EPIGENOMIC DYNAMICS OF CELLULAR DIFFERENTIATION

By

Kelly Ryan Stewart Barnett

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry

June 30, 2020

Nashville, Tennessee

Approved:

Emily C. Hodges, Ph.D.
David K. Cortez, Ph.D.
Scott W. Hiebert, Ph.D.
Manuel Ascano Jr., Ph.D.
John A. Capra, Ph.D.

Dedication

To my parents, Ernest and Rhonda, and their garden where

I learned hard work and love of living things

And

To my Brother, Christopher for not always understanding

But always believing in me

Acknowledgements

A thoughtful professional mentor, caring friend and scientific dreamer are what I found over the past years working with my advisor Emily Hodges. It's not easy to be all that at once and especially not when you are navigating the path for a new lab. I still remember the beginnings of working with Emily during my first year rotation. I was deeply shocked when Emily in "cowboy science" fashion allowed me to construct and submit for sequencing the lab's first ever ATAC-seq library. It was an abysmal failure with quite a price tag. Emily didn't even blink at the failure and was already scheming the beginnings of the work with ATAC-Me that would become the crown jewel of my graduate work, all before any measure of real success. I realize after more time has passed that incidents like these embody what Emily has: "vision". A vision beyond what thicket of experimental weeds I am in at any given moment, a vision of what her lab will and has grown to be. I am forever thankful that Emily allowed me what I consider the honored privilege of being her first graduate student and passed that "vision" for science on to me. We had to navigate the student advisor relationship as a first, it wasn't always easy. Usually due to a frequently stubborn and obsessive graduate student (me) and because we didn't really know what a "normal" student advisor relationship was. Thanks for taking a chance on me Emily. I loved getting to watch your lab grow to what it is now.

Thank you to all my committee members for their steady guidance over the years. Dave and Scott, your lessons in the advanced biochemistry topics course forever shaped my approach to science. Manny, thanks for always indulging "foodie" conversations with the same zeal as our chats about science. And thanks to Manny for a note of advice you once gave, "fail fast", it has really stuck with me and served me well working on method development. Tony, thanks for always being an encouraging voice even after nervous committee meetings.

So many new people have come to the Hodges Lab since I joined. I'm deeply thankful that all those new faces have become my friends not just colleagues. Thanks to Bob Chen for

helping chart the waters in the early days of lab, putting up with a stressed out student colleague, getting our first batch of Tn5 transposase in the freezer, and dreaming up science with me. Thanks to Jonathan for bringing order to our lab with your endless thoughtfulness. Tim, Tyler and Lindsey, thanks for laughing with me about the trials and tribulations of working in the lab whether that meant a gel-fail or an ACCRE fail. Thank you all for being one big Taco Bell loving crew.

A special thanks to the support I have received from the Program in Developmental Biology community and the T32 Stem Cell and Regenerative Developmental Biology training grant support. This was a warmly welcoming community that offered constant encouragement.

I've had so many mentors who took a chance on a me as a young scientist. Chris Raub, Sagar Kathuria, and Jozsef Gal your patience always seemed unending. Thanks for making time for a young scientist and giving me my first experiences in research.

Of course, I wouldn't be where I am without my friends. Thanks for always indulging my love for dinner parties where we could chat about science and vent our frustrations. Wine and cheese can solve any trouble.

To my parents and brother, thanks for always convincing me I really was good enough and that I could do anything I set my mind to. Thanks for teaching me what hard work was from the start.

And last, to Michael, thanks for staying with me all these years. You have always been a steadying force in my life frequently filled with anxiety and frustration brought on by science. We've had so many adventures together while I've been in grad school even going to South Korea. Thanks for always making me keep perspective on what is important. I look forward to many more journeys together.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

4-mC	4-methylcytosine
5-aza-2-dC	5-aza-2'-deoxycytidine
5-mC	5-methylcytosine
5hmC	5-hydroxymethylcytosine
6-mA	6-methyladenine
AP-1	Activator Protein 1 transcription factor
ARR	Allelically Regulated Region
ATAC-Me	Assay for Transposase Accessible Chromatin with Bisulfite Conversion
ATAC-seq	Assay for Transposase Accessible Chromatin sequencing
ATAC-STARR	Assay for Transposase Accessible Chromatin coupled with STARR-seq
CAGE	Capped Analysis Gene Expression
cDNA	Complementary DNA
CEBP	CCAAT-enhancer-binding proteins
CGI	CpG Island
ChIP-BS	Chromatin Immunoprecipitation coupled with Bisulfite sequencing
ChIP-seq	Chromatin Immunoprecipitation sequencing
ChrAcc	Chromatin accessibility
CpG	Cytosine-phosphate-Guanine
CTCF	CCCTC-binding factor
CXCL8	Interleukin-8 gene
DHS	DNase hypersensitive site
DNA	Deoxyribonucleic acid
DNAme	DNA methylation
DNase-I	Dexoyribonuclease-I protein

DNase-seq	DNase I hypersensitive site sequencing
DNMT	DNA (cytosine-5)-methyltransferase protein
DNMT1	DNA (cytosine-5)-methyltransferase 1 protein
DNMT2	DNA (cytosine-5)-methyltransferase 2 protein
DNMT3A	DNA (cytosine-5)-methyltransferase 3A protein
DNMT3B	DNA (cytosine-5)-methyltransferase 3B protein
DNMT3L	DNA (cytosine-5)-methyltransferase 3 Like protein
dNTP	deoxyNucleotide Triphosphate
DTT	Dithiothreitol
EBF1	EBF Transcription factor 1
Egr1	Early growth response protein 1
ERK	Extracellular signal-regulated kinases
FOS	Proto-oncogene c-Fos
FOXP3	Forkhead Box P3 Gene
FOXP3	Forkhead Box P3 protein
GEO	Gene Expression Omnibus database
GpC	Guanine-phosphate-Cytosine
GRE	Gene Regulatory Element
GRO-seq	Global Run-On sequencing
H3K27ac	Histone H3 Lysine 27 acetylation
H3K36me3	Histone H3 Lysine 36 trimethylation
H3K4me1/2	Histone H3 Lysine 4 mono- or di- methylation
Hi-C-seq	High resolution chromatin conformation sequencing
IAP	Intracisternal A Particle
IDAX	CXXC-type zinc finger protein 4 gene
Klf14	Krueppel-like factor 14

L-AA-2-P	L-Ascorbic acid 2-phosphate
LB	Lysogeny Broth
LINE	Long interspersed nuclear element
M.CviPI	GpC Methyltransferase
MACS2	Model-based Analysis of ChIP-Seq
MAP	Mitogen Activated Protein Kinase
MBD	Methyl-CpG-binding Domain
mCpG	5-methylCytosine-phosphate-Guanine
MeCP2	Methyl-CpG-binding protein 2
meDIP-seq	Methylated DNA Immunoprecipitation sequencing
Methyl-dCTP	5-methyldeoxycytosine triphosphate
MNase-seq	Micrococcal Nuclease sequencing
MspI	Restriction endonuclease MspI protein
NANOG	Homeobox transcription factor Nanog
NfKB complex	Nuclear Factor Kappa-light-chain-enhancer of activated B cells protein
NOMe-seq	Nucleosome Occupancy and Methylome sequencing
NRF1	Nuclear respiratory factor 1
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PEI	Polyethyleneimine
PMA	Phorbol 12-myristate 13-acetate
Pro-seq	Precision Run-On sequencing
PU.1	Transcription factor PU.1
RCF	Rotor Conversion Factor
RNA	Ribonucleic acid

RNA Pol II	RNA Polymerase II protein
RNA-seq	Ribonucleic acid sequencing
RPM	Rotations Per Minute
RRBS-seq	Reduced Representation Bisulfite sequencing
scNMT-seq	Single cell Nucleosome Methylation and Transcription sequencing
SELEX	Systematic evolution of ligands by exponential enrichment
SETD2	SET Domain Containing 2, Histone Lysine Methyltransferase protein
SINE	Short interspersed nuclear element
STARR-seq	Self-Transcribing Active Regulatory Region sequencing
STELLA	Developmental Pluripotency-Associated Protein 3
T-WGBS	Tn5 transposase based Whole Genome Bisulfite
TDG	Thymine DNA Glycosylase protein
TET	Ten-eleven translocation methylcytosine dioxygenase protein
TET1	Ten-eleven translocation methylcytosine dioxygenase 1 protein
TET2	Ten-eleven translocation methylcytosine dioxygenase 2 protein
TET3	Ten-eleven translocation methylcytosine dioxygenase 3 protein
TF	Transcription Factor
TRDMT1	tRNA (cytosine-5-)-methyltransferase 1 protein (DNMT2)
UCSC	University of California Santa Cruz
UHRF1	E3 ubiquitin-protein ligase UHRF1
WGBS	Whole Genome Bisulfite Sequencing

Chapter 1: Introduction

Epigenetics

The DNA (Deoxyribonucleic acid) sequence within each somatic cell of the human body is with rare exception identical. Yet, a vast multitude of functional cell types and states exist throughout our bodies. These different cell types and states are made possible through the differential suppression and activation of various segments of regulatory DNA sequence across the genome such as promoters, enhancers and insulators. These regulatory DNA sequences in turn modulate the activity at genes which are crucial DNA sequences encoding functional molecules within the cell, proteins and RNA. The differential utilization of gene regulatory together with gene activity mediates a wide spectrum of cells in terms of presence, absence and abundance of molecules such as enzymes, structural proteins, and regulatory molecules. Yet, the DNA sequence among cells in a human body are identical, so, what mediates this differential utilization of segments of gene regulatory DNA? How does a neuron and a fibroblast establish such different cellular functions with the same DNA sequence content?

The field of epigenetics aims to address these questions by exploring how additional information encoded upon DNA beyond the base pair sequence instructs the program of life. Epigenetics as a term can be traced back to its original coining in the field of developmental biology by Conrad Waddington in the 1940's (Waddington, 2012). However, even prior to Waddington's more modern usage the similar term "epigenesis" was utilized more vaguely as a competing theory to "preformationism". "Epigenesis" arguing that organismal development was directed by chemical reactions within cells while "preformationism" argued that components of the fully developed organism already existed within cells and simply grew larger as development proceeded. With the discovery that DNA is the encoding material of genetic information during organismal development there became a need to distinguish inherited changes due to DNA

sequence from those that were not explained by DNA sequence thus leading to Waddington's usage of the term "epigenetics" (Waddington, 1953). Ever since, epigenetics has undergone continual evolution across the past several decades to encompass many competing definitions with the common theme that the DNA associated modifier must be mitotically/meiotically inheritable (Berger et al., 2009; Russo, 1996). Epigenetics has grown to include a broad realm of modifiers to the base pair code of DNA, frequently with debate on fit into the formal definition (Berger et al., 2009). Such DNA modifiers include DNA methylation, chromatin accessibility, histone modifications, histone variants, TFs (Transcription factors) and more. Evidence for mitotic inheritance of epigenetic information has been suggested for many of these DNA modifiers including histone variants (Nekrasov et al., 2012) and histone modifications (Alabert et al., 2015; Ragunathan et al., 2015). While DNA methylation fits the more formal requirement of a true epigenetic mark the core functional importance of epigenetic marks lies in the additional information encoded beyond DNA sequence. However, only in the case of DNA methylation have mechanisms for base pair level maintenance, remodeling and inheritance for this epigenetic mark been well described.

DNA methylation is likely the most well studied epigenetic mark of the past several decades, primarily because of its functional implication in proper organismal development and cellular differentiation. Many studies in mice and human models have noted that genetic defects in proteins that shape DNA methylation landscapes yield prominent phenotypes such as embryonic lethality, impaired cellular differentiation and hyperproliferation of cells. Collectively this suggests that DNA methylation is crucial for proper cellular and organismal function. However, the direct mechanism that DNA methylation acts through are not always clear. In many cases the mode of action may be different depending on the genomic location of DNA methylation. The following sections explore the genomic patterns of DNA methylation, the molecules that establish this pattern, and functional roles throughout the genome in an effort to better appreciate the importance of DNA methylation.

DNA Methylation Landscape

Methylation is a covalent DNA modification that exists in many forms across the genomes of prokaryotes and eukaryotes such as 5-mC (5-methylcytosine), 4-mC (4-methylcytosine), and 6-mA (6-methyladenine) (Sanchez-Romero et al., 2015). Among eukaryotes, specifically mammals, 5-mC is by far the most prevalent covalent DNA modification present within the genome occurring almost exclusively in the CpG (Cytosine-phosphate-Guanine) dinucleotide context (Zemach et al., 2010). In a typical mammalian genome, methylated CpGs are the default state with 70- 80% of all CpG sites modified in non-tumorigenic settings (Li and Zhang, 2014). The dinucleotide frequency of CpGs is far below expected when compared to other dinucleotides in the human genome. The dinucleotide frequency discrepancy is driven by the mutagenic nature of 5-mC which is prone to spontaneous deamination yielding cytosine to thymine base transitions (Bird and Taggart, 1980; Cooper and Krawczak, 1989; Holliday and Grigg, 1993). In contrast, CpG frequency is near expected values among the genomes of species lacking DNA methylation. While the majority of CpGs in the human genome are methylated, these CpGs and their associated DNA methylation are not evenly distributed across the genome. The DNA methylation of these CpGs is strikingly interrupted across the genome by distinct DNA regions absent of DNA methylation coinciding with CGIs (CpG Islands), promoters, enhancers and generally locations of TF interaction (Meissner et al., 2008). The following is a survey of DNA methylation enzymes, genomic patterns, and dynamics in various contexts.

DNA Methylation Mechanisms

The mechanistic role of DNA methylation in gene regulation has been debated for many years, largely because of conflicting characterizations in the literature. It has been difficult to converge on a singular mechanistic theme for the diverse contexts at which DNA methylation is found. DNA methylation is thought to impact cellular function through modulation of TF factor binding, chromatin structure and transposable element suppression. The following briefly explores these different mechanistic roles of DNA methylation.

Disruption of TF binding can occur through multiple means both direct and indirect. Early *in-vitro* work testing the transcriptional output from methylated and unmethylated promoter constructs in Xenopus and mammalian cell extracts found evidence for an indirect block to transcription (Kass et al., 1997; Keshet et al., 1986). Initially after incubation with cell extracts the transcriptional output of methylated and unmethylated constructs are identical (Kass et al., 1997). Later however, it was found that the methylated construct recruits methyl-CpG binding proteins such as MBD (Methyl-CpG-binding Domain) proteins (Ng et al., 2000; Ng et al., 1999) and MeCP2 (Methyl-CpG-binding protein 2) (Nan et al., 1997). MBD and MeCP2 in turn recruit co-repressor complexes frequently containing HDACs which subsequently induces a compacted chromatin state not permissive to transcription (Jones et al., 1998; Nan et al., 1998). The compacted chromatin state consists of a dense nucleosome arrangement decorated with repressive histone marks that excludes TFs from binding. The end result is a block in TF binding but not mediated directly by the methyl mark itself on DNA.

In another well-known scenario, a direct block to binding due to DNA methylation has also been described for several TFs, most notably CTCF (CCCTC-binding factor) (Bell and Felsenfeld, 2000; Wang et al., 2012a) and NRF1 (Nuclear respiratory factor 1) (Domcke et al., 2015). Even this appears to be context dependent as only a subset of CTCF binding sites were perturbed by DNA methylation. Recent *in-vitro* SELEX (Systematic evolution of ligands by exponential enrichment) studies have characterized the sensitivity of TF binding to DNA

methylation revealing so called classes of “methyl-minus” (TF binding impeded by DNA methylation) and “methyl-plus” (TF binding enhanced by DNA methylation) TFs (Kribelbauer et al., 2017; Yin et al., 2017). However, many TFs lack a CpG in their consensus binding motif or the CpG is not within a critical DNA interface drawing into question if this mechanistic model is the exception or the norm.

Yet another proposed functional mechanism for the existence of DNA methylation within the human genome is the suppression of parasitic DNA sequences known as transposons. Transposons behave such that the DNA sequence is capable of propagating, copying and inserting itself throughout the genome possibly forming deleterious mutations. Such sequences make up greater than 40% of the human genome in the form of LINE (Long interspersed nuclear element), SINE (Short interspersed nuclear element), and retrotransposon sequences. Experiments which perturb levels of DNA methylation genome wide have been shown to see a reactivation of some of these viral genomic parasites. In particular a DNMT1 (DNA (cytosine-5)-methyltransferase 1 protein) knockout mouse model displayed a strong expression of IAP (Intracisternal A Particle) endogenous retroviruses (Walsh et al., 1998). A failure to suppress transposons with DNA methylation has also been connected with a sterility phenotype in mice (Barau et al., 2016). Collectively it is clear that DNA methylation is a crucial means to control parasitic DNA elements.

DNA Methylation Enzymes

The human genome contains 5 members of the DNMT (DNA (cytosine-5)-methyltransferase protein) family defined by sequence and protein structure homology (Bestor et al., 1988). Family members include DNMT1, DNMT2 (DNA (cytosine-5)-methyltransferase 2 protein) presently known as TRDMT1 (tRNA (cytosine-5-)-methyltransferase 1 protein), DNMT3A (DNA (cytosine-5)-methyltransferase 3A protein), DNMT3B (DNA (cytosine-5)-

methyltransferase 3B protein) and DNMTL (DNA (cytosine-5)-methyltransferase 3 Like protein). Only DNMT1, DNMT3A and DNMT3B are known to be catalytically active on DNA *in-vivo*. DNMT2 has only even been demonstrated to have catalytic activity on tRNA though there is some debate and interest on other possible substrates (Goll et al., 2006; Kaiser et al., 2017; Legrand et al., 2017). DNMTL has no catalytic methyltransferase activity on its own, however, it enhances the enzymatic activity of DNMT3 enzymes by forming a tetrameric structure which increases the affinity for DNA (Bourc'his et al., 2001; Jia et al., 2007). DNMT1 is primarily known as the maintenance methyltransferase due to its presence at the replication fork where DNA methylation patterns are copied onto the newly synthesized strand of DNA. As the new strand of DNA emerges from the replication fork it lacks DNA methylation at the symmetric CpG forming a hemi-methylated DNA substrate. Replication fork localization and hemi-methylated substrate preference (Hermann et al., 2004) are possible due to the interaction with the protein UHRF1 (Bostick et al., 2007). In contrast to the maintenance methyltransferase activity of DNMT1, DNMT3A and DNMT3B are known as the de novo methyltransferases due to their less selective activity on DNA and lack of preference for hemi-methylated DNA substrate (Okano et al., 1999; Okano et al., 1998). Structurally DNMT1 contains many more protein-protein interaction and regulatory domains than DNMT3A and DNMT3B some of which are of unknown functionality. While an abundance of data supports the canonical roles of DNMT1 as a maintenance methyltransferase and DNMT3 as a de-novo methyltransferase, this appears to be an oversimplification.

More recent studies have observed experimental outcomes and DNMT properties that are not consistent with perfectly segregated maintenance and de-novo methyltransferase roles. DNMT3 enzymes have been shown to have strong flanking sequence preference drawing into question the ability of these enzymes to act in broad contexts (Handa and Jeltsch, 2005; Lin et al., 2002). Further, DNMT3 enzymes engage with only one CpG in the double stranded symmetric context during DNA binding and methyltransferase activity ultimately yielding a hemi-

methylated DNA product (Jia et al., 2007). Hemi-methylated DNA is the preferred substrate for DNMT1, thus, it is thought that in actuality the combined actions of DNMT3 and DNMT1 enzymes are required for de-novo methylation events (Fatemi et al., 2002). Evidence also exists for DNMT1 occupying de-novo methylation roles. Studies have shown DNMT1 activity on unmethylated DNA templates *in-vitro* (Fatemi et al., 2001; Vilkaitis et al., 2005) and detection of newly established DNA methylation on an introduced template in DNMT3 double knockout cells (Lorincz et al., 2002). These results suggest that DNMT1 has a more promiscuous interaction with DNA than previously appreciated. Conversely, DNA methylation loss has been observed in many DNMT3 knockout models where one would expect preservation of DNA methylation profile due to DNMT1's intact maintenance activity (Chen et al., 2003; Dodge et al., 2005; Liang et al., 2002). While the major roles of DNMT1 and DNMT3 are maintenance and de-novo respectively, it is clear that there is far more cooperativity in these roles than previously described.

DNA Demethylation Enzymes

There are two main pathways to DNA demethylation involving active and passive mechanisms with frequent intersection between these two modes. Initially, only passive mechanisms for DNA demethylation were described involving inefficient maintenance methylation by DNMT1. Inhibition of DNMT1 activity is known to be achieved *in-vivo* by repression of DNMT1 gene transcription, nuclear exclusion of DNMT1 protein or by non-preferred DNA substrates. Passive demethylation occurs across multiple rounds of cell division accompanied by DNA replication which dilute the DNA methylation eventually leading to a fully hypomethylated symmetric CpG. Evidence points towards passive DNA demethylation as the major mechanism for demethylation in most biological contexts.

Active demethylation mechanisms remained elusive in mammals until the last decade. However, direct removal of 5-mC in plants was already identified (Agius et al., 2006; Morales-Ruiz et al., 2006). It was not until the discovery of the TET (Ten-eleven translocation methylcytosine dioxygenase protein) family of enzymes that DNA demethylation pathways became more clear (Tahiliani et al., 2009). The TET1 (Ten-eleven translocation methylcytosine dioxygenase 1 protein), TET2 (Ten-eleven translocation methylcytosine dioxygenase 2 protein) and TET3 (Ten-eleven translocation methylcytosine dioxygenase 3 protein) proteins are DNA dioxygenase enzymes that utilize iron and 2-oxoglutarate to iteratively oxidize 5-mC (Ito et al., 2010; Tahiliani et al., 2009). All three TET proteins are similar in their catalytic domain; however, only TET1 and TET3 possess a CXXC binding domain to mediate interaction with CpG cites (Xu et al., 2012; Zhang et al., 2010). The CXXC domain of TET2 exists as a separate gene called *IDAX* (CXXC-type zinc finger protein 4 gene) formed from a genomic inversion during evolution which interacts with TET2 and is thought to guide chromatin interactions (Ko et al., 2013). Overall these proteins are not known to have distinct binding preference in terms of DNA sequence other than CpG sites (Williams et al., 2011; Wu et al., 2011). TET oxidative products, 5-hmC (5-hydroxymethylcytosine) primarily, are subsequently capable of participating in both active and passive mechanisms of DNA demethylation.

In the active role 5-hmC is further oxidized by TET enzymes to 5fmC and 5caC which are recognized by base excision repair enzymes such as TDG (Thymine DNA Glycosylase) (He et al., 2011). Further processing of the resultant abasic site formed by TDG eventually yields a repaired unmodified cytosine (Maiti and Drohat, 2011; Weber et al., 2016). The full processing of the cytosine to its unmodified form involves formation of single stranded breaks and possibly double stranded breaks if both symmetrical CpGs are processed in quick succession (Weber et al., 2016). Work has demonstrated that this double stranded break scenario is quite rare, likely due to the rapid sequential enzyme activity mediated by co-localization of TET and TDG (Weber et al., 2016).

In the passive role, 5-hmC and further oxidized forms are DNA substrates inefficiently recognized by DNMT1 at the replication fork leading to disrupted maintenance methylation (Hashimoto et al., 2012; Ji et al., 2014; Otani et al., 2013). DNA replication is required for this mode of demethylation. Current evidence indicates that this mode of methylation is by far the most common making up the bulk of demethylation events (Jin et al., 2015). In particular this mode of demethylation is observed to be important in many systems where proliferating cells undergo differentiation (discussed in more detail in the following section).

Many recent works have described targeting of TET enzymes to specific areas of the genome particularly enhancers. TF interactions with TET are thought to act as a genomic guide bringing TET to targeted gene subsets and DNA sequences. Many TFs guiding TET to gene regulatory loci have been identified in the literature such as PU.1 (Transcription factor PU.1), NANOG (Homeobox transcription factor Nanog) and EBF1 (EBF Transcription factor 1) (Costa et al., 2013; Guilhamon et al., 2013; Lio et al., 2016). The pioneer factor and pluripotency classes of TFs in particular have been demonstrated as guides for TET localization throughout the genome suggesting possible cooperativity to relax condensed chromatin and mediated DNA methylation changes (Sardina et al., 2018). The interaction of TFs and TET underscores the importance of TET during cell differentiation that requires DNA methylome remodeling.

Promoter Methylation

DNA methylation at promoters has been the central focus of study within mammalian genomes for the past several decades. Primarily because of the well described coincidence of CGIs at most promoters and nearly all house-keeping gene promoters (Gardiner-Garden and Frommer, 1987; Larsen et al., 1992). High CpG densities at promoters have been attributed to both the constant engagement with DNA binding proteins offering protection from DNMTs and

the exceptionally low DNA methylation levels in male primordial germ cells; both avoiding evolutionary erosion due to 5-mC deamination (Molaro et al., 2011).

When a CGI promoter is methylated the genomic locus transforms into a state no longer permissive for transcription. *In-vitro* and cellular data largely indicates that the presence of DNA methylation is not inhibitory directly but rather the resultant inhibitory chromatin environment (Kass et al., 1997; Keshet et al., 1986). Dynamic methylation states at promoters can be observed across human development but this appears to be a rare and very specific case. The methylation state of CGIs and promoters in general is consistently hypomethylated (Bird et al., 1985). Promoter methylation mainly functions early in development to suppress germ cell specific (Borgel et al., 2010) and pluripotency related genetic programs (Farthing et al., 2008). Another well described incidence of promoter methylation occurs across almost all promoters of the inactivated X-chromosome during female development. However, well before DNA methylation is established on the inactivated X-chromosome most transcription has already been halted supporting that DNA methylation is more of a final transcriptional lock (Grant et al., 1992; Lock et al., 1987; Singer-Sam et al., 1990). It is far more common that when CGI associated genes are transcriptionally repressed the methylation state typically remains the same (hypomethylated) with repression mediated primarily by the repressive histone mark H3K27me3 (Riising et al., 2014). This repressed yet hypomethylated state is thought to offer more rapid reactivation of transcription at these promoter regions without requiring demethylation. Overall promoter methylation seems to be a specific case of gene regulation rather than a common mode of operation.

Gene Body Methylation

While DNA methylation at promoters is associated with transcriptional repression, somewhat paradoxically, DNA methylation in gene bodies is strongly associated with

transcriptional activity (Hellman and Chess, 2007; Lister et al., 2009). Due to the mutagenic nature of DNA methylation this would seem counter intuitive to preservation of protein coding genes, yet, this gene body methylation is one of the most highly conserved patterns of DNA methylation across eukaryotes (Zemach et al., 2010). Gene body methylation is partially explained by the affinity of DNMT3B via a PWWP domain for the histone mark H3K36me3 (Histone H3 Lysine 36 trimethylation) (Baubec et al., 2015; Morselli et al., 2015). The SETD2 (SET Domain Containing 2, Histone Lysine Methyltransferase protein) complex is known to deposit H3K36me3 during RNA Pol II (RNA Polymerase II protein) transcription elongation thus leaving a substrate to recruit DNMT3B to actively transcribed gene bodies. While this likely explains how DNA methylation is deposited at gene bodies it is debated if this is purely consequential or if a direct functional role exists. A study in a colorectal cancer cell line depleted of DNA methylation with 5-aza-2-dC (5-aza-2'-deoxycytidine) treatment followed by a recovery period showed that gene body methylation is restored more rapidly than promoters and was associated with restored transcription (Yang et al., 2014). Applying this same approach in DNMT3B deficient cells impaired the ability of cells to restore normal transcript levels after withdrawal of 5-aza-2-dC supporting a role in gene body methylation to support robust transcription (Yang et al., 2014). Though in a cancerous cell line context, this data suggests that there is indeed a functional role for DNA methylation at gene bodies.

In regards to function, there are two major theories for DNA methylation at gene bodies: splicing regulation and cryptic promoter inhibition. A role for splicing regulation has been described in the case of CTCF binding blocked by DNA methylation. In this case CTCF binding slows the rate of Pol-II extension leading to inclusion of an exon that without CTCF binding gets spliced out (Shukla et al., 2011). DNA methylation acts as the controlling switch within the gene body facilitating exon inclusion/exclusion. The theory of intragenic cryptic promoter inhibition has been studied via knock-out models of DNMT3B due to its described recruitment to gene bodies. In DNMT3B knock-out cells there was an increase in non-first exon usage, however, the

data indicates this was a rare occurrence which draws into doubt the importance of this mechanism (Neri et al., 2017). Further confounding these results, the phenomenon could not be recapitulated with RNA-seq (Ribonucleic acid sequencing) analysis of a triple DNMT knock-out cell line (DNMT1, DNMT3A, DNMT3B) (Teissandier and Bourc'his, 2017). While gene body methylation is species conserved, it doesn't seem essential for cellular function, rather, it may be a finer level of gene regulation or simply reflect the default methylation state of most non-genic CpGs. Further study is needed to explore the importance of gene body methylation given its broad evolutionary conservation across divergent species.

Enhancer Methylation

Enhancers are DNA sequences which increase the transcription of genes in spatial proximity or in the case of repressors decrease transcription. Enhancers in terms of linear sequence can act on their target genes from kilobases up to megabases away, mediated primarily by the 3-dimensional folding of chromatin DNA thus bringing distal sequences into close contact. Akin to promoters, enhancers are rich with TF binding sites and frequently lack DNA methylation at active loci. However, the CpG content is typically much lower at enhancer loci throughout the human genome. It is questionable if this low CpG density and concomitant DNA methylation is sufficient to have a measurable impact on the function of these locations as many well-known methylated-CpG binding proteins require dense CpG content for efficient binding. Even so, there is a growing body of evidence suggesting that enhancer methylation can have functional outcomes.

In contrast to promoter methylation states, there is considerably more variation at intergenic/intragenic enhancers across cell types and development (Figure 1, (Song et al., 2013)). On average DNA methylation at active enhancers possess lower methylation levels than at inactive loci. Several studies have observed that distinct regions of hypomethylation at

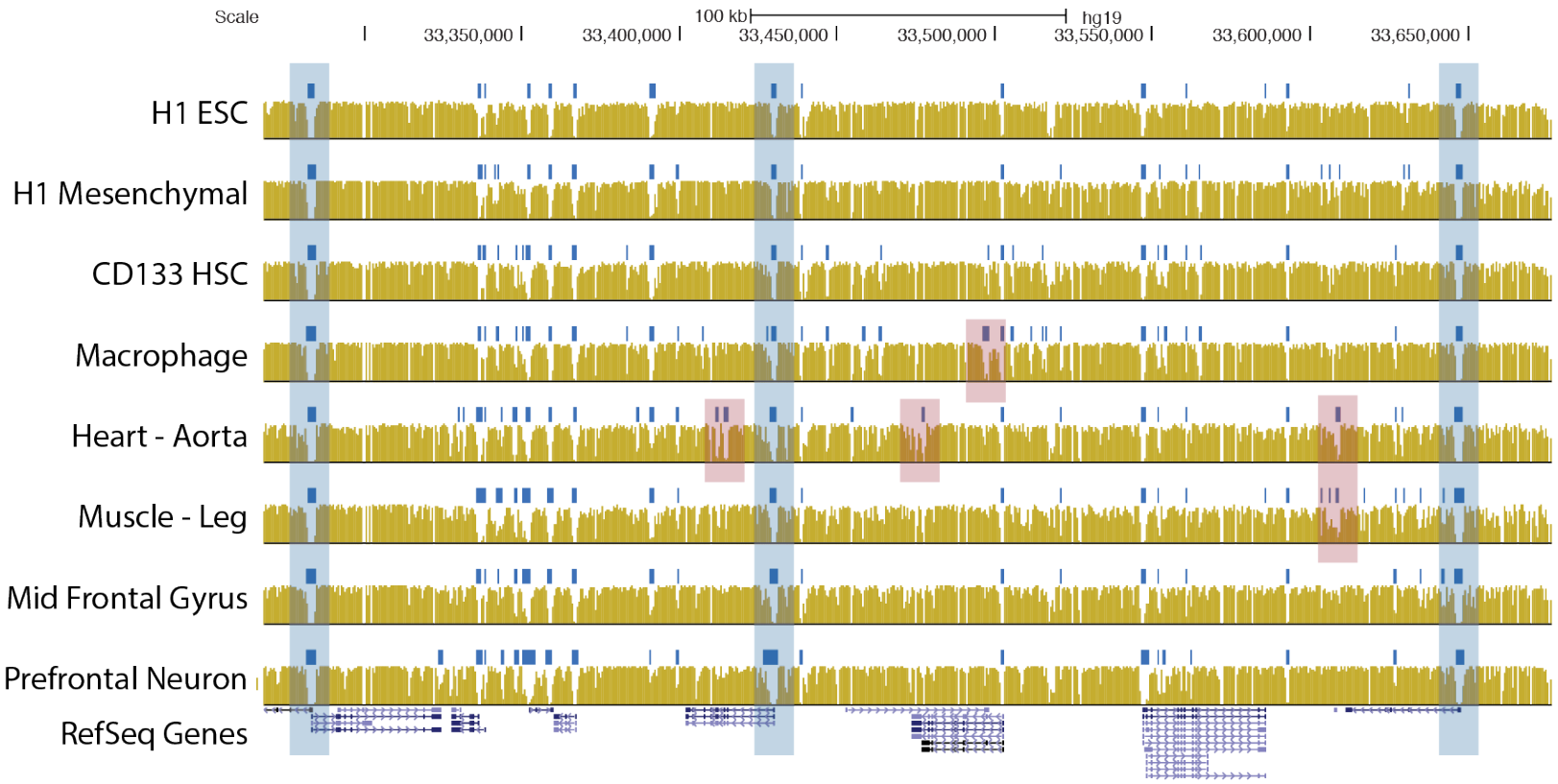

Figure 1. Genome browser snapshot of the DNA methylation landscape in humans. UCSC (University of California Santa Cruz) genome browser tracks of whole genome bisulfite data from various cell types and tissues. Vertical gold lines indicate the frequency of DNA methylation at CpG sites across a population of cells. Dark blue rectangles above vertical gold lines indicate a genomic hypomethylated region. Transparent blue rectangles indicate promoter regions with constitutive hypomethylation. Transparent red rectangles indicate intergenic or intragenic hypomethylation with cell type or tissue specificity. Lower track denotes a map of gene bodies as defined by RefSeq across the DNA locus. Locus coordinates are defined from hg19 human genome map. chr1:33,272,676-33,681,328

enhancer loci become newly established or expand/contract across a continuum of cell development stages (Hodges et al., 2011; Schlesinger et al., 2013). The binding of TFs in many cases has also been associated with hypomethylation at formerly methylated enhancer loci (Feldmann et al., 2013; Stadler et al., 2011; Wiench et al., 2011). Sustained TF binding is thought to protect enhancer regions from becoming remethylated by DNMTs and further stabilize the hypomethylated state (Brandeis et al., 1994). It is debated if DNA methylation dynamics at enhancers orchestrate TF binding or if methylation dynamics are a downstream consequence. The recent employment of targeted epigenome editing has allowed a more focused dissection of the importance DNA methylation has at enhancer loci. Targeted dCas9-Tet1 demethylation of the MyoD enhancer locus resulted in a phenotypic switch from fibroblast to myoblast (Liu et al., 2016). Collectively these and other studies indicate a role for methylation at enhancers but the non-trivial task of disentangling effect versus response at these loci merits further study.

Dynamics of DNA Methylation Across Development

Human development starts with multiple dramatic waves of DNA methylation changes. A multitude of knockout models support that properly orchestrated DNA methylation dynamics are crucial for organismal development. The following is an overview of DNA methylation dynamics in varying contexts across human development.

