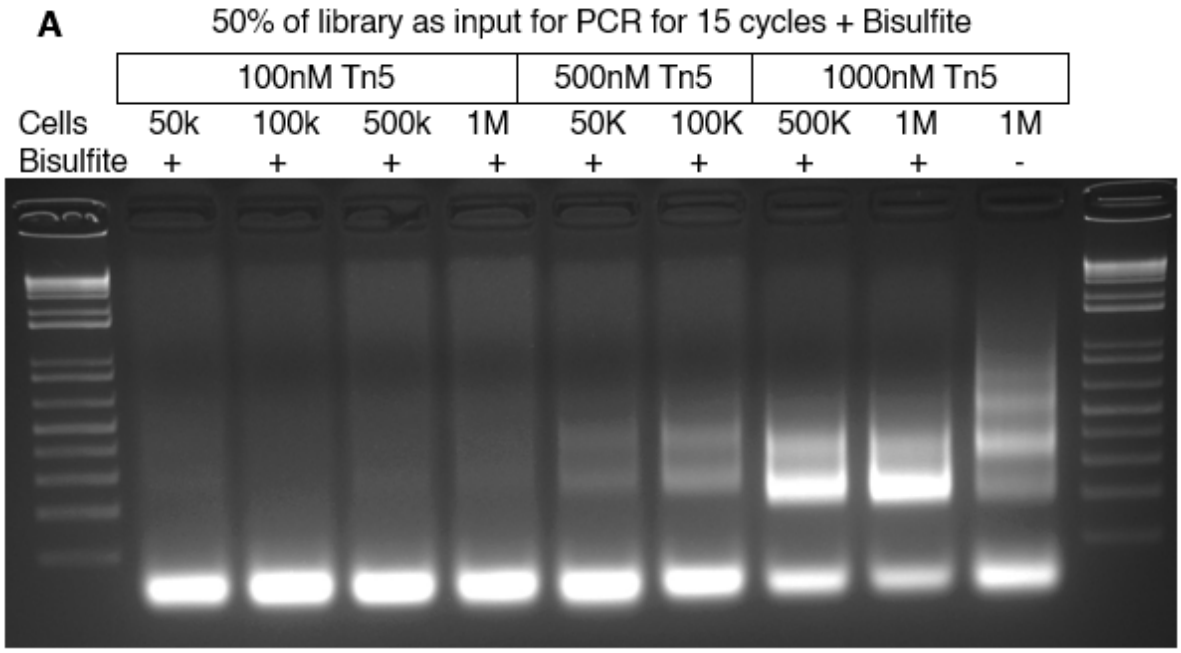


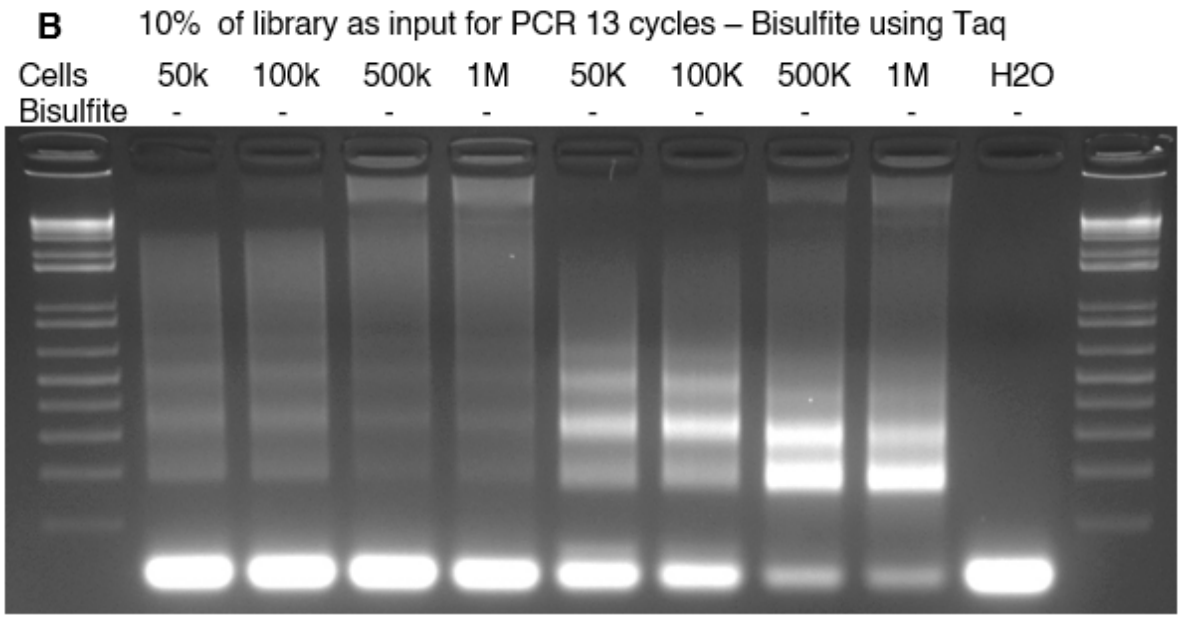


Supplemental Figure 3:
(A) Genomic features, and (B) HOMER annotations for clusters shown in Fig. 4.



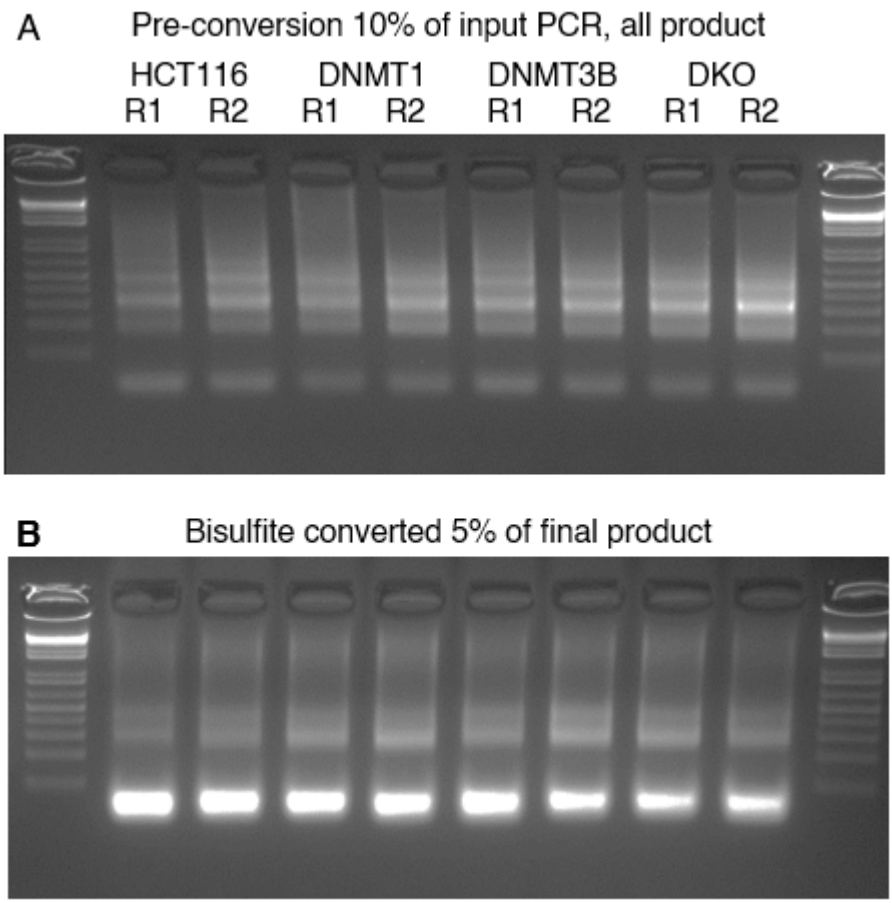
Supplemental Figure 4:
Genotyping by PCR to test for presence of *DNMT1* (A), *DNMT3B* (B), *GAPDH* (C), HeLa cell contamination (D), and Mycoplasma (E) in cells used for this study.





Supplemental Figure 5:
Titration of Tn5 transposase compared to input nuclei. An input range of nuclei and Tn5 were tested using libraries either (A) Bisulfite converted library amplified with PfuTurbo Cx (B) not converted library amplified with Taq to assay viable product for sequencing. Samples were run using a 3.5% Agarose gel in TAE; samples were

resuspended in loading buffer containing 1X SYBR Gold.



Supplemental Figure 6:
QC of final methyl-ATAC-seq libraries. PCR was performed on (A) 10% of pre-converted libraries using Taq (B) bisulfite converted libraries using PfuTurbo Cx. Samples were run using a 3.5% Agarose gel in TAE; samples were resuspended in loading buffer containing 1X SYBR Gold.

Supplemental tables:

Sample:	Trimmed sequences analyzed (M):	mapping:	duplication:	Conversion rate (Lambda mC/TotalC):	FRiP (5M reads):
mATAC-seq HCT116 Rep 1	32.6	45.40%	63.73%	99.76%	0.45
mATAC-seq HCT116 Rep 2	25.1	46.60%	38.42%	99.54%	0.49
mATAC-seq DKO Rep 1	24.6	44.70%	25.12%	98.93%	0.31
mATAC-seq DKO Rep 2	23.2	46.30%	26.22%	99.35%	0.27
Omni-ATAC-seq HCT116 Rep 1	15.0	89.30%	17.66%		0.26
Omni-ATAC-seq HCT116 Rep 2	14.3	89.20%	18.61%		0.28
Omni-ATAC-seq DKO Rep 1	16.5	88.50%	18.92%		0.27
Omni-ATAC-seq DKO Rep 2	22.7	88.20%	12.84%		0.15

Supplemental Table 1:
Samples sequenced in this study. Values show millions (M) trimmed sequences analyzed, mapping efficiency using Bismark, conversion rate of unmethylated Lambda-DNA spike-in after filtering, and Fraction of Reads in Peaks (FRiP) score in down-sampled libraries.

Genotyping oligonucleotides:

DNMT1-Gen-F1:	AAACTGGCAGGTGCTAACTG
DNMT1-Gen-R1:	AGATGTGATGGTGGTTTGCC
DNMT3B-Gen-F1:	TTGGTTTTGCTCAGAGCCAG
DNMT3B-Gen-R1:	ACGTGTGGGCAAGAGATTTC
GAPDH-Gen-F1:	GAAGGTGAAGGTCGGAGTC
GAPDH-Gen-R1:	GAAGATGGTGATGGGATTTC
GPO-3	GGGAGCAAACAGGATTAGATACCCT
MGSO:	TGCACCATCTGTCACTCTGTTAACCTC
VM164A:	TGCCTCCTAGATCGTATTCCC
VM-164B:	GCACTCTGTGGCATGAAGGT
RB164K2:	TGGCTCCTCCCTCTATTATCG

Supplemental Table 2:

Oligonucleotides used in Figure S4 to genotype and verify cell lines were ordered from Integrated DNA Technologies and purified by standard desalting.

Tn5 Oligos:

Tn5MErev:	/5Phos/CTGTCTCTTATACACATCT
Tn5ME-A:	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
Tn5ME-B:	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG
Tn5ME-A_mC:	T/iMe-dC/GT/iMe-dC/GG/iMe-dC/AG/iMe-dC/GT/iMe-dC/AGATGTGTATAAGAGA/iMe-dC/AG
Tn5ME-B_mC:	GT/iMe-dC/T/iMe-dC/GTGGG/iMe-dC/T/iMe-dC/GGAGATGTGTATAAGAGA/iMe-dC/AG
i7 Primers:	CAAGCAGAAGACGGCATACGAGAT[i7]GTCTCGTGGGCTCGG
i5 Primers:	AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC

Supplemental Table 3:

Oligonucleotides used to assemble Transposomes and PCR libraries were ordered from Integrated DNA Technologies and purified by standard desalting. '/iMe-dC/' represents an internal methylated dC. '/5Phos/' represents a 5' phosphate. '[i7]' represents Nextera i7 barcodes. '[i5]' represents Nextera i5 barcodes.

Chapter 3: Single cell ATAC-seq identifies broad changes in neuronal abundance and chromatin accessibility in Down Syndrome

Single cell ATAC-seq identifies broad changes in neuronal abundance and chromatin accessibility in Down Syndrome

Spektor R[1], Yang J[2], Lee S[3], Soloway PD[3,4]

1 Field of Genetics, Genomics, and Development, Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA.
2 Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, New York 14853, USA.
3 Division of Nutritional Sciences, College of Agriculture and Life Sciences, Cornell University, Ithaca, New York 14853, USA.
4 Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853, USA.

Author Contributions:
RS and PS conceived this project. RS generated libraries used in this study. RS and JY performed analysis. SL performed a PCR for the barnyard experiment.

Abstract:

Down Syndrome (DS) is caused by triplication of chr21 and is associated with cognitive impairment, Alzheimer's Disease, and other developmental alterations. The Ts65Dn mouse model for DS has triplication of sequences syntenic with human chr21, and traits resembling those seen in humans with DS. We performed single-cell combinatorial indexing assay for transposase accessible chromatin using sequencing (sci-ATAC-seq) on cortices of adult Ts65Dn mice and control littermates. Analyses of 13,766 cells revealed 26 classes of cells. The most abundant class of excitatory neurons was reduced by 17% in Ts65Dn mice, and three of the four most common classes of interneurons were increased by 50%. Ts65Dn mice display changes in

accessibility at binding motifs for transcription factors that are determinants of neuronal lineage, and others encoded within triplicated regions. These studies define previously uncharacterized cellular and molecular features of DS, and potential mechanisms underlying the condition.

Introduction:

Down Syndrome (DS), affecting one in 800 births in the United States, causes deficits in spatial, long-term, and short-term memory; and impairments in language development and new skill acquisition (reviewed in (Antonarakis, Lyle, Dermitzakis, Reymond, & Deutsch, 2004)). Alzheimer's Disease is common in DS, with a much earlier age of onset relative to the general population (Schupf et al., 1998). Several gross neuroanatomic abnormalities have been described that accompany DS, including reduced volumes of the hippocampus (Aylward et al., 1999; Pinter, Brown, et al., 2001), dentate gyrus (Guidi et al., 2008), parahippocampal gyrus (Guidi et al., 2008; Krasuski, Alexander, Horwitz, Rapoport, & Schapiro, 2002), cerebrum (Jernigan, Bellugi, Sowell, Doherty, & Hesselink, 1993; Pinter, Eliez, Schmitt, Capone, & Reiss, 2001), cerebellum (Jernigan et al., 1993), and amygdala (Aylward et al., 1999). There are larger subcortical gray matter volumes reported in DS (Pinter, Eliez, et al., 2001), and general cortical dysgenesis, particularly affecting the frontal lobe (Wisniewski, 1990).

Accompanying the gross neuroanatomic abnormalities are cellular changes.

GABAergic signaling is altered, which may impair synaptic plasticity, as well as learning and memory by altering the balance of excitatory and inhibitory signaling (reviewed in (Contestabile, Magara, & Cancedda, 2017)). In the hippocampus, there is evidence for less cell proliferation, as fewer cells stain positive for Ki67; and there is also evidence for increased apoptosis, as more cells stain positive caspase 3, and display pyknotic nuclei (Guidi et al., 2008). Efforts to document changes in specific cell-types have shown that neuronal populations are generally reduced in abundance and density in the hippocampus, including the dentate gyrus, lateral parahippocampal gyrus, entorhinal cortex, presubiculum (Guidi et al., 2008), and in cortical layers II and IV (Wisniewski, 1990); in contrast, astrocytes are more abundant in human fetal tissue (Guidi et al., 2008). Comprehensive assessments of changes in cellular composition of the DS brain have not been possible to date, owing to limitations of the available assays.

Studies of DS in human have been aided by mouse models, notably the widely used Ts65Dn mouse (Reeves et al., 1995). This model harbors a reciprocal translocation, and segmental trisomy, leading to triplication of 132 mouse orthologs of the 225 genes found on human chr21 (Reinholdt et al., 2011). Behavioral, anatomical, histological and molecular analyses of Ts65Dn mice identified traits in common with DS in human (reviewed in (Seregaza, Roubertoux, Jamon, & Soumireu-Mourat, 2006)). Notably, relative to control littermates, Ts65Dn perform worse in various learning and memory tasks, and mice have structural abnormalities in several brain

regions. Additionally, Ts65Dn mice exhibit hypomyelination of neurons, and have slower neocortical action potential transmission (Olmos-Serrano et al., 2016). They also share many of the gene expression changes found in humans with DS (Guedj et al., 2016).

Among the mechanisms proposed to explain how chr. 21 trisomy leads to DS is the gene dosage effect model. This posits that genes within triplicated regions are more highly expressed, owing to their increased copy number, and that the elevated expression levels of one or more of the 225 triplicated genes are responsible for DS traits. An alternative, and not mutually exclusive model to the gene dosage effect model is that the triplicated region binds gene regulatory factors that are present in limiting amounts, altering their levels in cells, and the chromatin and expression states of other critical genes outside of the triplicated region. In support of the gene dosage model are results from transcript profiling experiments, which have shown that chr. 21 gene expression is elevated for up to 93 transcripts in various tissues from individuals with DS, and 54 transcripts in three mouse models for DS; these findings are consistent with the gene dosage effect model (Guedj et al., 2016). Among the transcripts in the triplicated region that is upregulated in DS and mouse models for DS is TTC3, an inhibitor of neuronal differentiation (Guedj et al., 2016), which may explain the reductions in neuron abundance and density seen in DS. Triplicated genes also include several transcription- and chromatin-regulatory factors (BRWD1, HMGN1, PRDM15, DNMT3L, USP16, RUNX1, OLIG2, GABPA, ERG and ETS2),

although not all these are upregulated in DS. Chromatin state changes, notably DNA methylation, is altered in DS (Kerkel et al., 2010; Mendioroz et al., 2015). These changes may be due to triplication of the DNA methyltransferase, DNMT3L, however, cellular composition differences among the tissues compared could also be the source of DNA methylation changes. These intriguing findings, and the fact that transcription- and chromatin-regulatory factors are encoded within the triplicated region of DS, provide motivation to evaluate chromatin changes associated with DS.

In order to characterize the cellular and molecular changes associated with DS in greater detail, and to gain further insights into the mechanisms influencing DS traits, we subjected cortices of adult Ts65Dn mice, and their control littermates to single-cell combinatorial indexing assay for transposase accessible chromatin using sequencing (sci-ATAC-seq). This strategy eliminates the challenges of identifying or purifying specific cell-types using antibodies or other reagents, and instead enables their identification and quantification based on shared molecular features. It also enables the characterization of chromatin state changes associated with DS for the different cell populations identified, as well as the sequence features at domains undergoing changes in DS. The unbiased approach afforded by sci-ATAC-seq provides novel and unprecedented insights into the cellular and molecular features associated with DS.

Results:

sci-ATAC-seq library preparation, sequencing, and quality control

To prepare sci-ATAC-seq libraries using adult tissue we implemented several modifications embodied in Omni-ATAC-seq (Corces et al., 2017) to snATAC-seq (Preissl et al., 2018). These modifications increase coverage at peaks, remove additional cellular debris, and minimize mitochondrial contamination. Briefly, we disaggregated two mouse cortices each from control (2n) and Ts65Dn (Ts) littermate males using Dounce homogenization, and isolated nuclei using density centrifugation to remove myelin debris prior to tagmentation and sorting. We confirmed successful disaggregation and removal of debris via microscopy and during fluorescence activated nuclei sorting (FANS) (Fig. S1a).

The sci-ATAC-seq workflow relies on combinatorial indexing to acquire data from individual nuclei (Cusanovich et al., 2015). In this strategy, nuclei were equally distributed in a 96-well plate (~4,800 per well) and subjected to tagmentation of Tn5 transposase carrying 96 distinct oligonucleotide barcode combinations (Amini et al., 2014) in Omni-ATAC-seq buffers. After tagmentation, we pooled nuclei from all wells, and used FANS to sort 25 nuclei per-well into a second set of 96-well PCR plates and subjected the DNA to PCR using an additional 96 barcode combinations per PCR plate. This strategy produced 9,216 barcode combinations per PCR plate, containing libraries from 2,400 single nuclei (Fig. 1a).

We pooled DNAs from single cell library plates, and ascertained library fragment size distribution by Agilent Bioanalyzer, observing the size contributions expected from nucleosome DNA length, plus oligonucleotide lengths inserted by Tn5 and PCR (Fig. S1b). Libraries passing this quality measure were quantified by digital PCR (Fig. S1c) (Schweitzer et al., 2014). Sci-seq barcodes incorporated by tagmentation and PCR impart large stretches of identical sequences to each sequenced molecule, which can produce clustering and base calling problems on Illumina sequencers. In order to eliminate this problem, and increase the number of informative reads, we sequenced our libraries using dark cycle chemistry, skipping uninformative and identical bases (Vitak et al., 2017). We sequenced a total of 19,200 cells to saturation, with a read duplication rate of ~80 % (Table S1) on a NextSeq500.

Fig. 1b summarizes our sequencing results, showing the number of unique reads from each cell assayed for each replicate. The aggregate data had a bimodal distribution; we selected 820 reads/cell at the approximate inflection point as our threshold, carrying forward for further analysis the 72 % of cells with reads exceeding our cutoff (13,884 cells). We combined the resulting reads from each assigned sample and mapped them to the mouse genome to identify peaks from the aggregated data, which we define as pseudobulk data. As a further quality control criterion, we evaluated the fraction of reads from single cells that mapped to peaks (FRiP) and used only cells having a FRiP score greater than 20 % (Fig. S1d), as previously used in other sci-ATAC-seq analyses (Cusanovich, Reddington, et al.,

2018). 72 %, or 13,766 nuclei pass both quality control criteria and showed expected insert-size distributions (Fig S1e).

To assess the reproducibility of our sci-ATAC-seq libraries, we prepared Omni-ATAC-seq libraries from the same tissues in parallel and made two comparisons. First, we compared data from two replicate bulk libraries, and found them to be highly correlated (R2=0.9 for wild type (WT) cortices, Fig. 1c). Next, we aggregated the single cell data from the two corresponding WT sci-ATAC-seq libraries into pseudobulk datasets, and evaluated how well they correlated with each other, and with the true bulk libraries. R2 values for each comparison were 0.88 or greater (Fig. 1c). We also extended these analyses to the Ts65Dn libraries and found all R2 values were 0.87 or greater (Fig. S2). From these quality control tests, we concluded that our sci-ATAC-seq libraries provided reproducible results, and that the total single cell data accurately reflected the chromatin states observed in the bulk tissues.

As an additional data quality measure, we estimated the frequency with which single cell data were derived from more than one cell. For this test, we made a mixture of mouse and human cells, performed sci-ATAC-seq, and identified the collision rate (Cusanovich et al., 2015), defined as the frequency with which any identified cell barcode had both mouse- and human-specific sequences (Fig. 1d). This value was ~9 %, within previously reported expectations of ~12.5 % (Cusanovich et al., 2015; Cusanovich, Hill, et al., 2018).

As a final data quality measure of sequencing coverage, and to confirm proper assignment of samples to barcodes, we created pseudobulk datasets from demultiplexed samples, and performed Copy Number Analysis (Fig. S3a). We expected Ts65Dn pseudobulk data to show evidence for increased copy number of reads within the triplicated region of Mmu16 and Mmu17. This is precisely what we observed, a triplication of Mmu16 spanning chr16:81,000,000-98,200,000 and of a centromeric fragment of Mmu17 at approximately chr17:3,500,001-9,000,000 (Fig. S3b), in agreement with triplicated regions previously reported using Array-CGH (Duchon et al., 2011).