One of the most distinct and pronounced instances of DNA methylation dynamics occurs during primordial germ cell development which give rise to sperm and oocytes. Primordial germ cells initially have a DNA methylation profile very similar to other somatic cells. During maturation primordial germ cells undergo a combined migration and proliferation event that coincides with a two-stage demethylation process (Guibert et al., 2012; Seisenberger et al., 2012). Initially passive global demethylation dominates, mediated by exclusion of DNMT1 from

the nucleus due to a lack of replication fork anchoring by UHRF1 (E3 ubiquitin-protein ligase UHRF1) (Kagiwada et al., 2013). Later, TET1 expression is upregulated and facilitates more focal demethylation at parental specific genomic imprints (Hackett et al., 2013; Yamaguchi et al., 2013). After these two stages of demethylation primordial germ cells reach one of the most distinctly hypomethylated states known with average methylation levels across the population <10% (Gkountela et al., 2015; Guo et al., 2015; Wang et al., 2014). Both male and female gamete genomes later undergo remethylation mediated by DNMT3A (Hata et al., 2002; Kaneda et al., 2004), however, the extent of remethylation is quite different. Sperm genomes are typically ~70% methylated with only CGIs escaping this broad genomic remethylation (Molaro et al., 2011; Wang et al., 2014). In contrast, oocyte genomes only reach methylation levels of around 40% with most methylation existing in active gene bodies (Kobayashi et al., 2012; Smallwood et al., 2011). Parental specific DNA methylation imprints are established during this remethylation event. The lower level of methylation in oocytes has been attributed to the STELLA (Developmental Pluripotency-Associated Protein 3) mediated exclusion of UHRF1 and thus secondarily DNMT1 exclusion from the nucleus (Li et al., 2018c). STELLA is a maternal protein factor shown to have multiple roles in orchestrating proper gamete and zygote development (Han et al., 2019; Li et al., 2018c; Nakamura et al., 2007).

Primordial germ cell development represents the first major round of methylation dynamics in human development. The second major wave of methylation dynamics begins immediately after fertilization at the zygote stage. After fertilization both gamete genomes exist as pro-nuclei in physically separate parts of the single cell zygote forming two distinct methylation reprogramming environments. Immediately after fertilization the male pro-nucleus begins to undergo TET3 mediated demethylation detected initially by an accumulation of 5-hmC (Iqbal et al., 2011; Wossidlo et al., 2011). However, this demethylation is primarily mediated through passive DNA replication dependent means rather than direct active removal by TDG (Inoue et al., 2011; Inoue and Zhang, 2011). The maternal genome largely escapes this TET3

mediated demethylation due to the action of the STELLA protein carried by the oocyte which has been shown to exclude TET3 from the maternal pro-nucleus (Nakamura et al., 2007; Nakamura et al., 2012). The maternal genome undergoes primarily non-TET3 dependent passive demethylation (Cardoso and Leonhardt, 1999; Howell et al., 2001). Imprinted regions of the genome are spared from this wave of global demethylation during early development. Later, after implantation of the epiblast the developing embryo undergoes a wave of remethylation by way of the de novo DNA methyltransferases DNMT3B and DNMT3A (Auclair et al., 2014). These stages of demethylation and methylation may not be entirely segregated as evidence indicates that de novo methylation, primarily at transposons, is occurring during the global demethylation of the zygote (Amouroux et al., 2016). It is still unclear what orchestrates and specifies the DNA methylation patterns during this period of remethylation, many theorize this process is constrained by chromatin structure.

DNA methylation dynamics are not just restricted to embryonic development. A multitude of studies have observed DNA methylation differences across the continuum of cell types and state that occur as cell differentiate. Probably the most well-known is the remethylation of pluripotency related gene promoters and regulatory regions as stem cells commit to increasingly specific cellular phenotypes (Athanasiadou et al., 2010; Farthing et al., 2008; Gu et al., 2009). Conversely, differentiated somatic cells that are induced to pluripotency undergo demethylation at pluripotency loci, this event is crucial to stabilize the newly acquired stem cell identity (Koche et al., 2011; Mikkelsen et al., 2008; Polo et al., 2012). TET deficient stem cells demonstrate a distinct blockade to differentiation (Dawlaty et al., 2014). In contrast, somatic cells overexpressing TET enzymes show greatly enhanced frequency of pluripotency cell formation (Doege et al., 2012; Gao et al., 2013) underscoring the importance that DNA methylation dynamics play in regulation of cellular differentiation.

Methylation dynamics are not restricted to pluripotent cells. Studies examining distinct lineage intermediates within multipotent differentiation systems have long noted evidence for

continual DNA methylation dynamics beyond embryonic development. In particular, de-novo establishment and expansion of hypomethylated regions are prominent when comparing whole genome methylation profiles across the hematopoietic lineage (Hodges et al., 2011). Loci of newly formed DNA hypomethylation are strongly associated with functional enhancer marks such as H3K27ac (Histone H3 Lysine 27 acetylation), chromatin accessibility and intergenic transcription (Schlesinger et al., 2013). Similarly, DNA methylation was linked to intestinal stem cell expansion upon DNMT1 loss due to perturbed DNA methylation dynamics at enhancer loci (Sheaffer et al., 2014). The proper activation and differentiation of T-regulatory cells has also been shown to depend on the methylation status of enhancers within the *FOXP3* (Forkhead Box P3 Gene) gene locus (Floess et al., 2007). Naïve T-cells are fully methylated at these *FOXP3* locus enhancers and subsequently demethylated in T-regulatory cells (Floess et al., 2007). This switch in DNA methylation is a major factor in stabilizing FOXP3 (Forkhead Box P3 protein) expression, a TF important for T-regulatory cell identity (Floess et al., 2007; Yue et al., 2016).

DNA methylation dynamics occur across human development from embryogenesis to adult cellular maintenance. How and to what extent DNA methylation dynamics are crucial to these cell fate decisions is frequently unclear but may be related to TF binding, CpG density and nucleosome occupancy. Many more focused studies of differentiation are needed to fully understand the importance of DNA methylation in this context.

Defects of DNA Methylation

A multitude of model systems defective for DNA methylation related enzymes exists both in the lab and nature. Some of the earliest understandings of DNMTs comes from knockout models in mice. Knockout of either DNMT1, DNMT3A or DNMT3B in mouse models all lead to lethality early in development (Lei et al., 1996; Li et al., 1992; Okano et al., 1999). Peculiarly, even DNMT3L knockout mouse models exhibit a disease phenotype in spite of DNMT3L's

absent methyltransferase activity. DNMT3L knockout mice exhibit sterility in males and females fail to carry embryos to full term (Bourc'his and Bestor, 2004). In contrast to DNMT knockout mice, individual TET (TET1 or TET2) knockout mice display less dramatic phenotypes likely due to the functional redundancy between the three TET enzymes (Dawlaty et al., 2011; Ko et al., 2011). TET3 knockout mice also develop seemingly normal but in contrast to TET1/2 knockouts die perinatally (Gu et al., 2011). TET double (TET1, TET2) knockouts show some developmental lethality but adult mice that survive appear mostly normal other than focal epigenetic abnormalities(Dawlaty et al., 2013). TET triple knockouts (TET1, TET2, TET3) are completely embryonic lethal with isolated mouse embryonic stem cells unable to contribute to chimeric embryos (Dawlaty et al., 2014). The pronounced impact of these knockouts underscores the importance of proper orchestration of DNA methylation across the genome for mammalian development.

Considering the dramatic phenotypes resultant of disturbed DNA methylation during development one might assume the methylation changes in cancer are crucial for pathogenesis. Yet in reality it is unclear if DNA methylation is causative or correlative with cancer phenotypes. Cancer phenotypes typically present with a dramatic global hypomethylation of the genome accompanied by focal hypermethylation (Feinberg and Vogelstein, 1983; Gama-Sosa et al., 1983; Paz et al., 2003). Regions of global hypomethylation coincide mostly with intergenic portions of the genome frequently overlapping transposable elements and occasionally promoters of oncogenes (Gaudet et al., 2003). Hypermethylation frequently occurs at the promoters of tumor suppressor genes and cell cycle regulators with the assumption that this interferes with transcription at these genes (Paz et al., 2003). However, there is some question if the DNA methylation at these tumor suppressor genes is directly functional or a downstream consequence of chromatin changes in general. In most cases it has been observed that the genes that become hypermethylated in cancer phenotypes were already silenced, primarily by action of the polycomb complex (Sproul et al., 2011). Cancer associated DNA methylation

changes can also be detected within normal and pre-cancerous cells raising the question if these changes are indicative of increased proliferation capabilities or directly prime the cells for later pathogenic phenotypes (Maegawa et al., 2010; Raddatz et al., 2013; Sun et al., 2014).

Much of what we know about the role DNA methylation and DNMTs play in cancer is from studies utilizing the hypomethylating and DNA damaging drugs 5-aza-2-dC and decitabine. These drugs are clinically approved for the treatment of myelodysplastic syndrome (Kantarjian et al., 2006; Silverman et al., 2002). As a well-known DNA hypomethylating agent, naturally it has been thought to act on cancer through a DNA methylation mechanism. Yet, more and more studies indicate that the DNA hypomethylating action of 5-aza-2-dC and decitabine is more of a side effect than the driver of the clinical outcomes (Palii et al., 2008; Stresemann and Lyko, 2008). 5-aza-2-dC was originally created as a DNA damaging agent (Sorm et al., 1964). Both 5-aza-2-dC and decitabine act by incorporating into DNA during replication and when acted upon by DNMT1 leads to a covalent linkage of DNMT1's active site to DNA (Schermelleh et al., 2005). Effectively, these two drugs lead to formation of proteinaceous lesions on DNA which eventually lead to a hypomethylated state. It is difficult to assess the impact of DNA methylation on cancer formation and progression without disentangling the DNA damage and hypomethylation effects of 5-aza-2-dC and decitabine, more precision tools are needed.

DNA methylation related enzyme mutations are common amongst a wide group of cancers, particularly hematological cancers. Mutations in the methyltransferase domain of DNMT3A are found in around 30% of acute myeloid leukemia cases (Ley and Grant, 2011; Yan et al., 2011). Most typically the methyltransferase domain mutation either abolishes or severely impairs enzymatic activity of DNMT3A (Russler-Germain et al., 2014). These DNMT3A mutations are not direct driver mutations and frequently exist among hematopoietic stem cells with enhanced self-renewal capabilities before a true leukemic phenotype exists (Chan and Majeti, 2013). It is not until the co-occurrence with another mutation that a more distinct cancerous phenotype arises. In contradiction to the contribution DNMT mutations pay in forming

cancer phenotypes TET enzymes are also frequently mutated in cancers, particularly TET2. Impaired hmC levels throughout the genome and hypermethylation at enhancers caused by TET2 deficiency have been found to drive proliferative phenotypes in hematological cancers (Cimmino et al., 2017; Langemeijer et al., 2009). Mutations in several metabolic enzymes have also been connected with an indirect impairment of TET function. Isocitrate dehydrogenase under normal circumstances produces 2-oxoglutarate, a key co-substrate needed to oxidize 5-mC. Gain of function mutations in isocitrate dehydrogenate in some cancers render it capable of produce 2-hydroxyglutarate which inhibits the catalytic activity of TET enzymes thus further promoting gains of 5-mC (Xu et al., 2011). While it is clear that DNA methylation and demethylation enzymes are a hallmark of cancer we still do not fully comprehend the true impact of DNA methylation defects on cancer formation.

A common theme among examples of DNA methylation defects is a barrier to differentiation. Particularly in the context of hematological cancer which is known to be the result of a hyperproliferative progenitor cell phenotype. These progenitor cells have been observed as unable to proceed through normal differentiation instead favoring the self-renewal progenitor phenotype. Embryonic stem cells deficient for demethylation enzymes exhibit skewed and defective differentiation capabilities (Dawlaty et al., 2014). In contrast, restoration of DNA methylation machinery can allow for normal differentiation to proceed again (Cimmino et al., 2017). Collectively these studies suggest that epigenomic remodeling is necessary for proper cellular differentiation. However, exactly which epigenomic components and necessary and sufficient is not known.

Methods to Measure Chromatin Accessibility and DNA Methylation

Methods to study DNA methylation and chromatin accessibility extend across decades well before the high throughput genomics era. DNA methylation has been an epigenetic mark of

interest since its discovery for many decades. Prior tools utilized more traditional sanger sequencing (Frommer et al., 1992) and methylation sensitive restriction (Bestor et al., 1984; Bird and Southern, 1978; Cedar et al., 1979) endonucleases to determine the DNA methylation state of target sequences. Chromatin accessibility has long been noted to vary in the human genome from evidence gathered using nuclease digests of DNA coupled with agarose gel electrophoresis analysis (Hewish and Burgoyne, 1973). Scientists today have the luxury of assessing a vast array of epigenomic marks across the entire human genome, primarily made possible by high throughput sequencing technologies. The following is a comparison of current widely used methodologies for measuring DNA methylation and chromatin accessibility using high throughput genomics techniques.

Due to the expense of genomic sequencing and intensive data analysis many have developed enrichment procedures for a focused sampling of genomic DNA methylation. One of the most widely used methods for DNA enrichment, meDIP-seq (Methylated DNA Immunoprecipitation sequencing), involves antibody based DNA immunoprecipitation of methylated DNA fragments (Weber et al., 2005). The benefits of this methodology are the low cost and ease of protocol. Due to the antibody based enrichment the genomic resolution is relatively low because of DNA fragment size limitations. Additionally, meDIP-seq produces data inherently biased for methylated DNA fragments making it difficult to infer methylation states with low CpG density or low DNA methylation levels.

Alternatively, RRBS-seq (reduced representation bisulfite sequencing) can be employed to focus in on relevant portions of genomic DNA methylation (Meissner et al., 2005). This methodology utilizes methylation insensitive restriction endonucleases to enrich for CpG containing sites in the genome regardless of DNA methylation status. The restriction endonuclease MspI (Restriction endonuclease MspI protein) employed in RRBS-seq targets the palindromic double stranded DNA sequence of CCGG resulting in each DNA fragment containing a CpG at the termini. These digested fragments then go through sequencing library

construction and bisulfite conversion. The major benefits to this methodology are the bisulfite conversion which allows for base pair resolution measurement of DNA methylation among restriction endonuclease enriched fragments. Around 10-20% of CpGs can be assessed for DNA methylation by sequencing around 1% of the human genome allowing for significant reduction in sequencing costs. However, due to the DNA recognition sequence of the MspI enzyme regions that are CpG poor such as enhancers may not be interrogated at all with this method. Overall this method is highly focused at CpG rich promoters and CGIs.

The current gold standard to assess DNA methylation is WGBS (Whole Genome Bisulfite Sequencing) (Lister et al., 2009). Briefly, purified genomic DNA is sheared with sonication or enzymes, ligated with methylated DNA adapters, subjected to sodium bisulfite conversion and ultimately analyzed with high throughput sequencing. Sodium bisulfite conversion in this application serves to chemically convert all unmodified cytosines to uracil via a deamination reaction. During PCR (Polymerase Chain Reaction) amplification and sequencing these unmodified cytosines are detected as cytosine to thymine base conversions while sites of DNA methylation are detected as true cytosines due to chemical protection offered by 5-mC. The strengths of this methodology are the base pair level resolution that bisulfite conversion allows and the wide genomic coverage of cytosines in almost any genomic context. The whole genome nature of this methodology requires extensive sequencing to sample each cytosine at least 10 times for confidence in calculating DNA methylation frequencies across cell populations (Ziller et al., 2015). Vast whole genome coverage with WGBS allows for in depth analysis of DNA methylation far beyond what is possible with any other method. Nonetheless, the prohibitive sequencing costs have hampered the public availability of high quality WGBS datasets especially across complex experiments assaying many conditions.

Due to the rising interest in understanding complex relationships between epigenetic marks such as DNA methylation and chromatin accessibility several novel sequencing methods have been developed. NOME-seq (Nucleosome Occupancy and Methylome sequencing) is a

notable whole genome method for studying the DNA methylation-chromatin accessibility relationship. NOME-seq employs treatment of chromatin with a GpC (Guanine-phosphate-Cytosine) methyltransferase (M.CviPI) to map out nucleosomal and chromatin accessible portions of the genome (Kelly et al., 2012). Chromatin is treated with GpC methyltransferase enzyme to artificially methylated GpC nucleotides at enzyme accessible chromatin sites while CpG nucleotides keep their endogenous methylation levels. After subsequent bisulfite conversion and sequencing GpC methylated sites indicate enzyme accessible chromatin while CpG sites encode the genomic methylate state. In short, both accessibility and DNA methylation data can be extracted from the same DNA molecule. However, NOME-seq requires significant levels of DNA sequencing to reach appropriate read depth for DNA methylation determination which makes it prohibitory for many types of experiments requiring many samples. In addition, many intergenic regions coinciding with enhancers have very low CG/GC content in general and may not be interrogated well with this method. A major strength of this methodology is the direct distinction of both accessible and inaccessible chromatin from actual data as other methods frequently infer inaccessible chromatin from missing data. Overall the NOME-seq methodology has seen uptake in the literature but is likely held back due to the cumbersome whole genome sequencing required.

Aside from the bisulfite and GpC methyltransferase methodology employed by NOME-seq there are three primary genomic methods for mapping areas of low and high nucleosome density across the genome coinciding with euchromatin and heterochromatin respectively. The genomics methods known as ATAC-seq (Assay for Transposase Accessible Chromatin sequencing), DNase-seq (DNase I hypersensitive site sequencing) and MNase-seq (Micrococcal Nuclease sequencing) each employ treatment of extracted chromatin with an enzyme having nuclease properties. The portions of the genome best mapped by each methodology depend on the enzymatic properties of the nuclease or nuclease-like enzyme utilized.

DNase-seq was the most prevalent methodology prior to the past five years and employs treatment of chromatin with the endonuclease DNase-I (Dexoyribonuclease-I protein) (Boyle et al., 2008; Crawford et al., 2006; Hesselberth et al., 2009; Sabo et al., 2006). The endonuclease action of DNase-I on DNA is greatly impeded by DNA wrapped around nucleosome and in dense chromatin structures. Thus, when chromatin is treated with DNase-I nucleosome depleted regions of the genome are preferentially cleaved and can subsequently be isolated. Subsequent DNA isolation, adapter ligation and high throughput sequencing yields a map of the human genome where an abundance of reads indicates which genomic regions are likely nucleosome depleted and available to be occupied by TFs. Additionally, intermittent interruptions of high read counts at DNase sensitive loci can be used to infer sites of TF occupancy due to local interference of DNase-I cleavage by bound DNA. Overall DNase-seq is a reliable and well proven method to profile genomic areas of nucleosome depletion albeit with a somewhat more labor intensive protocol compared to other methodologies.

In recent years ATAC-seq has become the methodology of choice for profiling nucleosome depleted regions of the genome. Rather than using a traditional nuclease ATAC-seq treats chromatin with the enzyme Tn5 transposase that simultaneously fragments DNA and ligates sequencing adapters (Buenrostro et al., 2013). The unique fragmentation and ligation properties of Tn5 transposase give ATAC-seq several advantages over other methodologies including low cell input, a facile protocol, customizable adapter sequences and lower sequencing requirements. A downside of the ATAC-seq methodology is the abundance of sequencing reads mapping to mitochondria as mitochondrial DNA is entirely nucleosome free and thus preferentially acted upon by Tn5 transposase. Mitochondrial reads frequently account for 30-60% of sequencing reads, however, recent protocol advances have minimized this problem (Corces et al., 2017; Montefiori et al., 2017). ATAC-seq shows high concordance with DNase-seq but the easy protocol makes it likely to remain the workhorse methodology for profiling nucleosome free regions of the genome.

MNase-seq is a more specialized method yielding a slightly different data output than other methods. In contrast to DNase-seq or ATAC-seq, MNase-seq produces a profile of actual nucleosome positions rather than the intervening nucleosome free regions of DNA (Schones et al., 2008). This is primarily due to the enzyme properties of micrococcal nuclease enzyme used to digest the chromatin during DNA sequencing library preparation. Micrococcal nuclease is an endo-exo nuclease with strong preference to digest non-nucleosomal DNA including linker DNA between nucleosomes leading to an overall enrichment of DNA wrapped around nucleosomes. Though this methodology does not profile nucleosome depleted regions directly it can be used to discern regions of high and low nucleosome density corresponding with heterochromatin and euchromatin. MNase-seq is not frequently employed across studies in the literature, likely due to the specialized data type and the extensive sequencing required to profile the vast amount of the genome occupied by nucleosomes.

DNA methylation and chromatin accessibility relationship

Many studies have profiled epigenetic marks across the human genome including DNA methylation, chromatin accessibility, and histone modifications. In most cases these epigenomic marks are profiled on independent samples and integrated later as correlative overlays. Such integrations of different epigenetic data sets on the human genome are valuable because different epigenomic marks and associated molecular factors do not exist in exclusivity within living organisms (Ernst and Kellis, 2012; Kundaje et al., 2015). Different epigenomic marks influence and combine with each other to create the complete gene regulatory program within cells. One particular epigenetic relationship of interest connected with gene regulation is that of chromatin accessibility and DNA methylation. Both chromatin accessibility and DNA methylation are known to independently influence the interaction of TFs and molecular factors with DNA. DNA methylation has been described to both impede and enhance TF binding as found by *in-*

vitro SELEX studies (Kribelbauer et al., 2017; Yin et al., 2017) and more targeted *in-cellulo* studies (Domcke et al., 2015; Wang et al., 2012a). Further, DNA methylation is known to initiate formation of condensed chromatin states not permissive for transcription (Kass et al., 1997). Chromatin accessibility changes inherently control the availability of the DNA sequence underlying chromatin to TFs, DNA modifying enzymes and other DNA binding molecules. Much of what is already known about the relationship between chromatin accessibility and DNA methylation is in limited contexts frequently as static snapshots.

A correlation between chromatin accessibility and DNA methylation is already clear from many epigenomic profiling studies of the human genome. Chromatin accessible nucleosome depleted regions are almost always distinctly hypomethylated. Exceptions to this have been noted. Hypomethylated regions have been found within so called closed chromatin associated with a heterochromatin state dense with nucleosomes. These hypomethylated, closed chromatin regions are thought to be “decommissioned” gene regulatory elements, primarily enhancers, that were formerly utilized by the cell but are now no longer needed and simply represent a history of utilization (Jadhav et al., 2019). Alternatively, this may not be a “decommissioned” gene regulatory element but rather a “poised” epigenetic state with a lower energy barrier to reutilization in turn facilitating rapid responses. A similar phenomenon has been described for chromatin states that possess both activating and repressive histone marks (Bernstein et al., 2006). Nucleosomal DNA is typically methylated, however recent studies have demonstrated that in some contexts the nucleosome wrapped DNA is methylated while intervening linker DNA is hypomethylated (Kelly et al., 2012). Overall however, it is exceptionally rare to find methylated DNA that is chromatin accessible and nucleosome depleted.

While there are exceptions, chromatin accessibility and regions of DNA hypomethylation are tightly linked at both promoter and intergenic loci. These same regions of DNA hypomethylation and accessibility coincide with many epigenomic marks indicative of active gene regulatory elements such as the histone modification H3K27ac and transcriptional firing

(Schlesinger et al., 2013). However, the process and molecular timeline for establishment of this state hypomethylated and active gene regulatory element is unclear. Profiles of DNA methylation across hematopoietic lineage cell types revealed progressive expansions and an increase in the number of hypomethylated regions when comparing stem cell to differentiated cell types (Hodges et al., 2011). This progression of hypomethylated region formation suggests a dynamic process establishing these regions during cell differentiation. Further, Integration of separate DNase-seq and whole genome bisulfite data sets from embryonic stems cells found a large portion of DHS (DNase hypersensitive site) coincide with DNA methylation (Schlesinger et al., 2013) suggesting a transitional state leading to hypomethylated region formation and enhancer activation. Collectively this data suggests a step-wise ordered process for chromatin accessibility, hypomethylated region establishment and gene regulatory element activation. If the repressive role of DNA methylation is assumed in this scenario then gene regulatory element activation depends on progressive formation of a chromatin accessible and hypomethylated state to act on target genes (Figure 2).

Scope of dissertation

Gene regulatory elements such as enhancers form networks that are potent guides for cell specification and organismal development. Epigenetic changes such as DNA methylation and chromatin accessibility dynamics have been shown to reflect the modulation of these gene regulatory elements, yet little is known about the precise timing of these changes at the earliest stages of cell fate specification. A better understanding of the temporal relationship between chromatin accessibility, DNA methylation and gene expression dynamics across cellular differentiation is needed. It is unclear if chromatin accessibility and DNA methylation changes are interdependent or distinct dynamic molecular events. The subsequent impact of DNA methylation and chromatin accessibility dynamics on gene regulatory activity is also poorly

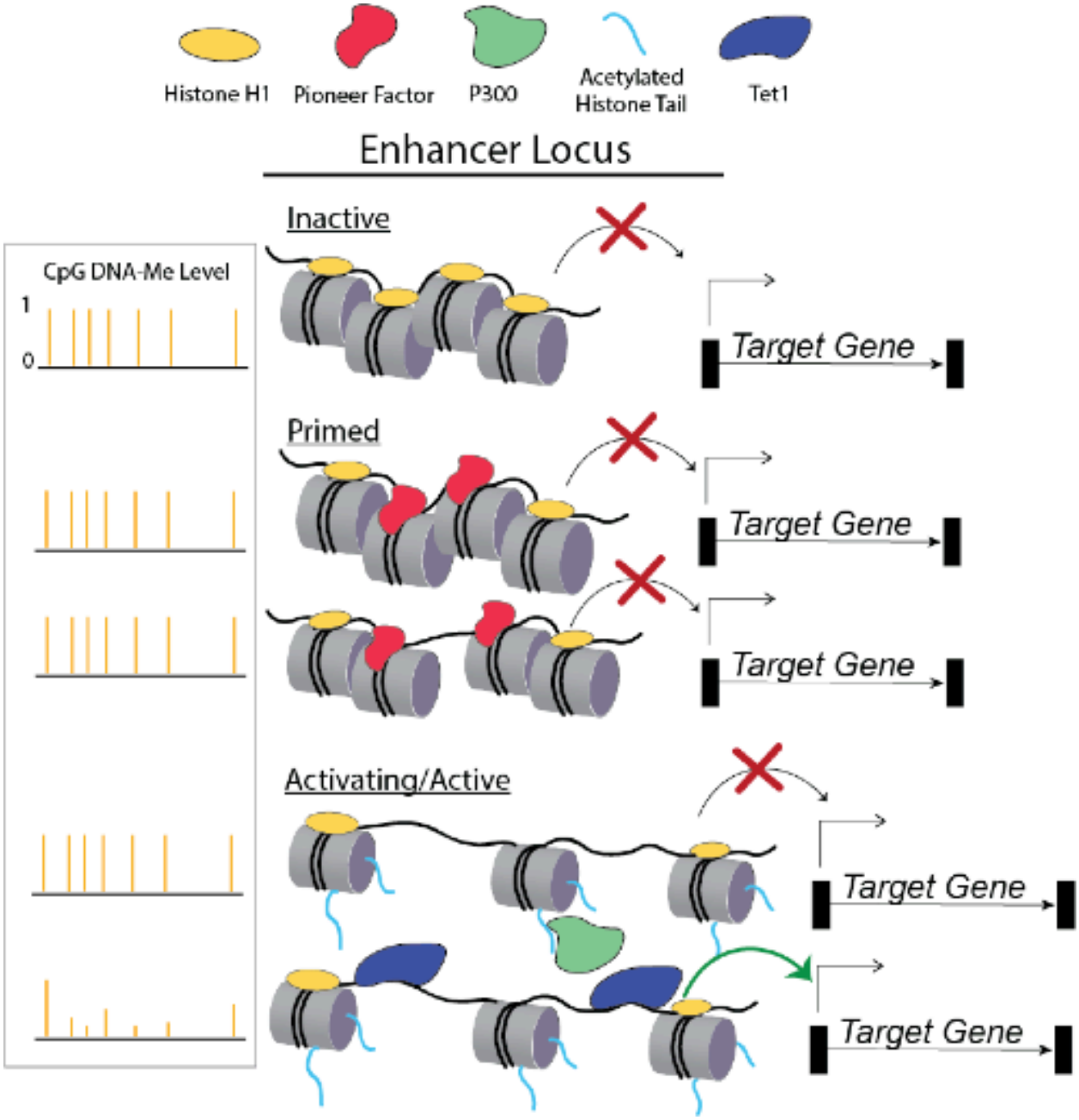

Figure 2. Graphical model for epigenomic dynamics at an enhancer Locus. Theoretical model of DNA methylation, chromatin accessibility and neighboring gene transcription at an enhancer locus. Vertical gold bars denote locations of CpG DNA methylation. Transcription remains impeded until the enhancer locus becomes fully demethylated after establishment of chromatin accessibility.

described. A proposed timeline of molecular events follows chromatin accessibility changes preceding DNA hypomethylation formation and ultimately gene regulatory element activation. This dissertation aims to assess this time line and the DNA methylation, chromatin accessibility and gene regulatory activity relationship in the context of cellular differentiation.

The following represents a general overview of the chapter by chapter progression of this dissertation work. Chapter 1 presents background and literature review to frame the current knowledge base of DNA methylation and its impact on gene regulation across human development. Materials and methods necessary for the studies contained herein are described within chapter 2. The greatest body of this dissertation work is represented by chapter 3 regarding epigenomic dynamics during a cell fate specification and represents research published in *Molecular Cell* (Barnett et al., 2020). Additional studies contained within chapter 4 were pursued in complement to this dissertation work for assessing gene regulatory element function within dynamics contexts. Discussions of limitations, future research directions and outstanding gaps in knowledge are covered within chapter 5.

Chapter 2: Materials and Methods

THP-1 monocytes

THP-1 cells (gift of Dr. Manuel Ascano Jr., Vanderbilt University) were cultured in RPMI Medium 1640 supplemented with 10% fetal bovine serum, 2mM GlutaMAX, 100 units/mL penicillin, 100 µg/mL streptomycin and 1mM sodium pyruvate (referenced as complete RPMI medium). Culture conditions were maintained at 5% CO2, 37°C and 80% humidity. During routine culture THP-1 cells were maintained at a density of 0.5-2E6 cells/mL with 50% media change every 72 hours.

Drosophila S2 Cells

S2 cells (gift of Dr. Andrea Page-McCaw, Vanderbilt University) were cultured in Schneider's *Drosophila* Medium supplemented with 10% heat inactivated fetal bovine serum, 100 units/mL penicillin and 100 µg/mL streptomycin. Culture conditions were maintained at 25°C, with atmospheric CO2 and humidity. During routine culture cells were maintained at a density of 2-10E6 cells/mL.

All cell lines have been authenticated by genome sequencing and were regularly screened for mycoplasma contamination using the MycoAlert kit (Lonza).

Cell Treatments

In preparation for the PMA (Phorbol 12-myristate 13-acetate) time course, 5x106 THP-1 monocytes were plated in 1mL complete RMPI medium/well of a flat bottom 6-well cell culture treated plate. Differentiation was subsequently induced by addition of 1mL complete RPMI

medium supplemented with 200ng PMA for a final PMA concentration of 100ng/mL PMA (162nM). Additions of PMA were temporally staggered to allow for simultaneous collection of time points. THP-1 L-AA-2-P (L-Ascorbic acid 2-phosphate) treatment conditions were initially plated identically to PMA only conditions in a final volume of 2mL complete RPMI medium with 100ng/mL PMA. Plated cells were then supplemented with 50mM L-AA-2-P to a final concentration of 350μM. L-AA-2-P was replenished with daily media changes to account for L-AA-2-P degradation over time until harvest at 72hrs. PMA stimulation was withdrawn due to toxicity after 24 hours for extended time points (48hr, 72hr). THP-1 cells were harvested by first collecting non-adherent cells suspended in media followed by a 5 minute incubation with TrypLE Express at 37°C to collect adherent cells. Non-adherent and adherent cells were pooled and pelleted at 4°C, 500 RCF (Rotor Conversion Factor) for 5 minutes in a 15mL conical tube. After pelleting cells were resuspended in ice cold PBS (Phosphate Buffered Saline) and counted using an automated cell counter with trypan blue to assess cell viability. Approximately 1x106 cells were reserved for TRIzol RNA extraction, 1x106 cells for ATAC reactions and 2x105 cells for genomic DNA extraction.

Transposome Preparation

Plasmid (pTXB1-Tn5, gift of Dr. Rickard Sandberg, Karolinska Institutet) for production of Tn5 transposase (Picelli et al., 2014) was transformed into *E. coli* strain C3013 (NEB) according to manufacturer supplied protocol on ampicillin plates. Resulting bacterial colonies were inoculated into 5mL LB (Lysogeny Broth) + 100μg/mL ampicillin and incubated at 37°C to OD600 = 1.0. 5mL cultures were subsequently used to inoculate 1L LB +100μg/mL ampicillin and grown at 37°C, shaking at 220RPM (Rotations Per Minute) in baffled flasks until OD600 was approximately 0.70. Cultures were then transferred to an ice water bath and cooled to 15°C. Protein expression was induced by adding IPTG to a final concentration of 250μM

followed by incubation at 23°C with shaking at 120RPM for approximately 4hrs or until OD600 = approximately 2.1. Bacteria cultures were harvested by centrifugation at 5,000 RPM (F9-4x1000y rotor), 4°C for 10 min in 1L centrifuge bottles. Supernatants were decanted and pellets stored overnight at −70°C. The following day pellets were thawed and resuspended in 60mL HEGX buffer (20 mM HEPES-KOH at pH 7.2, 0.8 M NaCl, 1 mM EDTA, 10% glycerol, 0.2% Triton X-100) supplemented with Complete EDTA-free protease inhibitor tablet (Roche). The thawed bacterial suspension was cooled on salt-ice to −1°C and subsequently lysed with 8 rounds of sonication for 40 s at 70% duty cycle. During sonication the lysate was cooled repeatedly on salt-ice to avoid warming above 10°C. Lysis progress was monitored by removing 200µL aliquots between rounds of sonication and comparing pellet size after low speed centrifugation at 2000 RCF for 1 min. Lysate was then pelleted at 15K RPM (SS-34 rotor), 4°C for 30 min. The pelleted lysate supernatant was cleared of contaminating *E. coli* genomic DNA by drop wise addition of 1.6mL 10% neutralized PEI (Polyethyleneimine) solution with slow stirring. The PEI lysate solution should begin to turn milky white as the DNA precipitate forms. Lysate was pelleted at 12,000 RPM (SS-34 rotor), 4°C for 10 min. Supernatant was loaded onto a prepared 5mL chitin column at 0.4mL/min flow rate, 4°C. Chitin column was washed with 300mL HEGX buffer at a flow rate of 0.4mL/min. Tn5 transposase was eluted by the addition of 14mL HEGX + 100mM DTT (Dithiothreitol) to the column, followed by a 48hr incubation at 4°C to induce cleavage from the column. Tn5 transposase was collected from the column in 1mL fractions and subsequently analyzed for protein concentration. Fractions with the highest protein concentration were pooled and dialyzed twice with 1L of 2x dialysis buffer (100mM HEPES, pH 7.2, 0.2M NaCl, 0.2mM EDTA, 2mM DTT, 0.2% TX100, 20% Glycerol). Dialyzed protein was concentrated with Amicon Ultracel 30 centrifugal filters (Millipore) in multiple rounds of centrifugation at 3000 RCF, 4°C for 20 min with intermittent mixing until a final OD280 of at least 3.0. Purified Tn5 transposase was aliquoted and stored at −20°C after the addition of 1 volume

100% glycerol. Oligos used for transposome assembly and DNA library barcoding are listed in (Table 1)

Standard ATAC transposome adapters were prepared by annealing two oligo mixes in separate PCR tubes (Mix A: 5µl 100µM Tn5MEREV oligo, 5µl 100µM Tn51 oligo, 40µL nuclease free water; Mix B: 5µl 100µM Tn5MEREV oligo, 5µl 100µM Tn5_2_ME_Comp oligo, 40µL nuclease free water). Oligo mix A and B were both incubated in a PCR thermocycler as follows: 95°C for 3 minutes, 65°C for 3 minutes, ramp down to 24°C at a rate of −1°C/second, hold at 24°C. After annealing, oligo mix A and B were combined along with 100µL glycerol to create a 5µM, 50% glycerol adaptor mixture. Transposomes were assembled by mixing equal parts purified Tn5 transposase enzyme and adaptor mixture followed by a 25°C incubation for 60 min.

ATAC-Me (Assay for Transposase Accessible Chromatin with Bisulfite Conversion) and T-WGBS (Tn5 transposase based Whole Genome Bisulfite) transposome adapters were prepared by annealing the oligonucleotides in PCR tubes (10µl 100µM Tn5mC-Apt1 oligo, 10µl 100µM Tn5mC1.1-A1block oligo, 80µL nuclease free water). Oligos were incubated in a PCR thermocycler as follows: 95°C for 3 minutes, 65°C for 3 minutes, ramp down to 24°C at a rate of −1°C/second, hold at 24°C. After annealing oligos were combined along with 100µL glycerol to create a 5µM, 50% glycerol adaptor mixture. Transposomes were assembled by mixing equal parts purified Tn5 transposase enzyme and adaptor mixture followed by a 25°C incubation for 60 min.

Extension based T-WGBS transposome adapters (Spektor et al., 2019) were prepared by annealing two oligo mixes in separate PCR tubes (Mix A: 5µl 100µM Tn5MEREV oligo, 5µl 100µM Tn5mC-Apt1 oligo, 40µL nuclease free water; Mix B: 5µl 100µM Tn5MEREV oligo, 5µl 100µM Meth_Tn5_2_ME_comp oligo, 40µL nuclease free water). Oligo mix A and B were both incubated in a PCR thermocycler as follows: 95°C for 3 minutes, 65°C for 3 minutes, ramp down to 24°C at a rate of −1°C/second, hold at 24°C. After annealing, oligo mix A and B were

Primer/Oligo name	Sequence (5' to 3')
Tn5mC-Apt1	T/mC/GT/mC/GG/mC/AG/mC/GT/mC/AGATGTGTATAAGAGA/mC/AG
Tn5mC1.1-A1block	/5Phos/CTGTCTCTTATACA/3ddC/
Tn5mC-RepIO1	/5Phos//mC/TGT/mC/T/mC/TTATA/mC/A/mC/AT/mC/T/mC //mC/GAG/mC//mC//mC/A/mC/GAGA/mC//3InvdT/
Meth_Tn5_2_ME_comp	GT/mC/T/mC/GTGGG/mC/T/mC/GGAGATGTGTATAAGAGA/mC/AG
Tn5_1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
Tn5_2_ME_comp	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG
TN5MEREV	/5Phos/CTGTCTCTTATACACATCT
Index N701	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGG
Index N702	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGG
Index N703	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGG
Index N704	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGG
Index N705	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGG
Index N706	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGG
Index N501	AATGATACGGCGACCACCGAGATCTACACTAGATCGCTCGTCGGCAGCGTC
Index N502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTCGGCAGCGTC
Index N503	AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGCAGCGTC
Index N504	AATGATACGGCGACCACCGAGATCTACACAGAGTAGATCGTCGGCAGCGTC
Index N505	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGTCGTCGGCAGCGTC
Index N506	AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGCAGCGTC

Table 1. DNA Oligonucleotides and primers used in this study. 5Phos indicates a 5' phosphate modification on the oligonucleotide. mC indicates a 5-methylcytosine at that oligonucleotide position. 3ddC indicates a 3' dideoxycytidine. 3InvdT indicates a 3' inverted deoxythymidylate. Highlighted orange nucleotides within index sequences indicate the variable barcode sequence.

combined along with 100µL glycerol to create a 5µM, 50% glycerol adaptor mixture. Transposomes were assembled by mixing equal parts purified Tn5 transposase enzyme and adaptor mixture followed by a 25°C incubation for 60 min.

ATAC-seq

ATAC-seq libraries were prepared similarly to previously reported methods (Buenrostro et al., 2013). Harvested THP-1 cells were centrifuged at 4°C, 500 RCF for 5 min and subsequently resuspended in 1mL ice cold PBS. Resuspended cells were counted and assayed for cell viability with trypan blue using an automated cell counter. Cell viability > 80% was required to proceed with library preparation. Cell suspension volume corresponding to 2x105 THP1 cells was pipetted into a 1.5mL eppendorf tube and pelleted at 4°C, 500 RCF for 5 min. Supernatant was aspirated and the cell pellet resuspended in 150µL cold ATAC lysis buffer (10mM Tris-HCl pH 7.4, 10mM NaCl, 3mM MgCl2, 0.1% IGEPAL-630). This suspension was agitated by gently pipetting up/down with a 200µL micropipette tip 10 times. Subsequently, this suspension was pelleted at 4°C, 500 RCF for 10 minutes. Supernatant was discarded and nuclei pellet immediately resuspended in 190µL transposition reaction mix (10mM Tris-HCl pH 7.5, 5mM MgCl2, 10% Dimethylformamide) by pipetting up/down with a 200µL micropipette tip 3 times. 10µL of pre-assembled Tn5 transposome with standard ATAC adapters was added. Tubes were gently agitated to mix all components and incubated at 37°C, 30 minutes, 700RPM in an Eppendorf Thermomixer. ATAC reactions were terminated by adding 1mL Zymo DNA binding buffer and vortexing. Reactions were purified according to manufacturer instructions in a DNA Clean and Concentrator-5 kit (Zymo) and eluted in 25µL nuclease free water. Eluted ATAC DNA was amplified and barcoded in 50µL PCR reactions (25µL 2x NEBNext High-Fidelity PCR Master Mix, 20µL eluted ATAC DNA, 2.5µL 10µM i5 index primer, 2.5µL 10µM i7 index primer) with the following PCR thermocycler program: 72°C 5 min; 98°C 30 s; 8 cycles of 98°C 10 s,

62°C 30 s, 72°C 30 s; final extension 72°C 5 min; hold at 12°C. Post-amplification PCR reactions were cleaned and concentrated with a Zymo DNA Clean and Concentrator-5 column kit. DNA was eluted in 22µL nuclease free H2O. Preliminary library analysis for concentration and size distribution was performed using an Agilent 2200 TapeStation with a D5000 screentape. ATAC-seq DNA libraries were sequenced using 2x75bp paired-end reads on the NextSeq500 instrument.

ATAC-Me

Harvested THP-1 cells were centrifuged at 4°C, 500 RCF for 5 min and subsequently resuspended in 1mL ice cold PBS. Resuspended cells were counted and assayed for cell viability with trypan blue using an automated cell counter. Cell viability > 80% was required to proceed with library preparation. Cell suspension volume corresponding to 2x105 THP1 cells was pipetted into a 1.5mL eppendorf tube and pelleted at 4°C, 500 RCF for 5 min. Supernatant was aspirated and the cell pellet resuspended in 150µL cold ATAC lysis buffer (10mM Tris-HCl pH 7.4, 10mM NaCl, 3mM MgCl2, 0.1% IGEPAL-630). This suspension was agitated by gently pipetting up/down with a 200µL micropipette tip 10 times. The suspension was pelleted at 4°C, 500 RCF for 10 minutes. Supernatant was discarded and nuclei pellet immediately resuspended in 190µL transposition reaction mix (10mM Tris-HCl pH 7.5, 5mM MgCl2, 10% Dimethylformamide) by pipetting up/down with a 200µL micropipette tip 3 times. 10µL of pre-assembled Tn5 transposome (containing methylated adaptors) was added. Tubes were gently agitated and incubated at 37°C, 30 minutes, 700RPM in an Eppendorf Thermomixer. ATAC-Me reactions were terminated by adding 1mL Zymo DNA binding buffer and vortexing. Reactions were purified according to manufacturer instructions in a DNA Clean and Concentrator-5 kit (Zymo) and eluted in 13µL nuclease free water.

DNA eluate was used as input into the gap repair reaction (11µL ATAC-Me DNA eluate, 2µL 10µM Tn5mC-Repl01 oligo, 2µL 10x ampligase buffer, 2µL dNTPs 2.5mM each). This gap repair reaction was assembled in a PCR tube and incubated as follows in a PCR thermocycler: 50°C for 1 minute, 45°C for 10 minutes, ramp down to 37°C at a rate of −0.1°C/second, hold at 37°C. Once reaching 37°C, 1µL T4 DNA polymerase and 2.5µL ampligase were added separately without removing the tube from the thermocycler. After final addition of enzymes, the reaction was mixed by pipetting up/down with a 20µL micropipette tip without removing tube from the thermocycler. The gap repair reaction was incubated as follows: 37°C for 30 minutes, hold at 4°C. 2µL of the gap repair reaction was reserved for a test PCR amplification to confirm successful ATAC nucleosomal laddering. 2µL of 250mM EDTA (pH = 8.0) was added to stop the gap repair reaction. Gap repaired ATAC-Me material was bisulfite converted according to manufacturer instructions using the Zymo Lightning EZ DNA Methylation-Lightning Kit (cat no D5030) with slight modification. 20µL gap repaired DNA and 130µL CT conversion reagent were mixed and split between three PCR tubes, 50µL/tube. Bisulfite conversion reactions were then incubated in PCR tubes as follows: 98°C for 8 minutes, 54°C for 60 minutes, hold at 4°C. The three bisulfite conversion reactions were re-pooled into a single tube. Final purification/desulfonation was performed as directed by the kit manufacturer manual instructions. Final elution was in 25µL of M-elution buffer supplied by the kit. Eluted ATAC-Me DNA was amplified and barcoded in 50µL PCR reactions (25µL 2x KAPA HiFi HotStart Uracil+ ReadyMix, 20µL eluted ATAC-Me DNA, 1.5µL 10µM i5 index primer, 1.5µL 10µM i7 index primer) with the following PCR thermocycler program: 98°C 45 s; 10 cycles of 98°C 15 s, 62°C 30 s, 72°C 30 s; final extension 72°C 2 min; hold at 12°C. Post-amplification PCR reactions were cleaned and concentrated with a Zymo DNA Clean and Concentrator-5 column kit. Elution was in 22µL nuclease free H2O. Preliminary library analysis for concentration and size distribution was performed using an Agilent 2200 TapeStation with a D5000 screentape. ATAC-

Me DNA libraries were sequenced using 2x150bp paired-end reads on the HiSeq4000 and NovaSeq6000 instruments.

THP-1 T-WGBS

THP-1 T-WGBS libraries were prepared similarly to previously reported methods (Adey and Shendure, 2012, Wang et al., 2013). Genomic DNA purified from THP-1 cells was diluted in a 50µL tagmentation reaction (100ng genomic DNA, 10mM Tris-HCl pH 7.5, 5mM MgCl2, 10% Dimethylformamide). 2.5µL of transposome assembled with T-WGBS adapters was added and the reaction incubated at 55°C for 8 min in a PCR thermocycler. Tagmentation reactions were immediately stopped with the addition of 250µL Zymo DNA binding buffer from the DNA Clean & Concentrator-5 kit (Zymo). Reactions were then purified according to manufacturer instructions in a DNA Clean and Concentrator-5 kit (Zymo) and eluted in 15µL nuclease free water. DNA eluate was used as input into the gap repair reaction (11µL ATAC-Me DNA eluate, 2µL 10µM Tn5mC-Repl01 oligo, 2µL 10x ampligase buffer, 2µL dNTPs 2.5mM each). This gap repair reaction was assembled in a PCR tube and incubated as follows in a PCR thermocycler: 50°C for 1 minute, 45°C for 10 minutes, ramp down to 37°C at a rate of −0.1°C/second, hold at 37°C. Upon reaching 37°C, 1µL T4 DNA polymerase and 2.5µL ampligase were added separately without removing the tube from the thermocycler. The reaction was mixed by pipetting up/down with a 20µL micropipette tip without removal from the thermocycler. The gap repair reaction was subsequently incubated as follows: 37°C for 30 minutes, hold at 4°C. 2µL of the gap repair reaction was reserved for a test PCR amplification for troubleshooting. 2µL of 250mM EDTA (pH = 8.0) was added to stop the reaction. Gap repaired, tagmented DNA was subsequently bisulfite converted according to manufacturer instructions using the Lightning EZ DNA Methylation-Lightning Kit (Zymo) with slight modification. 20µL gap repaired DNA and 130µL CT conversion reagent were mixed and split between three PCR tubes, 50µL/tube. Bisulfite

conversion reactions were then incubated in PCR tubes as follows: 98°C for 8 minutes, 54°C for 60 minutes, hold at 4°C. The three bisulfite conversion reactions were re-pooled into a single tube. Final purification/desulfonation was performed as directed by the kit manufacturer manual instructions. Final elution was in 25µL of M-elution buffer supplied by the kit. Eluted bisulfite converted DNA was amplified and barcoded in 50µL PCR reactions (25µL 2x KAPA HiFi HotStart Uracil+ ReadyMix, 20µL eluted ATAC-Me DNA, 1.5µL 10µM i5 index primer, 1.5µL 10µM i7 index primer) with the following PCR thermocycler program: 98°C 45 s; 8 cycles of 98°C 15 s, 62°C 30 s, 72°C 30 s; final extension 72°C 2 min; hold at 12°C. Post-amplification PCR reactions were cleaned and concentrated in a DNA Clean and Concentrator-5 kit (Zymo). Elution was in 22µL nuclease free water. Preliminary library analysis for concentration and size distribution was performed using an Agilent 2200 TapeStation with a D5000 screentape. THP-1 T-WGBS DNA libraries were sequenced using 2x150bp paired-end reads on the HiSeqX instrument.

Drosophila S2 cell T-WGBS

Drosophila S2 cell T-WGBS libraries were prepared similarly to previously reported methods (Lu et al., 2015, Spektor et al., 2019, Suzuki et al., 2018). Genomic DNA purified from *Drosophila* S2 cells was diluted in a 50µL tagmentation reaction (100ng genomic DNA, 10mM Tris-HCl pH 7.5, 5mM MgCl2, 10% Dimethylformamide). 2.5µL of transposome assembled with T-WGBS adapters was added and the reaction incubated at 55°C for 8 min in a PCR thermocycler. Tagmentation reactions were immediately halted with the addition of 250µL Zymo DNA binding buffer from the DNA Clean & Concentrator-5 kit (Zymo). Reactions were then purified according to manufacturer instruction in a DNA Clean and Concentrator-5 kit (Zymo) and eluted in 15µL nuclease free water. Eluted DNA was subsequently end-repaired with either Klenow exo- polymerase (NEB) or T4 DNA polymerase (NEB). Klenow extension

based libraries were end-repaired by Klenow exo- repair mix (2μL 10x NEBuffer 2.0, 5 units Klenow exo- [NEB], 4μL 2.5mM each 5-methyl-dCTP (5-methyldeoxycytosine triphosphate) substituted dNTP mix, 13.5μL eluted tagmented DNA). Klenow repair reactions were incubated at 37°C, 30 minutes followed by adding 2μL 250mM EDTA to terminate the reaction. T4 polymerase extension based libraries were end repaired by T4 polymerase repair mix (2μL 10x NEBuffer 2.1, 2μL 2.5mM each 5-methyl-dCTP substituted dNTP mix, 11μL eluted tagmented DNA). T4 polymerase repair mixture was incubated as follows: 50°C, 1 min; hold at 37°C; add 1μL T4 polymerase, 3000 units/mL (NEB); 37°C, 5 min. T4 polymerase repair reaction was subsequently terminated by adding 2μL 250mM EDTA. Klenow or T4 DNA polymerase end-repaired DNA was bisulfite converted, PCR amplified and sequenced as described above.

RNA-seq

RNA from approximately 1x106 THP-1 cells was harvested from each PMA stimulation time point by pelleting cells at 4°C, 500 RCF for 5 minutes. After removal of supernatant, cell pellet was homogenized with 1mL of TRIzol Reagent by repeatedly pipetting up/down with a 1mL micropipette tip. RNA was purified from Trizol homogenate according to recommended manufacturer instructions. RNA-seq libraries were prepared using the SMARTer® Stranded Total RNA Sample Prep Kit (Takara Bio). RNA-seq libraries were sequenced using 2x100bp paired-end reads on the NovaSeq6000 instrument.

Sequencing Library Processing

All sequencing library reads were trimmed of adapters using TrimGalore script wrapper for Cutadapt (Martin, 2011) and FastQC. Standard ATAC and ATAC-Me/WGBS reads were mapped with Bowtie2 (Langmead and Salzberg, 2012) or WALT (Chen et al., 2016),

respectively, to the hg19 genome assembly (Lander et al., 2001). Methylation analysis of ATAC-Me and WGBS reads was performed using the MethPipe suite of tools (Song et al., 2013). Preseq (Daley and Smith, 2013) was used to compare library complexity across protocols. RNA libraries were mapped with the STAR mapper (Dobin et al., 2013) and analyzed for differential RNA expression using DESeq2 (Love et al., 2014). Regions enriched for chromatin accessibility in standard ATAC and ATAC-Me data were identified using the MACS2 (Model-based Analysis of ChIP-Seq) (Zhang et al., 2008) peak caller suite of tools. Regions dynamic for chromatin accessibility were identified with the TCseq R-package (Wu and Gu, 2019). HOMER was used for all TF motif analysis of dynamic or static chromatin accessible regions. Annotation and gene association for dynamic and static chromatin accessible regions was performed with the ChIPseeker (Yu et al., 2015) and ClusterProfiler (Yu et al., 2012) R-packages. *De-novo* footprint analysis by identifying ATAC-seq read signal depressions was performed using the Wellington footprinting algorithm (Piper et al., 2013). Dual analysis of footprint depth and flanking accessibility at pre-identified TF DNA sequence motifs was performed using the bagfootr R-package (Baek et al., 2017). The samtools (Li et al., 2009), bedtools (Quinlan and Hall, 2010) and deeptools (Ramírez et al., 2014) suites of tools were used to aid in data manipulation and visualization.

Quantification and Statistical Analysis

Chromatin accessibility peaks were initially selected for analysis based upon filtering of the Benjamini-Hochberg corrected p value (q-value) reported by the MACS2 peak-calling algorithm (q-value < $1x10^{-10}$). Differentially chromatin accessible genomic loci across the time course were selected from FDR corrected p values produced by the likelihood ratio test implemented in the TCseq R-package (corr. p value < $5x10^{-3}$). Differentially expressed genes across the time course were selected from corrected p values produced by the likelihood ratio

test implemented in the DESeq2 R-package (corr. p value < $5x10^{-3}$). Statistical analyses were performed within the R computing environment and visualized with ggplot2 (Wickham, 2009). Details of statistical analyses can be found in figure legends and on our Github page (see below).

Data and Code Availability

Datasets utilized in this study may be accessed at GEO: GSE130096 and GEO: GSE96800. Detailed code and workflows associated with main figures may be accessed at: https://github.com/HodgesGenomicsLab/ATAC-Me

Chapter 3: ATAC-Me Capture Prolonged DNA Methylation of Dynamic Chromatin Accessibility Loci during Cell Fate Transitions

Introduction

Integrative analysis of epigenomic data has identified a range of enhancer states across diverse cell types and their developmental intermediates (Ernst and Kellis, 2012). Certain classes of enhancers (active, repressed, poised) feature distinct combinations of histone modifications (Rada-Iglesias et al., 2012; Zentner et al., 2011) and chromatin accessibility (ChrAcc) patterns (Thurman et al., 2012). Such combinations may represent stages in the stepwise process of gene regulation coordinated by TF DNA binding events. This intimate temporal relationship between the epigenome and TFs is thought to ensure proper orchestration of TF activity and is critical for normal specification of cellular function and phenotype (Heinz et al., 2010).

DNAme in particular is thought to stabilize cellular identity in the context of gene regulation (Schubeler, 2015). DNAme is typically associated with stable repression of promoters and enhancers, while hypo- or low DNAme reflects past or present occupancy of gene regulatory elements (GRE) by factors involved in transcriptional control (Hodges et al., 2011; Jadhav et al., 2019; Lister et al., 2009; Stadler et al., 2011). Despite being indispensable for proper cellular differentiation and gene activation (Dawlaty et al., 2014; Okano et al., 1999), current models of enhancer dynamics only weakly describe a role for DNAme within this ordered process. Further, the degree to which DNA methylation influences enhancer regulation of gene transcription is unclear.

Comparative analysis of genome-wide DNAme datasets from diverse tissues has shown that the degree of differential methylation between cell types is much greater in distal gene-regulatory elements outside of CGI promoters (Bock et al., 2012; Hodges et al., 2011; Stadler et

al., 2011). These patterns of DNAme are characterized by local, cell type specific hotspots of CpG hypomethylation often overlapping enhancer histone marks and DHSs. We have previously suggested that combined signatures of DNAme and ChrAcc may reflect stages of enhancer activation (Schlesinger et al., 2013). Our analysis of steady state data revealed that DNAme and ChrAcc co-occur across a surprising number of sites in pluripotent stem cells. In mature cells, this duality is seemingly resolved by further loss of methylation (active) or loss of accessibility (repressed), suggesting that DNAme changes are, perhaps not surprisingly, sequential to nucleosome repositioning during enhancer activation. Indeed, several recent studies support a model in which inaccessible enhancers are primed by pioneer TFs leading to chromatin remodeling, increased ChrAcc and concomitant loss of DNAme (Donaghey et al., 2018; Mayran et al., 2018). Based on this and other observations, we hypothesized that these dual states may indicate a unique stage along the enhancer regulation continuum, and that loss of methylation during cell fate transitions is a final, necessary step in the process of enhancer activation and thus, the establishment or stabilization of cell identity.

The ability to test this model, however, has been limited in part by a lack of molecular tools to measure spatio-temporal relationships at high resolution on a genome-wide scale, and most importantly, directly within the context of chromatin. Consequently, our genome-scale understanding of relationships between DNAme and other aspects of enhancer regulation are based on disconnected datasets derived from steady state, asynchronous cell populations or time point associated datasets too far removed from one another to capture transient cellular states. To better characterize the spatio-temporal relationship between ChrAcc and DNAme, we developed an approach that enables simultaneous measurement of these two distinct biochemical events from the same population of DNA molecules. Our approach, aptly named ATAC-Me, is an adaptation of two techniques combining transposase-assisted enrichment of ChrAcc fragments from native chromatin (Buenrostro et al., 2013) with subsequent sodium bisulfite conversion and deep sequencing (Adey and Shendure, 2012; Wang et al., 2013).

We performed ATAC-Me, RNA-seq and reference-point whole genome bisulfite sequencing (WGBS) on a densely sampled time course of PMA (Phorbol 12-myristate 13-acetate) induced THP1 monocytes differentiated to macrophages. Our approach reveals ChrAcc and DNAme may be quantitatively decoupled at myeloid enhancers as they appear to become poised, active or repressed during early monocyte to macrophage transitions. We observed persistent hypermethylation of an unexpected number of DNA fragments derived from nascent ChrAcc regions, corroborating earlier observations that potentially transient states coexist (Schlesinger et al., 2013). At early time points (0 to 24 hrs), dynamic ChrAcc regions were associated with a range of seemingly inert DNAme states, yet robust transcriptional responses of neighboring genes were observed, suggesting that DNAme does not immediately impact gene regulatory responses. Minimal loss of DNAme was observed at later time points (48 and 72hrs) and dramatic DNAme changes were only observed with the application of an artificial DNA demethylation stimulus (L-Ascorbic acid 2-phosphate, L-AA-2-P), suggesting a secondary consequential role for DNAme removal at GREs. Ultimately these studies are critical to disentangle the role of DNAme dynamics in normal cellular differentiation.

Design

To better elucidate spatiotemporal relationships between ChrAcc and DNAme, we required an approach that would capture both ChrAcc and DNAme from the same population of DNA molecules. Comparable approaches have been described (Kelly et al., 2012; Li and Tollefsbol, 2011), but these methodologies are technically prohibitive and limited in their ability to provide the resolution and coverage required to study dynamic enhancer regions. We required a sequence independent strategy (not reliant on specific recognition sites of enzymes), with the ability to directly probe the methylation status of accumulating ChrAcc fragments at transitioning enhancers. Because of its simple, cost effective protocol and low input

requirement, ATAC-seq (Assay for Transposase Accessible Chromatin) has become a widely adopted method for profiling ChrAcc regions genome-wide. In our adaptation of the approach, Tn5 enriched ChrAcc fragments are further subjected to sodium bisulfite conversion of unmethylated cytosines before library amplification and sequencing (Figure 3).

Preserving the fidelity of DNAme states during ATAC-Me library construction was crucial to this study. There are known opportunities during bisulfite library construction where the information content of the original DNA fragment can become artificially overwritten. Prior approaches for constructing WGBS libraries with Tn5 based methodologies have employed a conservative end-repair step involving gap filling and nick ligation to produce bisulfite converted, Illumina sequencing competent DNA libraries (Adey and Shendure, 2012; Wang et al., 2013). Alternative efforts to streamline bisulfite adaptation of Tn5 protocols have used an extension-based approach to create sequencing compatible libraries (Lu et al., 2015; Spektor et al., 2019; Suzuki et al., 2018). Specifically, an initial tagmentation with Tn5 transposomes assembled with asymmetric, methylated adapters is followed by a polymerase extension step with a 5-methyl-deoxycytidine nucleotide tri-phosphate blend. With fewer molecular steps required, this approach would simplify ATAC-Me library construction; however, several studies have reported detectable levels of exogenous cytosine methylation in the tagmented genomic DNA, effectively creating false methylation signals (Lu et al., 2015; Suzuki et al., 2018).

Taking advantage of Drosophila melanogaster genomic DNA (S2 cells), which is devoid of DNAme (Dunwell and Pfeifer, 2014), we compared the gap filling and nick ligation approach to the extension approach. Analysis of nucleotide base composition across reads from different bisulfite library construction conditions indicated that, in all replicates and conditions of extension-based approaches, cytosine signal was detected (Figure 4A,B). In contrast, a gap filling (using a 5-mC substituted dNTP (deoxyNucleotide Triphosphate) mix) and nick ligation approach produced the expected nucleotide base composition profile of bisulfite converted Drosophila DNA, virtually devoid of cytosine signal. Analysis of individual CpG methylation

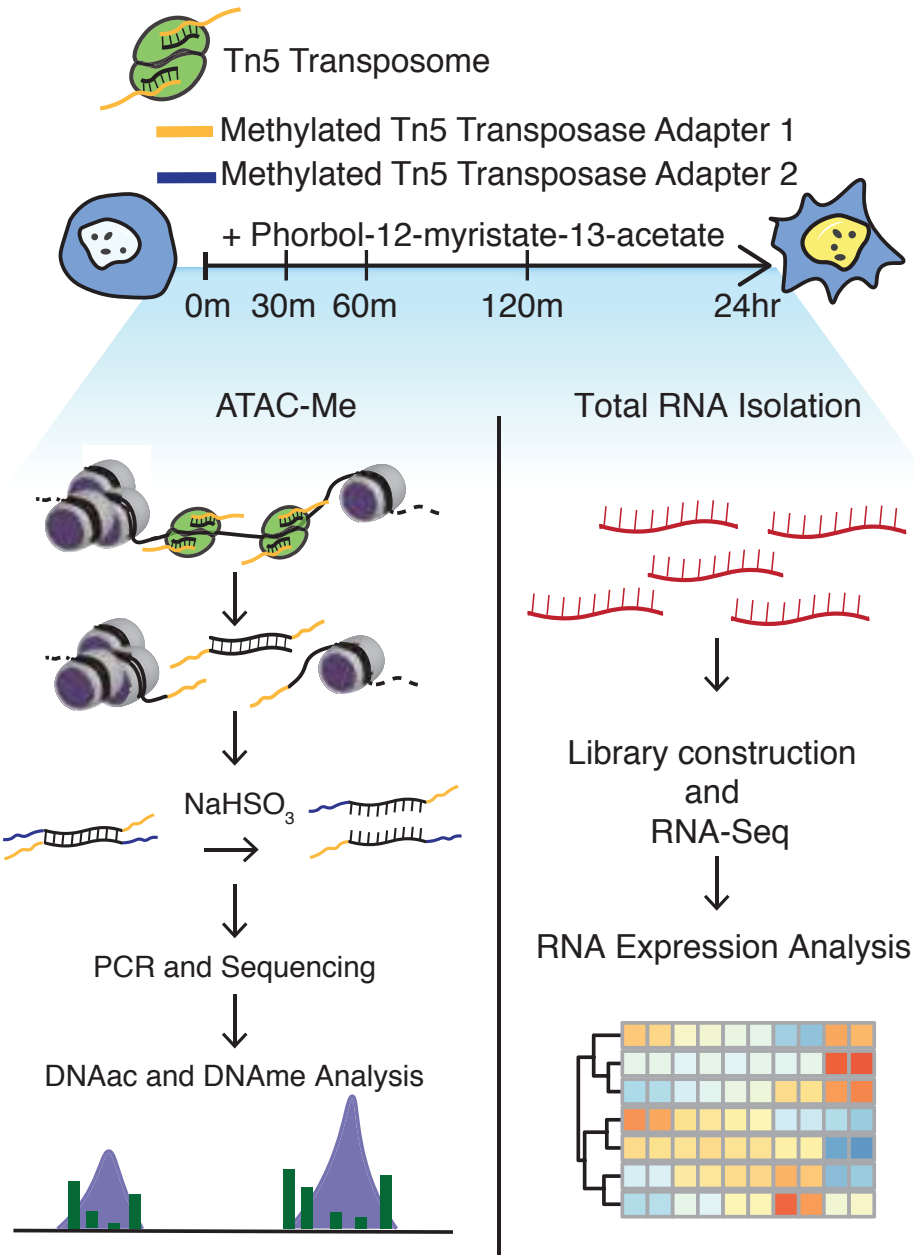

Figure 3. Methodology overview of ATAC-Me library construction and experimental design.

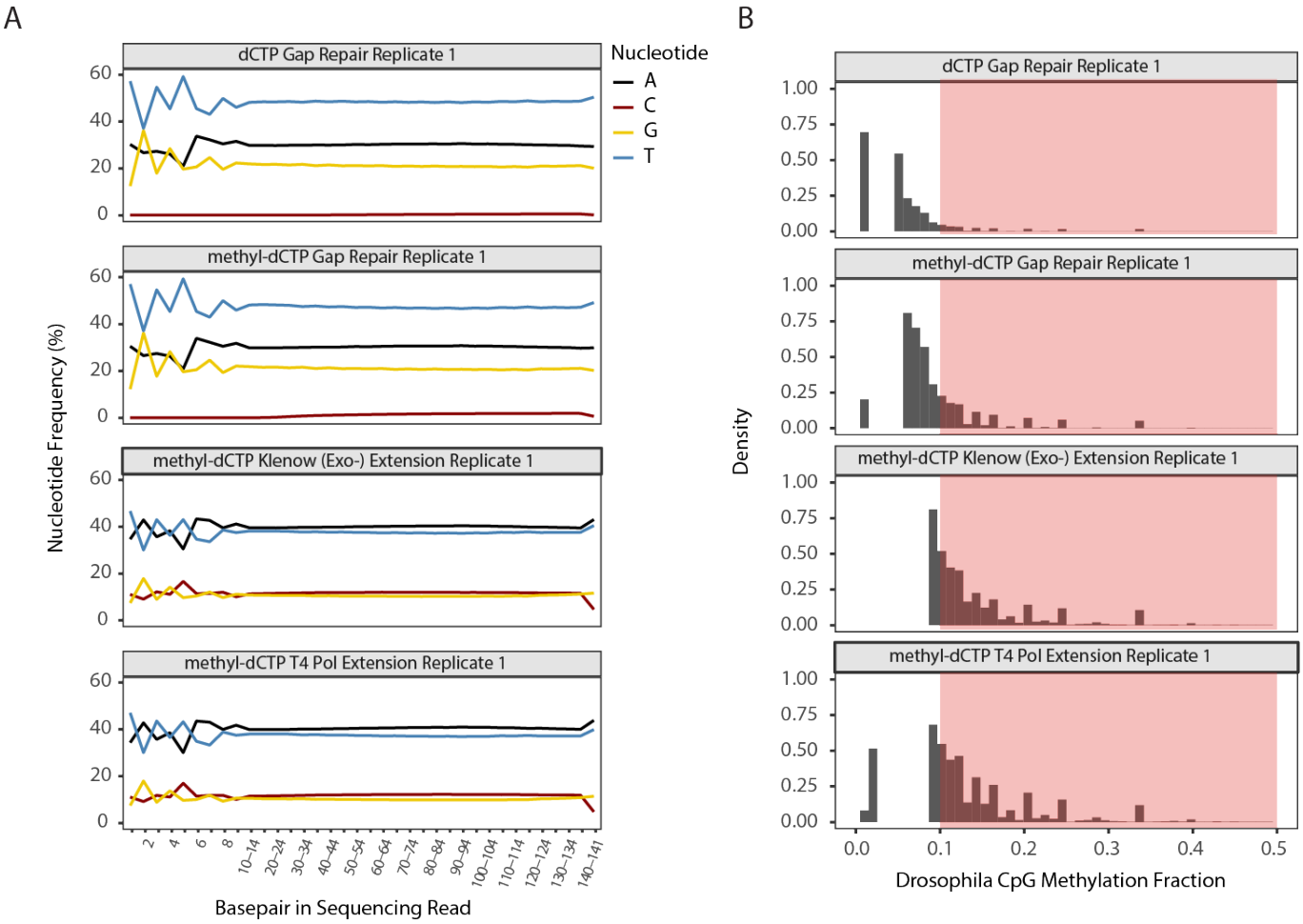

Figure 4. DNA methylation artifact analysis on Drosophila WGBS DNA libraries. A. Nucleotide base frequency across read-1 of 150 basepair illumina paired end sequencing for Tn5 based bisulfite converted libraries prepared from *Drosophila melanogaster* genomic DNA via various construction strategies. **B.** Histograms comparing CpG methylation fraction for Tn5 based bisulfite converted libraries prepared from *Drosophila melanogaster* genomic DNA via various construction strategies. The condition dCTP-Gap-Repair serves as experimental control for **A** and **B.**

levels revealed methylation frequencies greater than 10% for extension-based libraries compared to DNAme levels seen by gap filling and nick ligation with or without methyl-dCTP. Exogenous levels of methylation even at low levels could potentially confound our data; therefore, we opted to perform gap filling and nick ligation-based ATAC-Me.

Assessing the spatiotemporal relationship between ChrAcc and DNAme requires a dynamic, rather than steady state, cellular system. Approaches are not yet feasible that capture these relationships in real-time. Thus, we are required to sample DNA fragments at sequential time points, and the selection of those time points is an important component to our approach. To test our method, we chose THP1 cells, a well-established model system for studying inflammatory responses *in-vitro* (Tsuchiya et al., 1982). THP1s originate from a monocytic leukemia that, upon relatively short exposure to various stimuli including PMA, can be differentiated into naïve M(−) macrophages. In addition, these cells exhibit exceptionally low phenotypic heterogeneity, an important consideration when measuring epigenetic states from cell populations.