Cell-type identification and validation

We used the sci-ATAC-seq reads from 13,766 cells from both WT and Ts65Dn cortices to perform dimensionality reduction in Monocle to identify a total of 26 cell clusters (Fig. 2a). We assigned the 26 cell clusters identified to the six major cell-types known to be found in cortex: Excitatory neurons (EX), Interneurons (IN), Astrocytes (AC), Oligodendrocytes (OG), Endothelial cells (EC), and Microglia (MG) (Fig. 2b). To do so we aggregated reads from each cluster, preparing cluster- and replicate-specific pseudobulk datasets. We then performed hierarchical clustering of the average relative accessibility at gene bodies of marker genes in each cell-type for each cluster to identify their respective cell-types (Fig. 2c). As confirmation we visually inspected read coverage at marker genes for each cluster (Fig. 2d). We determined that 12 cell

clusters represent Excitatory neurons (EX), eight clusters represent Interneurons (IN), three clusters represent Oligodendrocytes (OG), and one cluster each represent Astrocytes (AC), Microglia (MG) and Endothelial cells (EC). To confirm these assignments, we performed the same clustering analysis using previously published bulk ATAC-seq data, generated using FACS-isolated neuronal subtypes (Gray et al., 2017; Mo et al., 2015) (Fig. S4).

To classify cells in our EX population, we looked at layer-specific markers. Clusters EXc1, EXc2, EXc4, EXc12, EXc13, and EXc23 are accessible at Cux1, Cux2, and Satb2, localizing to cortical layers II/III/IV (Leone, Srinivasan, Chen, Alcamo, & McConnell, 2008). Cluster EXc3, EXc5, EXc6, EXc17, EXc21, and EXc24 are accessible at Bcl11b while being inaccessible at layers II/III/IV markers indicating their localization to layers V/VI; of these, EXc5 and EXc17 are both more accessible at Foxp2, and Pcp4, indicating their localization to layer V (Molyneaux, Arlotta, Menezes, & Macklis, 2007). In our IN population, we assigned cell-types based on their accessibility at interneuron marker genes. INc8 and INc9 are dopamine D2 and dopamine D1 receptor accessible medium spiny neurons, respectively. Cluster INc14 contains Parvalbumin accessible interneurons, INc19 contain Somatostatin accessible interneurons, and INc15, and INc22 contain Vip accessible interneurons. We were not able to clearly identify the remaining two interneuron clusters; INc10, a class hyper-accessible at Frmd7, Sp8, Vipr2, Etv1, and Ntsr1 which appears to resemble Int16 in Zeisel et al (Zeisel et al., 2015), possibly localizing to layers V/VI of the cortex

(Gong et al., 2007), and INc11 as a class accessible at Lhx8, Zic4, and Myo3a. Of our three oligodendrocyte populations, OGc25 contained oligodendrocyte progenitor cells accessible at Pdgfra while inaccessible at Mog.

To further confirm our cell-type assignments we identified genic regions and their 2kbp flanks that are hyper- accessible in each cell-type and cluster relative to all other cell-types (Fig. S5a, S5b) and clusters (Fig. S5c-e). Gene Ontology (GO) analysis at hyper-accessible gene bodies in each cell-type showed findings were consistent with expectations for these cell-types (Fig 2e), providing further confidence in our assignment of cells comprising the 26 cell clusters to their corresponding six cell-types. Neurons showed GO enrichment for genes enriched in categories such as regulation of membrane potential, and modulation of synaptic transmission. Oligodendrocytes showed GO enrichment for escheatment of neurons and gliogenesis. Astrocytes showed less indicative GO categories such as arachidonic acid metabolic processes and Keratinocyte differentiation. Microglia showed enrichment in genes involved in cytokine production and leukocyte migration. Endothelial cells showed enrichment for genes involved in angiogenesis and epithelial cell proliferation.

Our assignments match previous reports quantifying cell-types in the cortex (Preissl et al., 2018). We observed most cells in WT mice to be excitatory neurons (64 %), followed by interneurons (21 %), oligodendrocytes (9 %), microglia (3 %), endothelial

cells (2 %), and astrocytes (1 %) (Fig. 2f). For clarity, we have prefixed all clusters to their assigned cell-types (Table S2a).

We then identified transcription factor (TF) motifs found at peaks of open chromatin that were enriched in each cell-type, relative to others. In EX population, hyper-accessible peaks were enriched for binding sites for TFs contributing to excitatory neuron specification such as NeuroD1 and NeuroG2. Similarly, they were enriched for Tbr1 motifs, a factor important to neuronal migration and axonal projection (Huang et al., 2014). Hyper-accessible peaks in our IN population were enriched Lhx3 motifs, a factor known to have a role in the development of interneurons (Thaler, Lee, Jurata, Gill, & Pfaff, 2002) (Fig. 2g). We observed an enrichment of Sox-family TF motifs in OGs and ETS-family transcription factors in MG populations. In ACs we observed an enrichment of NF1 sites, a regulator of astrocyte proliferation (Bajenaru et al., 2002). Within our ECs we observed an enrichment of COUP-TFII motifs, a regulator of endothelial identity (Aranguren et al., 2013). Such findings provided additional confidence in our assignments of cell clusters to known cell-types.

Changes in cell-type representation in Down Syndrome

Having identified 26 cell-types in adult cortices, and the six cell-types to which they belong, we asked how the Ts65Dn triplication affected these populations. We first evaluated cell abundances. Cortices of DS animals had significantly more interneurons; and specifically, among the eight cell clusters comprising interneurons,

five were significantly larger in Ts65Dn animals (Fig. 3a, Fig. 3b). This finding agrees with reports that interneurons are more abundant in DS (Chakrabarti et al., 2010; Hernandez-Gonzalez et al., 2015). Specifically, we observed a large change in INc8 and INc9, which are Drd2 and Drd1 positive medium spiny neurons, respectively, and INc10 and INc11. We did not observe a statistically significant change in Parvalbumin positive (INc14) or Somatostatin positive (INc19) interneurons (Fig. 3b).

Although the change in abundance of excitatory neurons was not statistically significant, among the 12 excitatory neuron clusters, three had significantly fewer cells in the Ts65Dn animals, including the most populous of the 12 clusters, and one had significantly more cells. Each of these clusters appears to localize to layers II/III/IV. This corresponds with known delays in the development of the cortex in prenatal Ts65Dn mice; though this change has not been measured in as great a detail in adults (Chakrabarti, Galdzicki, & Haydar, 2007). Interestingly, the single increased EX cluster showed the highest relative accessibility at Synj1, regulates levels of membrane phosphatidylinositol-4,5-bisphosphate; may affect synaptic transmission (Trapani, Obholzer, Mo, Brockerhoff, & Nicolson, 2009). Beside these changes in neuronal populations, we observed significantly fewer endothelial cells than their WT control littermates (Fig. 3a, 3b), a previously undocumented change that may influence impaired endothelial function reported in Down Syndrome patients (Cappelli-Bigazzi et al., 2004).

In addition to performing cell clustering in Monocle, we used a TF-IDF-based pipeline (Cusanovich, Hill, et al., 2018), identifying 25 cell clusters (Fig. S6). When we assigned these to the six cell-types using the above strategy, we found 92.5 % agreement in our identified cell-types and cell-types identities in clusters called using TF-IDF.

Gene ontology at differentially-accessible peaks

In order to assess biological processes changed in Ts65Dn mice, we performed Gene Ontology analysis on genes nearest to differentially-accessible peaks within each cell-type (Fig 3c) and cluster, (Fig. S7a) relative to WT mice. In our EX clusters, we observe significant enrichment in ontologies such as CNS development, cell-to-cell adhesion, neuron differentiation, and projection organization. In our IN cell-type, we see enrichment for neuron differentiation, axon development, and cell morphogenesis, forebrain development, and response to cyclic compounds. In our OG cell-type we observe enrichment in CNS differentiation, adhesion, and cell morphogenesis. In endothelial cells, we see enrichment of genes involved in the regulation of cell-cycle and phase transition; specifically, genes within these functional categories are hypo-accessible in Ts65Dn, which is likely to correlate with their diminished expression. We observe no enrichment of terms in astrocytes and microglia.

In order to further understand the contribution of differentially accessible peaks within each cell-type to pathology, we performed Disease Ontology analysis on human homologues of genes nearest to differentially-accessible peaks in each cell-

type (Figure 3d) and cluster (Fig. S7b). We observe genes nearest to differentially accessible peaks in neuronal clusters are enriched in mood and developmental disorders and epilepsy. Furthermore, in Ts65Dn mice we observe an increased representation of epilepsy-related genes, mood-disorder, and developmental-disorder-related genes in our IN cell-types, with a slight enrichment for genes related to developmental disorders of mental health in our EX cell-type.

Transcription factor motifs at differentially-accessible peaks

In addition to these measures of cell abundance, we identified TF binding motifs at differentially accessible peaks between Ts65Dn mice and their WT littermates for each of the six cell-types (Fig 3e) and 26 clusters (Fig. S8). Our first expectation was to observe increased binding in motifs associated with triplicated transcription factors (ERG, GABPA Runx1, and Bach1); which we observed within the top 20 enriched TFs at hyper-accessible sites in microglia, matching our expectations. Of the 91 hyper-accessible peaks in microglia, these transcription factors account for 31 %, 24 %, 20 %, and 4.4 %, of peaks observed, respectively.

The most radical change at hypo-accessible peaks interneurons of Ts65Dn mice occurred at bHLH-family TFs (NeuroD1, Olig2, Atoh1, NeuroG2, Tcf21), which have similar motifs, and dictate neuronal differentiation paths. Specifically, we observe enrichment of Olig2 motifs, which is encoded in the triplicated region, at both hyper- and hypo-accessible peaks in Ts65Dn interneurons. Complementary to this, we

observe hyper-accessibility at bZIP-family TFs (AP-1, Fra1, JunB, ATF3, Fra2, BATF, Fosl2) in Ts65Dn interneurons; these factors have been shown to be highly active in interneurons after hippocampal seizures (Dragunow, Yamada, Bilkey, & Lawlor, 1992). Outside of this enrichment in interneurons, we observed increased binding at MADS-family TFs (Mef2a, Mef2b, Mef2c), which have been shown to specify neural differentiation and is implicated in neurodevelopmental disorders (Harrington et al., 2016; Li et al., 2008; Mayer et al., 2018). Similarly, we observe an increase in accessibility at peaks containing motifs matching Eomes, a regulator of neurogenesis (Arnold et al., 2008), in Ts65Dn interneurons. Overall these results suggest to us a large change in cell-lineage commitment along with potentially aberrant localization of Interneurons.

We observe fewer significant changes outside of Interneurons, our most affected cell-type. Within excitatory neurons there is an enrichment in Homeobox-family Tfs (Lhx1, Lhx2, Lhx3, Isl1, Nkx6.1) at hyper-accessible sites, many of which are involved in interneuron/motor neuron differentiation (Hou et al., 2013; Sander et al., 2000; Sharma et al., 1998), and bHLH-family Tfs (Myf5, MyoD, MyoG), which have been demonstrated to inhibit neuronal differentiation (Delfini & Duprez, 2004).

In oligodendrocytes, we observed an enrichment of CTCF/BORIS motifs at hyper-accessible peaks. CTCF has been implicated as a marker of oligodendrocyte differentiation (Lourenco et al., 2016). This orthogonally matches GO enrichment at

terms relating to differentiation in oligodendrocytes. Defects in oligodendrocytes have previously been in Down Syndrome (Olmos-Serrano et al., 2016), both *in vivo* and when differentiated *in vitro*. Concordantly, we see bHLH-family TF motifs (MyoD, Tcf21, Ap3, ZBTB18, NeuroG2, Olig2) at hyper-accessible peaks. Interestingly, we note no change in either class of oligodendrocytes (OGc7, OGc18) or oligodendrocyte progenitors (OGc25).

Discussion:

Single cell sequencing has become an increasingly powerful method for measuring changes in cell populations, and their chromatin states, which are associated with disease. In this study, we performed a comprehensive pairwise comparison of chromatin accessibility in the Ts65Dn mouse model for Down Syndrome, and their wild-type littermates using sci-ATAC-seq. From 13,766 randomly sampled cells that met quality control criteria, we identified 26 discernible clusters representing distinct cell-types in the brain cortex, and quantitative changes in cell composition of the cortex. Additionally, we identified sites of altered chromatin accessibility caused by the Ts65Dn triplication. Because sci-ATAC-seq requires no cell purification, our identification of cell-types was unbiased, and provides a high resolution of cellular and chromatin state changes in Down Syndrome.

Consistent with previous studies, we observed a broad increase in the number of interneurons in Ts65Dn mice. Specifically, we observed that dopamine-receptor

expressing interneurons were affected disproportionately relative to other clusters. Interestingly, our data falls in closer agreement to the studies of Hernandez-Gonzalez et al (Hernandez-Gonzalez et al., 2015), which used similarly aged 4-month-old adult mice, than to those of Chakrabarti et al (Chakrabarti et al., 2010), which observed defects in Somatostatin+ interneurons in 1-4-week-old mice. Our data identified no changes in Somatostatin+ and Parvalbumin+ interneurons in adult Ts65Dn mice.

The increase in interneurons was accompanied by a significant decrease in abundance of four of the 12 classes of excitatory neurons we detected, most probably originating from the upper layers of the cortex. These findings agree with other reports that in Ts65Dn mice, there is an imbalance of inter, and excitatory neurons, with the former being elevated, and the latter being diminished (Chakrabarti et al., 2010). Such cellular changes are likely to contribute to cognitive deficits associated with Down Syndrome in Ts65Dn mice, as maternal choline supplementation during gestation and lactation - a treatment that partially ameliorate the deficits - also partially normalize cell population changes ((Velazquez et al., 2013).

In addition to these changes in cell abundances, we identified chromatin domains with altered accessibility in Ts65Dn mice relative to their wild-type controls. Transcription factor binding motifs present at those domains were consistent with previous assertions that increased gene dosage of transcription factors within the

triplicated region play a role in the mis-regulation in Down syndrome (Chakrabarti et al., 2010; Gardiner, 2004). Specifically, we observed that sites of altered chromatin accessibility in Ts65Dn microglia were enriched for binding motifs for RUNX1, GABPA, and ERG; and in oligodendrocytes, there was an enrichment of OLIG2 motifs at differentially accessible peaks.

Besides these motifs for transcription factors encoded by the triplicated regions, we observed that CTCF sites were hyper-accessible in oligodendrocyte chromatin from Ts65Dn mice, implying a subtle differentiation defect. Although we did not find changes in OG abundance between Ts65Dn mice and their WT littermates, the altered accessibility at these sites is likely to impart functional changes in cellular behavior. Our lack of statistically significant changes to any of the three oligodendrocyte clusters we observed contrasts with the results reported by Olmos-Serrano et al. However, those studies characterized 1-week old mice, whereas we used 10-week old adult mice in our study. We infer that for these cell populations, delayed differentiation is occurring in Ts65Dn mice, and this may be one of the functional changes caused by altered chromatin accessibility.

Our inferences regarding changes in low abundance cell-types such as endothelial cells, astrocytes, microglia, and minor clusters comprising excitatory and interneurons are limited by the number of cells we analyzed (13,766). More definitive conclusions will require profiling of larger cell numbers. Furthermore, by

characterizing mice at a single time point, we cannot define when during development regulatory changes, or changes in cell abundance occur. That will require analysis of mice at different ages. We anticipate that analyses of other brain domains, including hippocampus for example, will reveal additional cellular and molecular changes associated with Down Syndrome. A final challenge is discerning which of the changes detectable by sci-ATAC-seq are important to Down Syndrome traits. This can be facilitated by characterizing brains of Ts65Dn mice whose functional deficits are partially suppressed by perturbations such as maternal choline supplementation (Ash et al., 2014; Powers et al., 2017; Powers et al., 2016; Velazquez et al., 2013), transcription factor dosage compensation (Chakrabarti et al., 2010), and drug treatments (Latchney, Jaramillo, Rivera, Eisch, & Powell, 2015).

Methods:

Mice:

B6EiC3Sn.BLiA-Ts(1716)65Dn/DnJ (Jackson #001924) mice were purchased from Jackson and crossed to B6EiC3Sn.BLiAF1/J (Jackson #003647) for maintenance. Mice used for sci-ATAC-seq were crossed to B6.Cg-Tg(Thy1-YFP)16Jrs/J (Jackson #003709) to express YFP in excitatory neurons for future analysis (unpublished data). Genotyping PCR was performed on 50 ng genomic DNA using oligos from Table S3 from for 40 cycles using GoTaq (Promega #M3001) (94°C 2min, 40 cycles of: [94°C 30sec, 60°C 30sec, 72°C 30sec], 72°C 5min) and run on a 2 % agarose gel (Fig. S9).

Tissue disaggregation:

Tissue was disaggregated using the Omni-ATAC-seq protocol (Corces et al., 2017) with minor modifications. All steps were performed on ice or in a 4°C chilled centrifuge. Briefly, mice were sacrificed and cortices were rapidly dissected and transferred into a chilled 7 mL Dounce Homogenizer containing 2.5 mL HB (320 mM sucrose, 0.1 mM EDTA, 0.1 % NP40, 5 mM CaCl2, 3 mM Mg(Ac)2, 10 mM Tris pH 7.4, protease inhibitors (Pierce #88666), 0.016 mM PMSF). Tissue was homogenized using a Dounce Tissue Grinder set (Wheaton #357542) with 10 strokes using a loose pestle and filtered through a 100 µm nylon mesh (VWR #10199-658) followed by 20 strokes using a tight pestle and centrifuged for 1 minutes at 100g. 2 mL of supernatant was mixed with 2 mL 50 % iodixanol solution (50 % iodixanol in 1x homogenization buffer). 2 mL of a 29 % iodixanol solution (29 % iodixanol in 1x HB containing 480 mM sucrose) was layered under the 25 % iodixanol/tissue mixture. 2 mL of a 35 % iodixanol solution (35 % iodixanol in 1X homogenization containing 480 mM sucrose) was layered underneath the 29 % iodixanol solution. Nuclei were centrifuged for 20 minutes at 3,000g in a swing-bucket rotor. Nuclei at the 29 % and 35 % interface were transferred to a new tube and diluted 1:5 in ATAC-RSB (10 mM Tris-HCl pH 7.4, 10 mM NaCl, and 3 mM MgCl2 in water) and centrifuged for 5 minutes at 500g. Nuclei were resuspended in 1X Tagmentation Buffer (10 mM Tris pH 7.4, 5 mM MgCl2, 10 % DMF, 33 % 1X PBS (without Ca++ and Mg++), 0.1 % Tween-20, 0.01 % Digitonin (ThermoFisher #BN2006) and counted using Trypan Blue on a hemocytometer.

Tn5 Transposase:

Tn5 was produced exactly as described (Picelli et al., 2014) with no modifications. For Omni-ATAC-seq, Tn5 transposase was assembled using pre-Annealed ME-A and ME-B (Table S3). For sci-ATAC-seq, Tn5 transposomes were assembled using pre-Annealed ME-C and ME-D oligonucleotides (Table S3). Oligonucleotides were annealed in H2O by combining ME-A, ME-B, ME-C, or ME-D oligos at 25 µM to Tn5MErev and incubating for 2 minutes at 94°C followed by a 0.1°C/s ramp to 25°C. ME-A and ME-B hybridized oligos were combined in equal quantities at a final concentration of 2.5µM and incubated with Tn5 transposase at a final concentration of 1.625 µM. ME-C and ME-D hybridized oligos were each incubated separately at a final concentration of 2.5 µM with Tn5 transposase at a final concentration of 1.625 µM to make separated strip-tubes each containing a single barcode. The buffer in which hybridized Tn5 transposomes were stored consists of a mixture of 40 % glycerol and 43.5 % Tn5 dialysis buffer (100 mM HEPES-KOH pH 7.2, 200 mM NaCl, 20 mM EDTA, 2 mM DTT, 20 % Glycerol, 0.2 % Triton X-100 in DEPC H2O). Enzyme was stored at -80°C.

sci-ATAC-seq:

Sci-ATAC-seq was performed using a protocol based on snATAC-seq (Preissl et al., 2018). Nuclei were resuspended at 600,000 cells/mL in 1X Tagmentation Buffer. 8 µL of nuclei were aliquoted to each well of a 96-well plate (4,800 cells/well). 1 µL of

each ME-C or ME-D carrying barcoded transposome at ~1.5 µM was added to each well and gently vortexed. Tagmentation was performed at 37°C for 1 hour, briefly vortexing once at 30 minutes. 10 µL of 40 µM EDTA was added each well, briefly vortexed, and incubated at 37°C for 15 minutes to inactivate the Tn5. 20 µL sort buffer (2 % BSA and 2 mM EDTA in PBS (without Ca++ and Mg++)) was added to each well. Nuclei from each well were pooled, filtered through a 35 µm mesh (Corning #352235) and Draq7 (Abcam #ab109202) was added to a final concentration of 3 µM (35-38 µL). 25 single nuclei were sorted into each well of a 96-well plate using a BD FACSAria Fusion and transferred to a -80°C freezer.

Frozen plates containing sorted nuclei were thawed on ice. 2 µL of 0.2 % SDS was added to each well and incubated for 7 minutes at 55°C. 2.5 µL of 10 % Triton X-100 was added to each well. 2 µL of 25 µM Primer i5 and 2 µL of 25µM Primer i7 was added to each well. PCR was performed for using Q5 DNA polymerase (NEB #M0491S) with 1X GC buffer (72°C 5min, 98°C 30sec, 15 cycles of: [98°C 10sec, 63°C 30sec, 72°C 30sec], 72°C 5min). In order to minimize batch effect, all PCRs were performed sequentially on a single machine. All wells were then pooled and diluted 5:1 (~24 mL) in Buffer PB (Qiagen #19066) with 1/20th volume of 3 M NaOAc pH 5.2 (~1.2 mL). This solution was run through MinElute columns (Qiagen #28004) using a QIAvac apparatus (Qiagen #19413) and washed once using 750 µL Buffer PE (Qiagen #19065). Samples were eluted twice using EB (10 mM Tris pH 8) warmed to 55°C. Samples were size selected using a 0.5X Ampure XP cleanup to remove large

fragments followed by 3 consecutive 1.5X Ampure XP (Beckman Coulter # A63880) bead cleanup followed by a final 1.2X Ampure XP cleanup to completely remove all residual primers and resuspended in a final volume of 20 µL EB using the manufacturer's recommended protocol.

Omni-ATAC-seq:

Tagmentation was performed using the Omni-ATAC-seq protocol (Corces et al., 2017) with modifications to inactivation and size selection. Briefly, 100,000 nuclei were tagmented in parallel to sci-ATAC-seq in 1X Tagmentation Buffer (10 mM Tris pH 7.4, 5 mM MgCl2, 10 % DMF, 33 % PBS (without Ca++ and Mg++), 0.1 % Tween-20, 0.01 % Digitonin) using 100 nM of ME-A and ME-B bound Tn5 Transposase for 30 minutes at 37°C. Tagmentation was inactivated with the addition of 5 volumes SDS Lysis Buffer (100mM Tris pH 7.4, 50 mM NaCl, 10 mM EDTA, 0.5 % SDS in DEPC H2O) and 100 µg Proteinase K (Invitrogen #25530049) for 30 minutes at 55°C followed by Isopropanol Precipitation using GlycoBlue (Invitrogen #AM9516) as a carrier. Samples were resuspended in EB and size selected using a 0.5X Ampure XP cleanup to remove large fragments followed by a 1.8X Ampure XP cleanup using the manufacturer's recommended protocol. PCR was performed for using Q5 DNA polymerase (NEB #M0491S) with 1X GC buffer (72°C 5min, 98°C 30sec, 11 cycles of: [98°C 10sec, 63°C 30sec, 72°C 30sec], 72°C 5min) followed by a final 1.8X Ampure XP cleanup.

Cell-culture sci-ATAC-seq:

Briefly, human A549 cells and mouse Hepa1-6 cells were quickly thawed at 37°C from tubes containing 1X freezing medium and resuspended in 10 mL room temperature PBS (without Ca++ and Mg++). Cells were spun at 500g for 5 minutes at room temperature, resuspended in PBS, and counted using Trypan Blue on a hemocytometer. 450,000 A549 cells were mixed with 450,000 Hepa1-6 cells and lysed via the Omni-ATAC-seq protocol in 50 µL ice-cold ATAC-RSB-Lysis buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1 % NP-40, 0.1 % Tween20, 0.01 % Digitonin in DEPC H2O). Nuclei were washed in 1 mL (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1 % Tween20) and spun at 500g for 10 minutes. Nuclei were then processed identically to our sci-ATAC-seq protocol with no modifications. Reads were counted at each barcode in either the mm10 or hg19 genome. Any cell containing both human and mouse reads was identified as a collision. Collision rate was calculated as 2*(collisions/total cells) (Cao et al., 2017).

Library quality control, quantification, and sequencing:

Library fragment distribution was measured using a bioanalyzer for nucleosome patterning. Following this, libraries were subjected to digital PCR (Schweitzer et al., 2014) on a Bio-Rad QX200 droplet digital PCR system using oligos from Table S3. Libraries were loaded at 8pM on a Nextseq500 mid lane PE 150bp. Sequencing was performed using a custom recipe for the following, read 1: [36 imaged cycles], Index 1: [8 imaged cycles, 27 dark cycles, 8 imaged cycles], Index 2: [8 imaged cycles, 21

dark cycles, 8 imaged cycles], Read 2: [36 imaged cycles].

Preprocessing and alignment:

Libraries were preprocessed and aligned as in Preissl et al (Preissl et al., 2018) with minor modifications. Briefly, read names were labeled with combined 32bp index reads. Pair-end reads were aligned to mm10 (or mm10 and hg19 for collision tests) using Bowtie2 in pair-end mode with parameters, -p 5 -t -X2000–no-mixed–no-discordant and reads < MAPQ 30 and flag=1804 were removed and deduplicated. Barcode arrangement and orientation is altered in our Nextseq500 run compared to the HiSeq run used in Preissl et al. Each 8bp Barcode was corrected within a 2bp edit distance to their nearest barcode and separated into individual cells based on their barcode combination after which PCR duplicates and mitochondrial were removed. Cells were filtered by a minimum of 20 % fraction of reads in peaks (FRiP) score and an 820 read per cell cutoff. Peaks were called using MACS2 (Zhang et al., 2008) with parameters, -g mm -p 0.05 --nomodel --shift 150 --keep-dup all (Table S4). Reads were assigned to RefSeq (O'Leary et al., 2016) gene bodies flanked by 2kb on each side or to MACS2-called peaks using Rsubread featurecounts (Liao, Smyth, & Shi, 2013). Insert size metrics were plotted using ATACseqQC (Ou et al., 2018) using pseudobulks from aggregated samples. Copy number analysis was performed using HMMcopy as in Knouse et al (Knouse, Wu, & Hendricks, 2017) with no modification (Figure S3, Table S5).

Clustering:

Features were counted at gene body tiles extended by 2kb using Rsubread Featurecounts, covering ~65 % of reads per sample. Dimensional reduction was performed on the first 50 principal components using Monocle 3.0 Alpha (Trapnell et al., 2014) on the first 50 principal components on gene body reads using UMAP. Clusters were called using Louvain clustering (Table S2b). 3D plots were generated using plot3D (Soetaert, 2017). Cells were counted by dividing the number of cells from each replicate of each cluster by the total number of cells sequenced per replicate (% total) (Table S2c). Differential cell counts were visualized and compared using ggpubr (Kassambara, 2018). Clustering of the top 20,000 TF-IDF selected peaks in Figure S6 was performed as in Cusanovich et al (Cusanovich et al., 2018) with no modifications.

Cell-type identification:

Psuedobulk bam files from called clusters were merged using Samtools. Features were called at gene body tiles extended by 2kb using Featurecounts. Reads were filtered for a CPM > 1 in at least 4 samples. Cell-types were clustered based on their relative average accessibility at marker genes. Cell-type-specific markers (p.adj < 0.05 log2FC > 1) were called using Limma topTreat (Ritchie et al., 2015)by comparing to genes enriched in each cluster or cell-type compared to all other clusters (Table S6) or cell-types (Table S7). Plots for cell-type identification were generated using log2CPM values normalized per gene across all clusters on a scale of 0 to 1; average

scaled values were used for visualizations. Relative accessibility was visualized using ggplot2. Clustering was performed using hclust and visualized using ggdendro (de Vries, 2016). Cortical layer-specific (Gray et al., 2017) and interneuron (Mo et al., 2015) ATAC-seq data were downloaded from NCBI GEO database accessions GSE87548 and GSE63137, respectively.

Differential accessibility, Gene ontology and disease ontology:

Pseudobulk bam files from called clusters were combined using samtools. Features were called at all peaks using Featurecounts. Differential accessibility ($p.adj < 0.05$ and $|log2FC| > 0.585$) was called using Limma topTable (Table S7). Peaks were assigned to their nearest RefSeq gene; RefSeq files were downloaded from UCSC table browser (Karolchik et al., 2004). Gene ontology and disease ontology analysis was performed on differentially accessible gene bodies and peaks (Table S9) using ClusterProfiler (Yu, Wang, Han, & He, 2012) and DOSE (Yu, Wang, Yan, & He, 2015) respectively using clusterCompare with fdr correction. Dotplots were visualized using ClusterProfiler. Heatmaps were visualized using pheatmap.

Motif Analysis:

Motif analysis was performed on differentially accessible regions using HOMER (Heinz et al., 2010) against a background set of all called peaks (Table S10). HOMER data was aggregated using marge (Amezquita, 2018) for downstream analysis.

Comparison to Omni-ATAC-seq:

Omni-ATAC-seq samples single-end reads were aligned to mm10 using Bowtie2 in single-end mode with parameters, -p 5 and reads < MAPQ 30 and flag=1804 were removed and deduplicated. Read were assigned to peaks called in our sci-ATAC-seq libraries using Featurecounts. Libraries were downsampled to approximately 3.3M assigned reads using metaseqR downsample.counts (Moulos & Hatzis, 2015). Scatterplots and Pearson correlations were visualized LSD (Schwalb et al., 2018).

Genome browser tracks:

Bigwig files were generated using Deeptools 3.0.2 (Ramirez et al., 2016) bamCoverage --binSize 1 --normalizeUsing RPKM --ignoreForNormalization chrM using cell-clusters merged using from all samples and displayed using GViz (Hahne & Ivanek, 2016).

Data Access:

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE127257.

Acknowledgements:

Jennifer D. Mosher, Ann E. Tate, Jeff C. Mattison, and Peter A. Schweitzer from the Cornell University Biotechnology Resource Center (BRC) for genomic sequencing.

Cornell CARE staff. Andrew Adey for sharing their CPT-seq dark cycle chemistry. Danea F. Rebolini at Illumina for modification Adey lab dark cycle chemistry. Andrew Grimson for the gift of A549 and Hepa1-6 cells.

Figures:

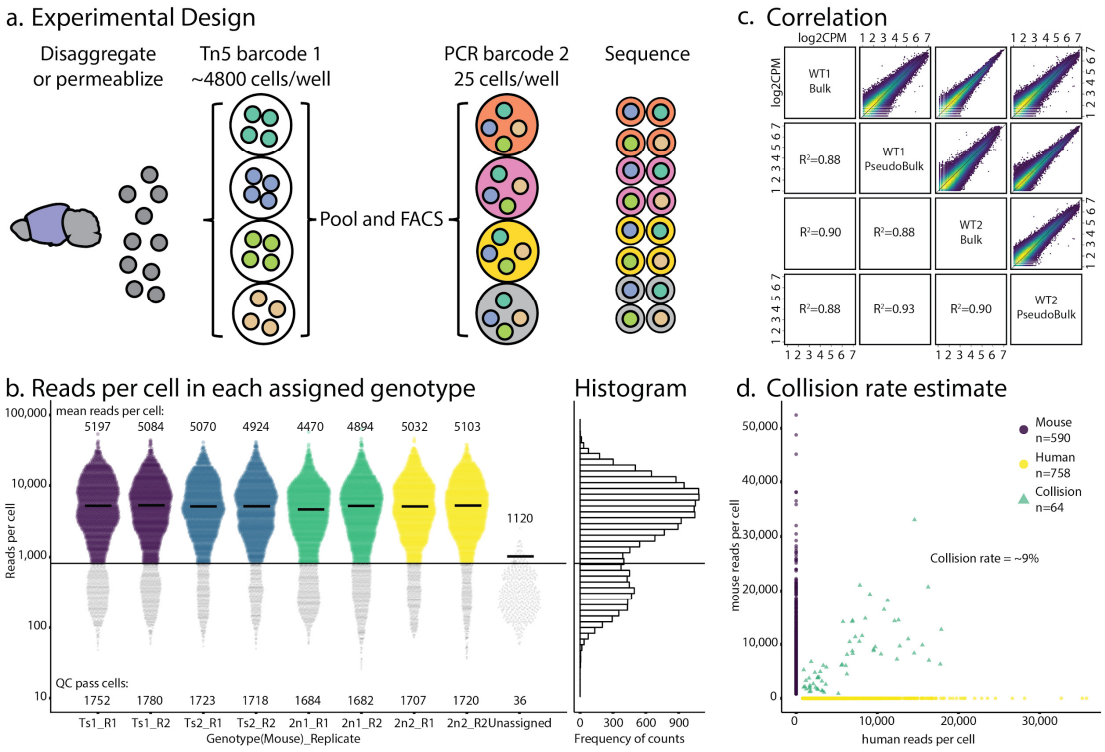


Figure 1:
Experimental design and QC
(a) The cortex from each brain is disaggregated and nuclei are distributed to 96 wells to be barcoded with Tn5 transposase. Wells are pooled and FACS sorted at 25 nuclei per well. Wells are barcoded via PCR. Sequenced cells are demultiplexed using combinatorial barcodes. **(b)** Reads per QC-passed nuclei per genotype and replicate. Lines denote medians, counts denote means. Line at 820 reads denotes minimum read cutoff. **(c)** log2CPM correlation of each WT brain vs bulk ATAC-seq showing Pearson correlation. **(d)** Collision rate estimates from mixed human and mouse samples.

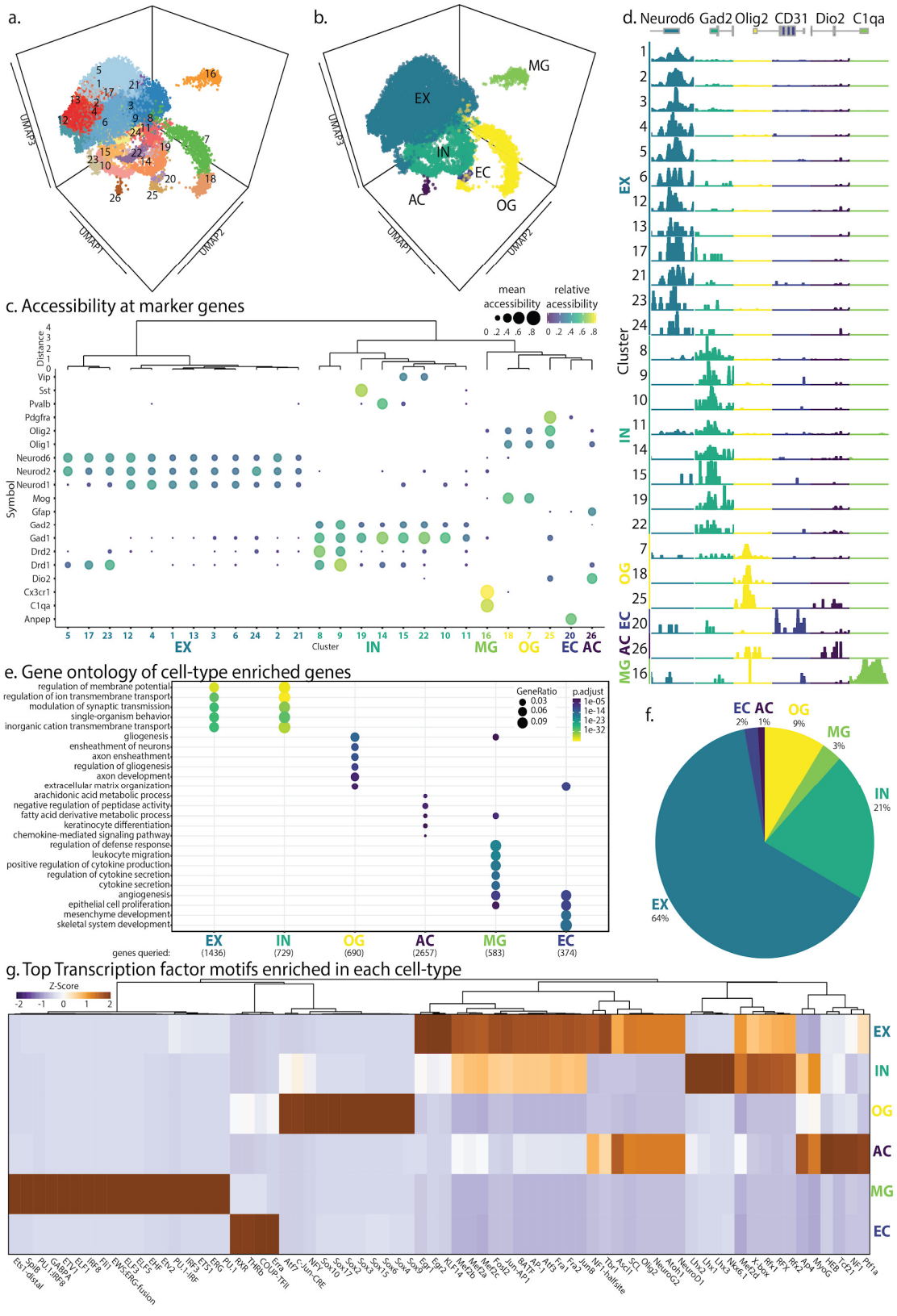

a.
b.
UMAP1
UMAP2
UMAP3
EX
IN
MG
EC
AC
OG
c. Accessibility at marker genes
mean accessibility
relative accessibility
Distance
Symbol
Vip
Sst
Pvalb
Pdgfra
Olig2
Olig1
Neurod6
Neurod2
Neurod1
Mog
Gfap
Gad2
Gad1
Drd2
Drd1
Dio2
Cx3cr1
C1qa
Anpep
Cluster
d. Neurod6 Gad2 Olig2 CD31 Dio2 C1qa
Cluster
e. Gene ontology of cell-type enriched genes
regulation of membrane potential
regulation of ion transmembrane transport
modulation of synaptic transmission
single-organism behavior
inorganic cation transmembrane transport
gliogenesis
ensheathment of neurons
axon ensheathment
regulation of gliogenesis
axon development
extracellular matrix organization
arachidonic acid metabolic process
negative regulation of peptidase activity
fatty acid derivative metabolic process
keratinocyte differentiation
chemokine-mediated signaling pathway
regulation of defense response
leukocyte migration
positive regulation of cytokine production
regulation of cytokine secretion
cytokine secretion
angiogenesis
epithelial cell proliferation
mesenchyme development
skeletal system development
GeneRatio
p.adjust
genes queried:
(1436)
(729)
(690)
(2657)
(583)
(374)
f.
EC AC
2%
1%
OG
9%
MG
3%
IN
21%
EX
64%
g. Top Transcription factor motifs enriched in each cell-type
Z-Score

Figure 2:
Cell-type identification
(a) UMAP of cells showing 26 clusters. **(b)** Clusters from (a) labeled by cell-type **(c)** Clusters are identified using cell-type specific markers. **(d)** Browser shots from clusters in (a) at cell-type specific markers: EX (Neurod6), IN (Gad2), OG (Olig1), EC (CD31), AC (Dio2), and MG (C1qa). **(e)** Genes ontology of genes enriched in each cell-type in (b). **(f)** Cell-type distribution of WT cortex as percent of total. **(g)** Top transcription factor motifs enriched at peaks in each cell-type ($p<1e-20$). Values denote Z-scores of -log10 p-values of transcription factor motifs.

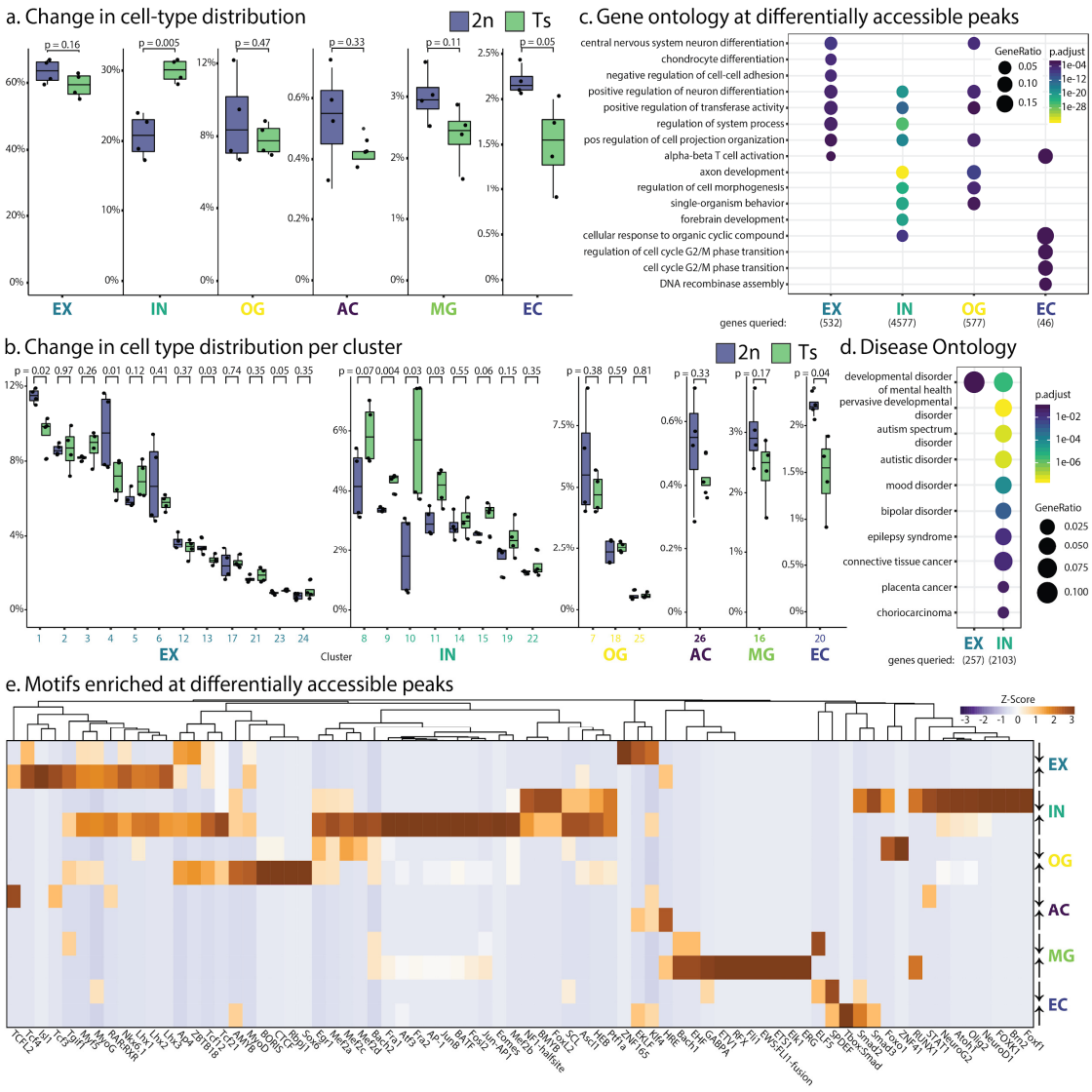
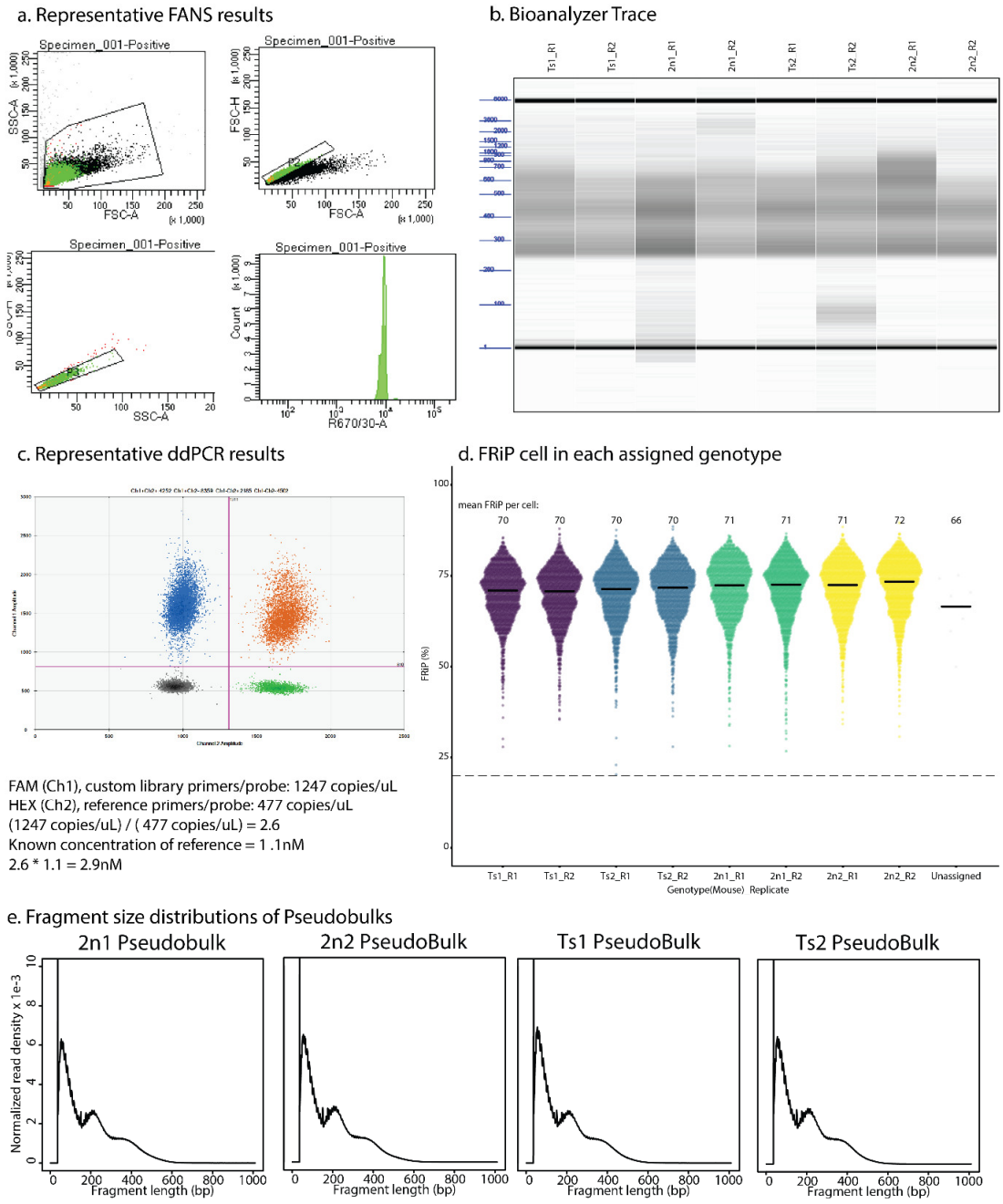