Nuclear run on sequencing studies, which map RNA polymerase activity at enhancers, have demonstrated that the immediate effects of TF signaling can occur within as little as 15 minutes upon stimulation (Hah et al., 2011). In addition, PMA is a robust stimulus at the applied concentration (100ng/mL, 162nM) as evident by a visual increase in cell adhesion in as little as 30 minutes. Therefore, we expect an immediate and primary gene regulatory response to PMA followed by a secondary more stable gene regulatory response specific to the macrophage transition. Given these considerations, we initially performed ATAC-Me on cells collected at 0, 30, 60, and 120 minutes, as well as 24 hours following PMA stimulation in order to capture both immediate and subsequent gene regulatory responses. In parallel, we collected RNA for each time point (Figure 3) in addition to standard ATAC-seq and WGBS data for key reference timepoints (0hr and 24hr). To validate our time course, we confirmed from RNA-seq data the up-

regulation of known markers of PMA-induced macrophage differentiation as early as 30 minutes post-PMA treatment (Figure 5).

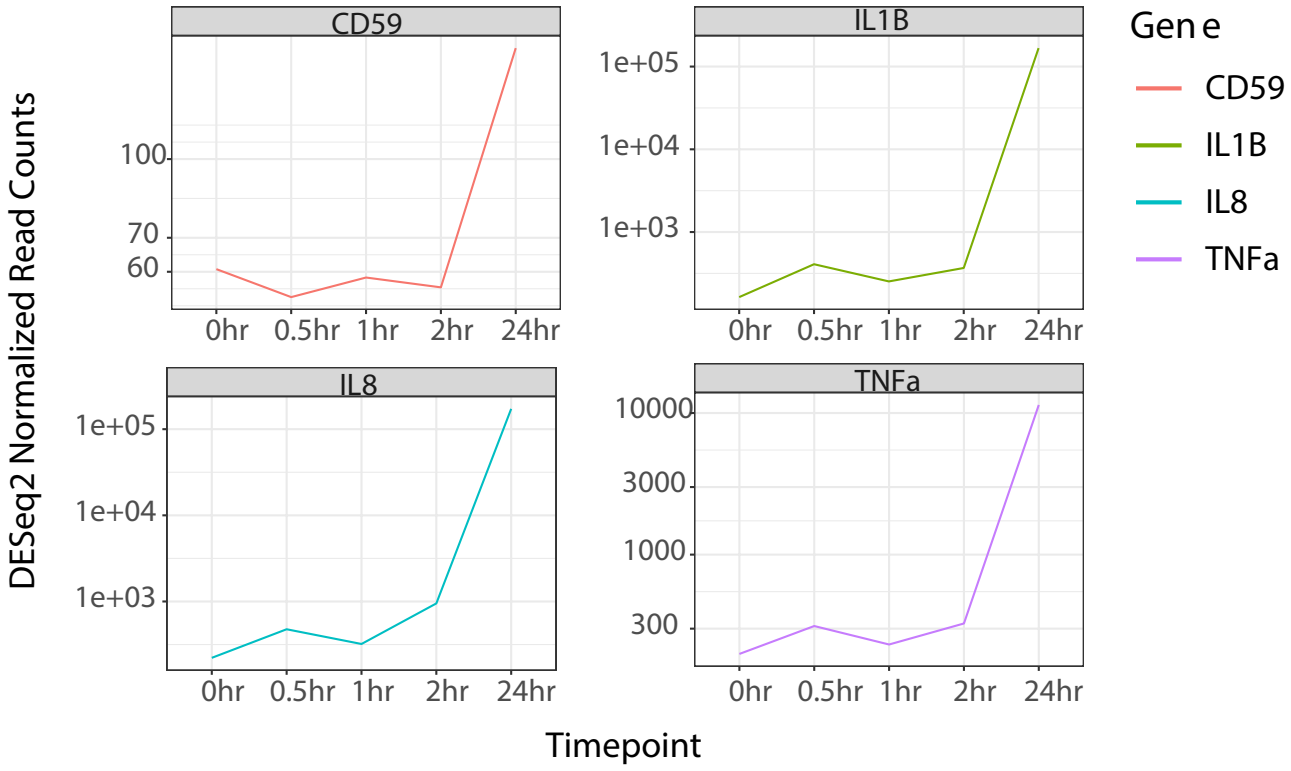

Figure 5. THP1 monocytes stimulated with PMA show characteristic markers of macrophage differentiation. Selected DESeq2 normalized transcript abundance measurements from RNA-seq data for known markers of THP1 monocyte to macrophage differentiation.

Results

ATAC-Me is reproducible and recapitulates standard ATAC-seq

Bisulfite treatment is known to be detrimental to DNA integrity, potentially leading to loss of material and the canonical ATAC-seq fragment size distribution. When comparing fragment size distributions for ATAC-Me with standard ATAC we observe a distinct loss of material, however, the DNA fragment size range (<130bp) most enriched for chromatin accessible fragments is preserved with the greatest fragment losses representing mono-, di-, and tri-nucleosome fragments (Figure 6A,B; Figure 7A,B). Nevertheless, distinct mono-nucleosome signal can be observed within the ATAC-Me size distribution, indicating nucleosome patterns proximal to open chromatin peaks are preserved to some extent.

We used MACS2 to determine ATAC-Me peak locations genome-wide, detecting on average 57,217 high confidence, reproducible peaks across all time points in accordance with ChrAcc peak estimates from earlier ATAC-seq studies (Buenrostro et al., 2013). Only peak regions reproducible in both individual and merged ATAC-Me replicates were selected for analysis. We compared the performance of ATAC-Me to standard ATAC-seq, generating libraries in parallel and in duplicate from the same population of time point specific nuclei (0 and 24hr). Initial comparisons of peak calls between the two methodologies demonstrated that over 75% of peak regions in standard ATAC-seq were replicated by the paired ATAC-Me data set (Figure 8A,B). Further, we observed high concordance between peak signals in both datasets as well as between replicates (Figure 9; Figure 10; Figure 11). Whereas twice the number of peaks is called in ATAC-Me versus standard ATAC-seq, manual and global comparison of read counts at ATAC-Me only peak regions reveals high correspondence between ATAC-Me and standard ATAC-seq (Figure 12; Figure 13), indicating the discrepancy in peak calling is likely

due to differences in local background estimates utilized by the peak calling algorithm, as discussed previously (Zhang et al., 2008).

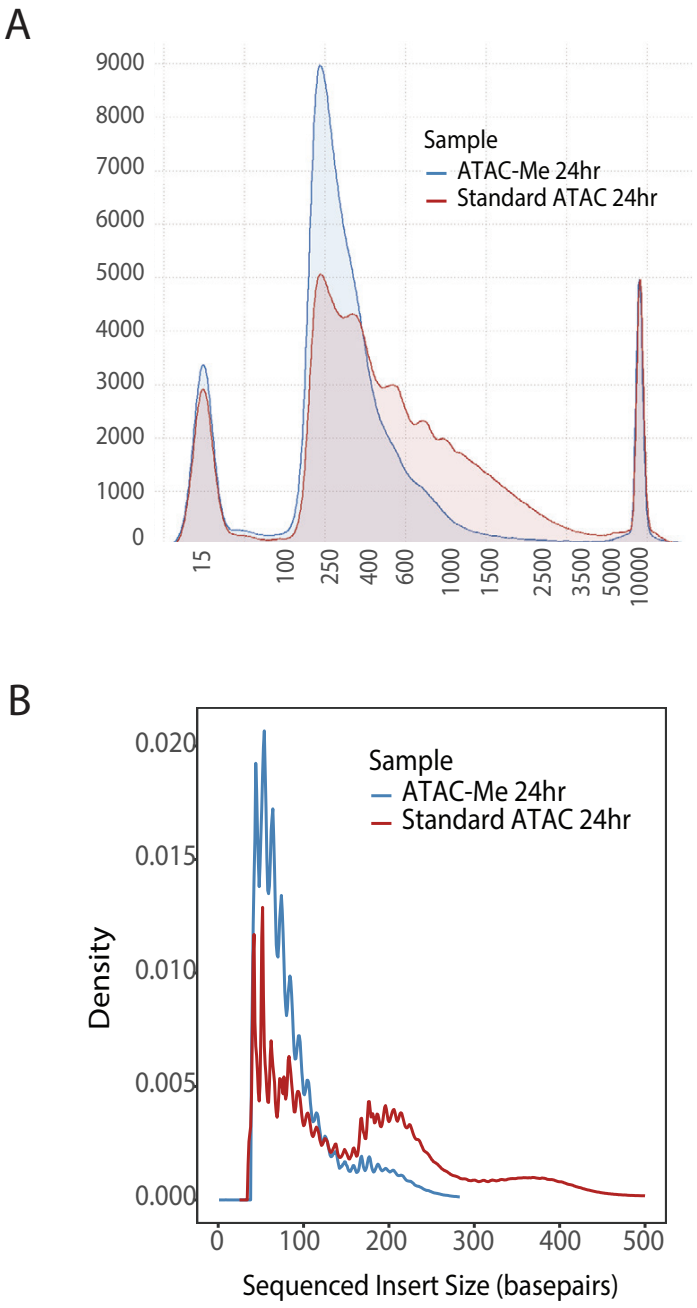

Figure 6. Comparison of DNA fragment sizes between ATAC and ATAC-Me DNA libraries. A & B. Genomic insert size density distribution of ATAC-Me or standard ATAC fragments pre- (A) or post- (B) sequencing.

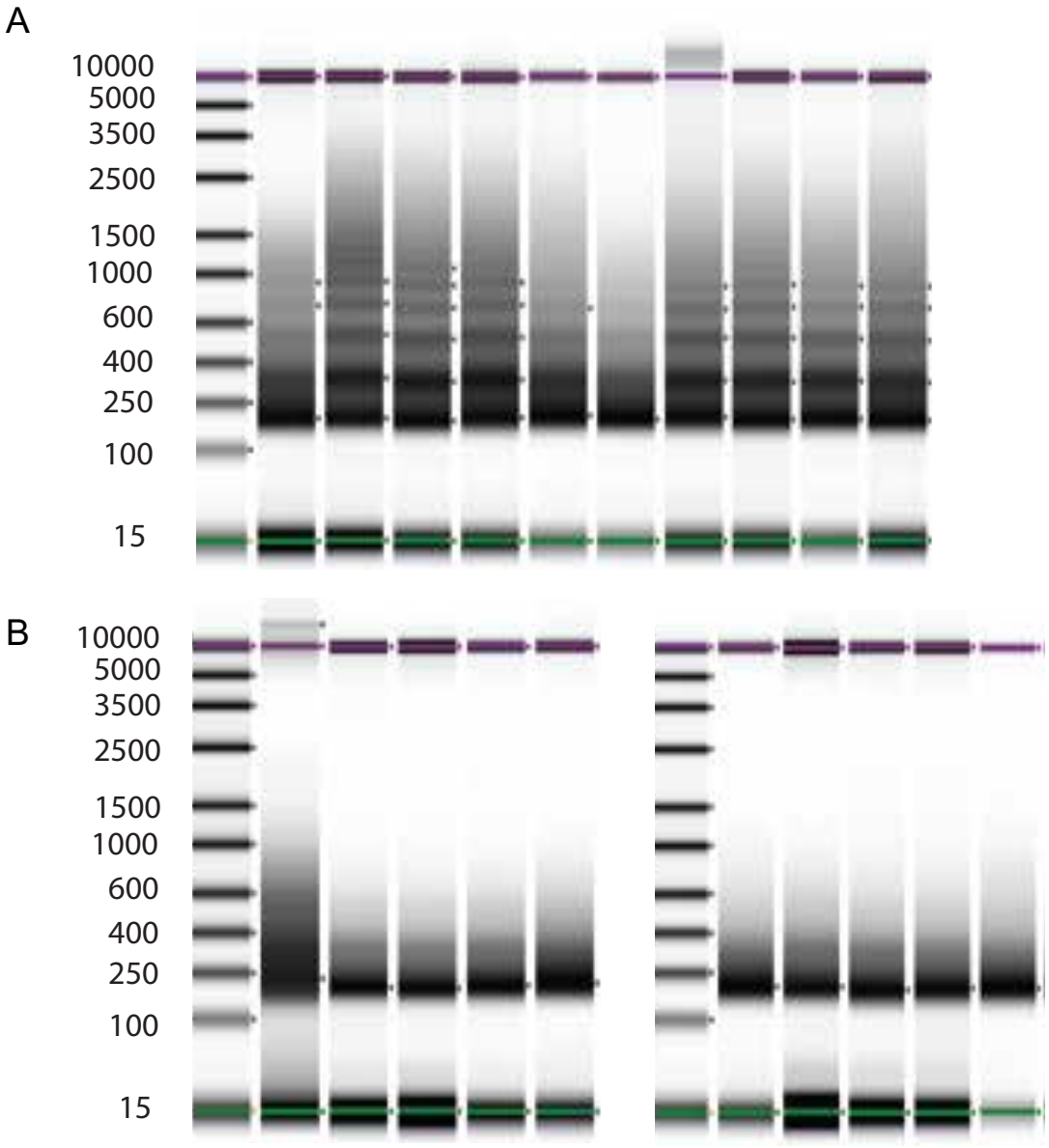

Figure 7. Capillary electrophoresis comparison of DNA fragment sizes between ATAC and ATAC-Me DNA libraries. A & B. Tapestation gel image of standard ATAC (**A**) or ATAC-Me (**B**) libraries pre-sequencing. Peaks/bands at approximately 15 and 10,000 base pairs are internal sizing markers.

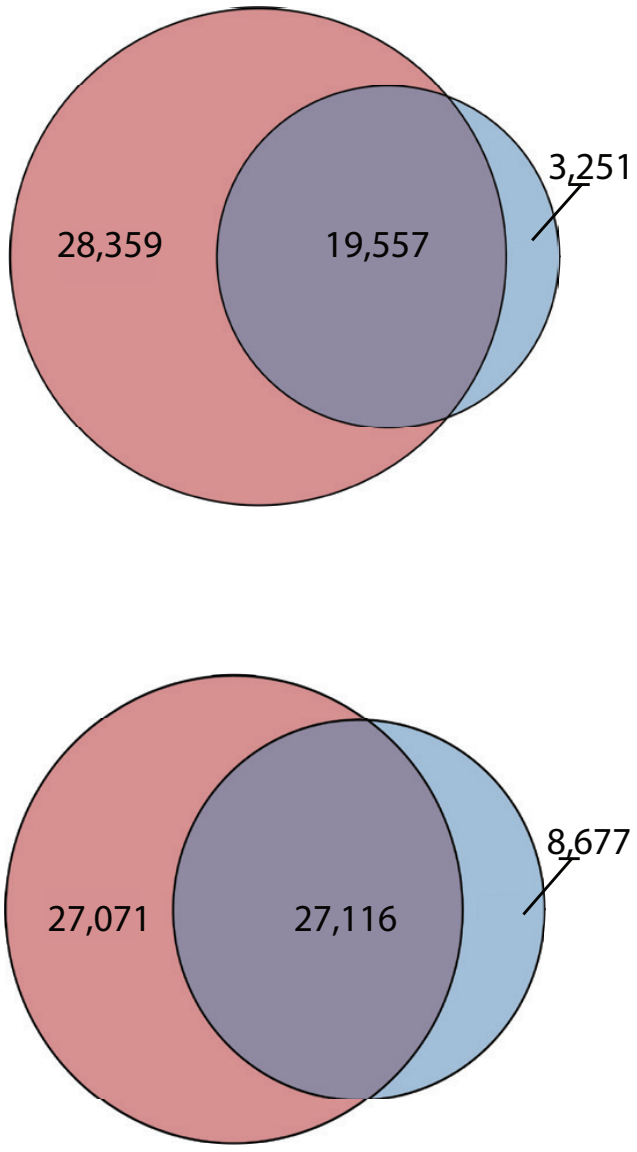

Figure 8. Comparison of MACS2 chromatin accessibility peak overlap between ATAC and ATAC-Me sequenced DNA libraries. A. & B. Venn diagram of broadpeak genomic intervals that overlap between standard ATAC and ATAC-Me for 0hr (A) and 24hr (B) datasets. Intervals must overlap at least 50% with standard ATAC to be counted.

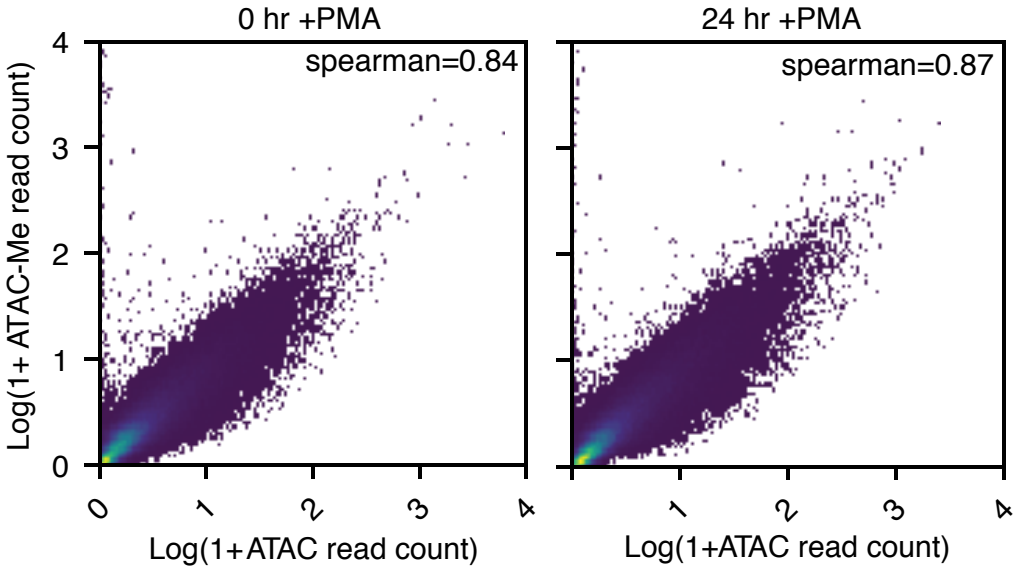

Figure 9. Spearman-rank comparison of standard ATAC and ATAC-Me read counts at 0hrs or 24hrs PMA stimulation within a common set of broadpeak accessible regions.

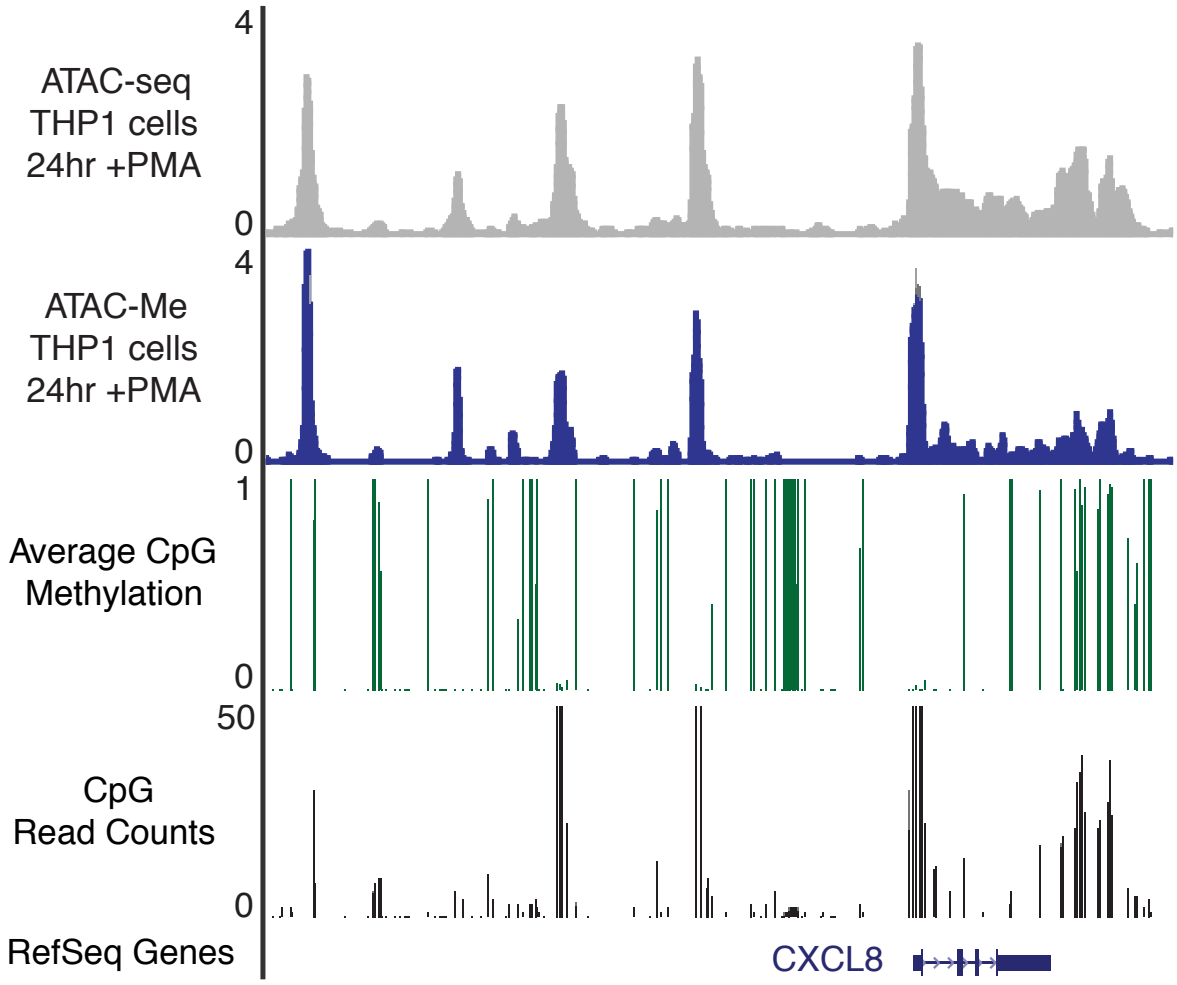

Figure 10. UCSC genome browser profiles comparing information content of ATAC and ATAC-Me sequenced DNA libraries. Genome browser profile of the *CXCL8* (Interleukin-8 gene) locus displaying ATAC-seq accessibility signal (gray), ATAC-Me accessibility signal (blue), CpG methylation frequency (green), and CpG read abundance (black). All data shown have been read depth normalized.

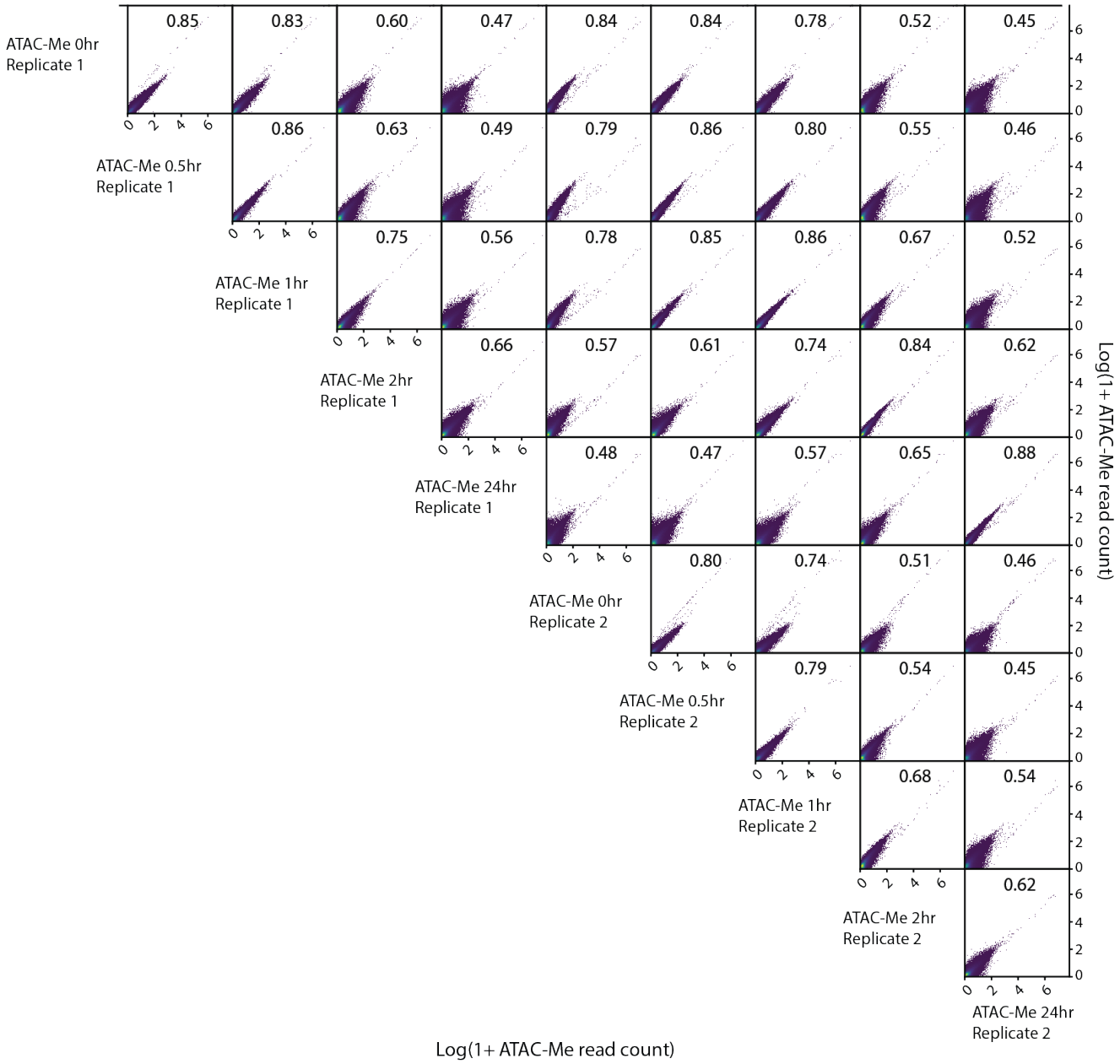

Figure 11. ATAC-Me biological replicate reproducibility comparison. Spearman-rank comparison of ATAC-Me PMA stimulation time course biological replicates within a common set of broadpeak accessible regions. Spearman correlation scores for each pairwise comparison are overlaid. Sequencing read counts were normalized for library depth prior to computation of correlation scores.

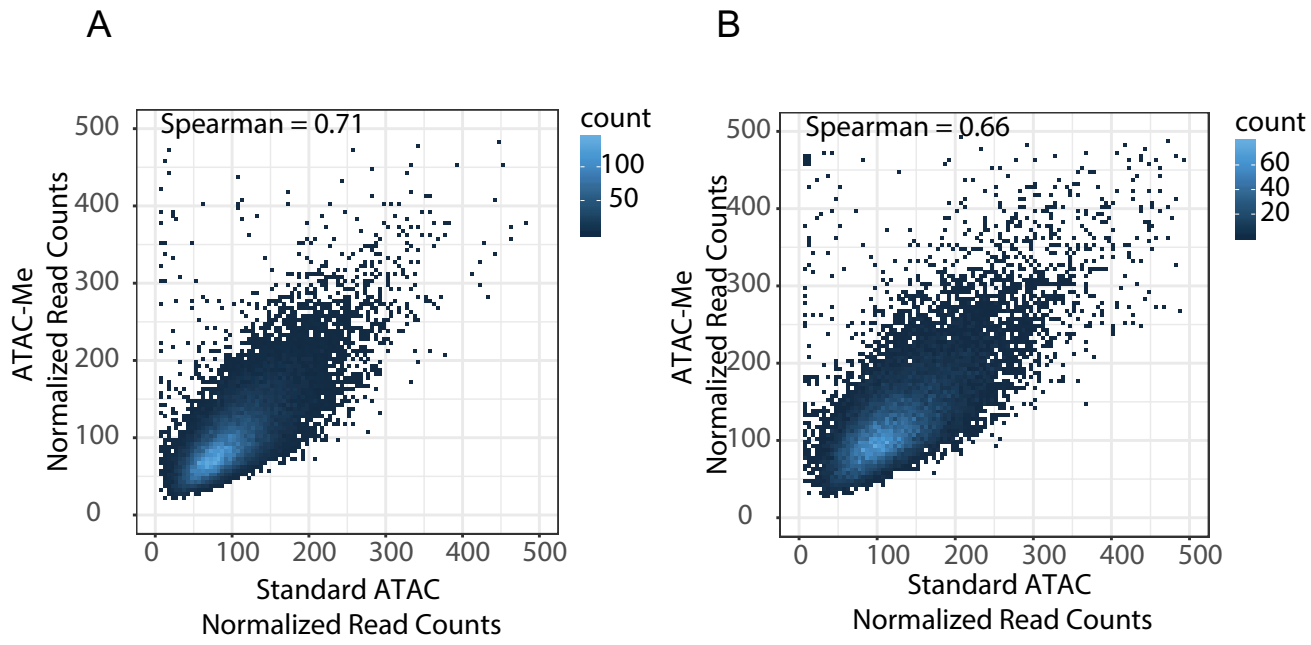

Figure 12. Correlation plot of read counts within chromatin accessible regions exclusive to ATAC-Me. A & B. Correlation plot of read counts within broadpeak regions exclusive to ATAC-Me for 0hr (A) and 24hr (B) datasets.

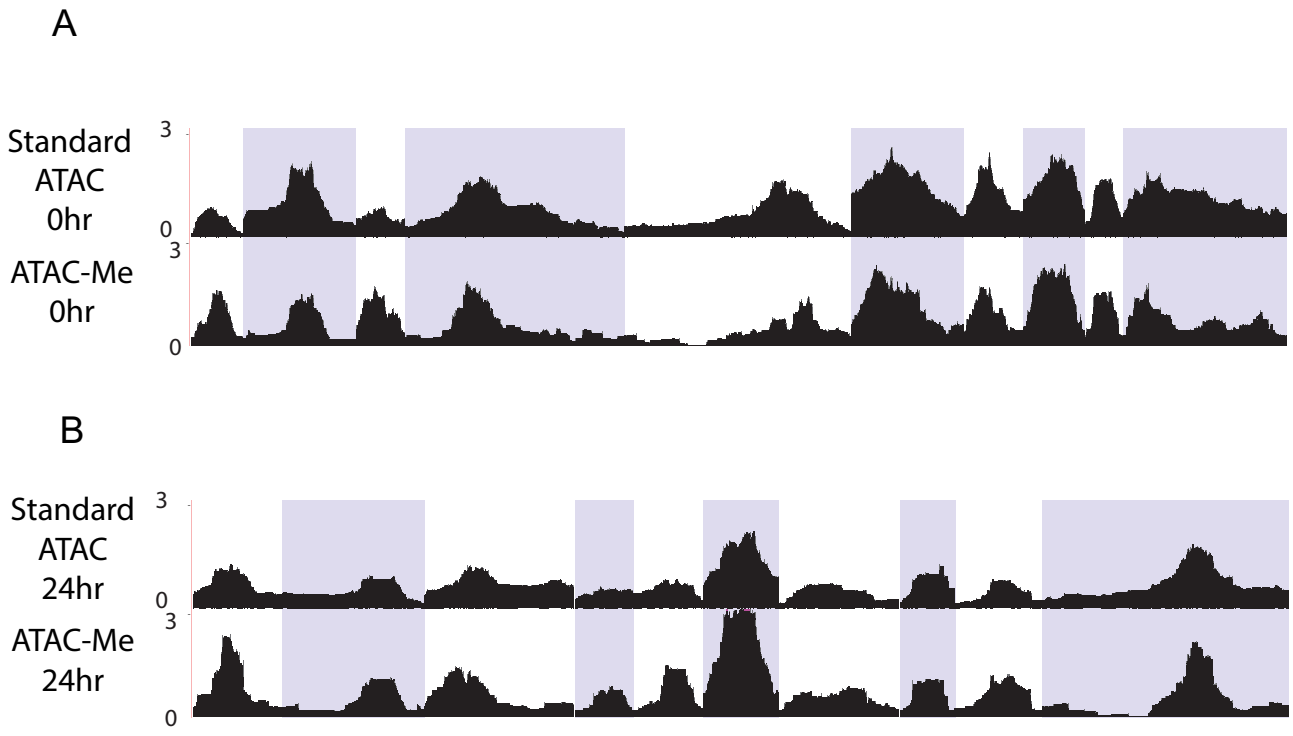

Figure 13. UCSC genome browser comparison of chromatin accessible regions exclusive to ATAC-Me. A & B. UCSC genome browser view of chromatin accessibility signal at 10 randomly selected ATAC-Me broadpeak regions that did not overlap sufficiently with standard ATAC broadpeaks.

ATAC-Me accurately profiles the methylome of open chromatin

Using ATAC-Me, we measured methylation at all CpG sites located within Tn5 enriched DNA fragments for every time point collected. In parallel, we performed tagmentation-based WGBS on genomic DNA collected at 0 minutes and 24 hours in order to create gold standard references for methylation of both end points (Figure 14A). We compared ATAC-Me and WGBS methylation distributions within ChrAcc peak regions identified at 0hr and 24hr. As expected, methylation levels of ATAC-Me and WGBS are largely concordant within peak regions and extensively hypomethylated compared to the bimodal distribution of methylation levels typically observed in WGBS (Figure 14A).

WGBS data represents an average measurement of DNAme without consideration of the actual *in-cellulo* accessibility of DNA fragments. For instance, allelically regulated regions (ARRs) are known to consist of one hypomethylated, chromatin accessible allele and one hypermethylated, inaccessible allele (Martos et al., 2017; Shibata et al., 1996; Stern et al., 2017), which would be 50% methylated according to WGBS. Accordingly, our ATAC-Me methodology should selectively enrich for the chromatin accessible, hypomethylated allele within these ARRs compared to WGBS. To determine the specificity of ATAC-Me to measure methylation of accessible DNA, we compared methylation levels of known ARRs, including imprinting control regions (Fang et al., 2012). Indeed, we detected significantly lower methylation levels of ARRs profiled by ATAC-Me compared to those measured by WGBS (Figure 14B). The implications of these observations are that 1) ATAC-Me is highly selective for accessible DNA fragments, offering a more incisive view of methylation states within the context of chromatin compared to WGBS, and 2) observed methylation of accessible fragments is not due to random noise.

In general ATAC-Me CpG methylation levels within steady state, ChrAcc regions are highly concordant with WGBS. However, outside of defined peak regions, low level noise resulting from molecular “breathing” of histone bound DNA as well as spurious exposure of DNA

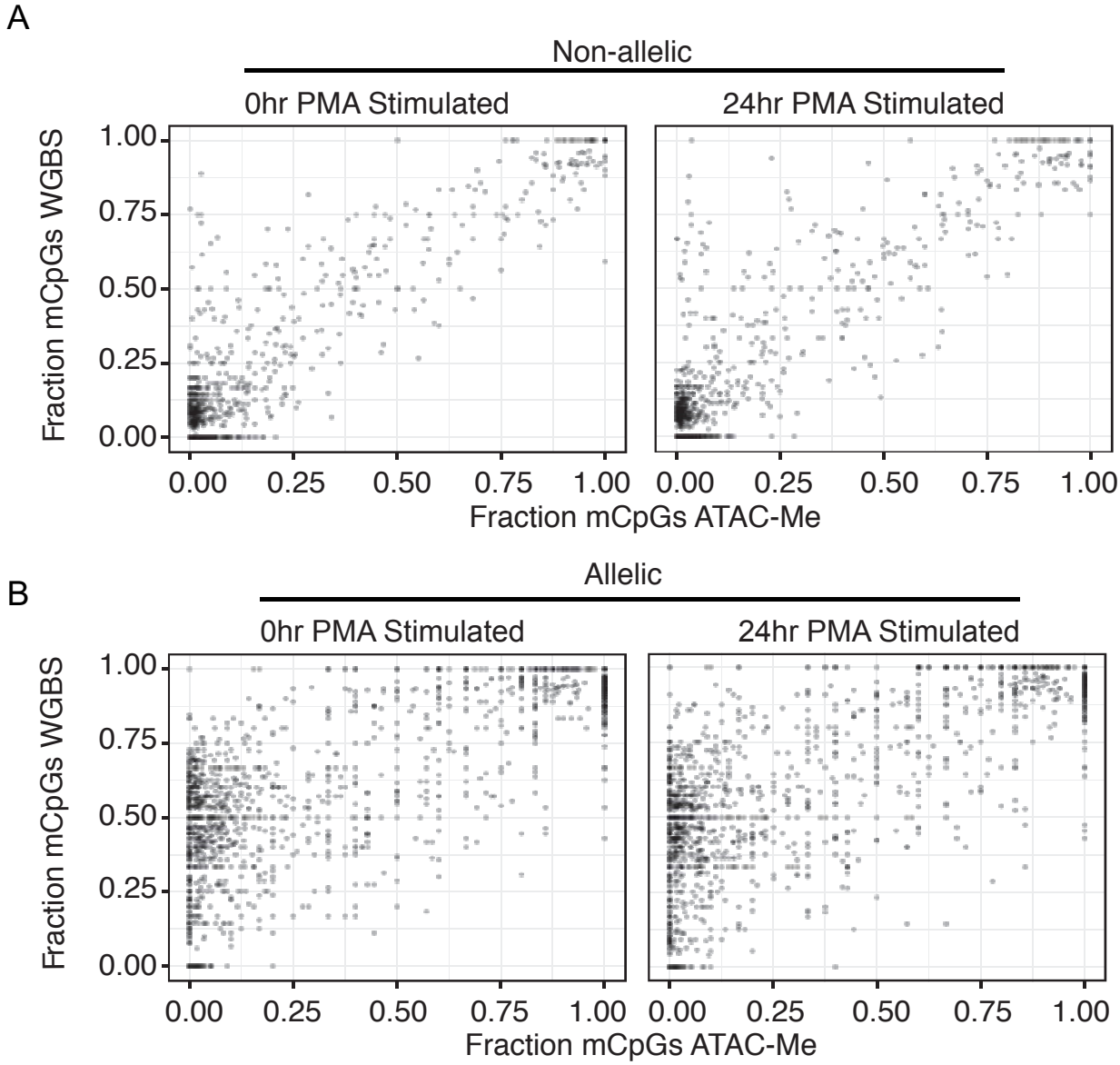

Figure 14. Methylation scatterplot comparison between WGBS and ATAC-Me for CpGs in non-allelic or allelic methylated regions. A. Scatterplot of mCpG (5-methylCytosine-phosphate-Guanine) fraction in THP1 cells calculated with WGBS or ATAC-Me data for non-allelic CpGs (Pearson = 0.92, $p < 0.0001$, 0hr; Pearson = 0.91, $p < 0.0001$, 24hr). **B.** CpGs within validated allele-specific methylated (Pearson = 0.80, $p < 0.0001$, 0hr; Pearson = 0.80, $p < 0.0001$, 24hr) regions. The CpG subset that displayed the expected intermediate methylation state for AMR CpGs as calculated from WGBS data was not correlated with ATAC-Me CpG methylation (Pearson = 0.02, $p = 0.59$, 0hr; Pearson = 0.02, $p = 0.64$, 24hr). See also Figure S1,S2,S3.

to Tn5 transposase results in low read coverage of select CpGs within inaccessible chromatin regions where DNAme may be determined with some accuracy (Figure 10). Comparison of ATAC-Me read “noise” outside of defined peak regions with WGBS data revealed a surprisingly high level of concordance despite locally minimal read depth in these “no peak” regions (Figure 15), thus demonstrating further utility for ATAC-Me to partially interrogate the DNAme state of long stretches of inaccessible chromatin.

Overall, ATAC-Me uniquely permits methylation of accessible chromatin regions to be determined, where TFs and other proteins are known to interact with DNA, thus providing enhanced biological context to DNAme patterns compared to traditional WGBS. As sequence coverage is highly focused on accessible DNA fragments, probing DNAme of relevant targets requires less sequencing than orthogonal methodologies. Similar to restriction enzyme based methods (ERRBS (Akalin et al., 2012), enhanced reduced representation bisulfite sequencing), the approach is essentially a selective representation of the methylome; yet, unlike RRBS (Meissner et al., 2005), ATAC-Me is agnostic to sequence context, thereby yielding better coverage of CpG sites in relevant GREs (namely enhancer regions with lower CpG densities). Additionally, requisite sequencing depths are equivalent to those of standard ATAC-seq, making ATAC-Me a cost-effective way to examine methylation of accessible DNA. As expected, the sequence complexity (or number of distinct reads) in ATAC-Me libraries is lower compared to standard ATAC-seq, yet substantially higher than the complexity of ERRBS libraries sequenced to corresponding read depths (Figure 16). Sequence and mapping statistics are provided in Table 2.

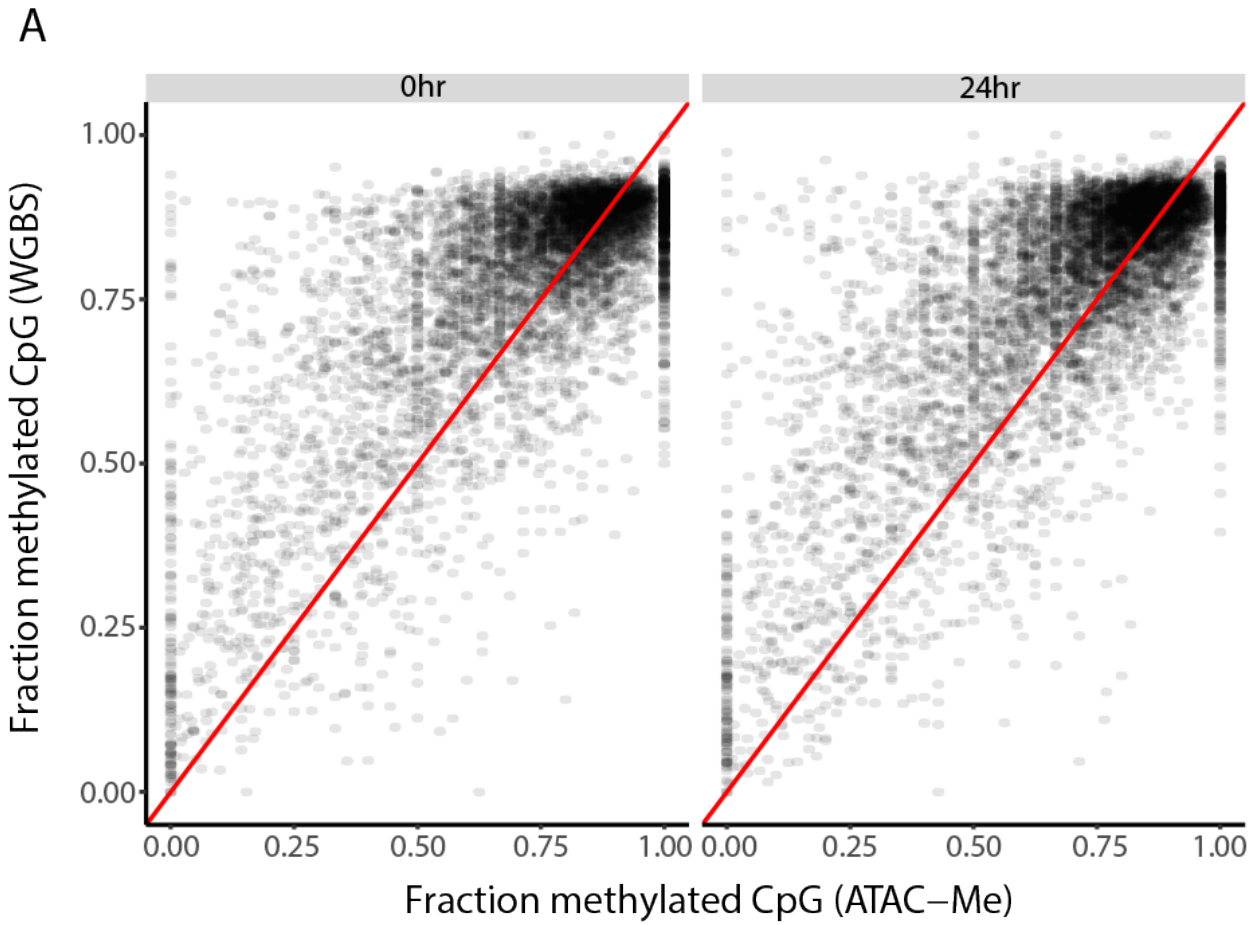

Figure 15. Methylation scatterplot comparison between WGBS and ATAC-Me for CpGs in chromatin inaccessible promoters. CpG methylation levels within inaccessible promoters computed with either ATAC-Me or WGBS data. Inaccessible promoters were determined by the absence of a MACS2 called peak region within ATAC-Me data. Red line represents y=x trend for comparison.

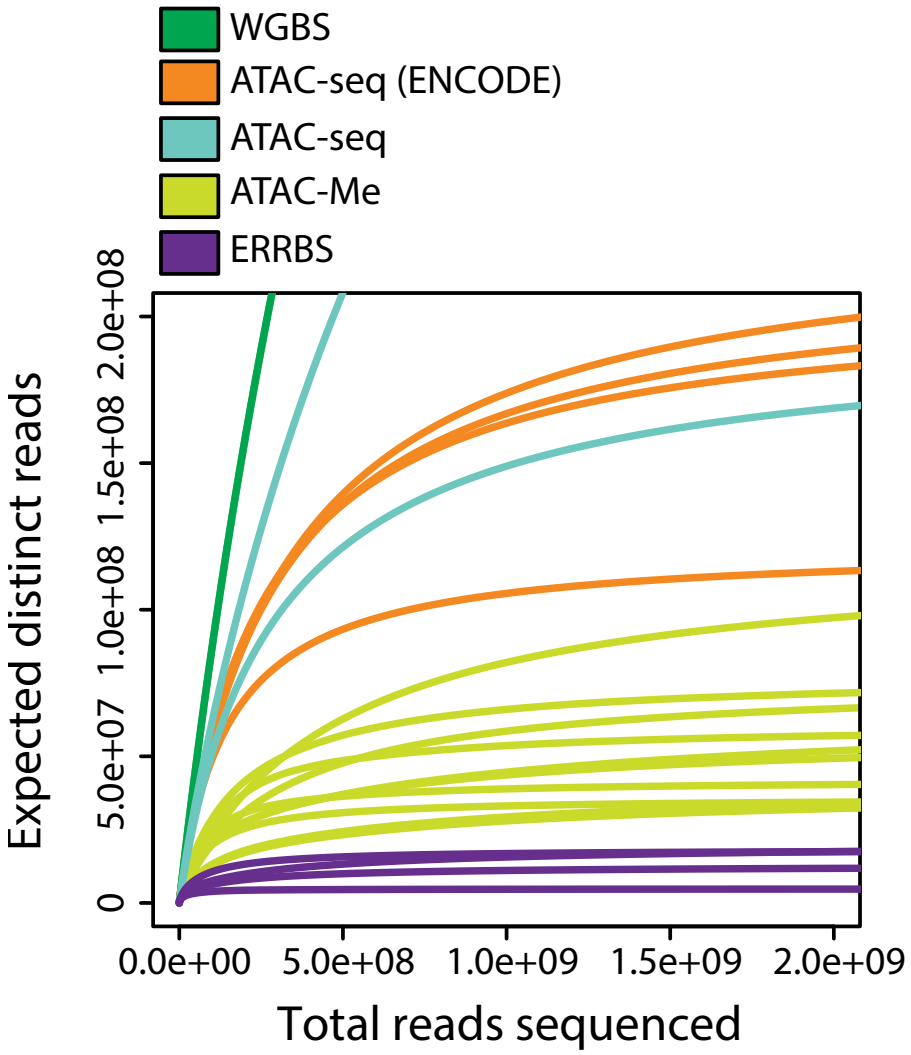

Figure 16. DNA library complexity analysis of various sequencing methodologies for DNA methylation profiling. Library complexity is measured as a function of distinct reads per total reads sequenced for various sequencing depths. Different sequencing depth datasets were generated by sub-sampling *in-silico* with the Pre-seq software algorithm.

Sample	Total Reads	Uniquely Mapped	% Uniquely Mapped	Symmetric CpG Frac Cov.	Symmetric CpG Mean Depth Cov.	Bisulfite Conversion Efficiency
ATAC-Me 0hr PMA Replicate 1	125037141	74904544	60	0.29	8	0.984
ATAC-Me 0hr PMA Replicate 2	245897016	124401123	51	0.23	9	0.985
ATAC-Me 0.5hr PMA Replicate 1	70640568	43379882	61	0.26	8	0.987
ATAC-Me 0.5hr PMA Replicate 2	183697986	94560875	51	0.27	9	0.988
ATAC-Me 1hr PMA Replicate 1	99960405	56033935	56	0.29	8	0.987
ATAC-Me 1hr PMA Replicate 2	77459408	47205838	61	0.24	8	0.988
ATAC-Me 2hr PMA Replicate 1	103135761	56003077	54	0.27	9	0.986
ATAC-Me 2hr PMA Replicate 2	229121882	119196888	52	0.30	8	0.988
ATAC-Me 24hr PMA Replicate 1	141269841	77056170	55	0.29	9	0.983
ATAC-Me 24hr PMA Replicate 2	186493086	99613032	53	0.12	8	0.985
WGBS 0hr PMA	445928612	316817178	71	0.96	14	0.989
WGBS 24hr PMA	457892152	325385211	71	0.96	15	0.989
ATAC-Me Average	146271309	79235536	55	0.26	8	0.986

Table 2. Read mapping, bisulfite Conversion, and coverage statistics for bisulfite sequencing DNA libraries. Total sequence reads acquired, uniquely mapped sequence reads, fraction of symmetric context CpGs covered genome wide, mean coverage depth for symmetric CpGs covered and bisulfite conversion efficiency for all bisulfite converted libraries including ATAC-Me and WGBS. Average across all ATAC-Me DNA libraries for each reported metric is included in the last table row.

Rapid changes in chromatin accessibility correspond to TF activity at putative monocyte and macrophage enhancers

Profiling DNAme in the context of chromatin under dynamic conditions is a cornerstone of our approach; therefore, we applied ATAC-Me to a time course of THP1 exposure to PMA, which induces terminal monocyte to macrophage differentiation. To understand temporal peak behaviors, we employed TCseq, an R package that enables quantitative, differential analysis and clustering of count-based epigenomic and transcriptomic datasets (Wu and Gu, 2019). Within the TCseq framework, we used C-means clustering to identify five specific groups of unique peak regions with different accessibility behaviors across the 5-point time course (Figure 17A). These groups display distinct behaviors from one another, where peak amplitudes reflect (1) Gradual Closing Response, (2) Transient Response, (3) Early Persistent Response, (4) Gradual Opening Response, And (5) Late Response of chromatin peak regions (Figure 17B; Figure18). Interestingly, only ~12,000 out of 90,000 unique peak regions demonstrated a dynamic temporal behavior, the majority of which are intergenic or intronic and significantly depleted in promoter proximal regions (Figure 19).

To assess the probability that similar clustering patterns are produced by random chance from our data, read counts were scrambled between time points and reanalyzed with time structure removed (Figure 20A). C-means max probabilities for clusters obtained from scrambled data were much lower than for non-scrambled data (Figure 20B), indicating the temporal specificity of our differential peak calls. We next examined if the temporal trends of different ChrAcc clusters determined entirely with ATAC-Me data would hold true if we substitute key time points (0hr and 24hr) with standard ATAC-seq data. Temporal trends among clusters remained strikingly similar even when recalculated with standard ATAC substitute data, again supporting the concordance of ATAC-Me with standard ATAC for profiling ChrAcc dynamics (Figure 21A,B).

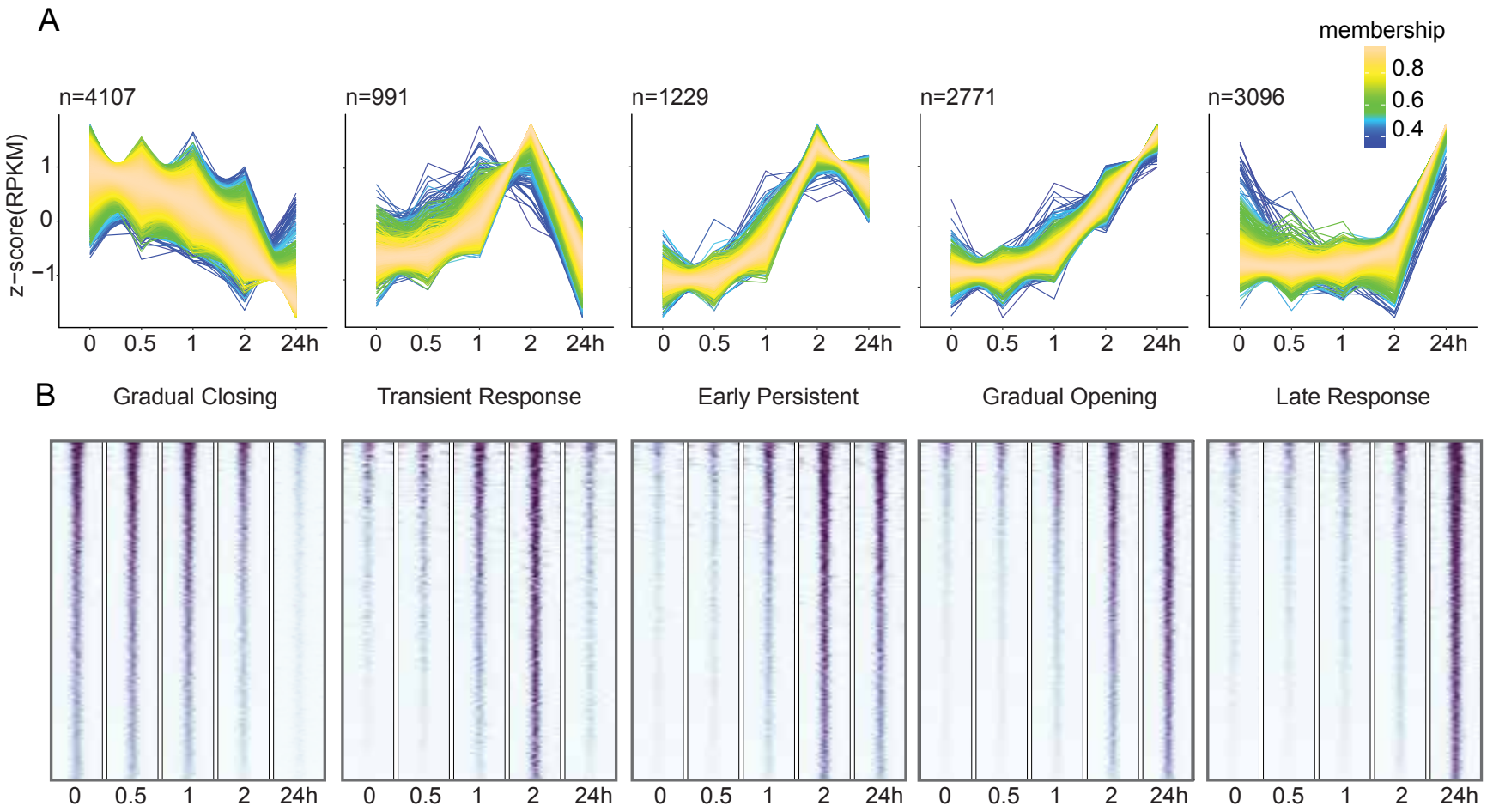

Figure 17. Dynamic chromatin accessibility identified during PMA stimulated gene activation. A. C-means clustering of ATAC-Me time course accessibility data reveals multiple categorizations of dynamic accessibility behavior. Membership indicates the goodness of fit for a region in a particular cluster. **B.** Heatmap of ATAC-Me accessibility signal at each time point for the dynamically accessible regions identified with corresponding C-means clustering.

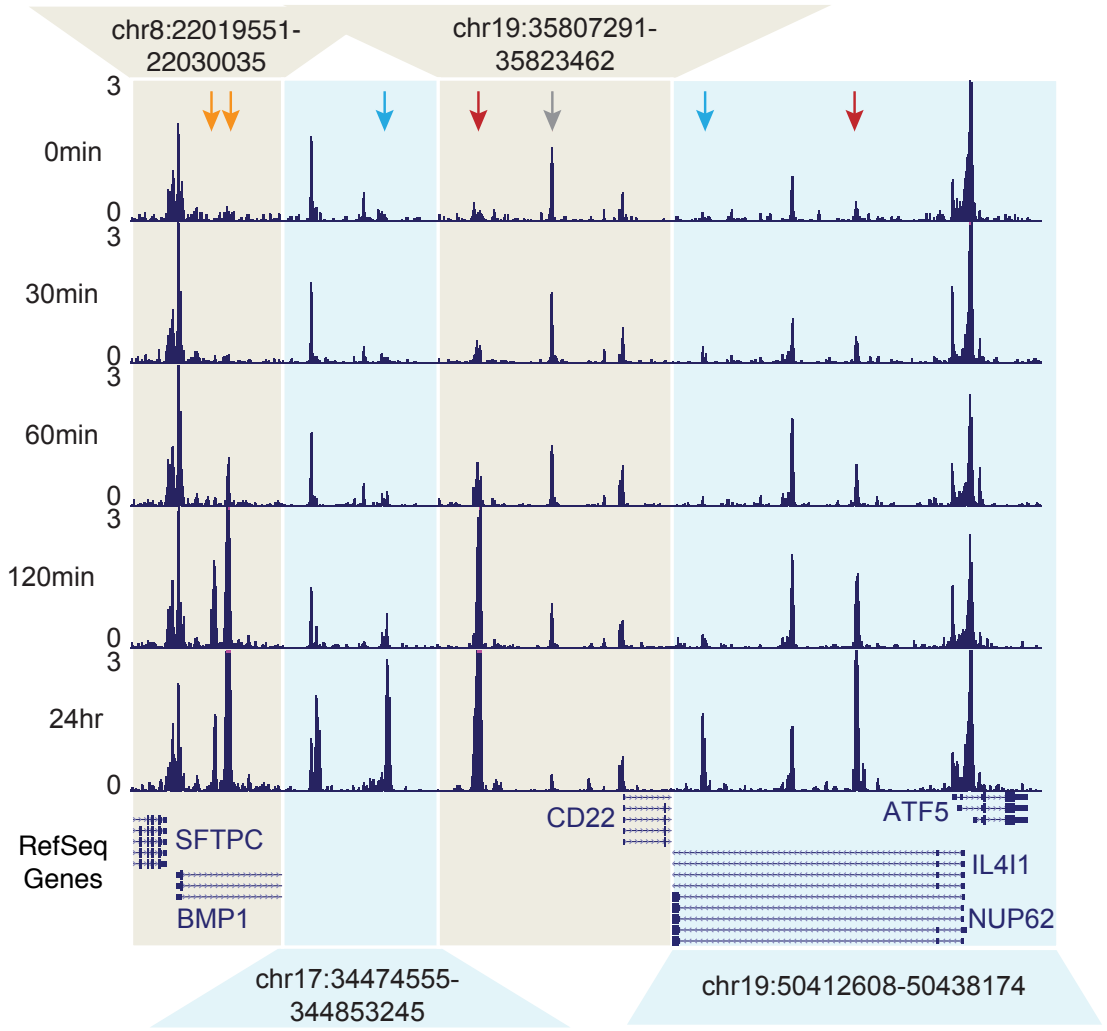

Figure 18. UCSC genome browser views of dynamic chromatin accessibility identified during PMA stimulated gene activation. UCSC genome browser non-contiguous, "Multi-region", view of multiple dynamic loci across time. Colored arrows indicate dynamic peak behaviors: Transient Response (yellow), Late Response (blue), Gradual Opening Response (red), Gradual Closing Response (gray).

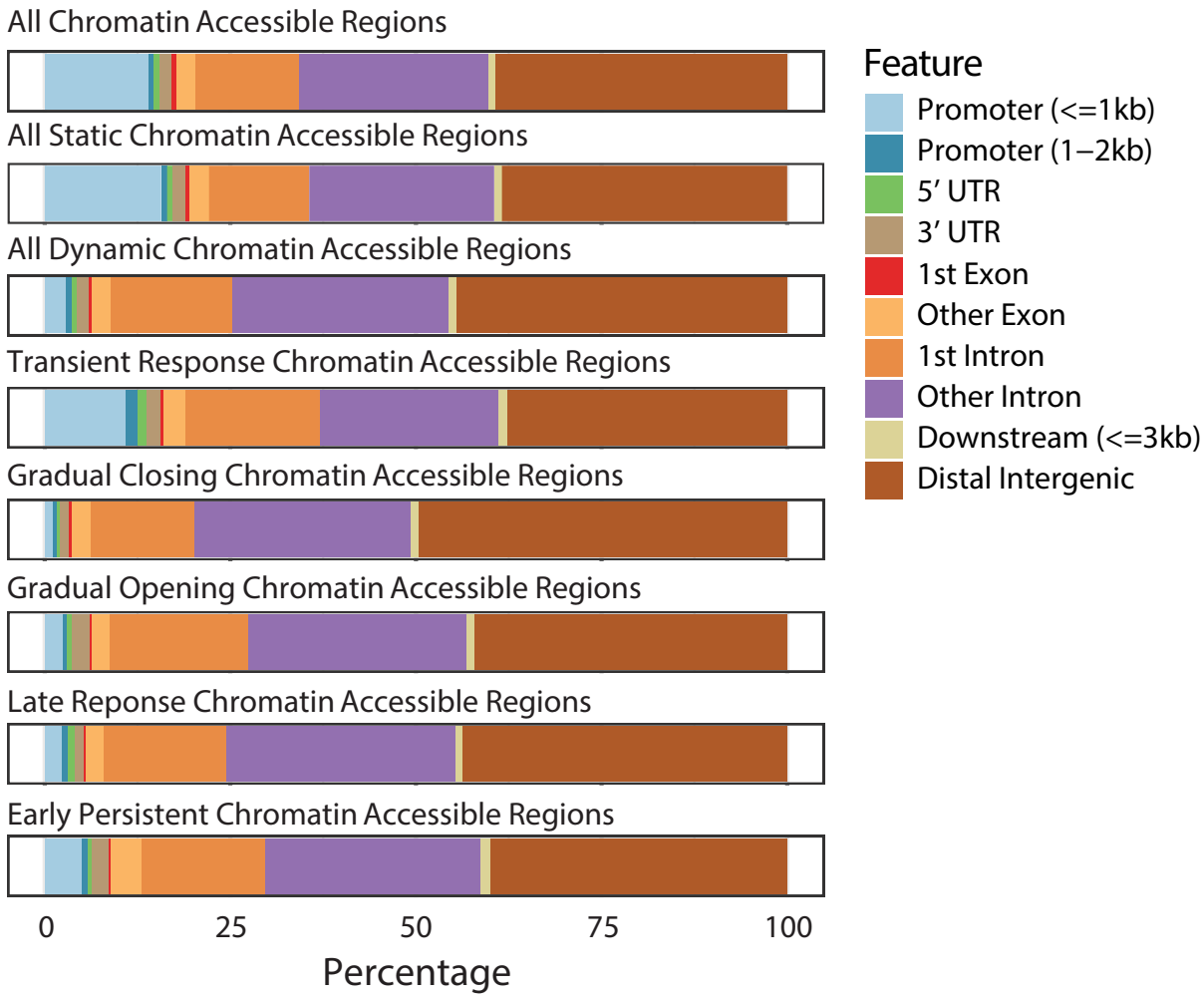

Figure 19. Genomic annotation of chromatin accessibility peak region sets. Annotation of regions was performed with the ChIPseeker R package. Promoters were defined as 1000bp upstream and 500bp downstream of previously reported transcription start sites.

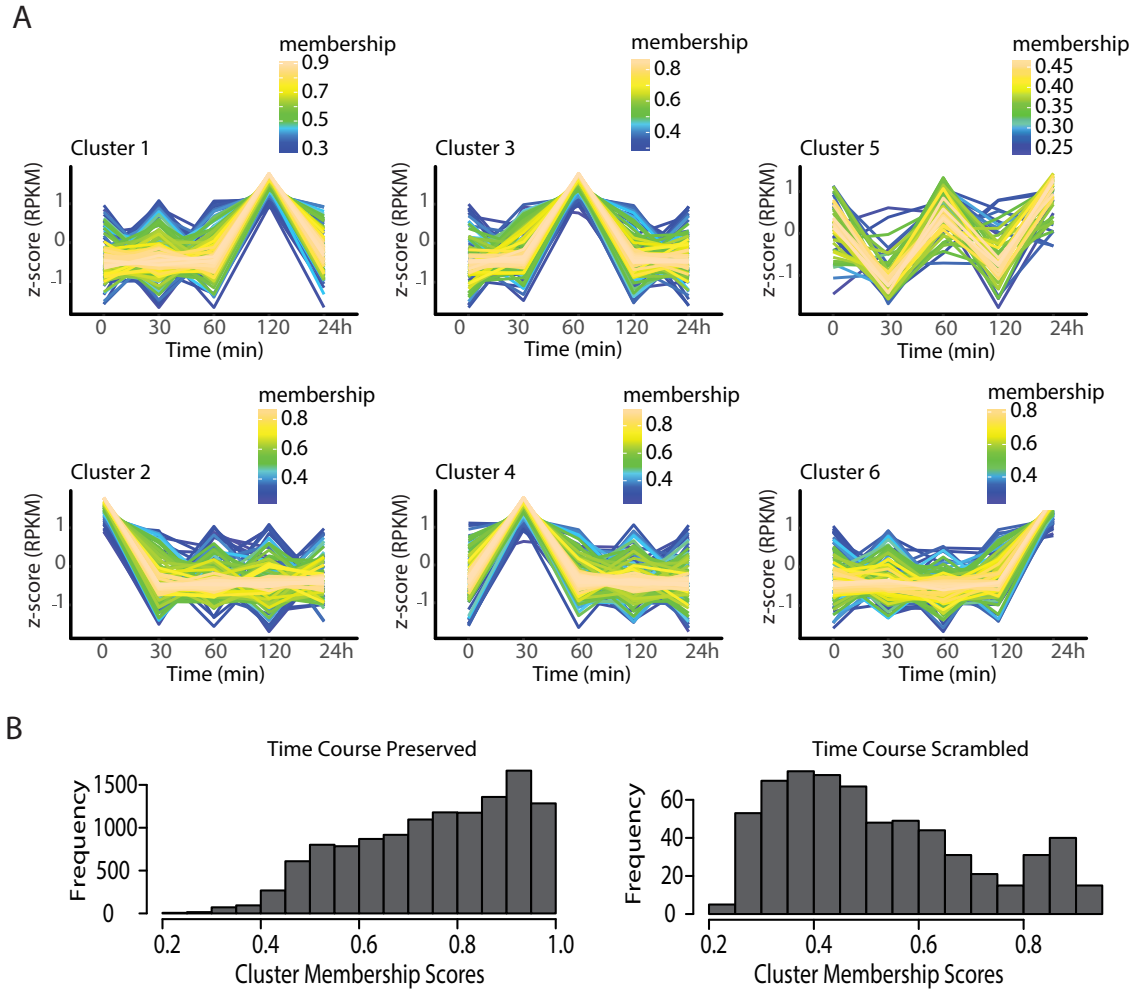

Figure 20. Chromatin accessibility dynamics identified with TCseq are not recapitulated with randomized data. A. TCseq C-means clustering of chromatin accessibility within identified peak regions after temporal randomization of read counts. **B.** Histogram distribution of membership grades for peak regions with either temporally ordered read counts (left) or temporally randomized read counts (right); significance of fit differences was determined by Wilcoxon rank-sum test, $p < 0.0001$).

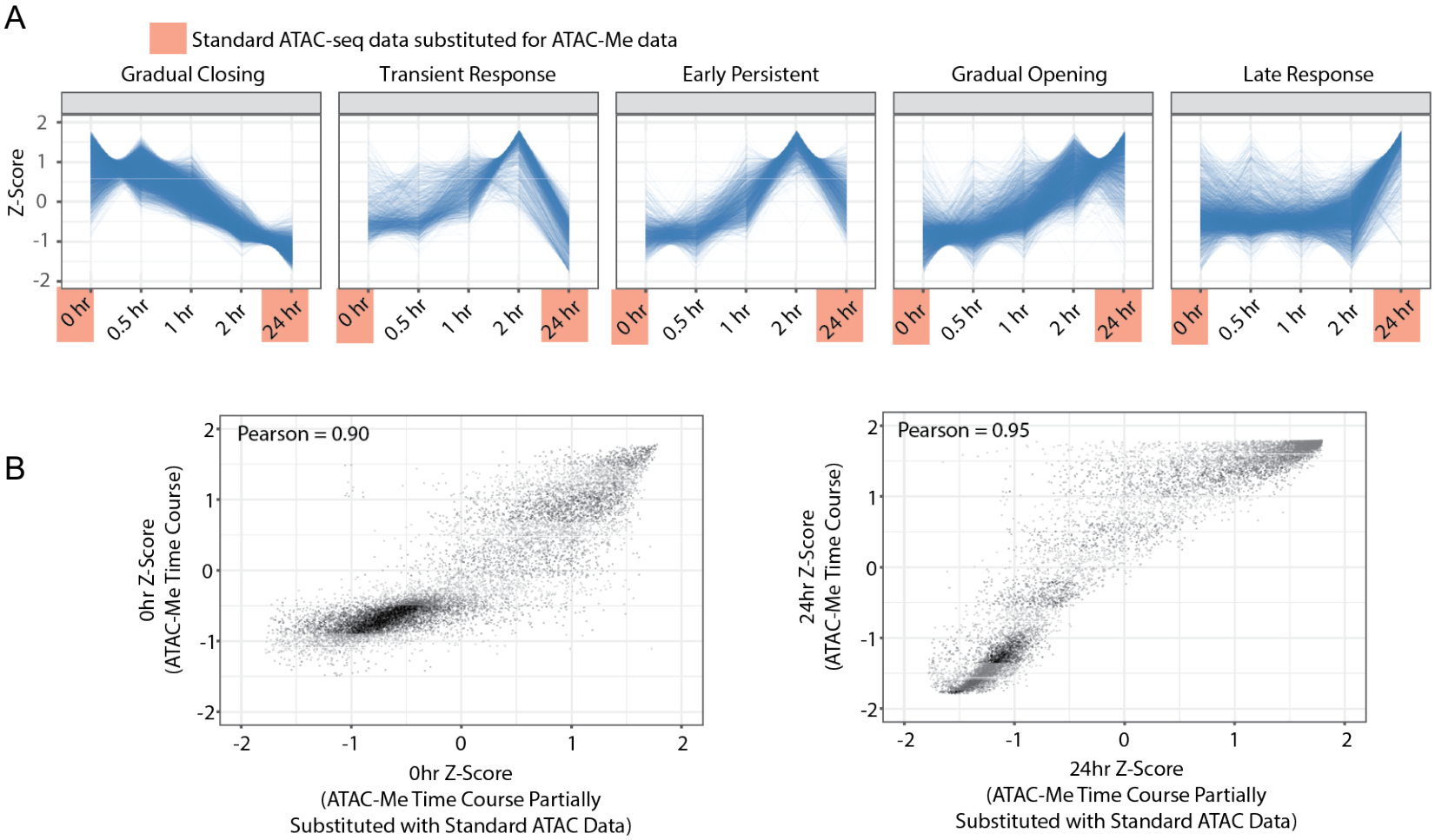

Figure 21. Chromatin accessibility dynamics calculated with mixed ATAC and ATAC-Me time course data. A. Z-score trends of C-means clustered temporal chromatin accessibility behaviors. Data from time points 0hr and 24hr are replaced with standard ATAC-seq data for recalculating z-scores. **B.** Pearson correlation of z-scores calculated with only ATAC-Me data or ATAC-Me data partially substituted with standard ATAC-seq data for 0hr or 24hr time points.