Figure 3:

Changes in Ts mice:

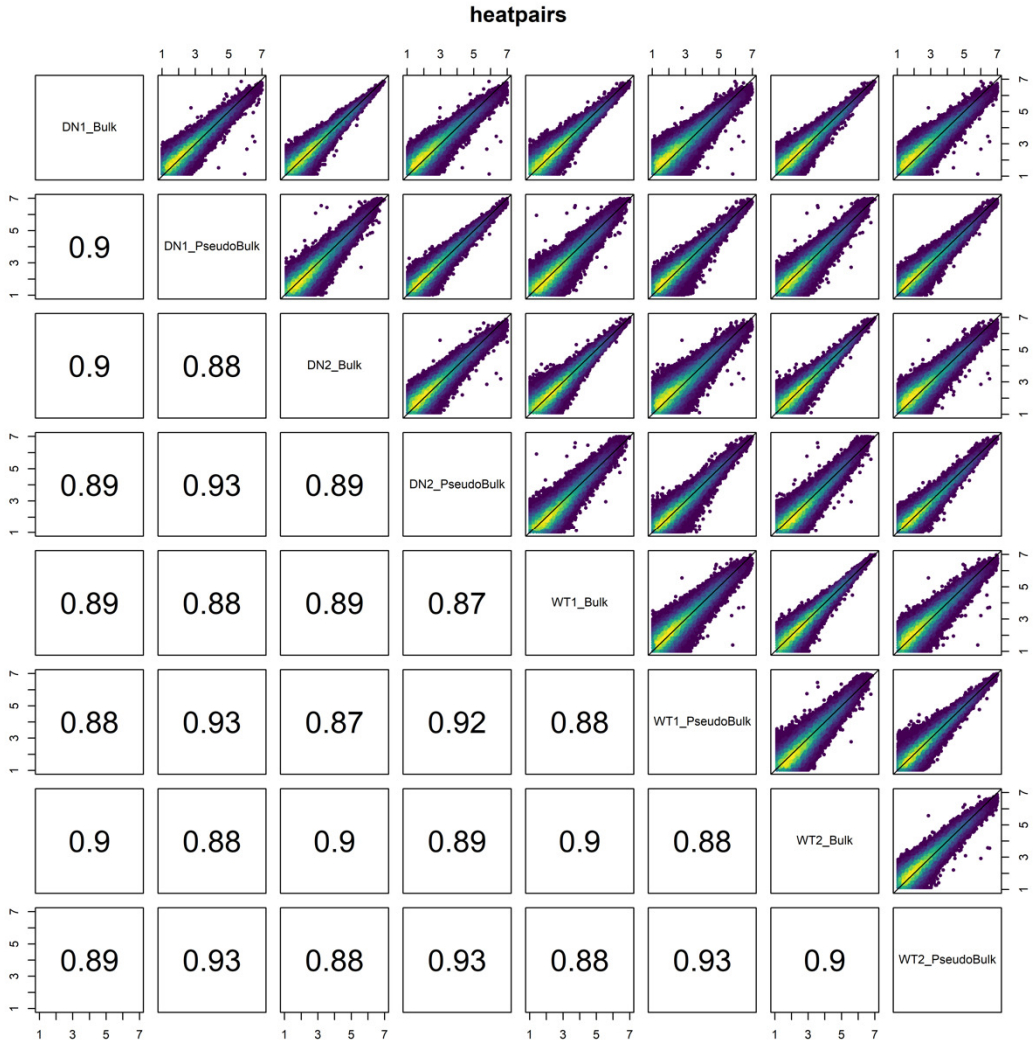
(a) Changes in cell-type (top) and **(b)** cell cluster (bottom) abundances in WT (blue) and Ts (green) mice. P values indicate Student's t-test. **(c)** Gene Ontology at genes nearest to differentially accessible peaks (p<0.05, FC>1.5) in Ts mice. **(d)** Significantly enriched Disease Ontology in Ts mice at genes from b. **(e)** Top 20 transcription factor motifs in Hyper-accessible (up arrow) or Hypo-accessible (down arrow) peaks in Ts mice in each cell-type. Values denote Z-scores of -log10 p-values of transcription factor motifs.

Supplemental figures:

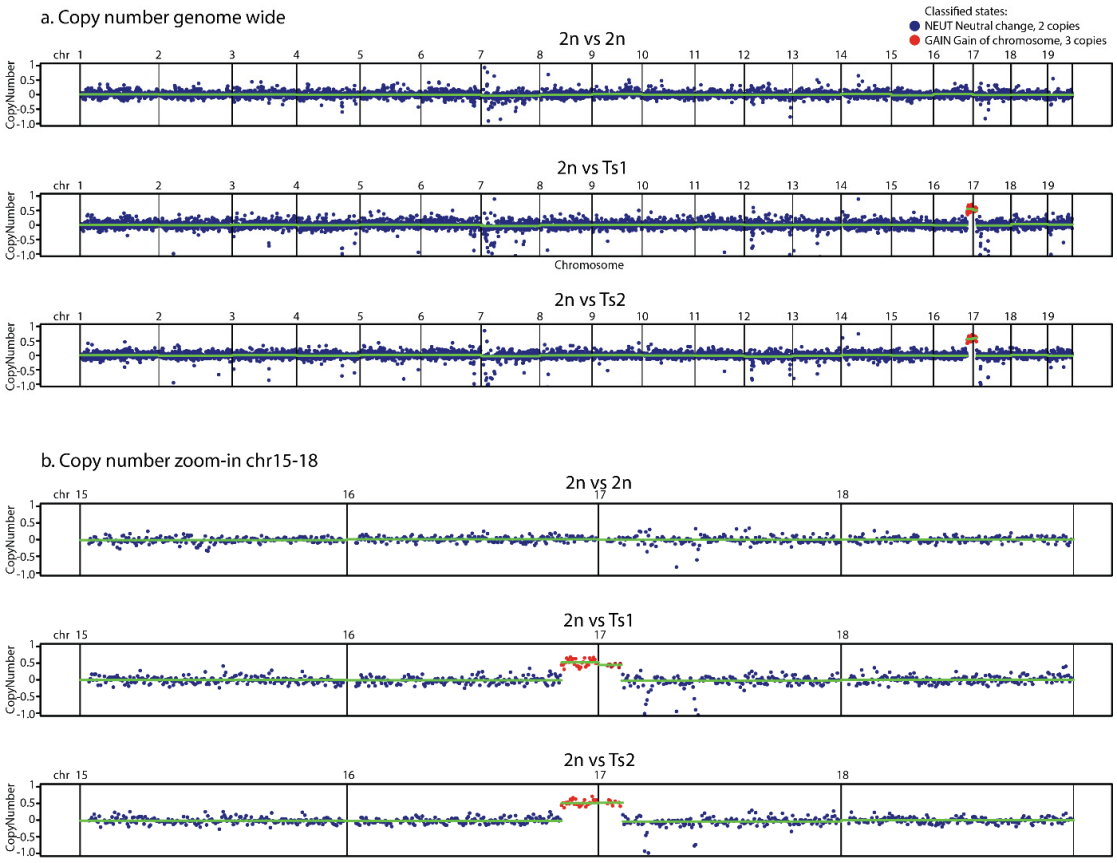


Supplemental Figure 1:

Library QC. (a) Representative FANS results showing FSC, SSC, and DRAQ7 gating (R670/30-A). **(b)** Bioanalyzer library sizes; figure has been globally contrast-adjusted for clarity. **(c)** Representative ddPCR results from a single library and control. **(d)** Distribution of FRiP per cells with reads > 820. **(e)** Fragment distribution of pseudobulks from QC-passed cells in each mouse.

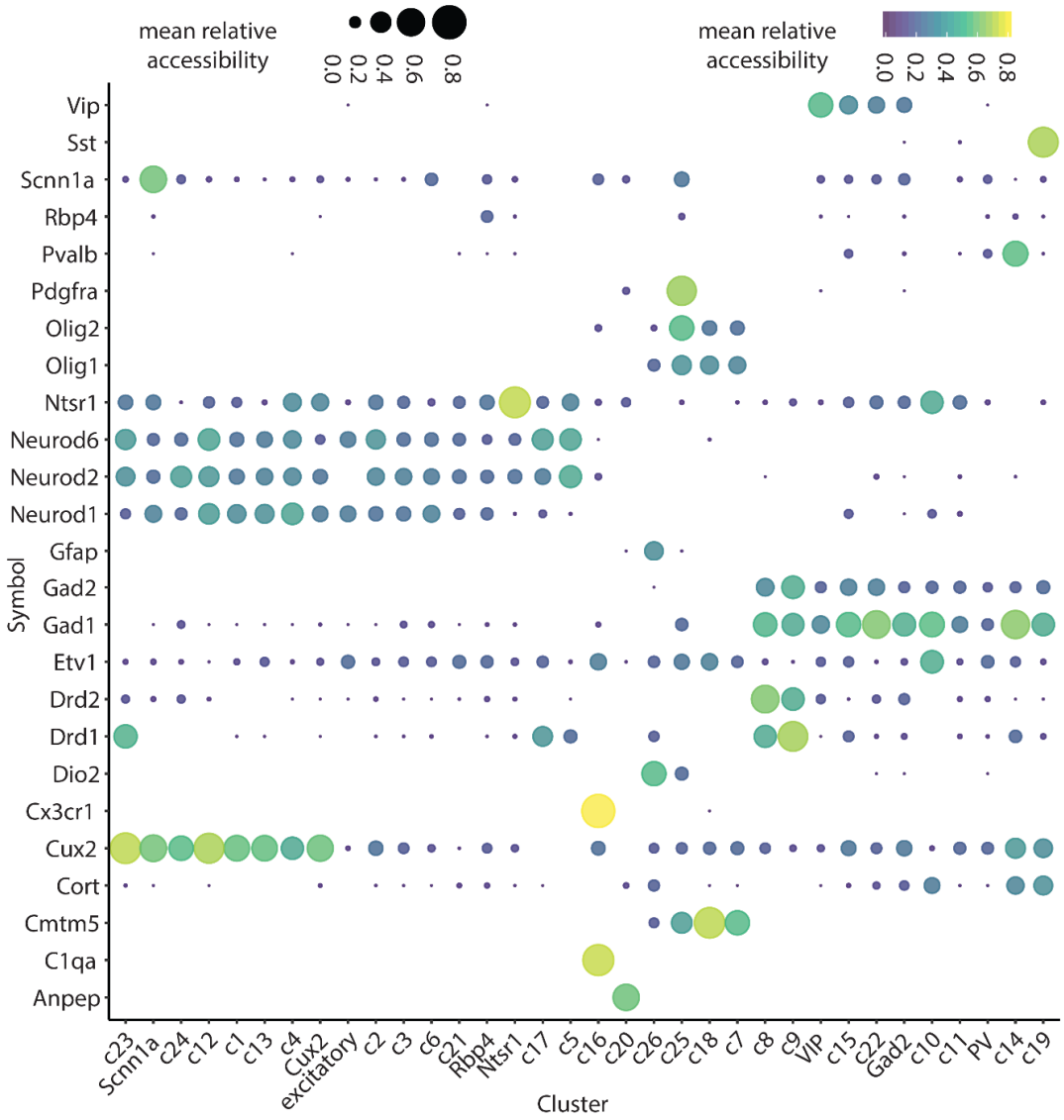


Supplemental Figure 2:
Sample correlations. Pearson correlation of all mice in bulk libraries (bulk) and pseudobulk libraries. Reads shown in log2CPM.



Supplemental Figure 3:

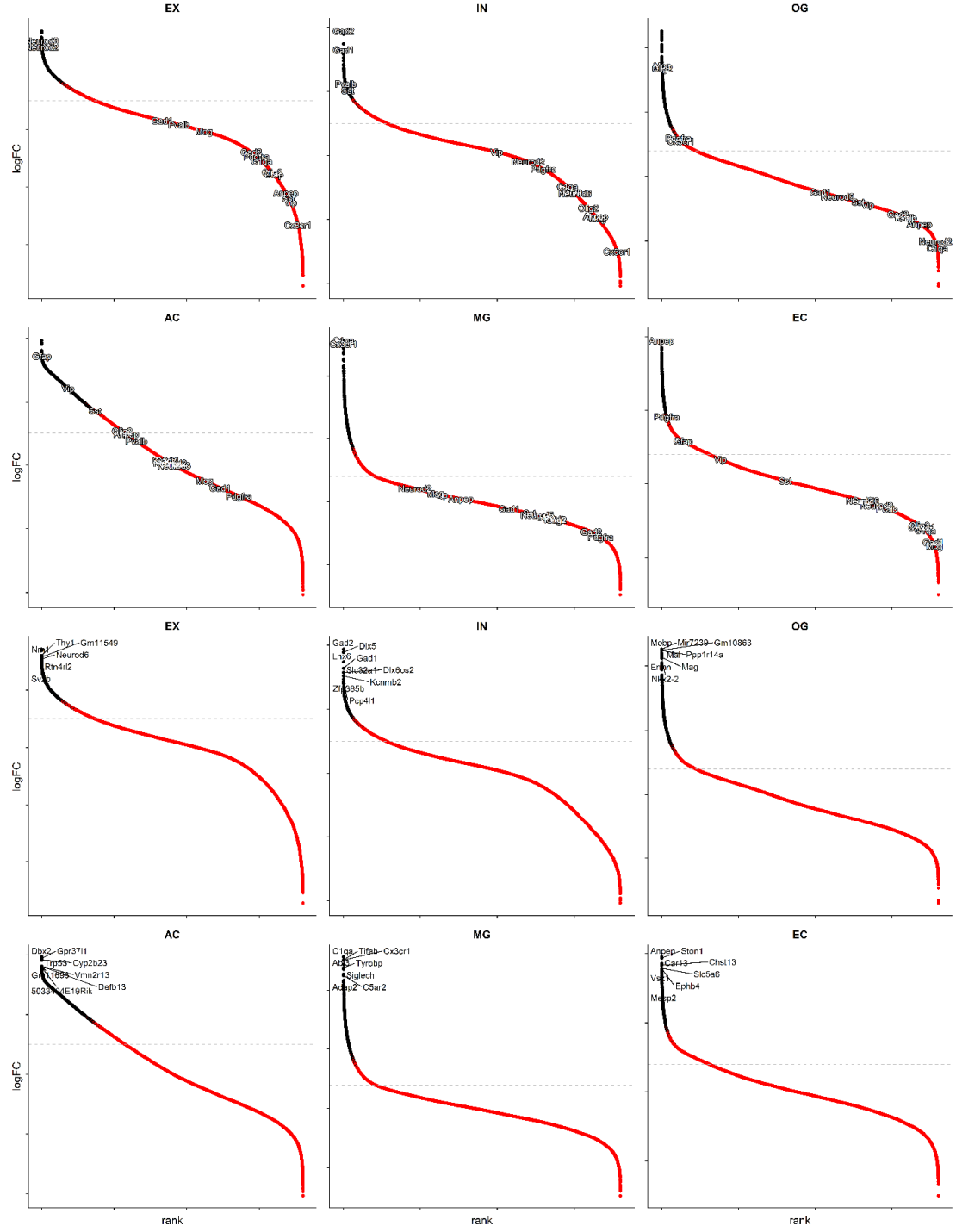
Copy number confirmation. (a) Copy number across genome compared to 2n animals at 500kb tiles using mm10. (b) Copy number zoomed in to chr15-18. Blue dots denote 2n state. Red dots denote 3n state at chr16:81,000,000-98,200,000 and chr17:3,500,001-9,000,000.



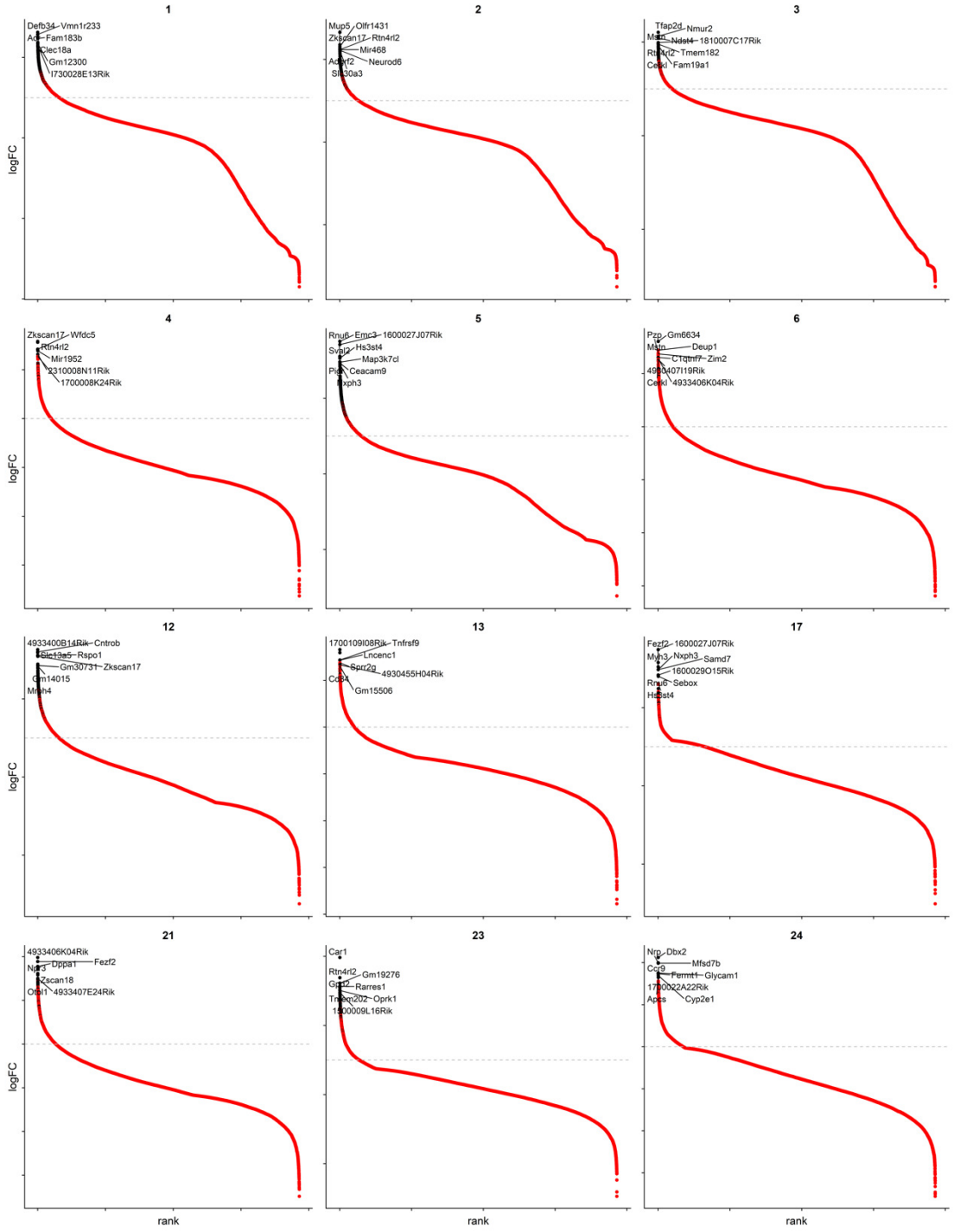
Supplemental Figure 4:

Comparison to FACS-sorted data. Cell-type specific markers at identified clusters and FACS sorted ATAC-seq data.

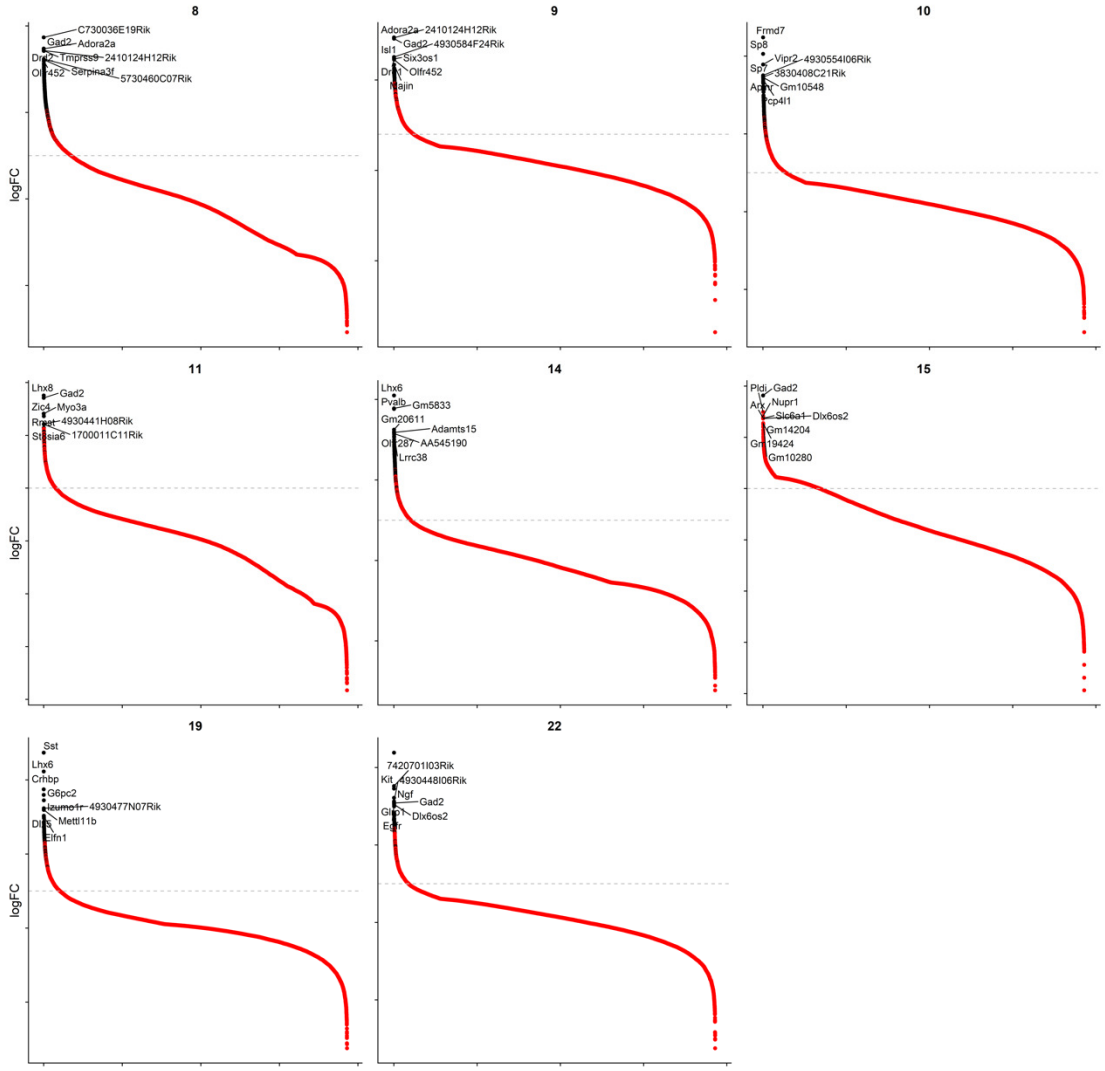
S5a (top) S5b (bottom), cell-types:



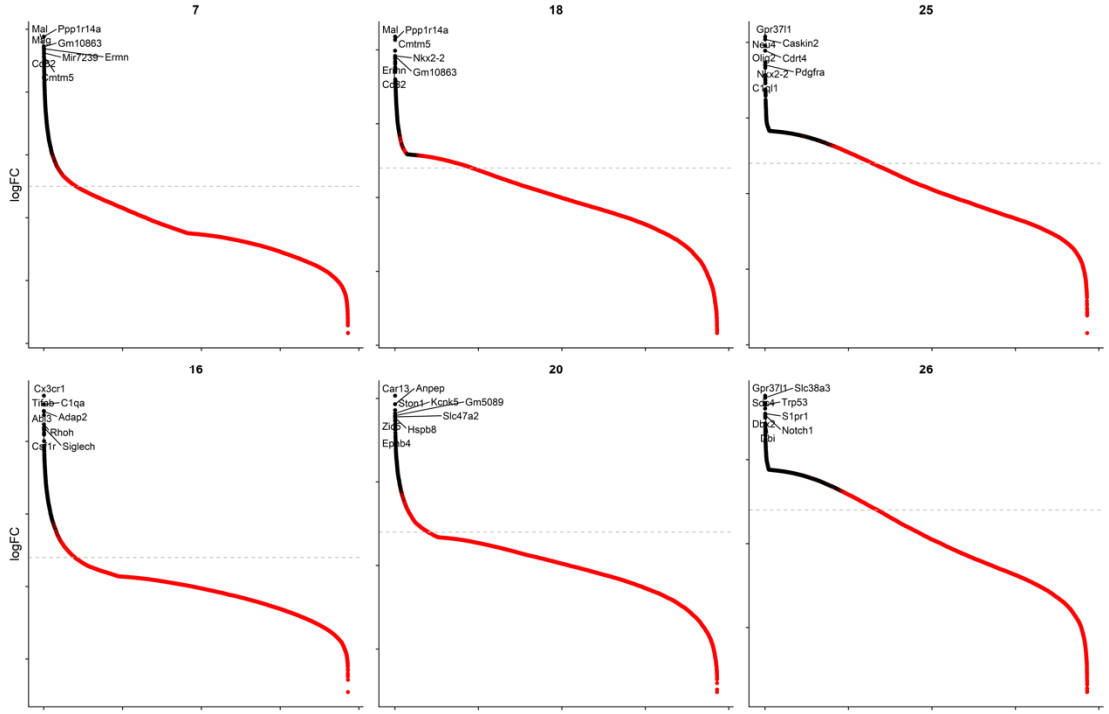
S5c EX clusters:



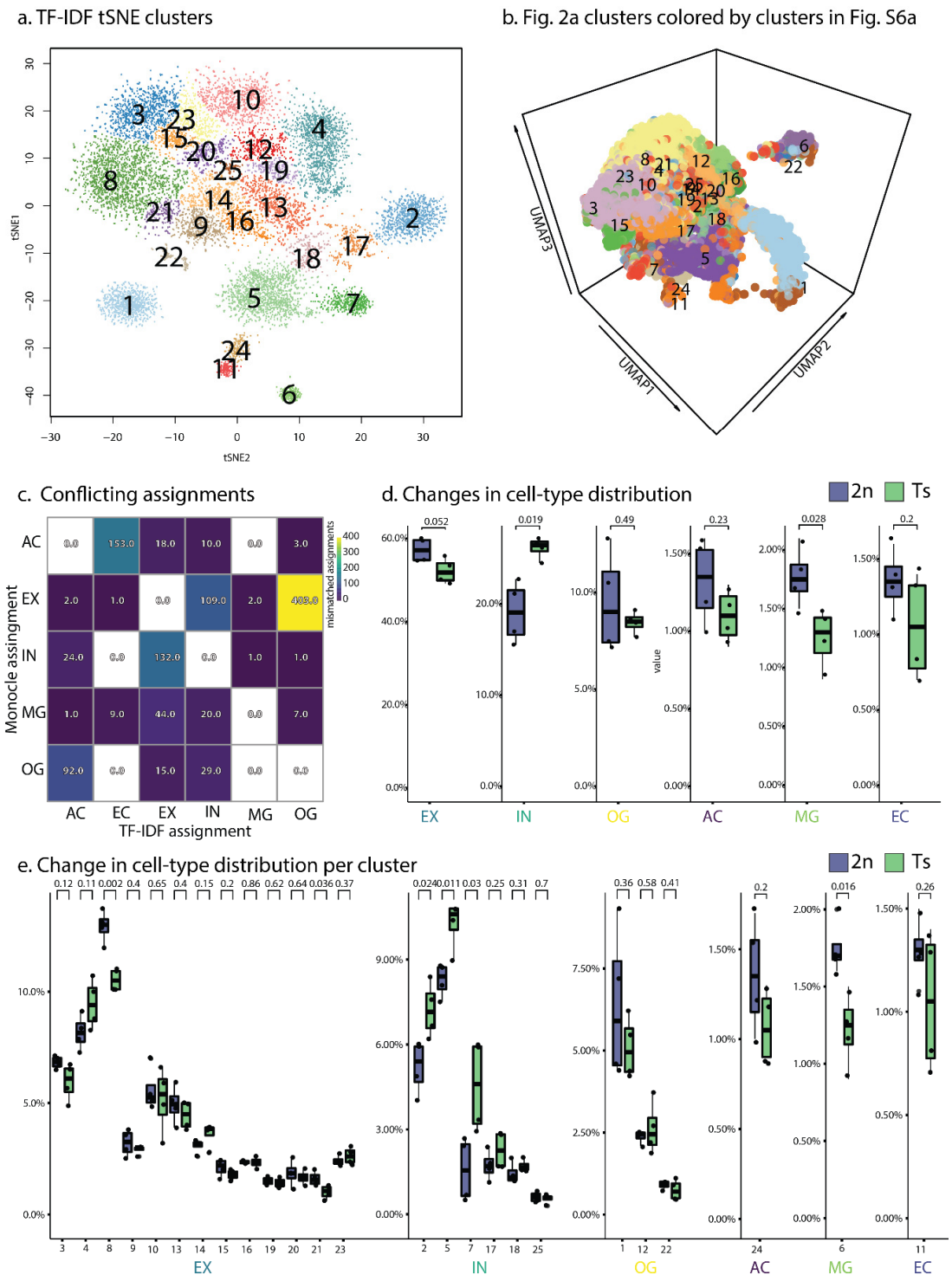
S5d: IN clusters


8
C730036E19Rik
Gad2
Adora2a
Tmprss9
2410124H12Rik
Serpina3f
5730460C07Rik
logFC
9
Adora2a
2410124H12Rik
Gad2
4930584F24Rik
Isl1
Six3os1
Olfr452
Najin
10
Frmd7
Sp8
Vipr2
4930554I06Rik
3830408C21Rik
Gm10548
Pcp4l1
11
Lhx8
Gad2
Zic4
Myo3a
4930441H08Rik
1700011C11Rik
14
Lhx6
Pvalb
Gm5833
Gm20611
Adamts15
AA545190
Lrrc38
15
Pldi
Gad2
Nupr1
Slc6a1
Dlx6os2
Gm14204
Gm19424
Gm10280
19
Sst
Lhx6
Cnbp
G6pc2
Ube2ql1
4930477N07Rik
Mettl11b
Elfn1
22
7420701I03Rik
Kit
4930448I06Rik
Ngf
Gad2
Dlx6os2
Egfr

S5e: glia clusters



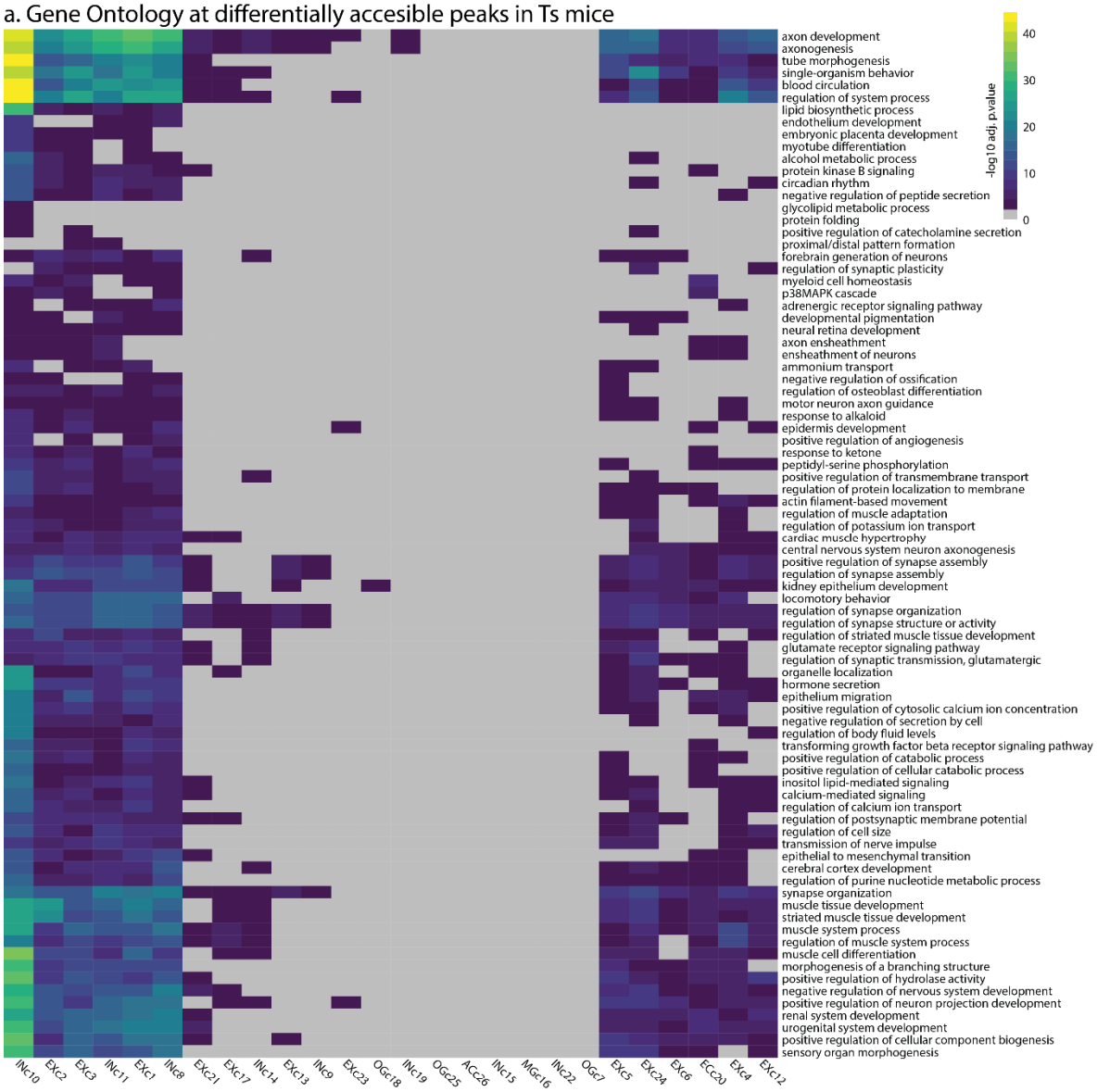
Supplemental Figure 5:
Cell-type and cluster-enriched genes. log2 Fold-change vs Rank at gene bodies in each cell-type showing **(a)** marker genes and **(b)** top 10 genes. Top10 hyper-accessible genes at each cluster sorted by EX (c), IN (d), and glia (e). Black dots denote p.adj < 0.05 and log2 fold-change > 1



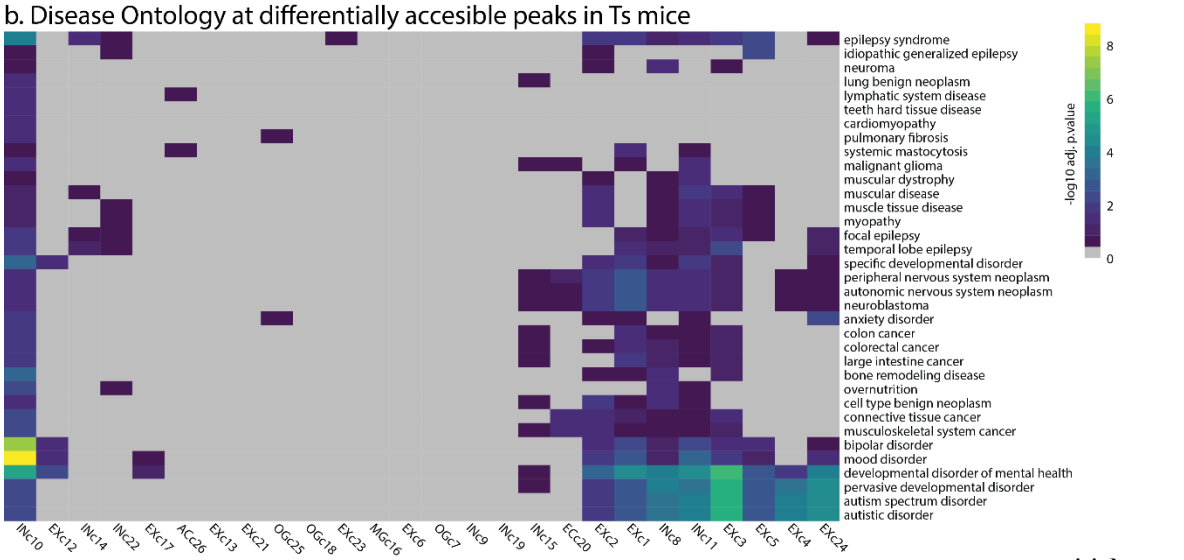
Supplemental Figure 6:

Comparison to TF-IDF. (a) t-SNE from 20,000 TF-IDF selected peaks. **(b)** Clusters in (a) overlaid on Fig. 2a UMAP. **(c)** Mismatched cell-type assignments. **(d)** Change in tell-type distribution. **(e)** Change in cell-type distribution per cluster. Blue bars denote 2n mice, green bars denote Ts mice.

a. Gene Ontology at differentially accesible peaks in Ts mice

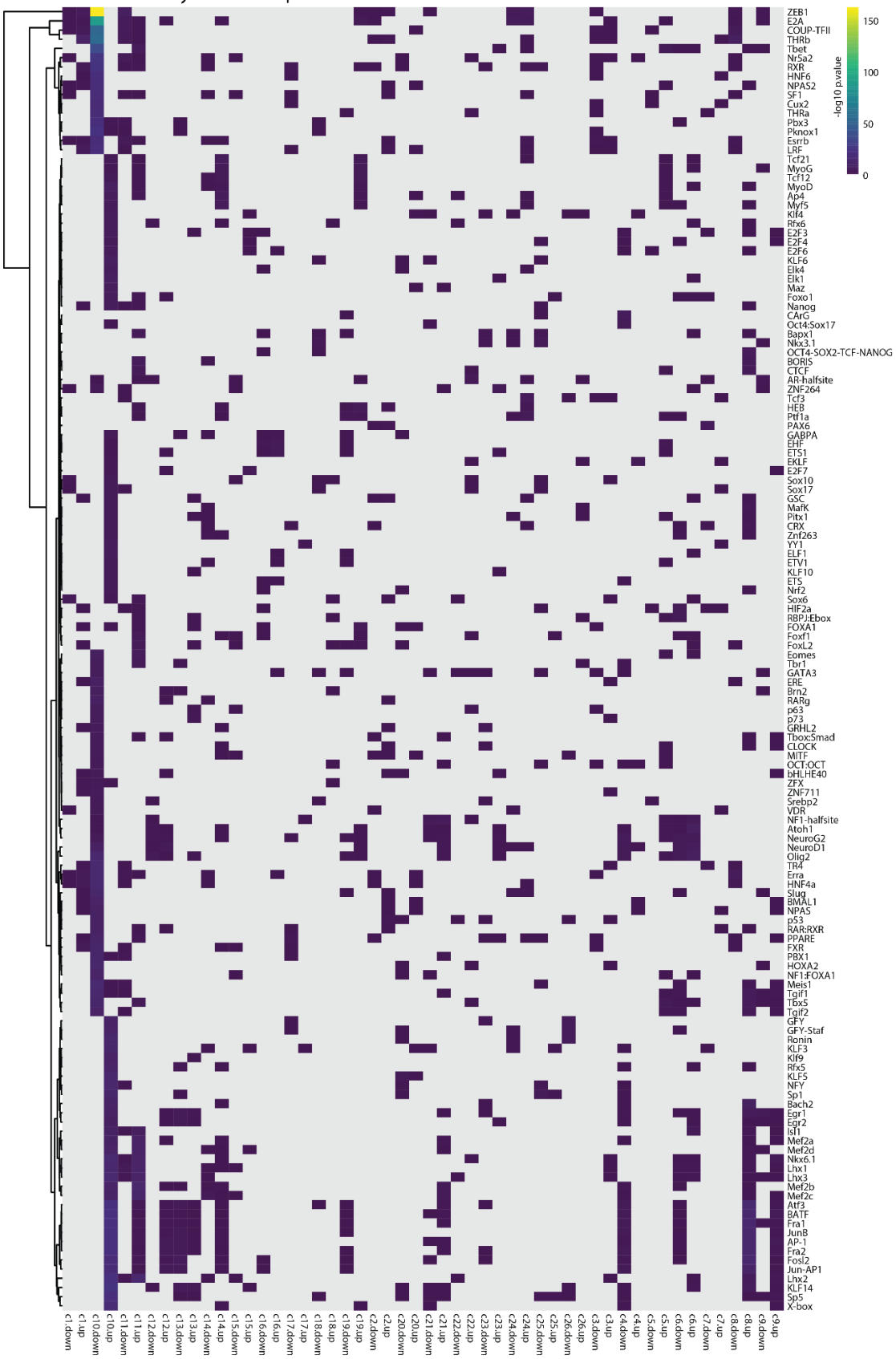


b. Disease Ontology at differentially accesible peaks in Ts mice

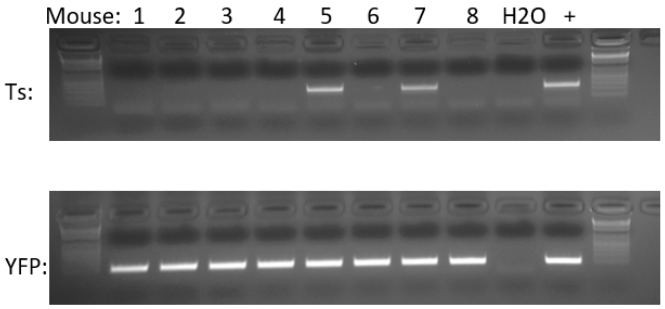


Supplemental Figure 7:
Gene ontology and disease ontology changes per clusters. (a) Gene ontology at nearest changed genes in Ts mice per cluster. Values denote fdr adjusted p. values filtered to a p value < 0.0001. **(b)** Disease ontology at nearest changed genes in Ts mice per cluster. Values denote fdr adjusted p. values filtered to a p value < 0.05.

Motifs at differentially accessible peaks



Supplemental Figure 8:
Motifs at differentially accessible peaks in clusters. Motifs changed at differentially accessible peaks in all clusters with a minimum p. value of 0.01 in at least one sample.



Supplemental Figure 9:
Genotype. PCR against the chr16-chr17 breakpoint in Ts65dn (top) and YFP (bottom) in Ts65Dn males used in sci-ATAC-seq. Mouse 5-8 were used in this study.

Supplemental tables:

Sequencing Lanes:	Read:	% Dups	% GC	M Reads	Samples:
Lane1	Read1	75.6%	50%	259.5	Ts1_R1
	Read2	82.2%	51%	259.5	2n1_R2
	Index1	99.3%	51%	259.5	Ts2_R1
	Index2	99.5%	46%	259.5	
Lane2	Read1	78.0%	51%	230.3	2n1_R2
	Read2	80.2%	52%	230.3	Ts2_R2
	Index1	99.3%	51%	230.3	2n2_R1
	Index2	99.6%	45%	230.3	
Lane3	Read1	79.9%	51%	219.8	2n1_R1
	Read2	82.9%	51%	219.8	2n2_R2
	Index1	99.4%	51%	219.8	Ts1_R2
	Index2	99.6%	45%	219.8	
Lane4	Read1	63.6%	49%	235.0	Collision
	Read2	66.5%	50%	235.0	Tests
	Index1	99.3%	52%	235.0	
	Index2	99.5%	44%	235.0	

sci-ATAC-seq samples:

Mouse	Fraction of reads in peaks (FRiP)	Deduplicated Reads (M)
WT_1	0.72	21.8
WT_2	0.72	24.1
Ts_1	0.70	25.4
Ts_2	0.71	24.0

Library Concentrations:

Name	Concentration (nM)	i7 PCR set (Table S3)	i5 PCR set (Table S3)
2n1_R1	1.3	I7_1	I5_3
2n1_R2	1.3	I7_1	I5_4
2n2_R1	1.3	I7_2	I5_3
2n2_R2	1.7	I7_2	I5_4
Ts1_R1	1.4	I7_1	I5_1
Ts1_R2	1.4	I7_1	I5_2
Ts2_R1	1.1	I7_2	I5_1
Ts2_R2	1.4	I7_2	I5_2
Collision		I7_2	I5_1

Supplemental Table 1:
Samples sequenced in this study. Values show millions (M) sequences analyzed, and Fraction of Reads in Peaks (FRiP) score in down-sampled libraries, and oligonucleotide sets used.