Significant and differential motif enrichment of specific TFs were observed for each TCseq cluster (Figure 22). Importantly, high correspondence is observed between cluster-associated TFs and the timeline of early PMA and late PMA response. These TFs include known factors downstream of the protein kinase C pathway, which is a direct target of PMA, and MAPK/ERK (Mitogen Activated Protein Kinase/Extracellular signal-regulated kinases) signaling (Traore et al., 2005). Within two hours, accessibility in peak regions enriched for myeloid pioneer factor PU.1 is significantly reduced, while Egr1 (Early growth response protein 1) and Klf14 (Krueppel-like factor 14) dominate peaks in which accessibility is greatest immediately following PMA addition. NFkB (Nuclear Factor Kappa-light-chain-enhancer of activated B cells protein complex), an important TF in macrophage differentiation and known target of immune regulation genes (Takashiba et al., 1999), emerges as the dominant TF motif within peak regions clustering at 24 hours. Methods of identifying TF footprints from DHS and ATAC-seq data have been described (Baek et al., 2017; Piper et al., 2013); thus, we asked whether these patterns are preserved in ATAC-Me data. Despite bisulfite conversion induced DNA damage, abundant TF footprints are detected (Figure 23). Accordingly, these footprints are consistent with time point specific TF binding events and differential analysis of TF footprints also reveals time point linked enrichment and depletion of specific TFs (Figure 24). To provide further support to these observations, we analyzed existing THP1 monocyte and macrophage ChIP-seq (Chromatin Immunoprecipitation sequencing) data for acetylated histone H3 lysine 27 (H3K27ac), a mark associated with active enhancers (Figure 25, (Phanstiel et al., 2017)). An increase or decrease in H3K27ac can be observed between opening and closing ChrAcc loci, respectively, when comparing unstimulated monocytes to macrophages exposed to PMA for 72hr. This is in contrast to static regions which overall show very little change in H3K27ac signal. Altogether, these data support that dynamic ChrAcc loci represent activating and responding enhancer elements.

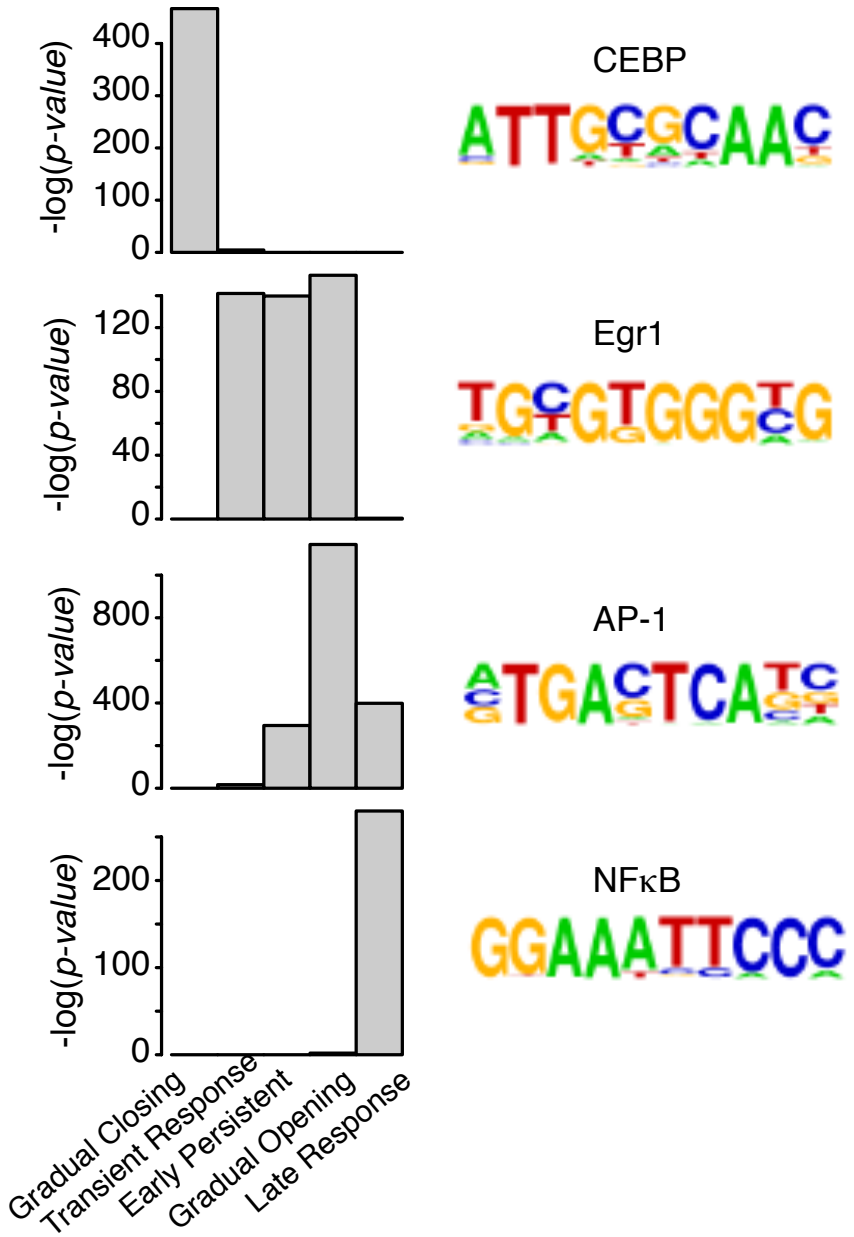

Figure 22. Transcription factor motif enrichment during PMA-stimulated gene activation. *P*-value significance scores for select transcription factor motif enrichment within accessibility clusters. Motif enrichment *p*-values were calculated with HOMER. Sequence logos for enriched motifs are depicted to the right of motif enrichment plots.

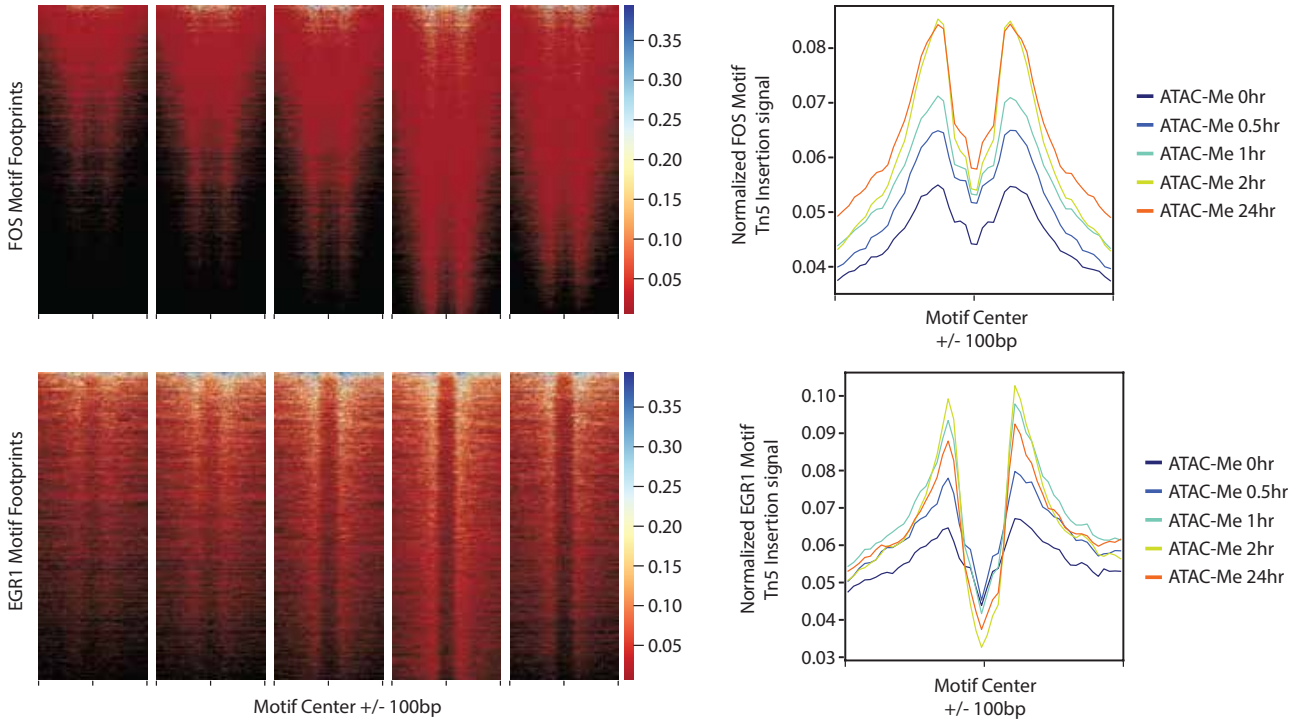

Figure 23. Transcription factor footprints of known PMA response factors. Footprints were identified *de novo* by searching for localized interruptions to Tn5 insertion frequency and then subsequently scanning for TF motifs contained within. Heatmaps represent Tn5 insertion signal centered around EGR1 or FOS (Proto-oncogene c-Fos) DNA motifs within a localized interruption to Tn5 insertion frequency.

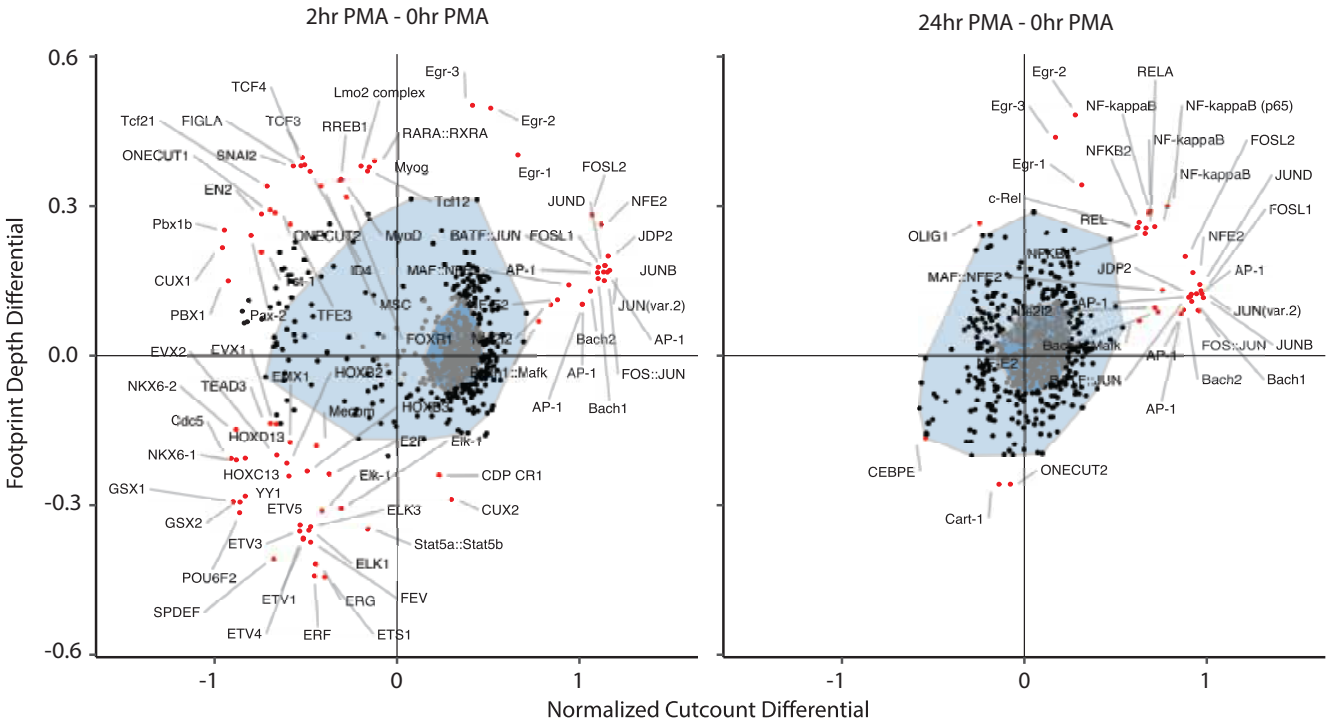

Figure 24. Temporal transcription factor footprint dynamics. Bivariate boxplots of differential footprint depth and flanking Tn5 insertion frequency around transcription factor motif occurrences for 715 motifs. Outlier motifs are indicated as red points. Analysis was performed using the BagFoot R package.

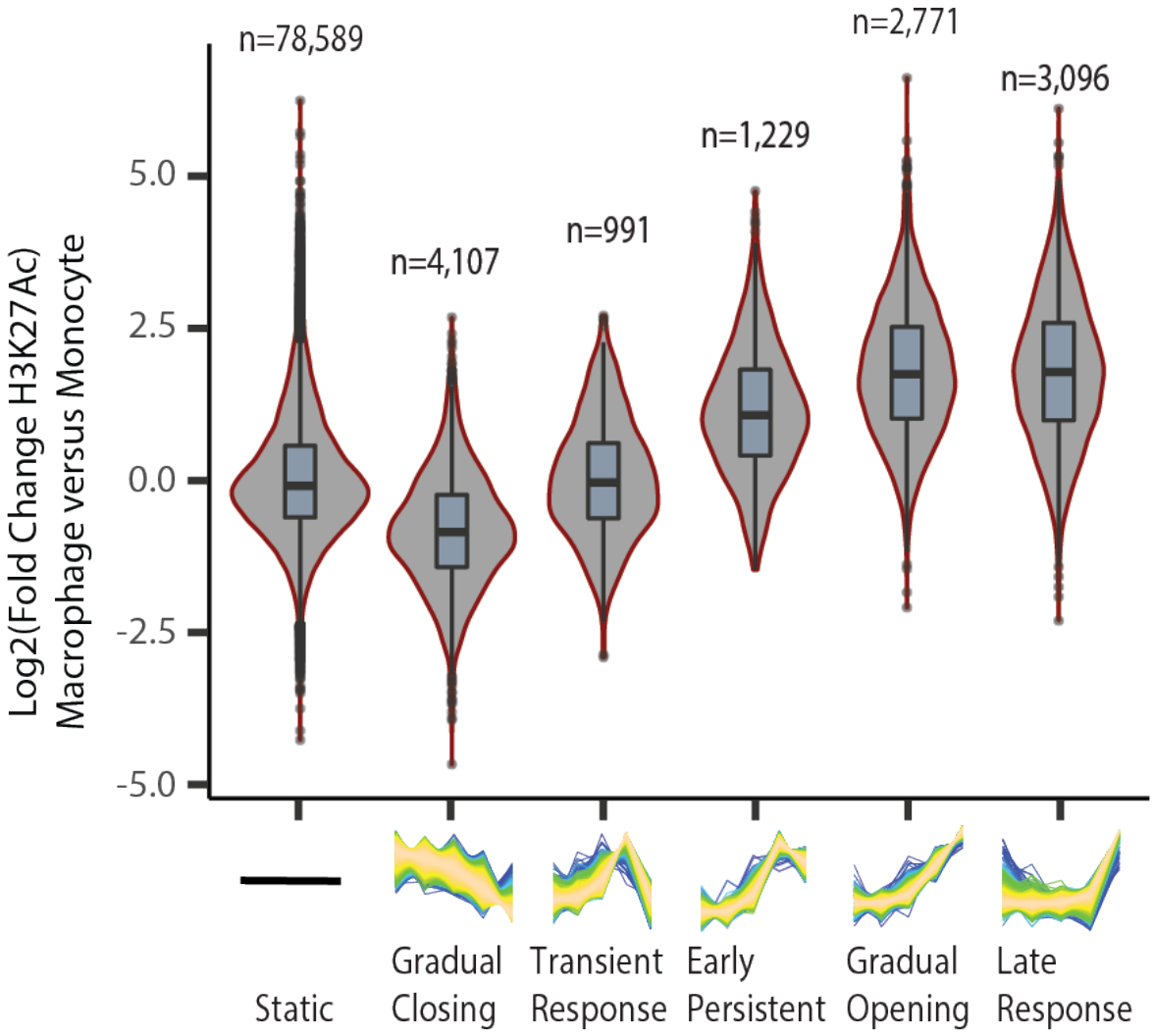

Figure 25. Gain or loss of H3K27Ac at genomic regions dynamic for chromatin accessibility. DEseq2 quantification of the log fold change in H3K27Ac ChIP-seq signal at genomic regions identified as dynamic for chromatin accessibility across a PMA stimulated THP1 monocyte to macrophage differentiation time course. H3K27Ac signal was calculated from publicly available data using time points of 0hr unstimulated THP1 cells and 72hr PMA stimulated cells.

Transitioning chromatin regions exhibit prolonged methylation states at early time points

The identification of highly dynamic ChrAcc loci consisting of newly accessible DNA is advantageous in determining whether or not the establishment of ChrAcc and loss of DNAme, or vice versa, are distinct events in the ordered process of enhancer regulation, while also capturing the temporal diversity of states that may exist genome-wide. Importantly, we did observe a significant number of accessible regions that were also highly methylated (>50% methylation), which supports our previous observation that "dichotomous" states do exist (Figure 26; Figure 27A). Comparing methylation levels within dynamic ChrAcc peaks (identified by TCseq) to static peaks, a greater number of hypermethylated regions are observed (Figure 26; Figure 27A). In other words, if all ChrAcc regions are considered, regardless of whether or not they change over time, the distribution of methylation is heavily biased toward hypomethylation. However, when we consider only dynamic regions, a substantial number of peaks are hypermethylated, and this hypermethylation persists over time despite increasing levels of accessibility (Figure 26; Figure 27A). This persistent hypermethylation frequently occurs at "nascent" GREs where little to no accessibility exists prior to PMA treatment (Figure 27A). Likewise, a subset of dynamic ATAC-Me peaks exhibits persistent hypomethylation despite increasing or decreasing peak signals. This phenomenon is more frequent among loci in which low levels of accessibility pre-exist, often corresponding to promoters and other high-density CGI regions, as exemplified by the "Transient Response" cluster, which is enriched for regions that contain higher CpG densities including promoters (Figure 27A,B, Figure 19). Additionally, the "Gradual Closing Response" peaks also feature persistent hypomethylation despite decreasing levels of accessibility. While the sampling nature of ATAC-Me can yield sparse read count data at early stages of chromatin opening or late stages of chromatin closing, end-point matched WGBS reference data supports these results (Figure 27A). These data suggest that DNAme and ChrAcc dynamics may be quantitatively de-coupled and that, while the

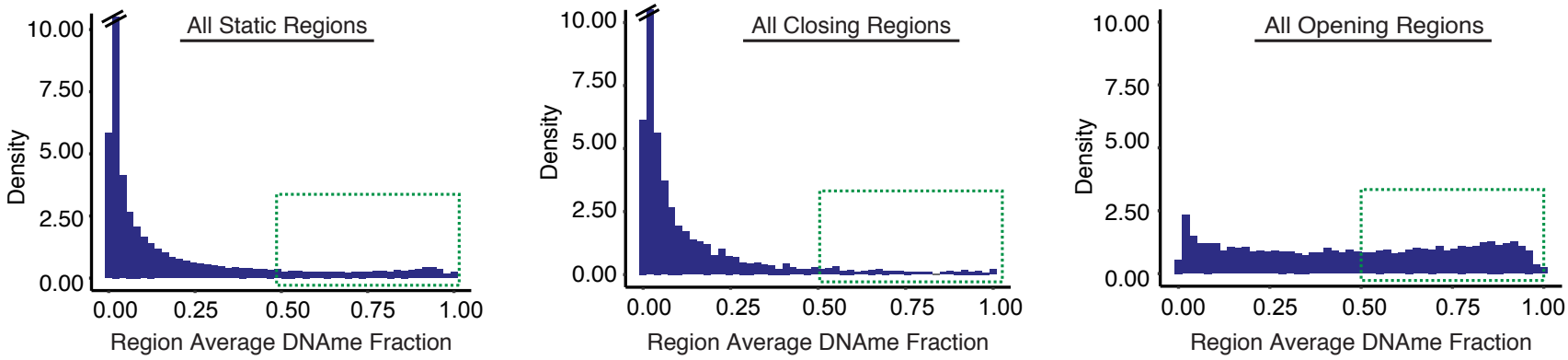

Figure 26. Density histograms of average methylation fraction for static and dynamic chromatin accessible genomic regions. Density histogram of average methylation fraction for static accessible regions (no membership to a TCseq cluster) across the time course, regions with decreasing accessibility signal across time (Gradual Closing Response), and regions with increasing accessibility signal (Late + Gradual Opening + Transient + Early Persistent Responses identified with TCseq).

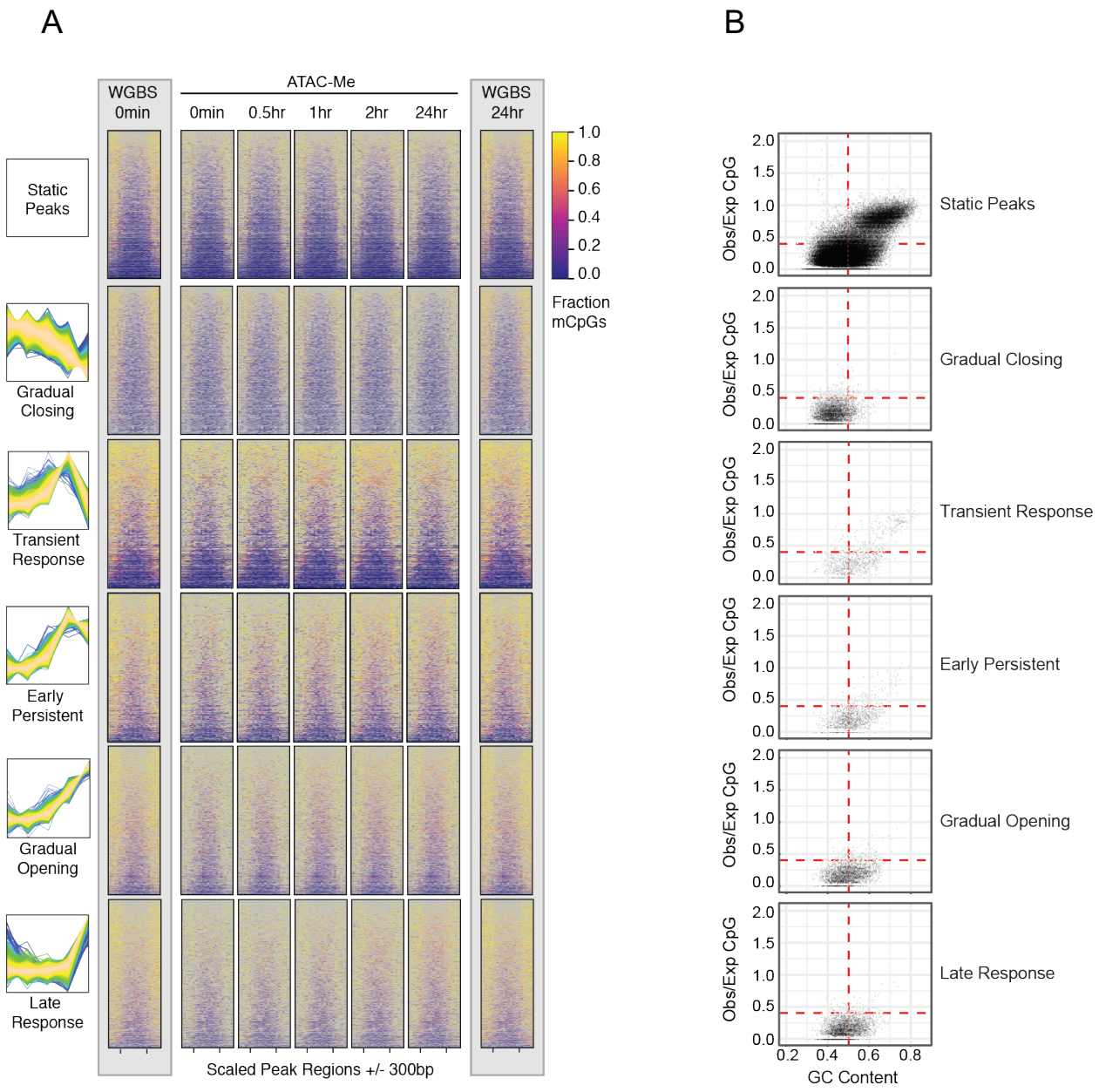

Figure 27. Genomic regions with rapid chromatin accessibility dynamics exhibit prolonged methylation. A. Heatmaps display mCpG fraction across individual peak regions in different accessibility groups (Static, Gradual Closing Response, Transient Response, Early Persistent Response, Gradual Opening Response, and Late Response). mCpG fraction is calculated in 50bp bins across regions scaled to 600bp with a flanking region of +/- 300bp. **B.** CpG density and GC content of TCseq cluster regions dynamic for chromatin accessibility across time. CpG density was calculated as observed/expected occurrence of CpG dinucleotides within TCseq cluster regions defined in Figure 2. Dashed lines represent CpG island thresholds defined by (Gardiner-Garden and Frommer, 1987) for CpG density/GC content.

two states are highly correlated, their spatiotemporal partnership is much less interdependent than previously appreciated.

Previous studies have established that regions with high CpG density are less dynamic and more hypomethylated than regions with lower CpG density. In our data, we detect notable differences in CpG content between clusters of peaks with different temporal ChrAcc behaviors and TF motif enrichments (Figure 22, Figure 27B). Moreover, the two peak clusters, “Transient Response” and “Early Persistent Response”, that rapidly gain accessibility upon PMA stimulation have the highest CpG content, whereas the more gradual, late responders feature significantly lower frequencies of CpGs. The “Transient Response” cluster also contains the majority of regions where gains in ChrAcc are characterized by pre-established and consistent hypomethylation. This may be important considering the necessity to invoke rapid control of gene expression without barriers that require additional enzymatic steps and/or DNA replication.

Transcriptional responses track closely with chromatin accessibility

Pathway enrichment analysis of genes in proximity to dynamic ATAC-Me peaks revealed gene sets with distinct functional associations consistent with PMA stimulation of THP1 cells (Figure 28). To probe transcriptional changes corresponding to ATAC-Me dynamics, we performed RNA-seq in parallel for each time point collected. Independent K-means clustering of RNA-seq data for the top 20% most variable genes across the time course demonstrates near identical group transcriptional responses to those temporal behaviors exhibited by clusters defined by accessibility (Figure 29). Importantly, these changes are detected as soon as a half hour post-induction, and include transcripts of TFs corresponding to TF motifs implicated among the divergent accessibility groups (Figure 30). Analysis of genes nearest to dynamic ATAC-Me peaks suggests that ChrAcc is a better predictor of transcriptional response than DNAme levels (Figure 31). Moreover, these dynamic transcriptional responses occur in spite of persistent

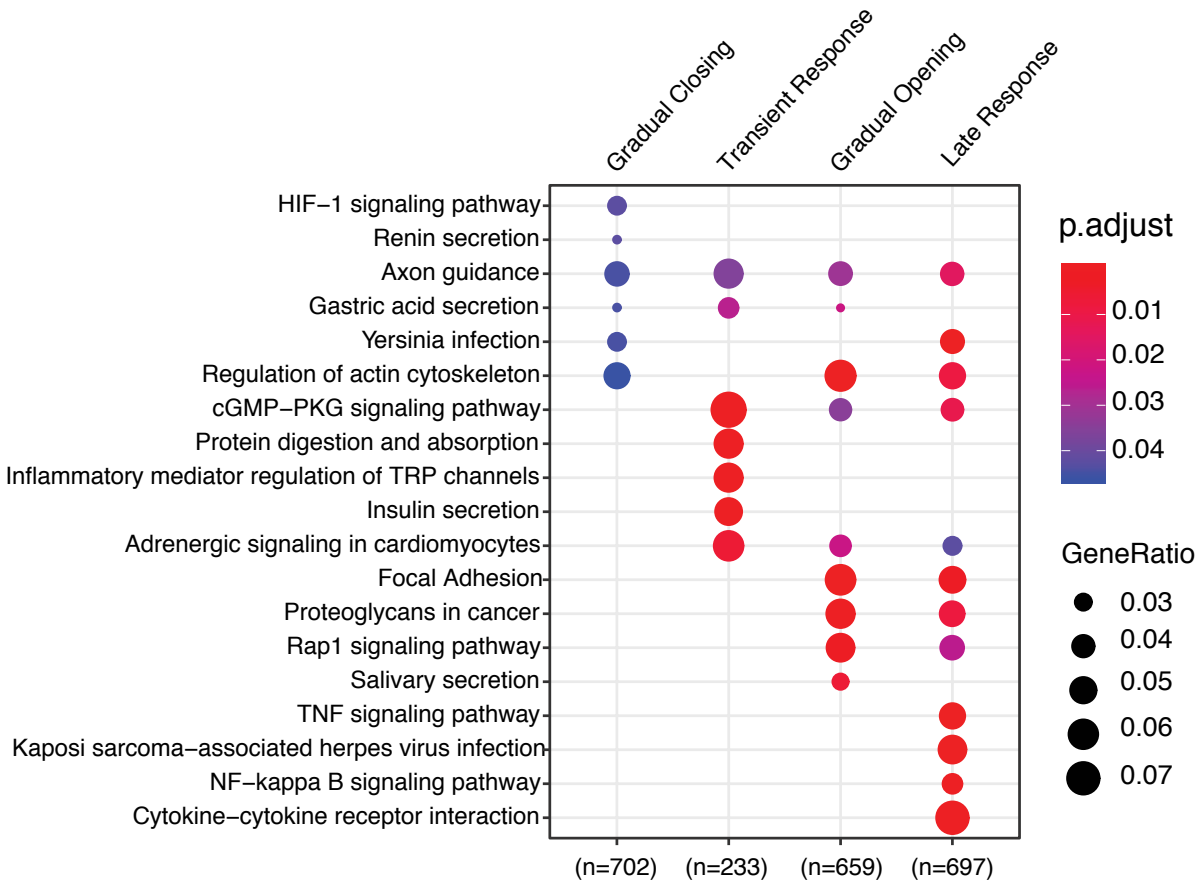

Figure 28. KEGG pathway enrichment of genes proximal to dynamic peaks in specific TC-seq clusters.

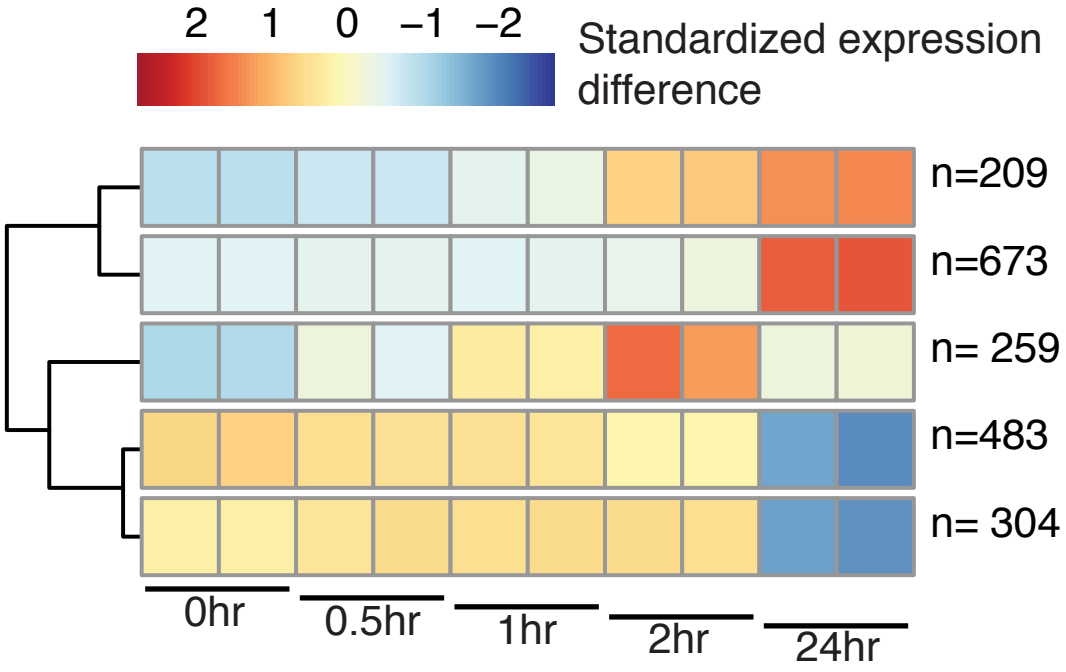

Figure 29. K-means clustering of the top 20% most variable genes identified with RNA-seq.

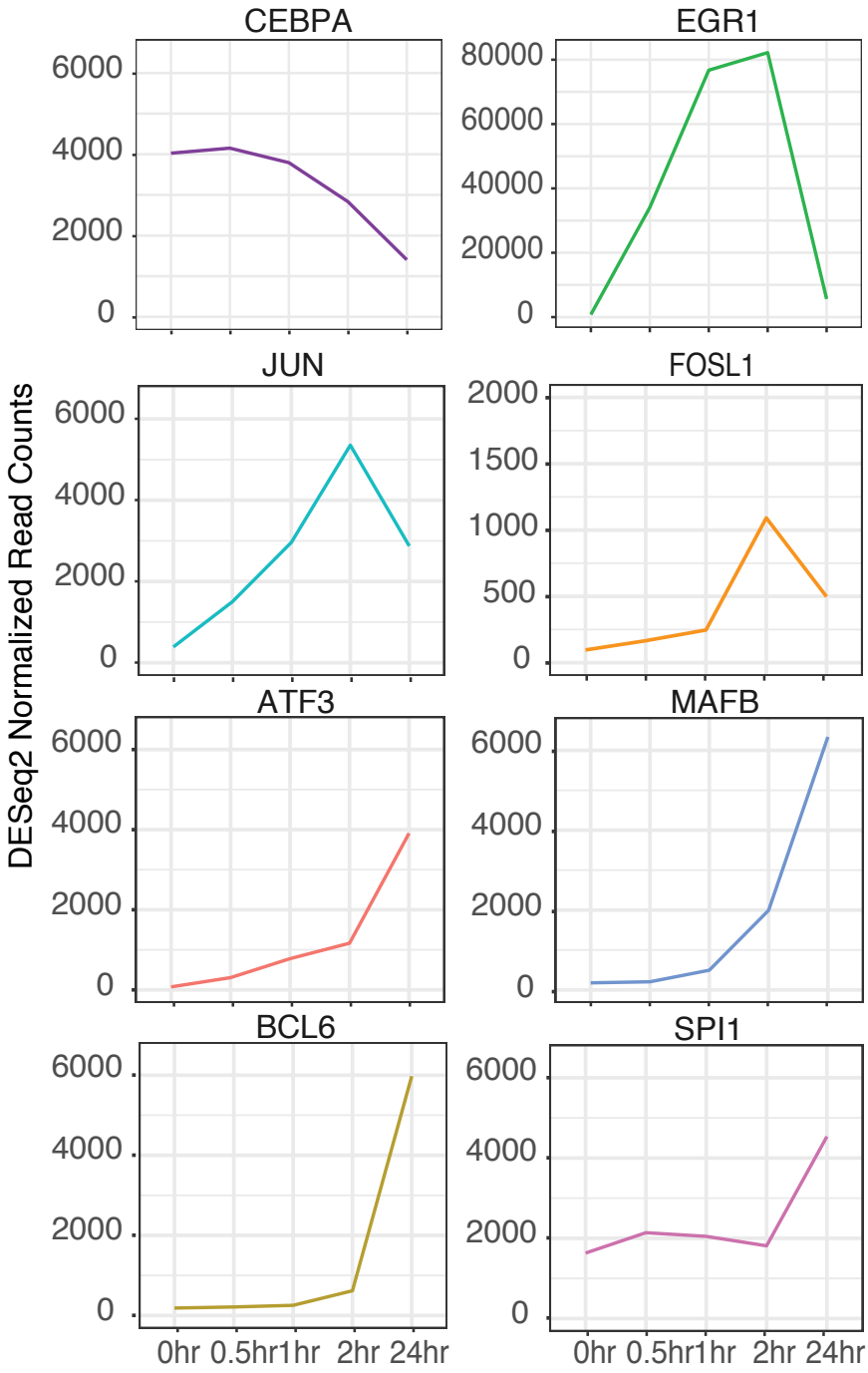

Figure 30. RNA transcript abundance dynamics for transcription factors important for PMA stimulation response in THP-1 cells. Read count plots across time for select TFs known to be related to the monocyte to macrophage transition. Depicted values are a replicate average of DESeq2 normalized read counts.