Cluster	Cell-type	Name	Total (nuclei/cluster)	Specifics
1	EX	EXc1	1440	Layer II/III/IV
2	EX	EXc2	1187	Layer II/III/IV
3	EX	EXc3	1164	Layer V/VI
4	EX	EXc4	1146	Layer II/III/IV
5	EX	EXc5	890	Layer V
6	EX	EXc6	872	Layer V/VI
7	OG	OGc7	739	Oligodendrocytes
8	IN	INc8	697	Drd2+ MSN
9	IN	INc9	526	Drd1+ MSN
10	IN	INc10	512	Interneurons (unknown)
11	IN	INc11	487	Interneurons (unknown)
12	EX	EXc12	480	Layer II/III/IV
13	EX	EXc13	414	Layer II/III/IV
14	IN	INc14	405	Parvalbumin +
15	IN	INc15	392	Vip +
16	MG	MGc16	369	Microglia
17	EX	EXc17	341	Layer V
18	OG	OGc18	339	Oligodendrocytes
19	IN	INc19	283	Somatostatin +
20	EC	ECc20	256	Endothelial cells
21	EX	EXc21	241	Layer V/VI
22	IN	INc22	191	Vip +
23	EX	EXc23	132	Layer II/III/IV
24	EX	EXc24	119	Layer V/VI
25	OG	OGc25	77	Oligodendrocyte progenitors
26	AC	ACc26	67	Astrocytes

Supplemental Table 2:
Characterized clusters and cell-types, their total number of nuclei queried per cluster, and their specific cell-type localization.*

*This table has been altered from submission for size constraints; it previously contained percentages per sample per cluster and top10 representative differentially accessible gene bodies per cluster.

ddPCR (Ordered IDT with HPLC purification):		**NOTES**
F	AGCAGAAGACGGCATACGAGAT	
R	ATACGGCGACCACCGAGATC	
Int	/56-FAM/TCTTATACACATCTGAGGCGG/3BHQ_1/	At Index 1

Genotyping (Ordered from IDT with standard desalting):		**NOTES:**
Ts65Dn_F	GTGGCAAGAGACTCAAATTCAAC	From JAX 275bp
Ts65Dn_R	TGGCTTATTATTATCAGGGCATTT	
YFP_F	CCACCTACGGCAAGCTGACC	
YFP_R	GGTAGCGGGCGAAGCACT	114bp Product

Sequencing Oligos (Ordered from IDT with PAGE purification):		**NOTES**
Read 1 Sequencing Primer	GCGATCGAGGACGGCAGATGTGTATAAGAGACAG	From: Vitak et al. doi: 10.1038/nmeth.4154
Read 2 Sequencing Primer	CACCGTCTCCGCCTCAGATGTGTATAAGAGACAG	
Index 1 Sequencing Primer	CTGTCTCTTATACACATCTGAGGCGGAGACGGTG	
Index 2 Sequencing Primer	CTGTCTCTTATACACATCTGCCGTCCTCGATCGC	

Transposome Reverse (Ordered from IDT with standard desalting):		
Tn5ME_Rev	/5Phos/CTGTCTCTTATACACATCT	
Transposome: Tn5 i5 Oligos:		**BARCODE**
C15_ME_1	TCGTCGGCAGCGTCTCCACGCTATAGCCT GCGATCGAGGACGGCAGATGTGTATAAGAGACAG	AGGCTATA
C15_ME_2	TCGTCGGCAGCGTCTCCACGCATAGAGGC GCGATCGAGGACGGCAGATGTGTATAAGAGACAG	GCCTCTAT
C15_ME_3	TCGTCGGCAGCGTCTCCACGCCCTATCCT GCGATCGAGGACGGCAGATGTGTATAAGAGACAG	AGGATAGG
C15_ME_4	TCGTCGGCAGCGTCTCCACGCGGCTCTGA GCGATCGAGGACGGCAGATGTGTATAAGAGACAG	TCAGAGCC
C15_ME_5	TCGTCGGCAGCGTCTCCACGCAGGCGAAG GCGATCGAGGACGGCAGATGTGTATAAGAGACAG	CTTCGCCT
C15_ME_6	TCGTCGGCAGCGTCTCCACGCTAATCTTA GCGATCGAGGACGGCAGATGTGTATAAGAGACAG	TAAGATTA
C15_ME_7	TCGTCGGCAGCGTCTCCACGCCAGGACGT GCGATCGAGGACGGCAGATGTGTATAAGAGACAG	ACGTCCTG
C15_ME_8	TCGTCGGCAGCGTCTCCACGCGTACTGAC GCGATCGAGGACGGCAGATGTGTATAAGAGACAG	GTCAGTAC

Tn5 i7 Oligos:		**BARCODE**
D15_ME_1	GTCTCGTGGGCTCGGCTGTCCCTGTCCCGAGTAAT CACCGTCTCCGCCTCAGATGTGTATAAGAGACAG	ATTACTCG
D15_ME_2	GTCTCGTGGGCTCGGCTGTCCCTGTCCTCTCCGGA CACCGTCTCCGCCTCAGATGTGTATAAGAGACAG	TCCGGAGA
D15_ME_3	GTCTCGTGGGCTCGGCTGTCCCTGTCCAATGAGCG CACCGTCTCCGCCTCAGATGTGTATAAGAGACAG	CGCTCATT
D15_ME_4	GTCTCGTGGGCTCGGCTGTCCCTGTCCGGAATCTC CACCGTCTCCGCCTCAGATGTGTATAAGAGACAG	GAGATTCC
D15_ME_5	GTCTCGTGGGCTCGGCTGTCCCTGTCCTTCTGAAT CACCGTCTCCGCCTCAGATGTGTATAAGAGACAG	ATTCAGAA
D15_ME_6	GTCTCGTGGGCTCGGCTGTCCCTGTCCACGAATTC CACCGTCTCCGCCTCAGATGTGTATAAGAGACAG	GAATTCGT
D15_ME_7	GTCTCGTGGGCTCGGCTGTCCCTGTCCAGCTTCAG CACCGTCTCCGCCTCAGATGTGTATAAGAGACAG	CTGAAGCT
D15_ME_8	GTCTCGTGGGCTCGGCTGTCCCTGTCCGCGCATTA CACCGTCTCCGCCTCAGATGTGTATAAGAGACAG	TAATGCGC
D15_ME_9	GTCTCGTGGGCTCGGCTGTCCCTGTCCCATAGCCG CACCGTCTCCGCCTCAGATGTGTATAAGAGACAG	CGGCTATG
D15_ME_10	GTCTCGTGGGCTCGGCTGTCCCTGTCCTTCGCGGA CACCGTCTCCGCCTCAGATGTGTATAAGAGACAG	TCCGCGAA
D15_ME_11	GTCTCGTGGGCTCGGCTGTCCCTGTCCGCGCGAGA CACCGTCTCCGCCTCAGATGTGTATAAGAGACAG	TCTCGCGC
D15_ME_12	GTCTCGTGGGCTCGGCTGTCCCTGTCCCTATCGCT CACCGTCTCCGCCTCAGATGTGTATAAGAGACAG	AGCGATAG

i7 PCR:

Name	**Total sequence**	**BARCODE**	**PCR SET**
N701	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGG	TAAGGCGA	I7_1
N702	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGG	CGTACTAG	I7_1
N703	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGG	AGGCAGAA	I7_1
N704	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGG	TCCTGAGC	I7_1
N705	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGG	GGACTCCT	I7_1
N706	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGG	TAGGCATG	I7_1
N707	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGG	CTCTCTAC	I7_1
N710	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGG	CGAGGCTG	I7_1
N711	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGG	AAGAGGCA	I7_1
N712	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGG	GTAGAGGA	I7_1
N714	CAAGCAGAAGACGGCATACGAGATTCATGAGCGTCTCGTGGGCTCGG	GCTCATGA	I7_1

N715	CAAGCAGAAGACGGCATACGAGATCCTGAGATGTCTCGTGGGCTCGG	ATCTCAGG	I7_1
N716	CAAGCAGAAGACGGCATACGAGATTAGCGAGTGTCTCGTGGGCTCGG	ACTCGCTA	I7_2
N718	CAAGCAGAAGACGGCATACGAGATGTAGCTCCGTCTCGTGGGCTCGG	GGAGCTAC	I7_2
N719	CAAGCAGAAGACGGCATACGAGATTACTACGCGTCTCGTGGGCTCGG	GCGTAGTA	I7_2
N721	CAAGCAGAAGACGGCATACGAGATGCAGCGTAGTCTCGTGGGCTCGG	TACGCTGC	I7_2
N722	CAAGCAGAAGACGGCATACGAGATCTGCGCATGTCTCGTGGGCTCGG	ATGCGCAG	I7_2
N723	CAAGCAGAAGACGGCATACGAGATGAGCGCTAGTCTCGTGGGCTCGG	TAGCGCTC	I7_2
N724	CAAGCAGAAGACGGCATACGAGATCGCTCAGTGTCTCGTGGGCTCGG	ACTGAGCG	I7_2
N726	CAAGCAGAAGACGGCATACGAGATGTCTTAGGGTCTCGTGGGCTCGG	CCTAAGAC	I7_2
N727	CAAGCAGAAGACGGCATACGAGATACTGATCGGTCTCGTGGGCTCGG	CGATCAGT	I7_2
N728	CAAGCAGAAGACGGCATACGAGATTAGCTGCAGTCTCGTGGGCTCGG	TGCAGCTA	I7_2
N729	CAAGCAGAAGACGGCATACGAGATGACGTCGAGTCTCGTGGGCTCGG	TCGACGTC	I7_2
X730	CAAGCAGAAGACGGCATACGAGATTACCAGAGGTCTCGTGGGCTCGG	CTCTGGTA	I7_2
X731	CAAGCAGAAGACGGCATACGAGATGGATGGAAGTCTCGTGGGCTCGG	TTCCATCC	I7_3
X732	CAAGCAGAAGACGGCATACGAGATATTGAGGCGTCTCGTGGGCTCGG	GCCTCAAT	I7_3
X733	CAAGCAGAAGACGGCATACGAGATCGGATAGAGTCTCGTGGGCTCGG	TCTATCCG	I7_3
X734	CAAGCAGAAGACGGCATACGAGATTGGTAGACGTCTCGTGGGCTCGG	GTCTACCA	I7_3
X735	CAAGCAGAAGACGGCATACGAGATACCTGGTTGTCTCGTGGGCTCGG	AACCAGGT	I7_3
X736	CAAGCAGAAGACGGCATACGAGATCAGTTCTGGTCTCGTGGGCTCGG	CAGAACTG	I7_3
X737	CAAGCAGAAGACGGCATACGAGATTCGAACGTGTCTCGTGGGCTCGG	ACGTTCGA	I7_3
X738	CAAGCAGAAGACGGCATACGAGATCGTTGCTTGTCTCGTGGGCTCGG	AAGCAACG	I7_3
X739	CAAGCAGAAGACGGCATACGAGATTACCGTTCGTCTCGTGGGCTCGG	GAACGGTA	I7_3
X740	CAAGCAGAAGACGGCATACGAGATTAGGTTGCGTCTCGTGGGCTCGG	GCAACCTA	I7_3
X741	CAAGCAGAAGACGGCATACGAGATGAGGCTAAGTCTCGTGGGCTCGG	TTAGCCTC	I7_3
X742	CAAGCAGAAGACGGCATACGAGATCGACCATAGTCTCGTGGGCTCGG	TATGGTCG	I7_3
X743	CAAGCAGAAGACGGCATACGAGATAGGCAGTAGTCTCGTGGGCTCGG	TACTGCCT	I7_4
X744	CAAGCAGAAGACGGCATACGAGATATCAAGCGGTCTCGTGGGCTCGG	CGCTTGAT	I7_4
X745	CAAGCAGAAGACGGCATACGAGATCATTGAAGGTCTCGTGGGCTCGG	CTTCAATG	I7_4
X746	CAAGCAGAAGACGGCATACGAGATCGACTTATGTCTCGTGGGCTCGG	ATAAGTCG	I7_4
X747	CAAGCAGAAGACGGCATACGAGATTCTATACGGTCTCGTGGGCTCGG	CGTATAGA	I7_4
X748	CAAGCAGAAGACGGCATACGAGATAGCATTAGGTCTCGTGGGCTCGG	CTAATGCT	I7_4
X749	CAAGCAGAAGACGGCATACGAGATAATTGGCAGTCTCGTGGGCTCGG	TGCCAATT	I7_4
X750	CAAGCAGAAGACGGCATACGAGATAGATTCGTGTCTCGTGGGCTCGG	ACGAATCT	I7_4
X751	CAAGCAGAAGACGGCATACGAGATTTCATGACGTCTCGTGGGCTCGG	GTCATGAA	I7_4
X752	CAAGCAGAAGACGGCATACGAGATTGAACTTGGTCTCGTGGGCTCGG	CAAGTTCA	I7_4
X753	CAAGCAGAAGACGGCATACGAGATATGGCATAGTCTCGTGGGCTCGG	TATGCCAT	I7_4
X754	CAAGCAGAAGACGGCATACGAGATCGTAATTCGTCTCGTGGGCTCGG	GAATTACG	I7_4

i5 PCR:

Name	Total sequence	BARCODE	PCR Set
S502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTCGGCAGCGTC	ATAGAGAG	I5_1
S503	AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGCAGCGTC	AGAGGATA	I5_1
S505	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGTCGTCGGCAGCGTC	CTCCTTAC	I5_1
S506	AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGCAGCGTC	TATGCAGT	I5_1
S507	AATGATACGGCGACCACCGAGATCTACACAAGGAGTATCGTCGGCAGCGTC	TACTCCTT	I5_1
S508	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTTCGTCGGCAGCGTC	AGGCTTAG	I5_1
S510	AATGATACGGCGACCACCGAGATCTACACCGTCTAATTCGTCGGCAGCGTC	ATTAGACG	I5_1
S511	AATGATACGGCGACCACCGAGATCTACACTCTCTCCGTCGTCGGCAGCGTC	CGGAGAGA	I5_1
X512	AATGATACGGCGACCACCGAGATCTACACTCGACTAGTCGTCGGCAGCGTC	CTAGTCGA	I5_2
X513	AATGATACGGCGACCACCGAGATCTACACTTCTAGCTTCGTCGGCAGCGTC	AGCTAGAA	I5_2
X514	AATGATACGGCGACCACCGAGATCTACACCCTAGAGTTCGTCGGCAGCGTC	ACTCTAGG	I5_2
X515	AATGATACGGCGACCACCGAGATCTACACGCGTAAGATCGTCGGCAGCGTC	TCTTACGC	I5_2
X516	AATGATACGGCGACCACCGAGATCTACACAAGGCTATTCGTCGGCAGCGTC	ATAGCCTT	I5_2
X517	AATGATACGGCGACCACCGAGATCTACACGAGCCTTATCGTCGGCAGCGTC	TAAGGCTC	I5_2
X518	AATGATACGGCGACCACCGAGATCTACACTTATGCGATCGTCGGCAGCGTC	TCGCATAA	I5_2
X519	AATGATACGGCGACCACCGAGATCTACACATCTGAGTTCGTCGGCAGCGTC	ACTCAGAT	I5_2
X520	AATGATACGGCGACCACCGAGATCTACACGGATACTATCGTCGGCAGCGTC	TAGTATCC	I5_3
X521	AATGATACGGCGACCACCGAGATCTACACTAAGATCCTCGTCGGCAGCGTC	GGATCTTA	I5_3
X522	AATGATACGGCGACCACCGAGATCTACACAAGAGATGTCGTCGGCAGCGTC	CATCTCTT	I5_3
X523	AATGATACGGCGACCACCGAGATCTACACAATGACGTTCGTCGGCAGCGTC	ACGTCATT	I5_3
X524	AATGATACGGCGACCACCGAGATCTACACGAAGTATGTCGTCGGCAGCGTC	CATACTTC	I5_3
X525	AATGATACGGCGACCACCGAGATCTACACATAGCCTTTCGTCGGCAGCGTC	AAGGCTAT	I5_3
X526	AATGATACGGCGACCACCGAGATCTACACTTGGAAGTTCGTCGGCAGCGTC	ACTTCCAA	I5_3
X527	AATGATACGGCGACCACCGAGATCTACACATTCGTTGTCGTCGGCAGCGTC	CAACGAAT	I5_3
X528	AATGATACGGCGACCACCGAGATCTACACAGGATAACTCGTCGGCAGCGTC	GTTATCCT	I5_4
X529	AATGATACGGCGACCACCGAGATCTACACTTCATCCATCGTCGGCAGCGTC	TGGATGAA	I5_4
X530	AATGATACGGCGACCACCGAGATCTACACAACGAACGTCGTCGGCAGCGTC	CGTTCGTT	I5_4
X531	AATGATACGGCGACCACCGAGATCTACACTGCCTTACTCGTCGGCAGCGTC	GTAAGGCA	I5_4
X532	AATGATACGGCGACCACCGAGATCTACACCGAATTCCTCGTCGGCAGCGTC	GGAATTCG	I5_4
X533	AATGATACGGCGACCACCGAGATCTACACGGTTAGACTCGTCGGCAGCGTC	GTCTAACC	I5_4
X534	AATGATACGGCGACCACCGAGATCTACACTCCGGTAATCGTCGGCAGCGTC	TTACCGGA	I5_4
X535	AATGATACGGCGACCACCGAGATCTACACTTACGACCTCGTCGGCAGCGTC	GGTCGTAA	I5_4

Supplemental Table 3:
All oligonucleotide sequences used in this study.

Supplemental Table 4-10:
Listed tables are not included in this thesis due to size constraints. It previously contained the following tables:
4. TableS4 snATAC234_peaks.narrowPeak.txt, all MACS2 called peaks

5. TableS5 HMMcopy states per mouse.xlsx, copy number per 500kb tile
6. TableS6 Enriched genes per cluster.xlsx, enriched genes per cluster vs all others
7. TableS7 Enriched genes per cell-type.xlsx, enriched genes per cell-type vs all others
8. TableS8 Tsvs2n Differential accessibility at peaks.xlsx, differentially accessible peaks in Ts mice
9. TableS9_GO and DO TsvsWT allclusters at peaks.xlsx, gene and disease ontology at genes nearest to peaks in Ts mice
10.TableS10_Homer Motifs.xlsx, representation of motif PWMs per cluster and cell-type

Chapter 4: Extended discussion

mATAC-Seq

In the second chapter of my thesis, I presented a modification to ATAC-seq, methyl-ATAC-seq, which queries accessible chromatin and its underlying DNA methylation. In this study, I performed mATAC-seq on a standard model studying depletion of DNA methylation, HCT116, and its respective DNMT knockouts. We observed an increase in accessibility that accompanies a depletion of DNA methylation. Hypo-methylated and hyper-accessible promoters produce an increased quantity mRNA. Transcription factors associated with these changes are methylation-sensitive, confirming methyl-SELEX (Yin et al., 2017) results *in vivo*. We can further cluster our peaks; segmenting them into different methylation-sensitive clusters, each showing a bias for methylation-sensitive transcription factor occupancy and differential expression. Thus far we have only published data studying DNA methylation at accessible chromatin in HCT116 lines in a highly simplified experiment and model. In the context of these cells, we have performed PRO-Seq (Mahat et al., 2016) to further characterize the transcriptional changes that occur at active enhancers and transposable elements; at transposable elements, we have observed a decrease in DNA methylation and an increase in accessibility preferentially at relatively young transposons (unpublished data).

Studies of HCT116 cells can be extended in several ways to assess chromatin state and the consequences of DNMT perturbation. Disrupting global DNA methylation in a controlled manner using 5-azacytidine, an inhibitor of DNA methylation to study

changes to accessibility and passive DNA methylation over time. From the opposing end of this query, using inducible overexpression of TET enzymes, we can query the consequences of active demethylation at accessible chromatin to provide us with further understanding of the time course of active, rather than passive, demethylation. Similarly, we can query changes and consequences of DNA methylation by using controlled expression of DNMT enzymes. To further query the interactions between remodeling machinery and 5mC we can disrupt histone remodeling machinery known to interact with 5mC such as LSD1 and SUV39H1.

Further topics of interest include the study of stem cells and embryogenesis which show a massive remodeling of their chromatin and DNA methylation landscape during differentiation (Messerschmidt, Knowles, & Solter, 2014) and the adaptation of sci-ATAC-seq to mATAC-seq. With mATAC-seq we can query these changes in high resolution. mATAC-seq is also highly amenable to 5hmC analysis using TET-enzyme based methods such as TAB-seq and APOBEC-based method such as ACE-seq to further modify ATAC-seq to measure 5hmC.

sci-ATAC-seq

In the third chapter of this thesis, I performed single cell ATAC-seq on the nuclei of Down syndrome model mice using a modified version of sci-ATAC-seq. Using optimizations from Omni-ATAC-seq we were able to recover highly informative single-cell data of exceptionally high quality, with a mean FRiP greater than 70%, compared to a reported mean FRiP of 22% reported in Preissl et al. using snATAC-

seq. Using this, we were able to show changes that occur in Down syndrome model mice with resolution that has not been previously available. The predominant result that we observed was an extreme enrichment of interneurons (~50%) in the cortex, with interneurons accessible at Drd1 and Drd2 showing the greatest change, and a corresponding depletion of excitatory neurons (~12%) localized to the upper layers of the cortex. At these cell-types, we frequently observed a change in motifs representing triplicated transcription factors, but also several cell-type specific changes.

Though this was a study performed at high resolution, we lack temporal resolution to track the changes that we observed as these experiments were performed in adult mice. In order to address each of these issues, we would have to perform sci-ATAC-seq on a time-course of animals to track where and how changes in population arise. Furthermore, we queried a single, large region of the brain, the cortex. Several regions of the brain have been shown to be dysfunctional in Down Syndrome models, and little is known about precisely what changes in each cell-type. The Soloway lab has expressed interest in therapeutic or preventative options in mitigating the effects of Down Syndrome. Such methods include gene therapies (Chakrabarti et al., 2010) and dietary supplementation (Ash et al., 2014). Both models have implied a change in cell-type composition and their veracity can be easily tested using sci-ATAC-seq. As we home in on precise changes that occur, it will become viable to test pharmaceutical interventions to more precisely assay treatments.

The weakness of this data is that although we query the regulatory changes in

accessible chromatin, we do not measure changes in gene expression. Combinatorial single-cell methods are currently being developed and show promise in correlating changes in accessibility and gene expression (Cao et al., 2018; Clark et al., 2018). In parallel, labs have begun to simultaneously analyze single cell accessibility and expression data, showing that changes in accessibility do not necessarily correlate to increased transcription (Jia et al., 2018). As costs continue to decrease and availability of data increases it will become important to integrate other single-cell datasets into future analysis.

The primary purpose of this study was to test the effectiveness of sci-ATAC-seq in identifying changes in a heterogenous population in disease states. In this sense it was a complete success. Currently, we have optimized our protocol to increase our total number of unique reads per cell (unpublished data) without compromising our FRiP score using increased temperatures during tagmentation. With this success, ease, and low-cost of this technique, we are now beginning to apply sci-ATAC-seq to other tissue-types, disease models, and organisms.