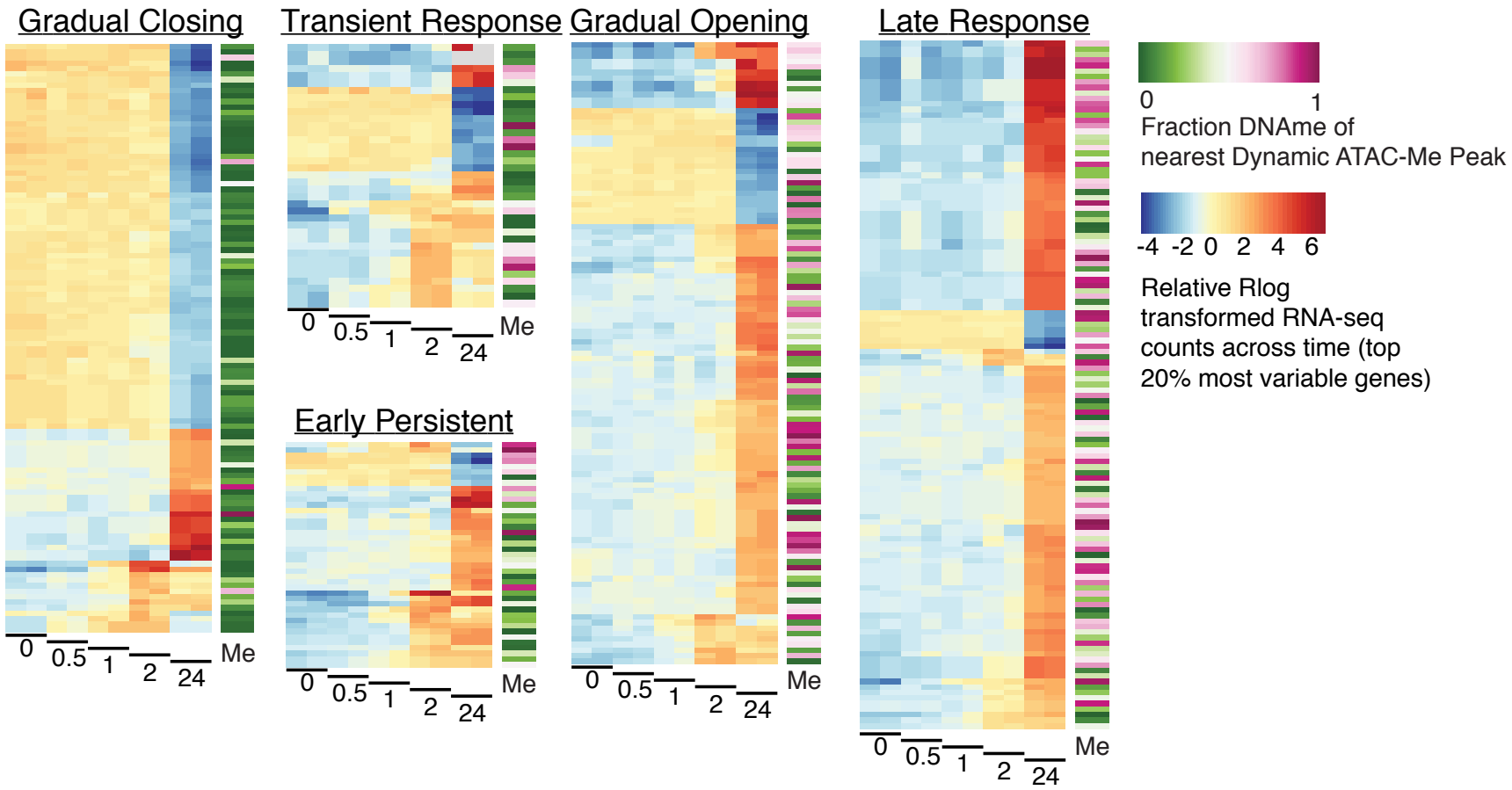

Figure 31. Integrating the dynamics of chromatin accessibility, DNA methylation and transcription. Hierarchical clustering of the top 20% most variable genes in nearby proximity to genomic loci identified as dynamic for chromatin accessibility. The mean mCpG fraction of all timepoints for neighboring ATAC-Me peaks is indicated to the right of each plot. Black bars below the plots indicate replicate pairs for each time point.

DNAme levels of paired peak regions. It is also worth noting that the directionality of change (i.e. accessibility and transcription) is not always uniform, as exemplified by gene pairs of “Gradually Closing” peaks that display increases in transcription. This may be explained by 1) inaccurate assignment of biological gene:peak pairs, or 2) the possibility of the GRE being either an enhancer or a repressor. It is also important to note that, while DNAme levels are less dynamic over time, gene associated peak regions consist of a range of methylation values (Figure 31).

To further investigate these relationships, we globally compared transcript abundances to either accessibility or DNAme levels for the top 20% most variable genes paired with nearest neighbor peaks, finding a stronger positive correlation with accessibility compared to a weaker negative correlation with DNAme (Figure 32). We considered that, due to the distinct characteristics of each data type, data standardization is required to describe the temporal relationships between these three molecular events. Therefore, a standardized difference from the mean across all time points was determined for every dynamic peak:gene pair (Figure 33). This analysis revealed a tight temporal relationship between ChrAcc and transcriptional dynamics but not DNAme. Related to these observations, a significant feature of the THP1 monocyte to macrophage system is that replicative potential is irreversibly lost upon PMA exposure, and replication may be required for both deposition and removal of DNAme (Figure 34) (Donaghey et al., 2018; Otani et al., 2013). Nonetheless, our data suggest that gain or loss of DNAme is not immediately required for gene or GRE activation (or repression) to occur.

Subtle loss of DNAme occurs downstream of chromatin and transcriptional changes

Thus far our analyses have focused on a relatively short-term time course of PMA induction. Within this short time frame, DNAme states do not appear to change among regions undergoing rapid chromatin accessibility changes. These results suggest that ChrAcc and DNAme changes do not occur on the same time scale, raising the possibility that DNAme changes occur as a secondary response to ChrAcc changes. Focusing specifically on ChrAcc regions that open by

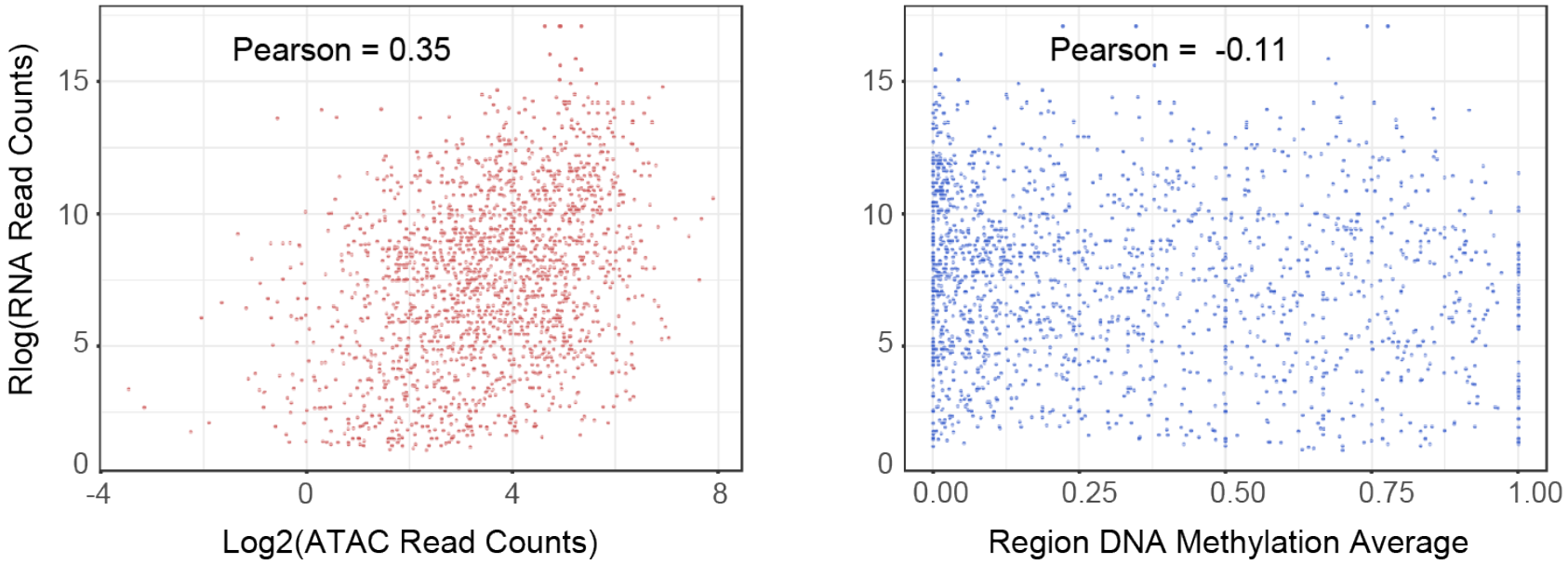

Figure 32. Scatterplot comparison of RNA transcript abundance with chromatin accessibility or DNA methylation. Scatterplot of RNA-seq read abundance for the top 20% most variable genes associated with a dynamic chromatin accessible locus plotted against neighboring dynamic chromatin accessible locus read counts (left, Pearson = 0.35, p-value < 0.0001) or average CpG DNA methylation (right, Pearson = -0.11, p-value < 0.0001). Scatterplots represent all time points (0hr, 0.5hr, 1hr, 2hr, 24hr) plotted simultaneously.

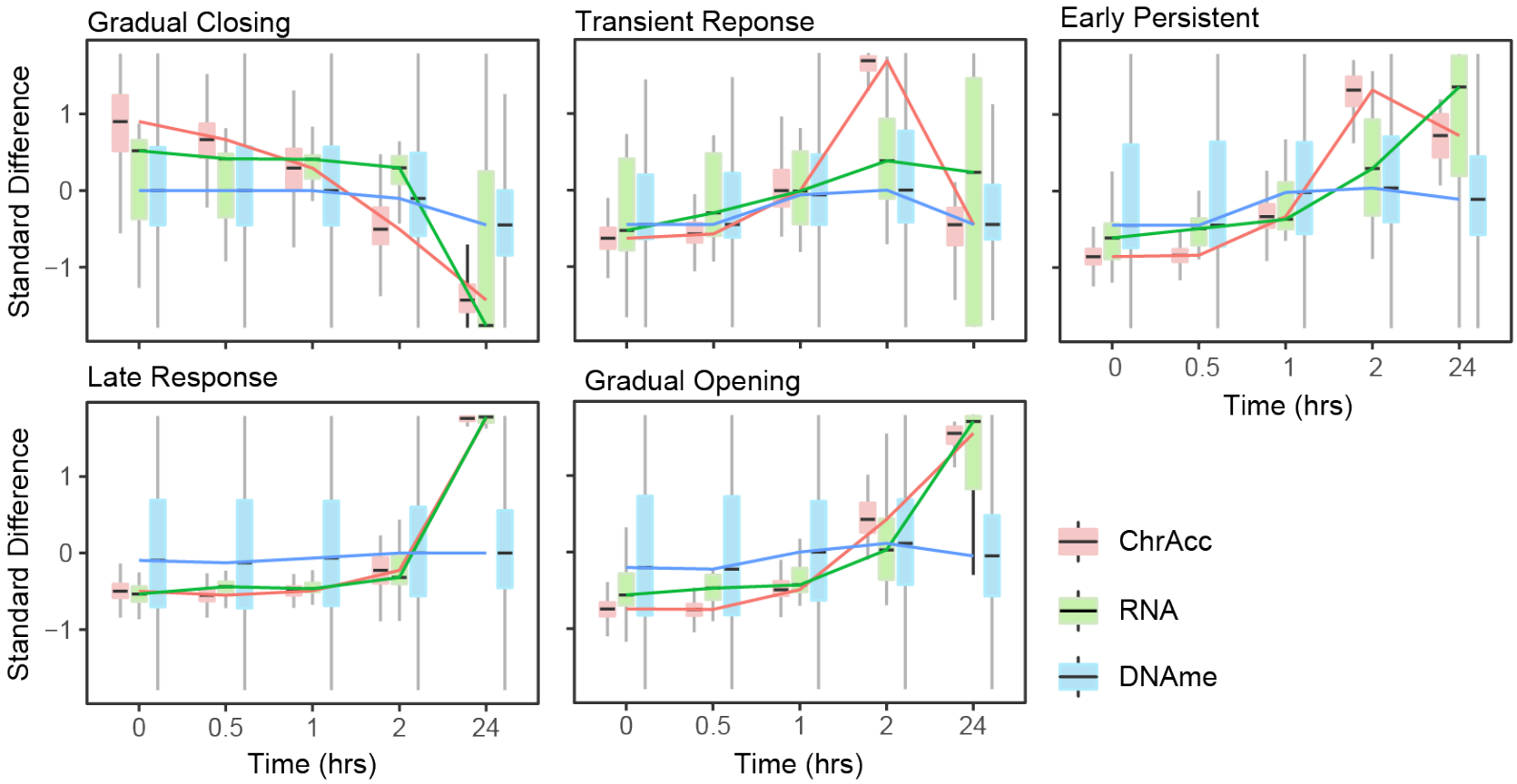

Figure 33. Boxplot distributions of standardized difference across time for all dynamic loci defined by TC-seq. Standardized difference across time was calculated for normalized ATAC read counts, rlog(read counts) of neighboring transcripts (top 20% most variable), and the DNA methylation state of the chromatin accessible DNA fragments. Lines represent median values.

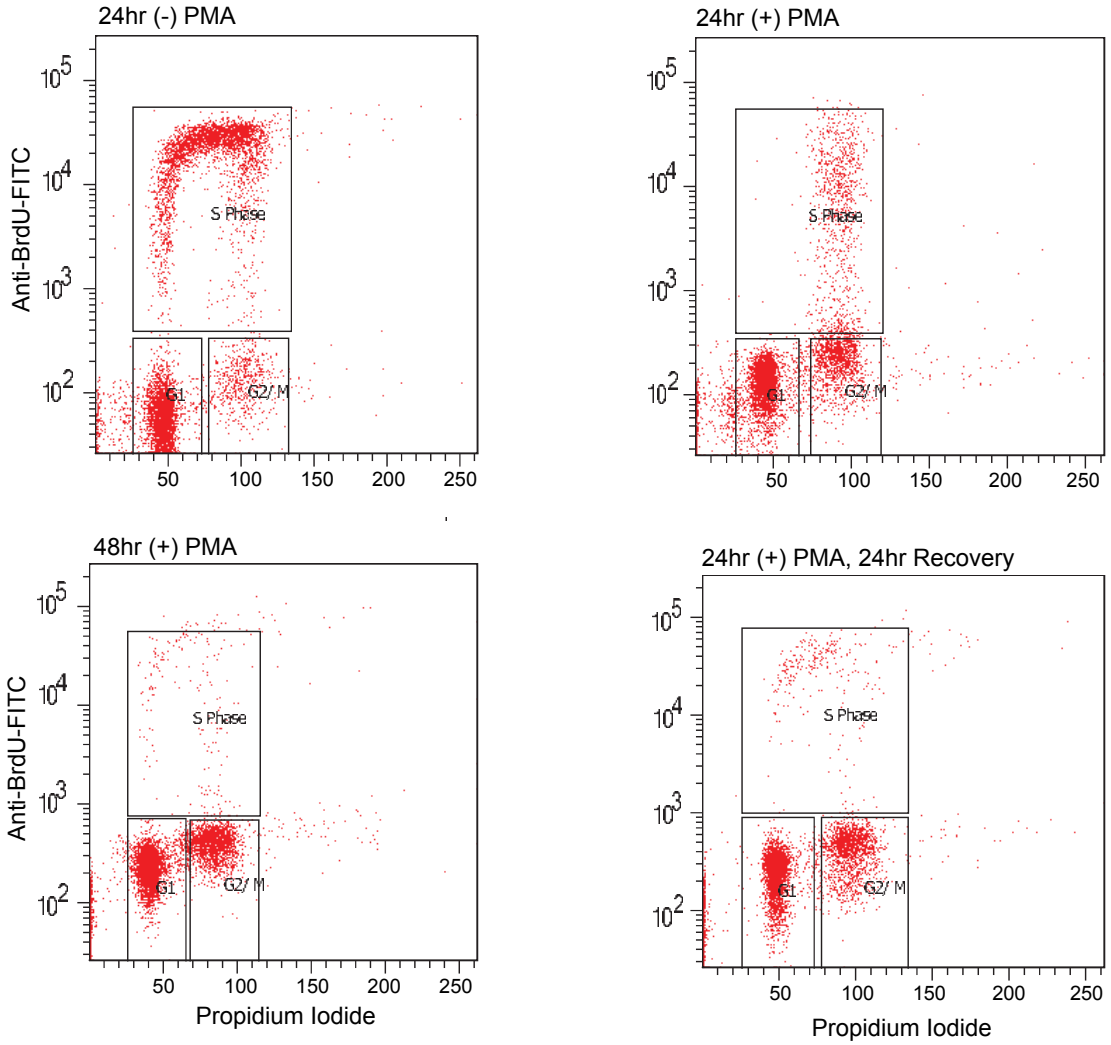

Figure 34. Cell cycle and DNA synthesis analysis of PMA stimulated THP1 monocytes. Flow cytometry cell cycle and DNA synthesis analysis of THP1 monocytes pulsed with BrdU prior to fixation and labeling with anti-BrdU-FITC antibody and propidium iodide. Prior to labeling cells were incubated with PMA for 0hrs, 24hrs, 48hrs or 24hrs followed by 24hrs of PMA-free recovery. Box overlays represent cell populations in S-phase, G1 phase or G2/M phase.

24hrs (~7000 loci), we performed an extended time course of PMA induction, collecting ATAC-Me data at 0, 24, 48, and 72hrs. For ChrAcc regions that begin to open by 24hrs, accessibility has largely plateaued by 48 and 72hrs (Figure 35A). Surprisingly, some small but significant loss of DNAme is observed within these regions by 72hrs (Figure 35A; Figure 36), further highlighting a molecular decoupling of ChrAcc and DNAme programs. Several recent lines of evidence support these data (discussed in further detail below); nonetheless, we sought to verify the competence of ATAC-Me to capture distinct changes in DNAme. At an applied concentration of 350μM, vitamin-C has been demonstrated to induce “artificial” hypomethylation of select genomic regions by enhancing the catalytic activity of TET family DNA dioxygenases involved in DNA demethylation pathways (Blaschke et al., 2013; Cimmino et al., 2017). In conjunction with PMA induction, we added stabilized vitamin-C (L-AA-2-P) to THP1 cells for a course of 72hrs. At the aforementioned “opening” regions we observe significant reduction in DNAme in the presence of vitamin C compared to control conditions (PMA alone, Figure 35A,B; Figure 36; Figure 37). Therefore, loss of DNAme is observed only after application of a considerable stimulus of a DNA demethylation pathway. Overall, these data confirm both the biological coincidence of DNAme within ChrAcc regions, and a spatiotemporal disconnect between DNAme and ChrAcc dynamics at enhancers during terminal cell fate transitions.

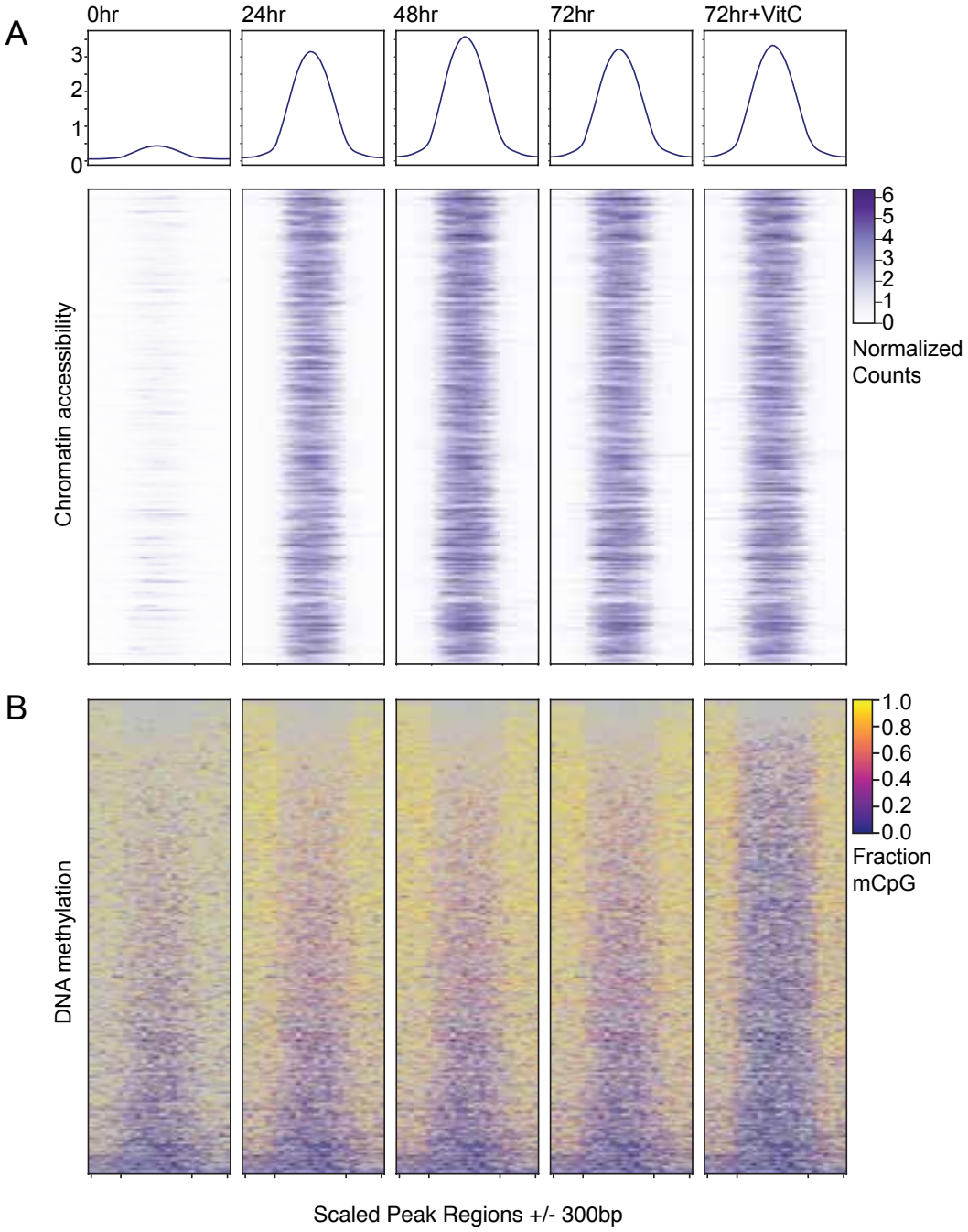

Figure 35. Loss of DNA methylation is delayed in nascent open chromatin regions. **A.** Heatmap of ATAC-Me accessibility signal for extended time points across all genomic exhibiting accessibility by 24hrs (Early Persistent Response + Gradual Opening Response + Late Response). **B.** Heatmaps display CpG methylation levels in 50bp bins across all regions in (A)

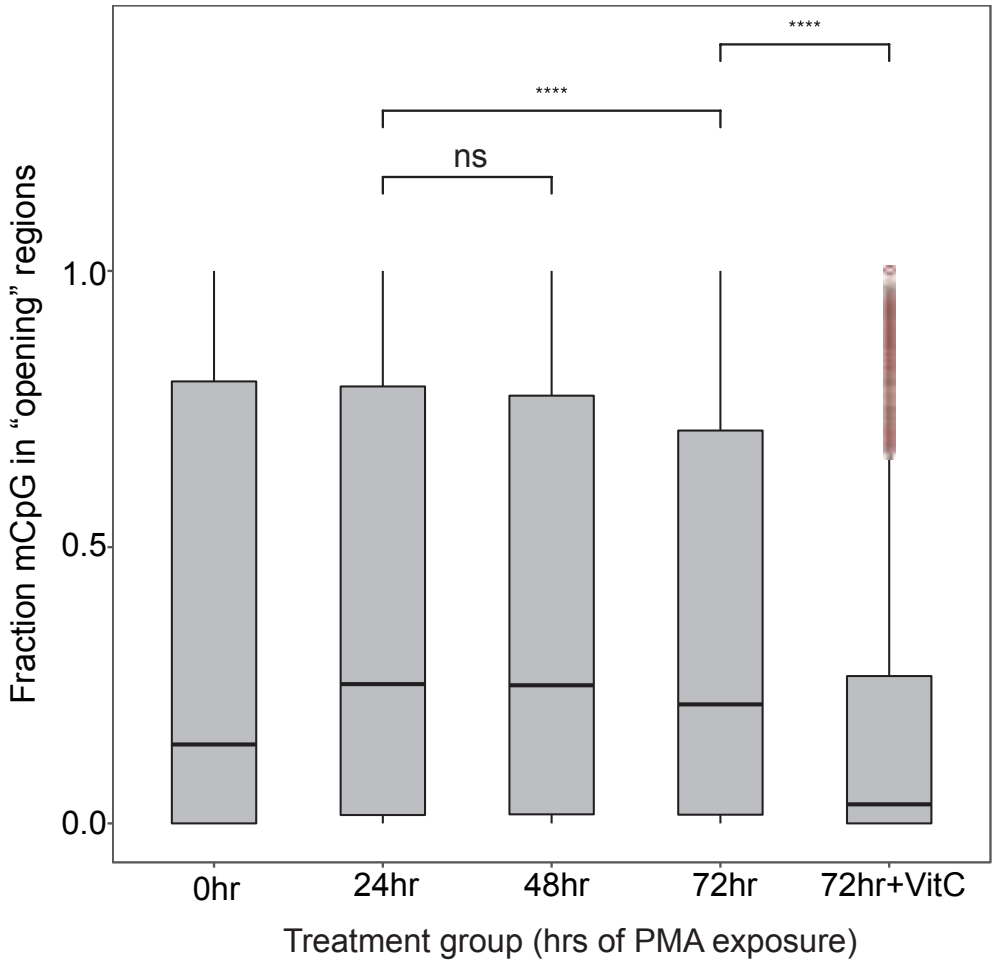

Figure 36. Boxplot distribution comparisons of methylated CpG fractions across extended time points and treatments. Wilcoxon rank sum test was used for statistical comparison (ns = not significant).

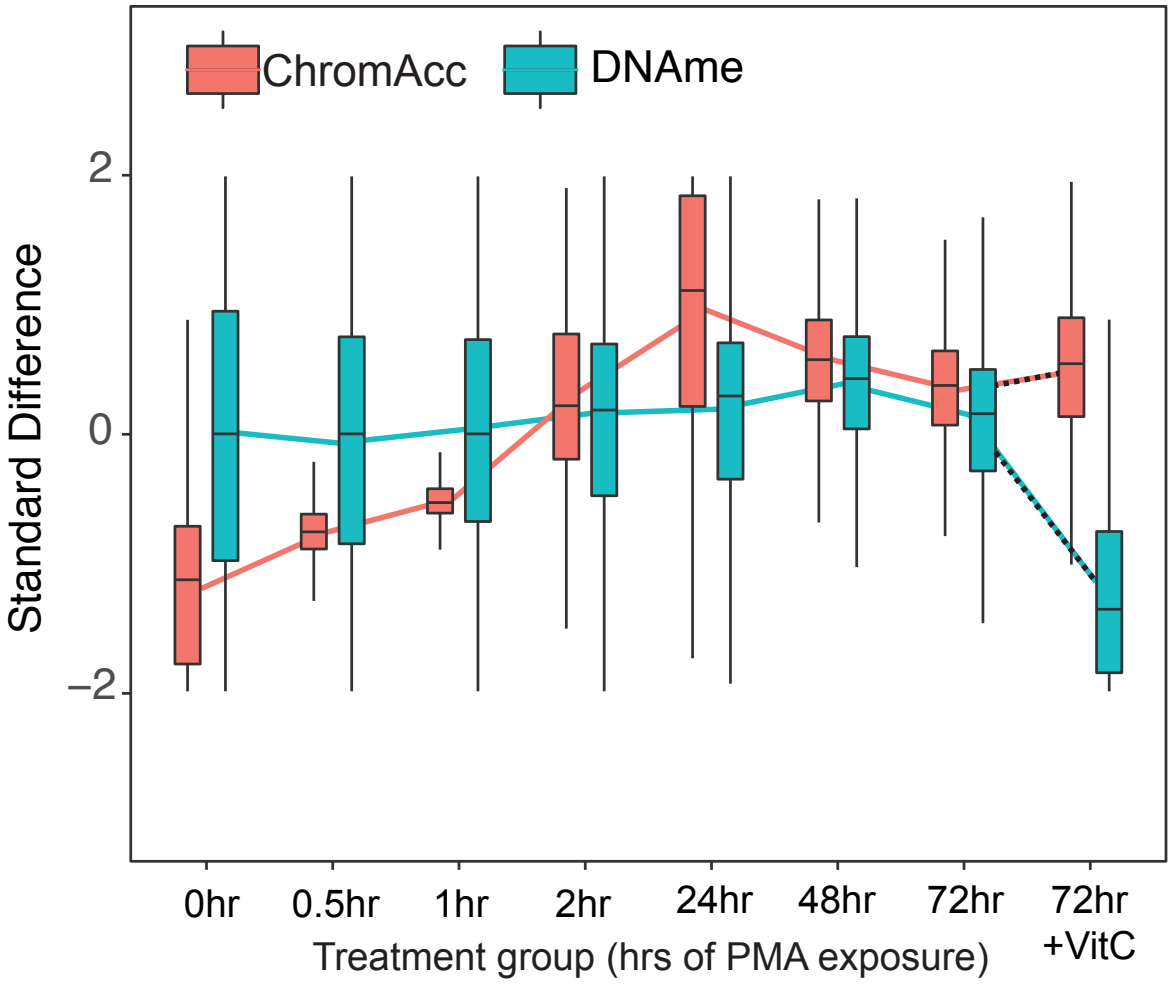

Figure 37. Chromatin accessibility and DNA methylation boxplot distributions across extended time points. Boxplot distributions of standardized difference across early and late time points for all genomic regions opening at 24hrs (A and B). Standardized difference across time was calculated for normalized ATAC read counts and CpG methylation fraction within chromatin accessible fragments. Lines represent median values.

Discussion

Many aspects of gene regulation are deeply conserved in eukaryotes and have been extensively studied in model organisms, but DNA methylation and its relationship with other critical aspects of enhancer regulation are not well understood. Genes are activated sequentially, and according to current models, enhancer chromatin states are modified by a series of biochemical events including pioneer factor initiated recruitment of chromatin remodelers and histone modifiers leading to the appearance of chromatin accessible DNA sites (Levine et al., 2014). DNAme is not typically permissive to transcription (Schubeler, 2015; Stein et al., 1982; Vardimon et al., 1982), and transcription at enhancers can be detected prior to promoter activation (Arner et al., 2015). So, loss of DNAme, whether immediate or eventual, is also expected at these sites. A variety of indirect evidence points to active removal of DNAme including the presence of hydroxymethylation (5-hmC) at cell fate determining enhancers (Li et al., 2018a; Serandour et al., 2012) and recruitment of Tet enzymes by pioneer factors (Li et al., 2018b; Lio et al., 2016). Furthermore, we and others have shown that enhancer hypomethylation is highly cell type specific and a better predictor of target gene transcription than promoter hypomethylation (Schlesinger et al., 2013). In light of these ideas, a key premise for performing ATAC-Me stemmed from the observation that an unexpected fraction of DHS sites ascertained by DHS-seq are methylated in WGBS datasets of the equivalent cell type, a duality which we reasoned may represent an intermediate state along a continuum of enhancer activation. Other studies support these observations, reporting “bivalent” regions of H3K27ac being marked by DNA methylation (Charlet et al., 2016). Accordingly, time is an equally vital

component to our approach, as we sought to capture both ChrAcc and DNAme changes during cell fate transitions.

ATAC-Me is a straightforward approach that detects DNAme of accessible DNA fragments from native chromatin, and requires minimal molecular adaptation of existing Tn5-assisted methods including ATAC-seq and tagmentation-based WGBS. ATAC-Me allows quantification of accumulating DNA fragments in accessible genomic regions (or in nascent opening regions); therefore, direct interrogation of the temporal relationship between ChrAcc and DNAme is possible in a rapidly changing chromatin environment. Additionally, ATAC-Me provides a focused view of the methylome most relevant to cellular function; the methylome of DNA molecules accessible to enzymes and TFs facilitating gene regulation. While approaches that enable joint profiling of chromatin features and DNAme, including NOME-seq (nucleosome occupancy and methylome sequencing), scNMT-seq (Single cell Nucleosome Methylation and Transcription sequencing) and ChIP-BS (chromatin immunoprecipitation coupled with bisulfite sequencing) have been described, ATAC-Me offers some key advantages (Brinkman et al., 2012; Clark et al., 2018; Kelly et al., 2012). First, ATAC-Me is not limited by cytosine frequency in specific dinucleotide contexts, so even the most CpG poor enhancers are evaluated. Notably, our method is readily able to identify TF footprints and the DNA methylation state of putative TF footprint sites can be calculated, a feature that distinguishes ATAC-Me from other approaches. Second, ATAC-Me selectively enriches Tn5-accessible genomic regions, so fewer sequence reads (50-60 million) are necessary to obtain high, focused coverage of accessible DNA, compared to whole genome approaches. Use of a transposome assembly (Picelli et al., 2014) with methylated adaptors also allows direct cloning of accessible fragments with minimal opportunities for input loss, and therefore many fewer cells (50,000-200,000) are required compared to other methods.