References: Chapter 1 & Chapter 4

Adey, A., Kitzman, J. O., Burton, J. N., Daza, R., Kumar, A., Christiansen, L., … Shendure, J. (2014). In vitro, long-range sequence information for de novo genome assembly via transposase contiguity. *Genome Research*, *24*(12), 2041–2049. https://doi.org/10.1101/gr.178319.114

Aguilar-Gurrieri, C., Larabi, A., Vinayachandran, V., Patel, N. A., Yen, K., Reja, R., … Panne, D. (2016). Structural evidence for Nap1-dependent H2A-H2B deposition and nucleosome assembly. *The EMBO Journal*, *35*(13), 1465–1482. https://doi.org/10.15252/embj.201694105

Arrowsmith, C. H., Bountra, C., Fish, P. V., Lee, K., & Schapira, M. (2012). Epigenetic protein families: a new frontier for drug discovery. *Nature Reviews Drug Discovery*, *11*(5), 384–400. https://doi.org/10.1038/nrd3674

Ash, J. A., Velazquez, R., Kelley, C. M., Powers, B. E., Ginsberg, S. D., Mufson, E. J., & Strupp, B. J. (2014). Maternal choline supplementation improves spatial mapping and increases basal forebrain cholinergic neuron number and size in aged Ts65Dn mice. *Neurobiology of Disease*, *70*, 32–42. https://doi.org/10.1016/j.nbd.2014.06.001

Booth, M. J., Branco, M. R., Ficz, G., Oxley, D., Krueger, F., Reik, W., & Balasubramanian, S. (2012). Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. *Science*, *336*(6083), 934–937. https://doi.org/10.1126/science.1220671

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, *10*(12), 1213–1218. https://doi.org/10.1038/nmeth.2688

Calo, E., & Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? *Molecular Cell*, *49*(5), 825–837. https://doi.org/10.1016/j.molcel.2013.01.038

Cancer Genome Atlas Research Network, Analysis Working Group: Asan University, BC Cancer Agency, Brigham and Women's Hospital, Broad Institute, Brown University, … Project Team: National Institutes of Health. (2017). Integrated genomic characterization of oesophageal carcinoma. *Nature*, *541*(7636), 169–175. https://doi.org/10.1038/nature20805

Cao, J., Cusanovich, D. A., Ramani, V., Aghamirzaie, D., Pliner, H. A., Hill, A. J., … Shendure, J. (2018). Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science (New York, N.Y.)*, *361*(6409), 1380–1385. https://doi.org/10.1126/science.aau0730

Cao, J., Packer, J. S., Ramani, V., Cusanovich, D. A., Huynh, C., Daza, R., … Shendure, J. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*, *357*(6352), 661–667. https://doi.org/10.1126/science.aam8940

Chakrabarti, L., Best, T. K., Cramer, N. P., Carney, R. S. E., Isaac, J. T. R., Galdzicki, Z., & Haydar, T. F. (2010). Olig1 and Olig2 triplication causes developmental brain

defects in Down syndrome. *Nature Neuroscience*, *13*(8), 927–934. https://doi.org/10.1038/nn.2600

Chen, X., Miragaia, R. J., Natarajan, K. N., & Teichmann, S. A. (2018). A rapid and robust method for single cell chromatin accessibility profiling. *Nature Communications*, *9*(1), 5345. https://doi.org/10.1038/s41467-018-07771-0

Christman, J. K. (2002). 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene*, *21*(35), 5483–5495. https://doi.org/10.1038/sj.onc.1205699

Clapier, C. R., Iwasa, J., Cairns, B. R., & Peterson, C. L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Reviews. Molecular Cell Biology*, *18*(7), 407–422. https://doi.org/10.1038/nrm.2017.26

Clark, S. J., Argelaguet, R., Kapourani, C.-A., Stubbs, T. M., Lee, H. J., Alda-Catalinas, C., … Reik, W. (2018). scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nature Communications*, *9*(1), 781. https://doi.org/10.1038/s41467-018-03149-4

Contestabile, A., Magara, S., & Cancedda, L. (2017). The GABAergic Hypothesis for Cognitive Disabilities in Down Syndrome. *Frontiers in Cellular Neuroscience*, *11*, 54. https://doi.org/10.3389/fncel.2017.00054

Cusanovich, D. A., Daza, R., Adey, A., Pliner, H. A., Christiansen, L., Gunderson, K. L., … Shendure, J. (2015). Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science*, *348*(6237), 910–914. https://doi.org/10.1126/science.aab1601

Cusanovich, D. A., Hill, A. J., Aghamirzaie, D., Daza, R. M., Pliner, H. A., Berletch, J. B., … Shendure, J. (2018). A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell*, *174*(5), 1309–1324.e18. https://doi.org/10.1016/j.cell.2018.06.052

Cusanovich, D. A., Reddington, J. P., Garfield, D. A., Daza, R. M., Aghamirzaie, D., Marco-Ferreres, R., … Furlong, E. E. M. (2018). The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature*, *555*(7697), 538–542. https://doi.org/10.1038/nature25981

Erlich, Y., Chang, K., Gordon, A., Ronen, R., Navon, O., Rooks, M., & Hannon, G. J. (2009). DNA Sudoku--harnessing high-throughput sequencing for multiplexed specimen analysis. *Genome Research*, *19*(7), 1243–1253. https://doi.org/10.1101/gr.092957.109

Felle, M., Hoffmeister, H., Rothammer, J., Fuchs, A., Exler, J. H., & Längst, G. (2011). Nucleosomes protect DNA from DNA methylation in vivo and in vitro. *Nucleic Acids Research*, *39*(16), 6956–6969. https://doi.org/10.1093/nar/gkr263

Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., … Paul, C. L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(5), 1827–1831. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1542678

Fuks, F., Hurd, P. J., Deplus, R., & Kouzarides, T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Research*, *31*(9), 2305–2312. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12711675

Furlan-Magaril, M., Rincón-Arano, H., & Recillas-Targa, F. (2009). Sequential Chromatin Immunoprecipitation Protocol: ChIP-reChIP. In *Methods in molecular biology (Clifton, N.J.)* (Vol. 543, pp. 253–266). https://doi.org/10.1007/978-1-60327-015-1_17

Gardiner-Garden, M., & Frommer, M. (1987). CpG islands in vertebrate genomes. *Journal of Molecular Biology*, *196*(2), 261–282. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3656447

Gawad, C., Koh, W., & Quake, S. R. (2016). Single-cell genome sequencing: current state of the science. *Nature Reviews Genetics*, *17*(3), 175–188. https://doi.org/10.1038/nrg.2015.16

Geisberg, J. V., & Struhl, K. (2004). Quantitative sequential chromatin immunoprecipitation, a method for analyzing co-occupancy of proteins at genomic regions in vivo. *Nucleic Acids Research*, *32*(19), e151–e151. https://doi.org/10.1093/nar/gnh148

Geisberg, J. V, & Struhl, K. (2005). Analysis of protein co-occupancy by quantitative sequential chromatin immunoprecipitation. *Current Protocols in Molecular Biology*, *Chapter 21*, Unit 21.8. https://doi.org/10.1002/0471142727.mb2108s70

Gilmour, D. S., & Lis, J. T. (1985). In vivo interactions of RNA polymerase II with genes of Drosophila melanogaster. *Molecular and Cellular Biology*, *5*(8), 2009–2018. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3018544

Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R., & Lieb, J. D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Research*, *17*(6), 877–885. https://doi.org/10.1101/gr.5533506

Guedj, F., Pennings, J. LA, Massingham, L. J., Wick, H. C., Siegel, A. E., Tantravahi, U., & Bianchi, D. W. (2016). An Integrated Human/Murine Transcriptome and Pathway Approach To Identify Prenatal Treatments For Down Syndrome. *Scientific Reports*, *6*(1), 32353. https://doi.org/10.1038/srep32353

Haque, A., Engel, J., Teichmann, S. A., & Lönnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Medicine*, *9*(1), 75. https://doi.org/10.1186/s13073-017-0467-4

He, Q., Johnston, J., & Zeitlinger, J. (2015). ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. *Nature Biotechnology*, *33*(4), 395–401. https://doi.org/10.1038/nbt.3121

Hennig, B. P., Velten, L., Racke, I., Tu, C. S., Thoms, M., Rybin, V., … Steinmetz, L. M. (2018). Large-Scale Low-Cost NGS Library Preparation Using a Robust Tn5 Purification and Tagmentation Protocol. *G3: Genes|Genomes|Genetics*, *8*(1), 79–89. https://doi.org/10.1534/g3.117.300257

Hwang, B., Lee, J. H., & Bang, D. (2018). Single-cell RNA sequencing technologies and

bioinformatics pipelines. *Experimental & Molecular Medicine*, *50*(8), 96. https://doi.org/10.1038/s12276-018-0071-8

Hyun, K., Jeon, J., Park, K., & Kim, J. (2017). Writing, erasing and reading histone lysine methylations. *Experimental & Molecular Medicine*, *49*(4), e324. https://doi.org/10.1038/emm.2017.11

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., … Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science*, *333*(6047), 1300–1303. https://doi.org/10.1126/science.1210597

Jia, G., Preussner, J., Chen, X., Guenther, S., Yuan, X., Yekelchyk, M., … Braun, T. (2018). Single cell RNA-seq and ATAC-seq analysis of cardiac progenitor cell transition states and lineage settlement. *Nature Communications*, *9*(1), 4877. https://doi.org/10.1038/s41467-018-07307-6

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., & Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark "nucleosome-free regions" of active promoters and other regulatory regions. *Nature Genetics*, *41*(8), 941–945. https://doi.org/10.1038/ng.409

Johnson, D. S., Mortazavi, A., Myers, R. M., & Wold, B. (2007). Genome-Wide Mapping of in Vivo Protein-DNA Interactions. *Science*, *316*(5830), 1497–1502. https://doi.org/10.1126/science.1141319

Jonkers, I., & Lis, J. T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, *16*(3), 167–177. https://doi.org/10.1038/nrm3953

Kalisky, T., Blainey, P., & Quake, S. R. (2011). Genomic Analysis at the Single-Cell Level. *Annual Review of Genetics*, *45*(1), 431–445. https://doi.org/10.1146/annurev-genet-102209-163607

Kelly, T. K., Liu, Y., Lay, F. D., Liang, G., Berman, B. P., & Jones, P. A. (2012). Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Research*, *22*(12), 2497–2506. https://doi.org/10.1101/gr.143008.112

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., … Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. *Nature*, *518*(7539), 317–330. https://doi.org/10.1038/nature14248

Lawrence, M., Daujat, S., & Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends in Genetics : TIG*, *32*(1), 42–56. https://doi.org/10.1016/j.tig.2015.10.007

Lay, F. D., Liu, Y., Kelly, T. K., Witt, H., Farnham, P. J., Jones, P. A., & Berman, B. P. (2015). The role of DNA methylation in directing the functional organization of the cancer epigenome. *Genome Research*, *25*(4), 467–477. https://doi.org/10.1101/gr.183368.114

Lea, A. J., Vockley, C. M., Johnston, R. A., Del Carpio, C. A., Barreiro, L. B., Reddy, T. E., & Tung, J. (2018). Genome-wide quantification of the effects of DNA methylation on human gene regulation. *ELife*, *7*.

https://doi.org/10.7554/eLife.37513
Li, H., Rauch, T., Chen, Z.-X., Szabó, P. E., Riggs, A. D., & Pfeifer, G. P. (2006). The Histone Methyltransferase SETDB1 and the DNA Methyltransferase DNMT3A Interact Directly and Localize to Promoters Silenced in Cancer Cells. *Journal of Biological Chemistry*, *281*(28), 19489–19500. https://doi.org/10.1074/jbc.M513249200
Liao, J., Karnik, R., Gu, H., Ziller, M. J., Clement, K., Tsankov, A. M., … Meissner, A. (2015). Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nature Genetics*, *47*(5), 469–478. https://doi.org/10.1038/ng.3258
Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., & Ecker, J. R. (2008). Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. *Cell*, *133*(3), 523–536. https://doi.org/10.1016/j.cell.2008.03.029
Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., … Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, *462*(7271), 315–322. https://doi.org/10.1038/nature08514
Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, *389*(6648), 251–260. https://doi.org/10.1038/38444
Luo, C., Keown, C. L., Kurihara, L., Zhou, J., He, Y., Li, J., … Ecker, J. R. (2017). Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science*, *357*(6351), 600–604. https://doi.org/10.1126/science.aan3351
Mahat, D. B., Kwak, H., Booth, G. T., Jonkers, I. H., Danko, C. G., Patel, R. K., … Lis, J. T. (2016). Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nature Protocols*, *11*(8), 1455–1476. https://doi.org/10.1038/nprot.2016.086
Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S., & Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Research*, *33*(18), 5868–5877. https://doi.org/10.1093/nar/gki901
Messerschmidt, D. M., Knowles, B. B., & Solter, D. (2014). DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes & Development*, *28*(8), 812–828. https://doi.org/10.1101/gad.234294.113
Mulqueen, R. M., Pokholok, D., Norberg, S. J., Torkenczy, K. A., Fields, A. J., Sun, D., … Adey, A. C. (2018). Highly scalable generation of DNA methylation profiles in single cells. *Nature Biotechnology*, *36*(5), 428–431. https://doi.org/10.1038/nbt.4112
Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., … Wigler, M. (2011). Tumour evolution inferred by single-cell sequencing. *Nature*, *472*(7341), 90–94. https://doi.org/10.1038/nature09807

Patel, D. J., & Wang, Z. (2013). Readout of epigenetic modifications. *Annual Review of Biochemistry*, *82*(1), 81–118. https://doi.org/10.1146/annurev-biochem-072711-165700

Piccolo, F. M., & Fisher, A. G. (2014). Getting rid of DNA methylation. *Trends in Cell Biology*, *24*(2), 136–143. https://doi.org/10.1016/j.tcb.2013.09.001

Picelli, S., Björklund, Å. K., Reinius, B., Sagasser, S., Winberg, G., & Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Research*, *24*(12), 2033–2040. https://doi.org/10.1101/gr.177881.114

Rasmussen, K. D., & Helin, K. (2016). Role of TET enzymes in DNA methylation, development, and cancer. *Genes & Development*, *30*(7), 733–750. https://doi.org/10.1101/gad.276568.115

Rhee, I., Bachman, K. E., Park, B. H., Jair, K.-W., Yen, R.-W. C., Schuebel, K. E., … Vogelstein, B. (2002). DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature*, *416*(6880), 552–556. https://doi.org/10.1038/416552a

Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D., & Klug, A. (n.d.). Structure of the nucleosome core particle at 7 A resolution. *Nature*, *311*(5986), 532–537. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6482966

Rossi, M. J., Lai, W. K. M., & Pugh, B. F. (2018). Simplified ChIP-exo assays. *Nature Communications*, *9*(1), 2842. https://doi.org/10.1038/s41467-018-05265-7

Roulet, E., Busso, S., Camargo, A. A., Simpson, A. J. G., Mermod, N., & Bucher, P. (2002). High-throughput SELEX–SAGE method for quantitative modeling of transcription-factor binding sites. *Nature Biotechnology*, *20*(8), 831–835. https://doi.org/10.1038/nbt718

Sakaue, M., Ohta, H., Kumaki, Y., Oda, M., Sakaide, Y., Matsuoka, C., … Okano, M. (2010). DNA Methylation Is Dispensable for the Growth and Survival of the Extraembryonic Lineages. *Current Biology*, *20*(16), 1452–1457. https://doi.org/10.1016/j.cub.2010.06.050

Sauer, P. V, Gu, Y., Liu, W. H., Mattiroli, F., Panne, D., Luger, K., & Churchill, M. E. (2018). Mechanistic insights into histone deposition and nucleosome assembly by the chromatin assembly factor-1. *Nucleic Acids Research*, *46*(19), 9907–9917. https://doi.org/10.1093/nar/gky823

Schones, D. E., Cui, K., Cuddapah, S., Roh, T.-Y., Barski, A., Wang, Z., … Zhao, K. (2008). Dynamic Regulation of Nucleosome Positioning in the Human Genome. *Cell*, *132*(5), 887–898. https://doi.org/10.1016/j.cell.2008.02.022

Schutsky, E. K., DeNizio, J. E., Hu, P., Liu, M. Y., Nabel, C. S., Fabyanic, E. B., … Kohli, R. M. (2018). Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase. *Nature Biotechnology*, *36*(11), 1083–1090. https://doi.org/10.1038/nbt.4204

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I. K., … Widom, J. (2006). A genomic code for nucleosome positioning. *Nature*, *442*(7104), 772–778. https://doi.org/10.1038/nature04979

Skene, P. J., & Henikoff, S. (2017). An efficient targeted nuclease strategy for high-

resolution mapping of DNA binding sites. *ELife*, *6*. https://doi.org/10.7554/eLife.21856

Smallwood, S. A., Lee, H. J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., … Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nature Methods*, *11*(8), 817–820. https://doi.org/10.1038/nmeth.3035

Smith, Z. D., & Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nature Reviews Genetics*, *14*(3), 204–220. https://doi.org/10.1038/nrg3354

Song, L., & Crawford, G. E. (2010). DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harbor Protocols*, *2010*(2), pdb.prot5384. https://doi.org/10.1101/pdb.prot5384

Stevens, T. J., Lando, D., Basu, S., Atkinson, L. P., Cao, Y., Lee, S. F., … Laue, E. D. (2017). 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature*, *544*(7648), 59–64. https://doi.org/10.1038/nature21429

Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, *403*(6765), 41–45. https://doi.org/10.1038/47412

Suzuki, H., Takatsuka, S., Akashi, H., Yamamoto, E., Nojima, M., Maruyama, R., … Toyota, M. (2011). Genome-wide Profiling of Chromatin Signatures Reveals Epigenetic Regulation of MicroRNA Genes in Colorectal Cancer. *Cancer Research*, *71*(17), 5646–5658. https://doi.org/10.1158/0008-5472.CAN-11-1076

Teissandier, A., & Bourc'his, D. (2017). Gene body DNA methylation conspires with H3K36me3 to preclude aberrant transcription. *The EMBO Journal*, *36*(11), 1471–1473. https://doi.org/10.15252/embj.201796812

Thoma, F., Koller, T., & Klug, A. (1979). Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *The Journal of Cell Biology*, *83*(2 Pt 1), 403–427. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/387806

van Attikum, H., & Gasser, S. M. (2009). Crosstalk between histone modifications during the DNA damage response. *Trends in Cell Biology*, *19*(5), 207–217. https://doi.org/10.1016/j.tcb.2009.03.001

Wijesinghe, P., & Bhagwat, A. S. (2012). Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G. *Nucleic Acids Research*, *40*(18), 9206–9217. https://doi.org/10.1093/nar/gks685

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., … Taipale, J. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*, *356*(6337), eaaj2239. https://doi.org/10.1126/science.aaj2239

Yong, W.-S., Hsu, F.-M., & Chen, P.-Y. (2016). Profiling genome-wide DNA methylation. *Epigenetics & Chromatin*, *9*, 26. https://doi.org/10.1186/s13072-016-0075-3

Yu, M., Hon, G. C., Szulwach, K. E., Song, C.-X., Zhang, L., Kim, A., … He, C. (2012).

Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. *Cell*, *149*(6), 1368–1380. https://doi.org/10.1016/j.cell.2012.04.027

Zaret, K. S., & Mango, S. E. (2016). Pioneer transcription factors, chromatin dynamics, and cell fate control. *Current Opinion in Genetics & Development*, *37*, 76–81. https://doi.org/10.1016/j.gde.2015.12.003

Zhang, Z., Wippo, C. J., Wal, M., Ward, E., Korber, P., & Pugh, B. F. (2011). A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science (New York, N.Y.)*, *332*(6032), 977–980. https://doi.org/10.1126/science.1200508

Zhao, Y., & Garcia, B. A. (2015). Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harbor Perspectives in Biology*, *7*(9), a025064. https://doi.org/10.1101/cshperspect.a025064

References: Chapter 2

Adey A, Shendure J. 2012. Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing. Genome Res 22: 1139-1143.

Akalin A, Franke V, Vlahovicek K, Mason CE, Schubeler D. 2015. Genomation: a toolkit to summarize, annotate and visualize genomic intervals. Bioinformatics 31: 1127-1129.

Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE. 2012. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol 13: R87.

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125: 315-326.

Blattler A, Yao L, Witt H, Guo Y, Nicolet CM, Berman BP, Farnham PJ. 2014. Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. Genome Biol 15: 469.

Booth MJ, Ost TWB, Beraldi D, Bell NM, Branco MR, Reik W, Balasubramanian S. 2013. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. Nat Protocols 8: 1841-1851.

Boyle P, Clement K, Gu H, Smith ZD, Ziller M, Fostel JL, Holmes L, Meldrim J, Kelley F, Gnirke A et al. 2012. Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. Genome Biol 13: R92.

Brinkman AB, Gu H, Bartels SJ, Zhang Y, Matarese F, Simmer F, Marks H, Bock C, Gnirke A, Meissner A et al. 2012. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Res 22: 1128-1138.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 10: 1213-1218.

Chatterjee A, Rodger EJ, Stockwell PA, Weeks RJ, Morison IM. 2012. Technical considerations for reduced representation bisulfite sequencing with multiplexed libraries. Journal of biomedicine & biotechnology 2012: 741542.

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34: i884-i890.

Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, Snyder MP, Pritchard JK, Kundaje A, Greenleaf WJ et al. 2016. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat Genet 48: 1193-1203.

Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B et al. 2017. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods 14: 959-962.

Cusanovich DA, Daza R, Adey A, Pliner H, Christiansen L, Gunderson KL, Steemers FJ,

Trapnell C, Shendure J. 2015. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 348: 910-914.
Das S, Chadwick BP. 2016. Influence of Repressive Histone and DNA Methylation upon D4Z4 Transcription in Non-Myogenic Cells. PLoS ONE 11: e0160022.
Garrett-Bakelman FE, Sheridan CK, Kacmarczyk TJ, Ishii J, Betel D, Alonso A, Mason CE, Figueroa ME, Melnick AM. 2015. Enhanced reduced representation bisulfite sequencing for assessment of DNA methylation at base pair resolution. Journal of visualized experiments : JoVE doi:10.3791/52246: e52246.
Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, Panhuis TM, Mieczkowski P, Secchi A, Bosco D et al. 2010. A map of open chromatin in human pancreatic islets. Nat Genet 42: 255-259.
Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW et al. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459: 108-112.
Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38: 576-589.
Horard B, Eymery A, Fourel G, Vassetzky N, Puechberty J, Roizes G, Lebrigand K, Barbry P, Laugraud A, Gautier C et al. 2009. Global analysis of DNA methylation and transcription of human repetitive sequences. Epigenetics 4: 339-350.
Jorda M, Diez-Villanueva A, Mallona I, Martin B, Lois S, Barrera V, Esteller M, Vavouri T, Peinado MA. 2017. The epigenetic landscape of Alu repeats delineates the structural and functional genomic architecture of colon cancer cells. Genome Res 27: 118-132.
Karimi MM, Goyal P, Maksakova IA, Bilenky M, Leung D, Tang JX, Shinkai Y, Mager DL, Jones S, Hirst M et al. 2011. DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. Cell Stem Cell 8: 676-687.
Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA. 2012. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res 22: 2497-2506.
Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12: 357-360.
Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27: 1571-1572.
Lay FD, Liu Y, Kelly TK, Witt H, Farnham PJ, Jones PA, Berman BP. 2015. The role of DNA methylation in directing the functional organization of the cancer epigenome. Genome Res 25: 467-477.
Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26: 589-595.
Liang G, Chan MF, Tomigahara Y, Tsai YC, Gonzales FA, Li E, Laird PW, Jones PA. 2002. Cooperativity between DNA methyltransferases in the maintenance methylation

of repetitive elements. Mol Cell Biol 22: 480-491.
Liao Y, Smyth GK, Shi W. 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res 41: e108.
Lindroth AM, Park YJ, McLean CM, Dokshin GA, Persson JM, Herman H, Pasini D, Miro X, Donohoe ME, Lee JT et al. 2008. Antagonism between DNA and H3K27 Methylation at the Imprinted Rasgrf1 Locus. PLoS Genetics 4: e1000145.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15: 550.
Lu H, Yuan Z, Tan T, Wang J, Zhang J, Luo HJ, Xia Y, Ji W, Gao F. 2015. Improved tagmentation-based whole-genome bisulfite sequencing for input DNA from less than 100 mammalian cells. Epigenomics 7: 47-56.
Maurano MT, Wang H, John S, Shafer A, Canfield T, Lee K, Stamatoyannopoulos JA. 2015. Role of DNA Methylation in Modulating Transcription Factor Occupancy. Cell Rep 12: 1184-1195.
Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res 33: 5868-5877.
Mulqueen RM, Pokholok D, Norberg SJ, Torkenczy KA, Fields AJ, Sun D, Sinnamon JR, Shendure J, Trapnell C, O'Roak BJ et al. 2018. Highly scalable generation of DNA methylation profiles in single cells. Nat Biotechnol 36: 428-431.
Picelli S, Björklund ÅK, Reinius B, Sagasser S, Winberg G, Sandberg R. 2014. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Research 24: 2033-2040.
Rahbari R, Sheahan T, Modes V, Collier P, Macfarlane C, Badge RM. 2009. A novel L1 retrotransposon marker for HeLa cell line identification. Biotechniques 46: 277-284.
Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res 44: W160-165.
Rhee I, Bachman KE, Park BH, Jair KW, Yen RW, Schuebel KE, Cui H, Feinberg AP, Lengauer C, Kinzler KW et al. 2002. DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature 416: 552-556.
Rhee I, Jair KW, Yen RWC, Lengauer C, Herman JG, Kinzler KW, Vogelstein B, Baylin SB, Schuebel KE. 2000. CpG methylation is maintained in human cancer cells lacking DNMT1. Nature 404: 1003-1007.
Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J et al. 2015. Integrative analysis of 111 reference human epigenomes. Nature 518: 317-330.
Rodriguez JM, Maietta P, Ezkurdia I, Pietrelli A, Wesselink JJ, Lopez G, Valencia A, Tress ML. 2013. APPRIS: annotation of principal and alternative splice isoforms. Nucleic Acids Res 41: D110-117.
Schmidl C, Rendeiro AF, Sheffield NC, Bock C. 2015. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat Methods 12: 963-965.

Song CX, Szulwach KE, Dai Q, Fu Y, Mao SQ, Lin L, Street C, Li Y, Poidevin M, Wu H et al. 2013. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. Cell 153: 678-691.
Stark R, Brown G. 2018. DiffBind: Differential binding analysis of ChIP- Seq peak data. In Bioconductor.
Statham AL, Robinson MD, Song JZ, Coolen MW, Stirzaker C, Clark SJ. 2012. Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. Genome Res 22: 1120-1127.
Suzuki M, Liao W, Wos F, Johnston AD, DeGrazia J, Ishii J, Bloom T, Zody MC, Germer S, Greally JM. 2018. Whole-genome bisulfite sequencing with improved accuracy and cost. Genome Res 28: 1364-1371.
Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B et al. 2012. The accessible chromatin landscape of the human genome. Nature 489: 75-82.
Walter M, Teissandier A, Perez-Palacios R, Bourc'his D. 2016. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. eLife 5.
Wu H, Wu X, Zhang Y. 2016. Base-resolution profiling of active DNA demethylation using MAB-seq and caMAB-seq. Nature protocols 11: 1081-1100.
Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F et al. 2017. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science 356.
Young L, Sung J, Stacey G, Masters JR. 2010. Detection of Mycoplasma in cell cultures. Nature protocols 5: 929-934.
Yu G, Wang LG, He QY. 2015. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31: 2382-2383.
Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B et al. 2012. Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. Cell 149: 1368–1380.
Zaret KS, Carroll JS. 2011. Pioneer transcription factors: establishing competence for gene expression. Genes Dev 25: 2227-2241.
Zaret KS, Mango SE. 2016. Pioneer transcription factors, chromatin dynamics, and cell fate control. Curr Opin Genet Dev 37: 76-81.
Zhu LJ, Gazin C, Lawson ND, Pages H, Lin SM, Lapointe DS, Green MR. 2010. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC bioinformatics 11: 237.
Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S. 2008. Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature 456: 125-129.

References: Chapter 3

Amezquita, R. A. (2018). marge: An API for Analysis of Motifs Using HOMER in R. bioRxiv, 249268. doi:10.1101/249268

Amini, S., Pushkarev, D., Christiansen, L., Kostem, E., Royce, T., Turk, C., . . . Steemers, F. J. (2014). Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing. Nat Genet, 46(12), 1343-1349. doi:10.1038/ng.3119

Antonarakis, S. E., Lyle, R., Dermitzakis, E. T., Reymond, A., & Deutsch, S. (2004). Chromosome 21 and down syndrome: from genomics to pathophysiology. Nat Rev Genet, 5(10), 725-738. doi:10.1038/nrg1448

Aranguren, X. L., Beerens, M., Coppiello, G., Wiese, C., Vandersmissen, I., Lo Nigro, A., . . . Luttun, A. (2013). COUP-TFII orchestrates venous and lymphatic endothelial identity by homo- or hetero-dimerisation with PROX1. J Cell Sci, 126(Pt 5), 1164-1175. doi:10.1242/jcs.116293

Arnold, S. J., Huang, G. J., Cheung, A. F., Era, T., Nishikawa, S., Bikoff, E. K., . . . Groszer, M. (2008). The T-box transcription factor Eomes/Tbr2 regulates neurogenesis in the cortical subventricular zone. Genes Dev, 22(18), 2479-2484. doi:10.1101/gad.475408

Ash, J. A., Velazquez, R., Kelley, C. M., Powers, B. E., Ginsberg, S. D., Mufson, E. J., & Strupp, B. J. (2014). Maternal choline supplementation improves spatial mapping and increases basal forebrain cholinergic neuron number and size in aged Ts65Dn mice. Neurobiol Dis, 70, 32-42. doi:10.1016/j.nbd.2014.06.001

Aylward, E. H., Li, Q., Honeycutt, N. A., Warren, A. C., Pulsifer, M. B., Barta, P. E., . . . Pearlson, G. D. (1999). MRI volumes of the hippocampus and amygdala in adults with Down's syndrome with and without dementia. Am J Psychiatry, 156(4), 564-568. doi:10.1176/ajp.156.4.564

Bajenaru, M. L., Zhu, Y., Hedrick, N. M., Donahoe, J., Parada, L. F., & Gutmann, D. H. (2002). Astrocyte-specific inactivation of the neurofibromatosis 1 gene (NF1) is insufficient for astrocytoma formation. Mol Cell Biol, 22(14), 5100-5113.

Cao, J., Packer, J. S., Ramani, V., Cusanovich, D. A., Huynh, C., Daza, R., . . . Shendure, J. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. Science, 357(6352), 661-667. doi:10.1126/science.aam8940

de Vries, A. (2016). Retrieved from https://CRAN.R-project.org/package=ggdendro

Cappelli-Bigazzi, M., Santoro, G., Battaglia, C., Palladino, M. T., Carrozza, M., Russo, M. G., . . . Calabro, R. (2004). Endothelial cell function in patients with Down's syndrome. Am J Cardiol, 94(3), 392-395. doi:10.1016/j.amjcard.2004.04.047

Chakrabarti, L., Best, T. K., Cramer, N. P., Carney, R. S., Isaac, J. T., Galdzicki, Z., & Haydar, T. F. (2010). Olig1 and Olig2 triplication causes developmental brain defects in Down syndrome. Nat Neurosci, 13(8), 927-934. doi:10.1038/nn.2600

Chakrabarti, L., Galdzicki, Z., & Haydar, T. F. (2007). Defects in embryonic neurogenesis and initial synapse formation in the forebrain of the Ts65Dn mouse model of Down syndrome. J Neurosci, 27(43), 11483-11495. doi:10.1523/JNEUROSCI.3406-07.2007

Contestabile, A., Magara, S., & Cancedda, L. (2017). The GABAergic Hypothesis for Cognitive Disabilities in Down Syndrome. Frontiers in Cellular Neuroscience, 11, 54. doi:10.3389/fncel.2017.00054

Corces, M. R., Trevino, A. E., Hamilton, E. G., Greenside, P. G., Sinnott-Armstrong, N. A., Vesuna, S., . . . Chang, H. Y. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods, 14(10), 959-962. doi:10.1038/nmeth.4396

Cusanovich, D. A., Daza, R., Adey, A., Pliner, H., Christiansen, L., Gunderson, K. L., . . . Shendure, J. (2015). Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. Science, 348, 910-914. doi:10.1126/science.aab1601

Cusanovich, D. A., Hill, A. J., Aghamirzaie, D., Daza, R. M., Pliner, H. A., Berletch, J. B., . . . Shendure, J. (2018). A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. Cell, 174(5), 1309-1324 e1318. doi:10.1016/j.cell.2018.06.052

Cusanovich, D. A., Reddington, J. P., Garfield, D. A., Daza, R. M., Aghamirzaie, D., Marco-Ferreres, R., . . . Furlong, E. E. M. (2018). The cis-regulatory dynamics of embryonic development at single-cell resolution. Nature, 555(7697), 538-542. doi:10.1038/nature25981

Delfini, M. C., & Duprez, D. (2004). Ectopic Myf5 or MyoD prevents the neuronal differentiation program in addition to inducing skeletal muscle differentiation, in the chick neural tube. Development, 131(4), 713-723. doi:10.1242/dev.00967

Dragunow, M., Yamada, N., Bilkey, D. K., & Lawlor, P. (1992). Induction of immediate-early gene proteins in dentate granule cells and somatostatin interneurons after hippocampal seizures. Brain Res Mol Brain Res, 13(1-2), 119-126.

Duchon, A., Raveau, M., Chevalier, C., Nalesso, V., Sharp, A. J., & Herault, Y. (2011). Identification of the translocation breakpoints in the Ts65Dn and Ts1Cje mouse lines: relevance for modeling Down syndrome. Mamm Genome, 22(11-12), 674-684. doi:10.1007/s00335-011-9356-0

Gardiner, K. (2004). Gene-dosage effects in Down syndrome and trisomic mouse models. Genome Biol, 5(10), 244. doi:10.1186/gb-2004-5-10-244

Gong, S., Doughty, M., Harbaugh, C. R., Cummins, A., Hatten, M. E., Heintz, N., & Gerfen, C. R. (2007). Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. J Neurosci, 27(37), 9817-9823. doi:10.1523/JNEUROSCI.2707-07.2007

Gray, L. T., Yao, Z., Nguyen, T. N., Kim, T. K., Zeng, H., & Tasic, B. (2017). Layer-specific chromatin accessibility landscapes reveal regulatory networks in adult mouse visual cortex. Elife, 6. doi:10.7554/eLife.21883

Guedj, F., Pennings, J. L. A., Massingham, L. J., Wick, H. C., Siegel, A. E., Tantravahi, U., & Bianchi, D. W. (2016). An Integrated Human/Murine Transcriptome and Pathway Approach To Identify Prenatal Treatments For Down Syndrome. Sci Rep, 6, 32353. doi:10.1038/srep32353

https://www.nature.com/articles/srep32353#supplementary-information

Guidi, S., Bonasoni, P., Ceccarelli, C., Santini, D., Gualtieri, F., Ciani, E., & Bartesaghi, R. (2008). Neurogenesis impairment and increased cell death reduce total neuron number in the hippocampal region of fetuses with Down syndrome. Brain Pathol, 18(2), 180-197. doi:10.1111/j.1750-3639.2007.00113.x

Hahne, F., & Ivanek, R. (2016). Visualizing Genomic Data Using Gviz and Bioconductor. Methods Mol Biol, 1418, 335-351. doi:10.1007/978-1-4939-3578-9_16

Harrington, A. J., Raissi, A., Rajkovich, K., Berto, S., Kumar, J., Molinaro, G., . . . Cowan, C. W. (2016). MEF2C regulates cortical inhibitory and excitatory synapses and behaviors relevant to neurodevelopmental disorders. Elife, 5. doi:10.7554/eLife.20059

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., . . . Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell, 38(4), 576-589. doi:10.1016/j.molcel.2010.05.004

Hernandez-Gonzalez, S., Ballestin, R., Lopez-Hidalgo, R., Gilabert-Juan, J., Blasco-Ibanez, J. M., Crespo, C., . . . Varea, E. (2015). Altered distribution of hippocampal interneurons in the murine Down Syndrome model Ts65Dn. Neurochem Res, 40(1), 151-164. doi:10.1007/s11064-014-1479-8

Hou, P. S., Chuang, C. Y., Kao, C. F., Chou, S. J., Stone, L., Ho, H. N., . . . Kuo, H. C. (2013). LHX2 regulates the neural differentiation of human embryonic stem cells via transcriptional modulation of PAX6 and CER1. Nucleic Acids Res, 41(16), 7753-7770. doi:10.1093/nar/gkt567

Huang, T. N., Chuang, H. C., Chou, W. H., Chen, C. Y., Wang, H. F., Chou, S. J., & Hsueh, Y. P. (2014). Tbr1 haploinsufficiency impairs amygdalar axonal projections and results in cognitive abnormality. Nat Neurosci, 17(2), 240-247. doi:10.1038/nn.3626

Jernigan, T. L., Bellugi, U., Sowell, E., Doherty, S., & Hesselink, J. R. (1993). Cerebral morphologic distinctions between Williams and Down syndromes. Arch Neurol, 50(2), 186-191.

Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., & Kent, W. J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res, 32(Database issue), D493-496. doi:10.1093/nar/gkh103

Kassambara, A. (2018). Retrieved from https://cloud.r-project.org/web/packages/ggpubr/index.html

Kerkel, K., Schupf, N., Hatta, K., Pang, D., Salas, M., Kratz, A., . . . Tycko, B. (2010). Altered DNA methylation in leukocytes with trisomy 21. PLoS Genet, 6(11), e1001212. doi:10.1371/journal.pgen.1001212

Knouse, K. A., Wu, J., & Hendricks, A. (2017). Detection of Copy Number Alterations Using Single Cell Sequencing. J Vis Exp(120). doi:10.3791/55143

Krasuski, J. S., Alexander, G. E., Horwitz, B., Rapoport, S. I., & Schapiro, M. B. (2002). Relation of medial temporal lobe volumes to age and memory function in nondemented adults with Down's syndrome: implications for the prodromal phase of Alzheimer's disease. Am J Psychiatry, 159(1), 74-81. doi:10.1176/appi.ajp.159.1.74

Latchney, S. E., Jaramillo, T. C., Rivera, P. D., Eisch, A. J., & Powell, C. M. (2015). Chronic P7C3 treatment restores hippocampal neurogenesis in the Ts65Dn mouse

model of Down Syndrome [Corrected]. Neurosci Lett, 591, 86-92. doi:10.1016/j.neulet.2015.02.008
Leone, D. P., Srinivasan, K., Chen, B., Alcamo, E., & McConnell, S. K. (2008). The determination of projection neuron identity in the developing cerebral cortex. Curr Opin Neurobiol, 18(1), 28-35. doi:10.1016/j.conb.2008.05.006
Li, H., Radford, J. C., Ragusa, M. J., Shea, K. L., McKercher, S. R., Zaremba, J. D., . . . Lipton, S. A. (2008). Transcription factor MEF2C influences neural stem/progenitor cell differentiation and maturation in vivo. Proc Natl Acad Sci U S A, 105(27), 9397-9402. doi:10.1073/pnas.0802876105
Liao, Y., Smyth, G. K., & Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res, 41(10), e108. doi:10.1093/nar/gkt214
Lourenco, T., Paes de Faria, J., Bippes, C. A., Maia, J., Lopes-da-Silva, J. A., Relvas, J. B., & Graos, M. (2016). Modulation of oligodendrocyte differentiation and maturation by combined biochemical and mechanical cues. Sci Rep, 6, 21563. doi:10.1038/srep21563
Mayer, C., Hafemeister, C., Bandler, R. C., Machold, R., Batista Brito, R., Jaglin, X., . . . Satija, R. (2018). Developmental diversification of cortical inhibitory interneurons. Nature, 555(7697), 457-462. doi:10.1038/nature25999
Mendioroz, M., Do, C., Jiang, X., Liu, C., Darbary, H. K., Lang, C. F., . . . Tycko, B. (2015). Trans effects of chromosome aneuploidies on DNA methylation patterns in human Down syndrome and mouse models. Genome Biol, 16, 263. doi:10.1186/s13059-015-0827-6
Mo, A., Mukamel, E. A., Davis, F. P., Luo, C., Henry, G. L., Picard, S., . . . Nathans, J. (2015). Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. Neuron, 86(6), 1369-1384. doi:10.1016/j.neuron.2015.05.018
Molyneaux, B. J., Arlotta, P., Menezes, J. R., & Macklis, J. D. (2007). Neuronal subtype specification in the cerebral cortex. Nat Rev Neurosci, 8(6), 427-437. doi:10.1038/nrn2151
Moulos, P., & Hatzis, P. (2015). Systematic integration of RNA-Seq statistical algorithms for accurate detection of differential gene expression patterns. Nucleic Acids Res, 43(4), e25. doi:10.1093/nar/gku1273
O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., . . . Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res, 44(D1), D733-745. doi:10.1093/nar/gkv1189
Olmos-Serrano, J. L., Kang, H. J., Tyler, W. A., Silbereis, J. C., Cheng, F., Zhu, Y., . . . Sestan, N. (2016). Down Syndrome Developmental Brain Transcriptome Reveals Defective Oligodendrocyte Differentiation and Myelination. Neuron, 89(6), 1208-1222. doi:10.1016/j.neuron.2016.01.042
Ou, J., Liu, H., Yu, J., Kelliher, M. A., Castilla, L. H., Lawson, N. D., & Zhu, L. J. (2018). ATACseqQC: a Bioconductor package for post-alignment quality assessment of ATAC-seq data. BMC Genomics, 19(1), 169. doi:10.1186/s12864-018-4559-3

Picelli, S., Björklund, Å. K., Reinius, B., Sagasser, S., Winberg, G., & Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Research, 24(12), 2033-2040. doi:10.1101/gr.177881.114

Pinter, J. D., Brown, W. E., Eliez, S., Schmitt, J. E., Capone, G. T., & Reiss, A. L. (2001). Amygdala and hippocampal volumes in children with Down syndrome: a high-resolution MRI study. Neurology, 56(7), 972-974.

Pinter, J. D., Eliez, S., Schmitt, J. E., Capone, G. T., & Reiss, A. L. (2001). Neuroanatomy of Down's syndrome: a high-resolution MRI study. Am J Psychiatry, 158(10), 1659-1665. doi:10.1176/appi.ajp.158.10.1659

Powers, B. E., Kelley, C. M., Velazquez, R., Ash, J. A., Strawderman, M. S., Alldred, M. J., . . . Strupp, B. J. (2017). Maternal choline supplementation in a mouse model of Down syndrome: Effects on attention and nucleus basalis/substantia innominata neuron morphology in adult offspring. Neuroscience, 340, 501-514. doi:10.1016/j.neuroscience.2016.11.001

Powers, B. E., Velazquez, R., Kelley, C. M., Ash, J. A., Strawderman, M. S., Alldred, M. J., . . . Strupp, B. J. (2016). Attentional function and basal forebrain cholinergic neuron morphology during aging in the Ts65Dn mouse model of Down syndrome. Brain Struct Funct, 221(9), 4337-4352. doi:10.1007/s00429-015-1164-y

Preissl, S., Fang, R., Huang, H., Zhao, Y., Raviram, R., Gorkin, D. U., . . . Ren, B. (2018). Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. Nat Neurosci, 21(3), 432-439. doi:10.1038/s41593-018-0079-3

Ramirez, F., Ryan, D. P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., . . . Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res, 44(W1), W160-165. doi:10.1093/nar/gkw257

Reeves, R. H., Irving, N. G., Moran, T. H., Wohn, A., Kitt, C., Sisodia, S. S., . . . Davisson, M. T. (1995). A mouse model for Down syndrome exhibits learning and behaviour deficits. Nat Genet, 11(2), 177-184.

Reinholdt, L. G., Ding, Y., Gilbert, G. J., Czechanski, A., Solzak, J. P., Roper, R. J., . . . Davisson, M. T. (2011). Molecular characterization of the translocation breakpoints in the Down syndrome mouse model Ts65Dn. Mamm Genome, 22(11-12), 685-691. doi:10.1007/s00335-011-9357-z

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res, 43(7), e47. doi:10.1093/nar/gkv007

Sander, M., Paydar, S., Ericson, J., Briscoe, J., Berber, E., German, M., . . . Rubenstein, J. L. (2000). Ventral neural patterning by Nkx homeobox genes: Nkx6.1 controls somatic motor neuron and ventral interneuron fates. Genes Dev, 14(17), 2134-2139.

Schupf, N., Kapell, D., Nightingale, B., Rodriguez, A., Tycko, B., & Mayeux, R. (1998). Earlier onset of Alzheimer's disease in men with Down syndrome. Neurology, 50(4), 991-995.

Schwalb, B., Tresch, A., Torkler, P., Duemcke, S., Demel, C., Ripley, B., & Venables, B. (2018). Retrieved from https://CRAN.R-project.org/package=LSD

Schweitzer, P., Harris, A., Mandelman, D., Jackson, S., Cifuentes, F., & Degoricija, L. (2014). Precise Quantification of Next Generation Sequencing Ion Torrent™ and Illumina Libraries using the QuantStudio™ 3D Digital PCR Platform. Journal of Biomolecular Techniques : JBT, 25(Suppl), S15-S15.

Seregaza, Z., Roubertoux, P. L., Jamon, M., & Soumireu-Mourat, B. (2006). Mouse models of cognitive disorders in trisomy 21: a review. Behav Genet, 36(3), 387-404. doi:10.1007/s10519-006-9056-9

Sharma, K., Sheng, H. Z., Lettieri, K., Li, H., Karavanov, A., Potter, S., . . . Pfaff, S. L. (1998). LIM homeodomain factors Lhx3 and Lhx4 assign subtype identities for motor neurons. Cell, 95(6), 817-828.

Soetaert, K. (2017). Retrieved from https://CRAN.R-project.org/package=plot3D

Thaler, J. P., Lee, S. K., Jurata, L. W., Gill, G. N., & Pfaff, S. L. (2002). LIM factor Lhx3 contributes to the specification of motor neuron and interneuron identity through cell-type-specific protein-protein interactions. Cell, 110(2), 237-249. doi:10.1016/S0092-8674(02)00823-1

Trapani, J. G., Obholzer, N., Mo, W., Brockerhoff, S. E., & Nicolson, T. (2009). Synaptojanin1 is required for temporal fidelity of synaptic transmission in hair cells. PLoS Genet, 5(5), e1000480. doi:10.1371/journal.pgen.1000480

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., . . . Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol, 32(4), 381-386. doi:10.1038/nbt.2859

Velazquez, R., Ash, J. A., Powers, B. E., Kelley, C. M., Strawderman, M., Luscher, Z. I., . . . Strupp, B. J. (2013). Maternal choline supplementation improves spatial learning and adult hippocampal neurogenesis in the Ts65Dn mouse model of Down syndrome. Neurobiol Dis, 58, 92–101. doi:10.1016/j.nbd.2013.04.016

Vitak, S. A., Torkenczy, K. A., Rosenkrantz, J. L., Fields, A. J., Christiansen, L., Wong, M. H., . . . Adey, A. (2017). Sequencing thousands of single-cell genomes with combinatorial indexing. Nat Methods, 14(3), 302-308. doi:10.1038/nmeth.4154

Wisniewski, K. E. (1990). Down syndrome children often have brain with maturation delay, retardation of growth, and cortical dysgenesis. Am J Med Genet Suppl, 7, 274-281.

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS, 16(5), 284-287. doi:10.1089/omi.2011.0118

Yu, G., Wang, L. G., Yan, G. R., & He, Q. Y. (2015). DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. Bioinformatics, 31(4), 608-609. doi:10.1093/bioinformatics/btu684

Zeisel, A., Munoz-Manchado, A. B., Codeluppi, S., Lonnerberg, P., La Manno, G., Jureus, A., . . . Linnarsson, S. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science, 347(6226),

1138-1142. doi:10.1126/science.aaa1934
Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., . . . Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol, 9(9), R137. doi:10.1186/gb-2008-9-9-