Using ATAC-Me, we observe clear patterns of rapid ChrAcc that are tightly associated with the time points measured. While this time course sampled both very early (0min, 30min,

1hr, 2hr) and later (24hr, 48hr, 72hr) time points, our results indicate that ChrAcc dynamics responses seem to plateau by 24hrs. It remains to be seen if additional waves of ChrAcc are observed at time points beyond 72hrs (for example with stimulation of different activation sates of macrophages). ATAC-Me detected TF footprints and distinct pattern specific TF motifs that tracked closely with time – even for time points separated by less than two hours, reflecting the rapid responses of ChrAcc within this timeframe. Transcriptional changes also follow similar patterns to those observed with accessibility. By contrast, we observe a range of DNAme states associated with accessibility changes. Importantly, DNAme showed only minimal changes within this timeframe in our THP1 monocyte to macrophage model. This surprising temporal disconnect between DNAme, ChrAcc, and transcription suggests that loss of DNAme is not immediately required for transcription or for TF binding, an observation supported by several lines of evidence. In particular, a recent study showed that DNA demethylation is not required for pioneer FoxA2 occupancy and its concomitant effects on ChrAcc (Donaghey et al., 2018). Moreover, in dendritic cells gene upregulation is observed ahead of DNAme changes in response to bacterial infection (Pacis et al., 2015).

Enhancer regions that remain methylated despite increasing accessibility may be marked by 5-hmC, which would not be distinguished from 5MeC in our assay. However, application of high levels of vitamin-C induced considerable loss of DNAme by 72hr, indicating that non-physiological levels of vitamin-C accelerate methylation removal. These data also imply that the higher levels of DNAme we observe in dynamic ChrAcc peaks is not likely a result of failure to discriminate 5-hmC from 5-mC. The slight loss of DNAme we observe in the absence of vitamin-C suggests that, with enough time, the DNAme in these regions will eventually be lost, though recent evidence suggests that replication is required for extensive DNA demethylation (Barwick et al., 2016; Donaghey et al., 2018; Otani et al., 2013). THP1 macrophages, like dendritic cells, are terminally differentiated and no longer replicate. Indeed, the question remains if DNA methylation would be more dynamic in a cell fate specification

system that maintains cell division status. Nevertheless, our data demonstrate the importance of evaluating DNA methylation in the context of a changing chromatin environment during cellular differentiation. ATAC-Me may also provide the resolution to catalogue precise epigenetic events that lead to sequential cell fate decisions in normal and abnormal cell fate models.

Some noteworthy limitations of ATAC-Me are that the approach relies on the assumption that inaccessible loci, and therefore missing data, are methylated, so a complementary WGBS dataset may be useful in some cases. ATAC-inspired methodologies are also plagued by mitochondrial DNA contamination, which can be minimized by use of modified lysis protocols (Corces et al., 2017) that include digitonin or depletion by Cas9 paired with a library of mitochondrial guide RNAs (Montefiori et al., 2017).

Chapter 4: Development of ATAC-STARR for Enhancer Activity Identification and Monitoring Across Cell Differentiation

Introduction

Epigenomic analysis of cellular differentiation can be limited by a lack of tools to validate gene regulatory elements such as enhancers. Enhancers and promoters after all are what drive the gene expression and phenotypic changes required for cell differentiation. Most studies to date rely heavily on using combinatorial epigenomic marks to identify gene regulatory elements. For instance, the histone mark H3K27Ac is often used to indicate an "active" regulatory element and the coincidence with H3K4me1/2 (Histone H3 Lysine 4 mono- or di- methylation) signal indicates an enhancer rather than promoter (Creyghton et al., 2010). This is just one example as there are many other usages of DNA methylation, chromatin accessibility and histone marks to identify gene regulatory elements. However, all of these are simply suggestive of activity and frequently inaccurate when assayed functionally (Kwasnieski et al., 2014). These approaches do not assess actual transcriptional activity of gene regulatory elements. It is paramount to monitor the functionality of gene regulatory elements to truly understand the context and impact of epigenomic dynamics on cell fate specification.

There are two central approaches for functionally detecting gene regulatory elements such as enhancers, assays for transcriptional initiation and reporter plasmid methods. Transcriptional initiation at enhancers (Andersson et al., 2014; Core et al., 2014) has been noted by several methods such as CAGE (Capped Analysis Gene Expression) (Kodzius et al., 2006), GRO-seq (Global Run-On sequencing) (Core et al., 2008), and PRO-seq (Precision Run-On sequencing) (Mahat et al., 2016). Transcription initiation based methods assay for enhancer signatures (bidirectional transcription) within the endogenous chromatin context. However, these experimental methodologies are difficult to apply across a wide variety of

conditions/contexts. In contrast, the traditional approach for validation of DNA sequence for enhancer activity is the luciferase reporter plasmid assay. In this assay a DNA fragment of interest in cloned upstream of a minimal promoter and luciferase gene within a circular plasmid construct. DNA fragments that have enhancer activity will be able to increase the transcriptional output of the luciferase gene compared to plasmid constructs with no DNA fragment inserted. However, this approach is primarily performed on predefined DNA fragments and is low throughput.

In the past decade more high throughput approaches have been developed to apply principles of the luciferase assay to the entire human genome. These approaches in general are referred to as massively parallel reporter assays. One such assay, STARR-seq (Self-Transcribing Active Regulatory Region sequencing) (Arnold et al., 2013), validates putative enhancers and their transcriptional strength via their placement downstream of a core promoter sequence, which results in self-transcription if the insert is a *bona fide* enhancer. However, STARR-seq was developed in the context of the *Drosophila* genome, which is some twenty-seven fold smaller than the human genome. Applying STARR-seq to the human genome represents considerable difficulty due to the large human genome size. Additionally, the exogenous plasmid nature of STARR-seq offers more flexibility to assay DNA fragments in differing cellular environments compared to CAGE/GRO-seq/PRO-seq.

In an effort to circumvent the logistic issues posed by assaying the whole genome via STARR-seq we turned to approaches that allowed for a focused sampling of likely gene regulatory elements. Nucleosome depletion (open chromatin) is a well-established property of gene regulatory elements including promoters and enhancers. Considering this property of gene regulatory elements, we sought to utilize an ATAC based strategy for a focused sampling of gene regulatory elements while avoiding assaying likely inert DNA sequences. Effectively isolated permeabilized nuclei are treated with Tn5 transposase which preferentially fragments and inserts sequencing adapters into nucleosome depleted regions of the genome. These

genomic fragments representing nucleosome depleted regions of the genome are then cloned into a reporter plasmid to assay for the capability to drive self-transcription (Figure 38). Our approach (ATAC-STARR, Assay for Transposase Accessible Chromatin coupled with STARR-seq) allows us to easily apply STARR-seq to the human genome, map waves of enhancer activity, validate regions as true functional enhancers and analyze TF activity. While the ultimate goal is to apply ATAC-STARR across a time course of cell fate specification the work presented within this chapters represents a proof of principle testing for the basic method.

Results

As initial proof-of-principle, the ATAC methodology captured nucleosome free regions from HEK293 cells as input to establish feasibility. ATAC origin open chromatin fragments were cloned into a linearized STARR-seq GFP reporter plasmid with *en masse* gibson cloning. Enhancers in this plasmid context increase the expression of GFP beyond that of plasmids containing no nucleosome free region insert. Flow cytometry analysis of HEK293 cells transfected with either a control GFP plasmid library containing no ATAC fragment inserts or a GFP plasmid library with ATAC fragments inserted revealed GFP fluorescence well beyond that of control (Figure 39). These results suggest that at least some population of ATAC fragments in the HEK293 ATAC-STARR plasmid library are functioning as enhancers to increase the output of GFP when transfected into HEK293 cells.

As further validation of the technique we tested an ATAC-STARR plasmid library for the capability to drive self-transcription that could be detected with RNA-seq analysis. An ATAC-STARR library was prepared from HEK293 origin open chromatin fragments and cloned into the previously reported STARR-seq plasmid vector for self-transcription assessment. This reporter plasmid library of putative enhancers was then transfected back into the source cell type (HEK293) as well as a distinct alternate cell type (U2OS). RNA-seq analysis of cDNA

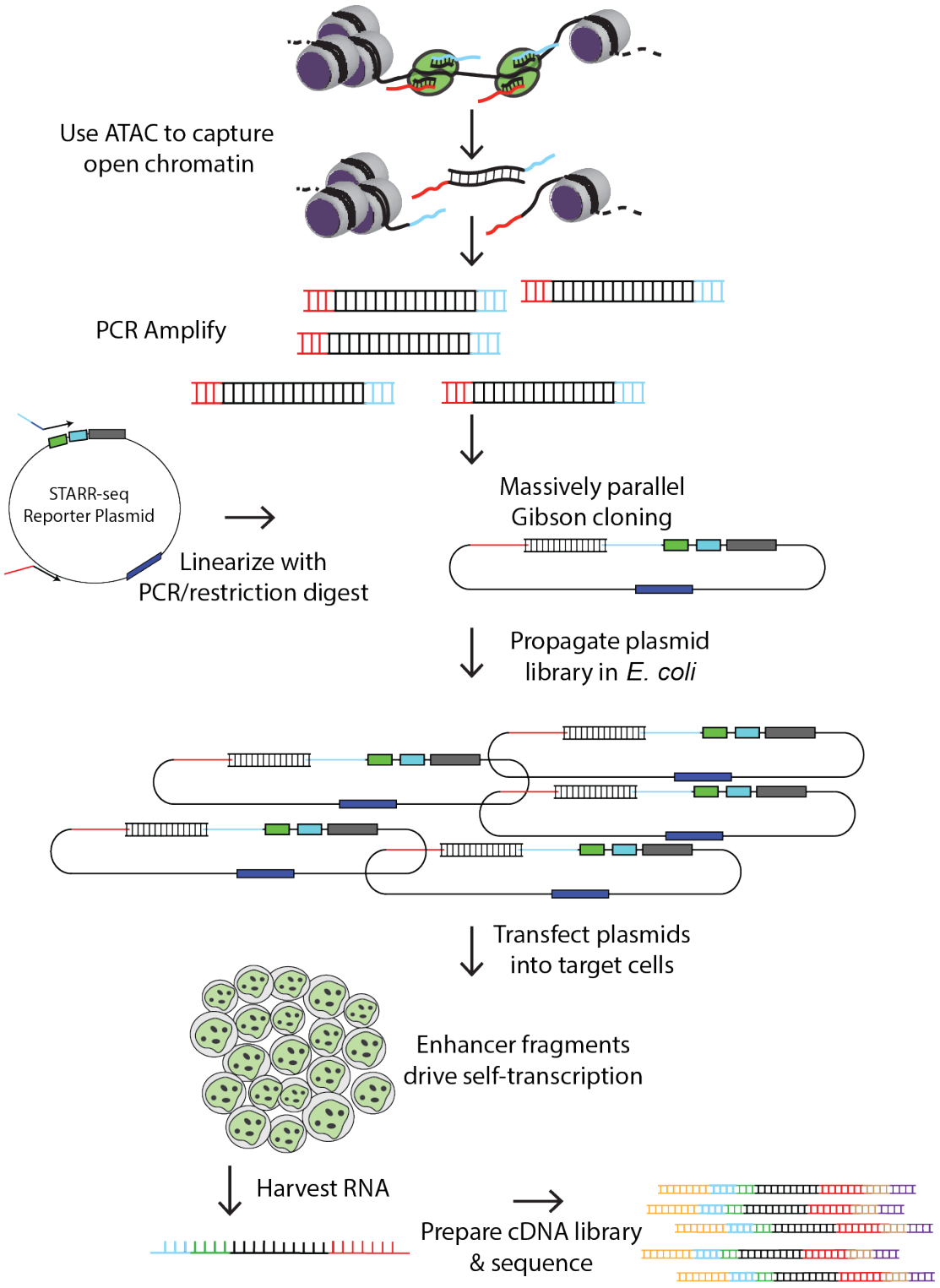

Figure 38. Methodology overview for ATAC-STARR sequencing library construction.

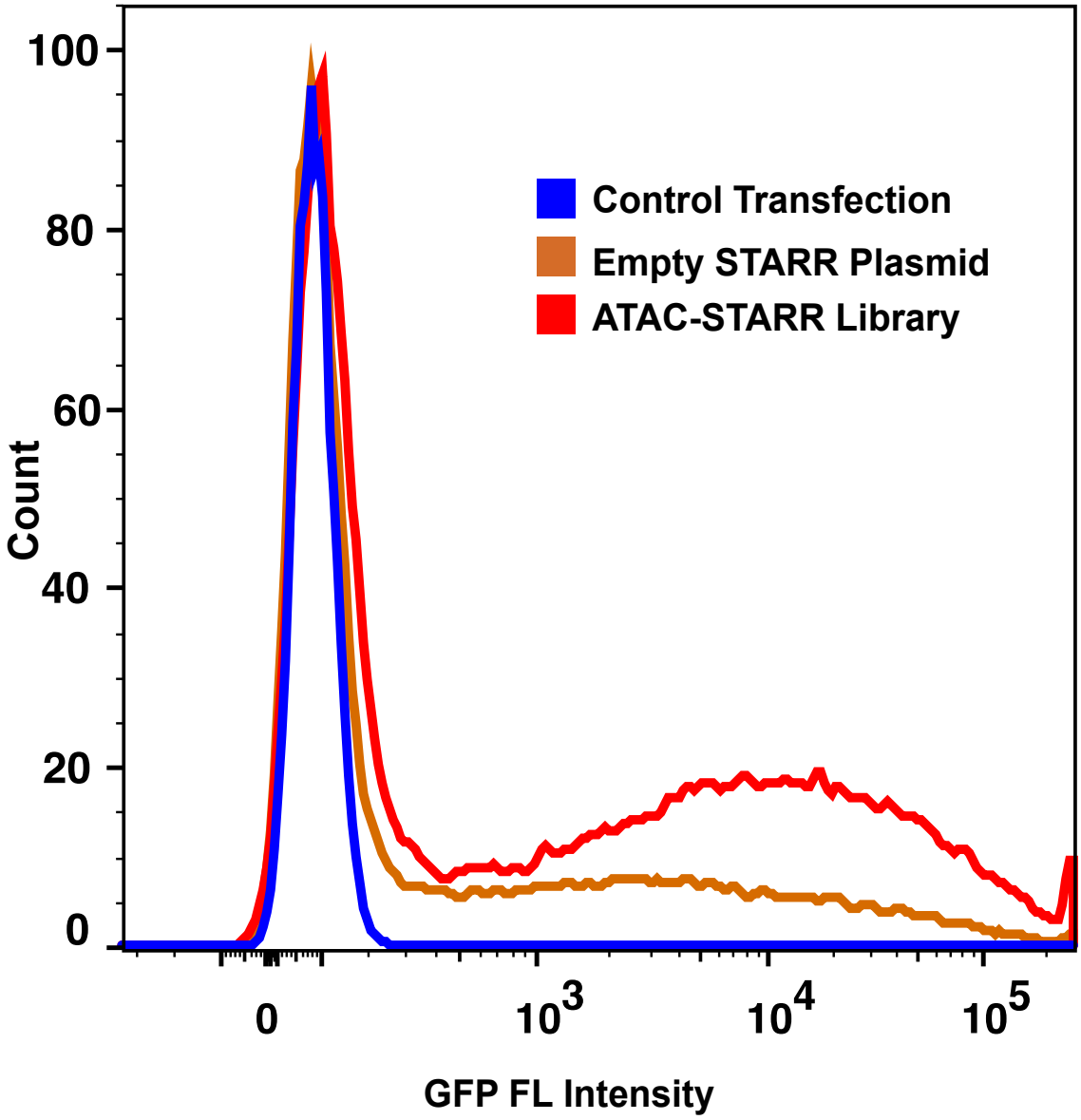

Figure 39. ATAC-STARR plasmid libraries drive transcription of a GFP cassette. Flow cytometry analysis of GFP fluorescence intensity measurements of HEK293 cells transiently transfected with either no plasmid (control), empty plasmid or plasmids containing open chromatin fragments from HEK293 cells.

(complementary DNA) prepared from ATAC-STARR plasmid transfected cells revealed transcripts originating from plasmids carrying open chromatin fragments and an overall preservation of ATAC regions. In addition, two quality control DNA libraries (ATAC fragments and plasmid library) were sequenced to assess the preservation of DNA fragment diversity across the stages of ATAC-STARR library construction. ATAC DNA fragment diversity was generally preserved across the stages of STARR-seq processing and DNA fragments appear to be acting as enhancers in the reporter plasmid context (Figure 40).

Discussion

The data presented here establishes ATAC-STARR as a feasible technique for a focused sampling of the human genome in the context of a massively parallel reporter assay. However, many improvements still remain to be implemented and data analysis complexities resolved. While the mass cloning procedure used to insert ATAC DNA fragments into linearized plasmid appears to have overall preserved DNA complexity there was still information loss. Maximal complexity at the earliest stages of plasmid library construction is crucial for the success of the ATAC-STARR technique. The higher the diversity of ATAC DNA fragments represented by the plasmid library the more confident later analysis of enhancer activity will be. Further, coverage of the same genomic locus by multiple plasmids increases the resolution for determination of key DNA base pairs for gene regulatory activity rather than making conclusions only on a locus assayed by a single DNA fragment. Different modes of plasmid transfection and cell types likely impact the true plasmid diversity existing within cells. A more detailed analysis of how these parameters influence enhancer identification is needed. Only those plasmids which are successfully introduced into a viable cell can be assayed for possible enhancer activity. Increased input from ATAC reactions, further optimized cloning, optimized cell transfection and

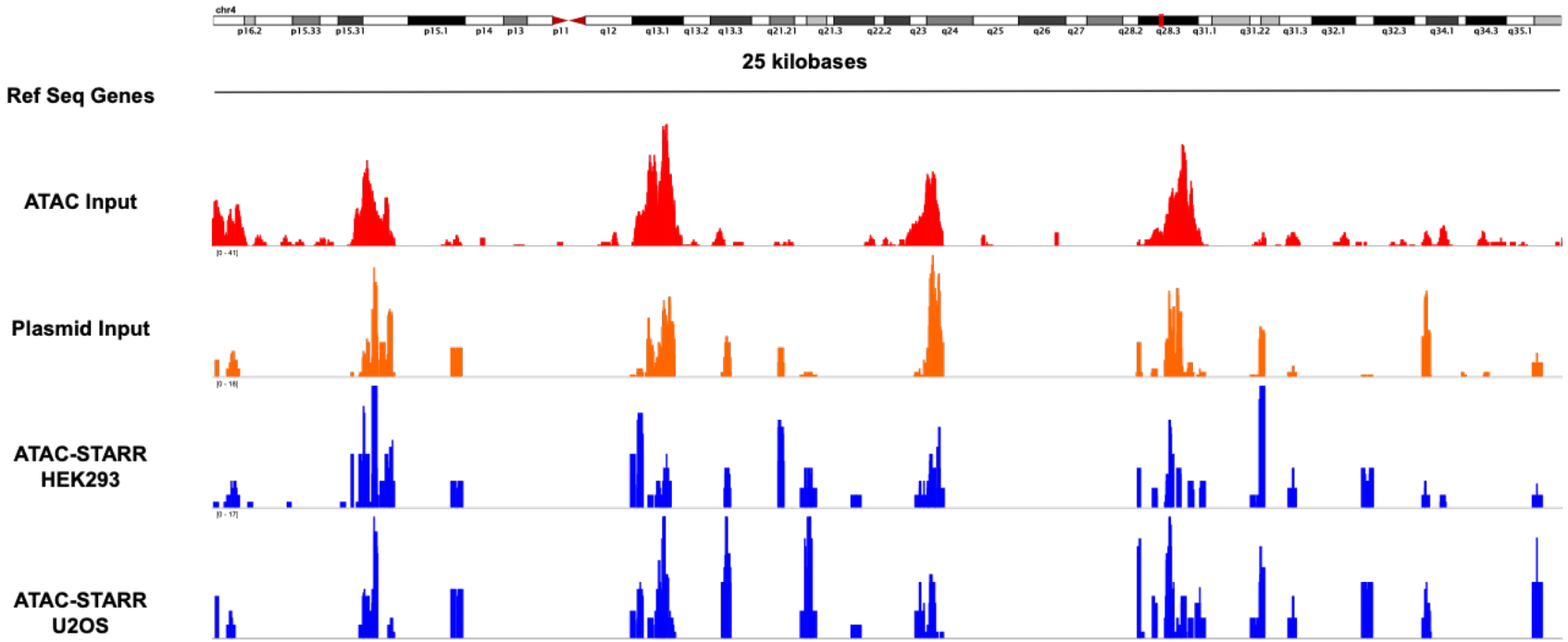

Figure 40. UCSC genome browser views of ATAC-STARR libraries. Genome browser tracks of sequencing libraries originating from either HEK293 ATAC-seq library (red), HEK293 ATAC-STARR plasmid library (orange), cDNA from HEK293 cells transfected with HEK293 ATAC-STARR plasmid library, or cDNA from U2OS cells transfected with HEK293 ATAC-STARR plasmid.

efficient RNA purification are all crucial components for the robust success of ATAC-STARR sequencing library construction.

Computational analysis of the resultant data also represents a considerable challenge for ATAC-STARR. Unlike methods that prepare a plasmid library from purified genomic DNA, ATAC-STARR is already pre-enriched for nucleosome depleted genomic loci. This enrichment is a boon for more manageable technical construction of the ATAC-STARR plasmid library but difficulty arises at the analysis stage. Specifically, ATAC-STARR forces the determination of further enrichment in the form of self-transcription originating from the plasmid library which is itself already enriched for these nucleosome depleted regions. Simply put, one must determine a region of enriched signal (self-transcription) from a region of enriched signal (chromatin accessibility/nucleosome depletion). Future development of the ATAC-STARR technique must address technical issues in tandem with computational difficulties if the methodology is set to be feasible for application across complex experiments.

The ultimate vision has always been to integrate ATAC-STARR and ATAC-Me data across a time course of cell differentiation. Sampling of ATAC DNA fragments at different stages of cell differentiation would allow capture of the full range of activating enhancer elements. Separate differentiation time points of an ATAC-STARR plasmid library could then be pooled together to create a “master” plasmid library containing DNA fragments of all possible gene regulatory elements. This “master” ATAC-STARR plasmid library can then be reintroduced into cells which are subsequently differentiated to monitor waves of gene regulatory element utilization. Tandem collection of ATAC-Me datasets across this differentiation offers useful inferences about the endogenous loci that ATAC-STARR plasmids originate from. ATAC-STARR is a complementary assay that offers a validated network of enhancers upon which to focus ATAC-Me analysis.

Much work is to be done before ATAC-STARR can be transitioned into a cell differentiation context. Current efforts by a Hodges Lab graduate student, Tyler Hansen, have

made significant progress to the technical and data analysis difficulties. ATAC-STARR is a promising technique for enhancing our identification of functional enhancers and their constituent most crucial base pair DNA sequences.

Chapter 5: Concluding Discussions

The research described within this dissertation work contributed to outstanding questions of epigenomic dynamics and cell fate specification. Specifically, questions regarding the relationship between DNA methylation, chromatin accessibility and gene regulation during cell differentiation. Limitations of current tools for tracking epigenomic dynamics drove the development of two promising methodologies. ATAC-Me simultaneously measures chromatin accessibility and DNA methylation offering a focused DNA binding factor relevant view of the methylome. Application of ATAC-Me across a time course of cell fate specification led to several key findings for continued investigation. Namely, uncoupled chromatin accessibility and DNA methylation dynamics, delayed loss of DNA methylation in nascent open chromatin, and transcriptional changes driven absent of DNA methylation dynamics. As complementary work ATAC-STARR was validated as an early development methodology mediating a focused massively parallel reporter assay. Validating putative gene regulatory elements with ATAC-STARR within dynamic cell differentiation contexts will assist in narrowing the focus of tandem epigenomic studies. However, no scientific endeavor is without flaw or cannot be improved upon by further study. The following is a discussion of improvements, limitations and directions for future study of epigenomic dynamics across cell differentiation.

The genomics methods used in this study employed bisulfite conversion to monitor DNA methylation at base pair resolution. While bisulfite conversion is considered the current gold standard for base pair resolution measurement of DNA methylation this technique is in actuality a hinderance to the field of epigenomics. The bisulfite conversion reaction is severely damaging to input DNA making it far from amenable to low input levels and single cell approaches. Though several works exploring usage of bisulfite based genomics methodologies for low input and single cell approaches exist they are not without considerable data limitations (Clark et al., 2018; Hui et al., 2018; Smallwood et al., 2014). The study of DNA methylation must move

beyond the limitations of bisulfite based methodologies. Several new promising methodologies can already be found within the literature. A recent publication described an alternative less damaging chemical conversion using pyridine borane that does not degrade input DNA in the same way as bisulfite conversion (Liu et al., 2019). Further, methodologies employing enzymatic treatments of DNA without use of bisulfite are available from manufacturers as prepared kits (Vaisvila et al., 2019). Both the pyridine borane and enzymatic approaches still offer base pair resolution analysis of DNA methylation but without the severe DNA damage and unwanted fragmentation of bisulfite conversion methods. Undoubtedly these new methods will open up more difficult systems to the study of DNA methylation, particularly in the case of single cell applications. Combining such new approaches with a focused sampling of open chromatin as in the ATAC-Me methodology would further enhance data quality, ease of library preparation and lower cellular input requirements.

The assays and data employed in this dissertation work are based on a homogenization of cell populations. While certainly useful in many contexts there are distinct limitations to the use of so-called bulk assays. The assumption when using a bulk assay is that the output data is representative of the population, however, that may not always be true. An array of cell phenotypes may exist across cell populations or a continuum of cell differentiation as has been shown in many systems to be the case. Within this dissertation work the THP1 system of cell fate specification in response to PMA was chosen in part to circumvent this issue of heterogeneity across the cell population as this model is known to be phenotypically homogeneous in PMA response. The bulk ATAC-Me and RNA-seq data in this dissertation work cannot with absolute certainty be said to be sampling a population of cells that are perfectly homogeneous. Perfect homogeneity is rarely the case in biology. The solution to this is to employ single cell based genomic approaches which allow one to assess the individual phenotype of each cell among a population. However, such methods are not trivial and require considerable effort to employ. Single cell RNA-seq is the most well-developed of these

methodologies in use today but the same cannot be said for single cell DNA methylation and chromatin accessibility analysis. Unlike frequently abundant transcriptional copies within cells the human genome exists as only two double stranded DNA molecules per cell. Efficient sampling of the full genome for each cell is impossible with current single cell methods assessing DNA methylation and chromatin accessibility. DNA methylation analysis in single cells is particularly plagued by allelic drop-out due to the further issue of damaging bisulfite conversion. A proposed coupling of single cell ATAC technologies with the aforementioned non-bisulfite genomics methods could achieve a clearer image of epigenomic heterogeneity across cell populations.

Chromatin accessibility and DNA methylation genomic analysis frequently infer TF binding by presence of enriched TF binding motifs among populations of genomic loci. However, this is only suggestive of actual TF binding as TF motifs are frequently known to be available for binding without actual engagement by the cognate TF (Wang et al., 2012b). A type of analysis long employed in biochemistry (Galas and Schmitz, 1978) and first pioneered on the genomic scale with DNase-seq (Hesselberth et al., 2009) data referred to as TF footprinting offers a partial solution. Effectively interruptions within regions enriched for ATAC (Buenrostro et al., 2013) or DNase sensitivity signal imply the occupancy of a DNA binding factor interfering with the enzymatic activity of DNase-I or Tn5 transposase. While inferior to ChIP-seq which directly assays binding of molecular factors across the genome, TF footprinting offers a more realistic profile of molecular DNA binding dynamics. TF footprinting also does not require *a priori* selection of TFs of interest, rather, TF footprints in general are matched with the TF motif DNA sequence upon which they rest. This manner of analysis was briefly explored as part of analysis within the THP1 ATAC-Me time course but can certainly be advanced and integrated further. In particular, ATAC-Me can profile TF footprints and the DNA methylation status of those footprints. An analysis of this manner is particularly useful to make inferences about the sensitivity of TF binding to DNA methylation *in-cellulo*. The TFs more prevalent in the THP1 TF

time course frequently did not contain sites for DNA methylation as such methylation footprint analysis was largely irrelevant. However, the same cannot be said in the context of other cell differentiation models. SELEX based analysis noted an enrichment for embryonic pluripotency TFs among TFs found to be sensitive to DNA methylation (Yin et al., 2017). This suggests applying ATAC-Me across an embryonic cell differentiation system would be particularly interesting in terms of methylation TF footprinting analysis.

The epigenomic dynamics described within this dissertation work are notably within a terminal cell fate model. THP1 cells upon stimulation with PMA no longer proceed through DNA replication and cannot be de-differentiated. The absence of DNA replication is a crucial parameter that must be explored further. Many studies have implicated the importance of DNA replication for removal of DNA methylation through both TET enzyme dependent and independent means. It is plausible that this lack of DNA replication explains the absence of any dramatic DNA methylation dynamics across the THP1 differentiation time course. For this reason, the ATAC-Me methodology must be applied across a cell differentiation model that maintains DNA replication capabilities. Human embryonic stem cells offer an ideal system for this study as these cells can be stimulated to differentiate *in-vitro* to a wide variety of cells. Ideally one could differentiate embryonic stem cells to a terminal state, however, current differentiation protocols frequently fail in this regard. Protocols which differentiated stem cells to traditionally terminal cell fates frequently continue to undergo cell division and DNA replication. In most cases this is due to occupying a more fetal-like cell fate rather than the truly terminal cell fate of the adult cell phenotype (Rossi et al., 2018; Snir et al., 2003). Nonetheless, it will be insightful to compare the epigenomic dynamics of a replicating cell differentiation system to a terminal system such as the THP1 PMA stimulation model. Improvement of *in-vitro* differentiation techniques may allow for capture of dynamics across the full breadth of replicating cell to non-replicating cell.

Associating intergenic/intronic epigenomic dynamics with specific gene transcriptional consequences is difficult without knowing the enhancer:gene relationship. This dissertation employed a nearest-neighbor strategy to associate distal intergenic/intronic genomic loci with actual genes. It is well known that this approach is frequently erroneous and results in mis-pairings between intergenic enhancers and the genes they control (Yao et al., 2015). Future study should assuredly attempt to incorporate more advanced data to more accurately partner intergenic loci with genes. Chromatin conformation capture techniques such as Hi-C-seq (High resolution chromatin conformation sequencing) offer a solution to this problem (Lieberman-Aiden et al., 2009). Physical molecular interaction of linearly distal enhancers with the gene promoters they control has been described multiple times as a hallmark of enhancer function (Pombo and Dillon, 2015). In this manner chromatin conformation capture techniques aim to map out the physical 3-dimensional DNA interactions between such intergenic loci and genes they likely control. While not without limitations, incorporating this type of data would greatly enhance the accuracy of enhancer:gene pairings across an ATAC-Me cell differentiation time course making epigenomic and transcription relationship all the more relevant. Current approaches for chromatin conformation capture type genomic sequencing have issues with reproducibility and data quality (Schmitt et al., 2016), however, improvements to approaches are rapidly arising making this a promising data source to guide future work.

Our current assessments of DNA methylation even in seemingly static systems may not be what they seem. Assumptions at regions of persistent DNA methylation are likely too simplistic. While these genomic regions may indeed be stable in regard to the overall methylation state, the same may not be true for individual CpG sites. Observations in the literature have frequently suggested more stochastic dynamics at what appear to be stable 5-mC sites (Laird et al., 2004; Liang et al., 2002; Metivier et al., 2008; Okano et al., 1999). As with any biochemical reaction there must be a bidirectional flow between substrate and products. In the case of DNA methylation, the substrate (CpG sites) and product (5-mC) can transition

between these two states by the actions of methylation mediated by DNMTs and demethylation mediated by TET, TDG and DNA replication. How rapid then is the switch in state from unmodified cytosine to modified cytosine and vice versa at what we observe to be stable DNA methylation loci? What level of constant maintenance must DNMTs carryout at 5-mC sites to counteract the constant biochemical action of DNA demethylation mechanisms? Unfortunately, current techniques for measuring DNA methylation genome wide cannot measure the stochasticity of 5-mC in this manner. A true understanding of DNA methylation "turn-over" is required for a more complete image of true DNA methylation dynamics. Any approach that aids in this understanding would be warmly welcomed by the field of epigenomics.

DNA methylation has been studied for decades. Deceivingly, many assume that in this length of time an endpoint conclusion has been reached regarding the functional role of DNA methylation. A functional role most commonly described as simply a repressive epigenetic mark deposited by maintenance and de novo DNMTs. It is far too simplistic to describe DNA methylation in this manner. It has been demonstrated many times that DNA methylation has many context specific roles and mechanisms. In particular, a crucial role for DNA methylation during cell differentiation is strongly suggested by the array of developmental defects when DNA methylation is artificially or pathogenically perturbed. This dissertation sheds light on the seemingly minimal role DNA methylation plays on gene regulation within the context of a terminal cell differentiation. Only further study of epigenomic and gene regulatory dynamics across more varieties of cell differentiation models can allow us to reach a more firm conclusion. With hope, the genomics methods developed as part of this dissertation will aid in reaching this goal.

References

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

Agius, F., Kapoor, A., and Zhu, J. (2006). Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation. Proceedings of the National Academy of Sciences of the United States of America *103*, 11796-11801.

Akalin, A., Garrett-Bakelman, F.E., Kormaksson, M., Busuttil, J., Zhang, L., Khrebtukova, I., Milne, T.A., Huang, Y., Biswas, D., Hess, J.L.*, et al.* (2012). Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. PLoS Genet *8*, e1002781.

Alabert, C., Barth, T., Reveron-Gomez, N., Sidoli, S., Schmidt, A., Jensen, O., Imhof, A., and Groth, A. (2015). Two distinct modes for propagation of histone PTMs across the cell cycle. Genes & Development *29*, 585-590.

Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P., D'Souza, Z., Nakayama, M., Matsuda, M., Turp, A., Ndjetehe, E.*, et al.* (2016). De novo DNA methylation drives 5hmC accumulation in mouse zygotes. Nature Cell Biology *18*, 225-+.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T.*, et al.* (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455-+.

Arner, E., Daub, C.O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drablos, F., Lennartsson, A., Ronnerblad, M., Hrydziuszko, O., Vitezic, M.*, et al.* (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. Science *347*, 1010-1014.

Arnold, C., Gerlach, D., Stelzer, C., Boryn, L., Rath, M., and Stark, A. (2013). Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. Science *339*, 1074-1077.

Athanasiadou, R., de Sousa, D., Myant, K., Merusi, C., Stancheva, I., and Bird, A. (2010). Targeting of De Novo DNA Methylation Throughout the Oct-4 Gene Regulatory Region in Differentiating Embryonic Stem Cells. Plos One *5*.

Auclair, G., Guibert, S., Bender, A., and Weber, M. (2014). Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. Genome Biology *15*.

Baek, S., Goldstein, I., and Hager, G.L. (2017). Bivariate Genomic Footprinting Detects Changes in Transcription Factor Activity. Cell Rep *19*, 1710-1722.

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Herault, Y., Guillou, F., and Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science *354*, 909-912.

Barnett, K.R., Decato, B.E., Scott, T.J., Hansen, T.J., Chen, B., Attalla, J., Smith, A.D., and Hodges, E. (2020). ATAC-Me Captures Prolonged DNA Methylation of Dynamic Chromatin Accessibility Loci during Cell Fate Transitions. Mol Cell *77*, 1350-1364.e1356.

Barwick, B.G., Scharer, C.D., Bally, A.P.R., and Boss, J.M. (2016). Plasma cell differentiation is coupled to division-dependent DNA hypomethylation and gene regulation. Nat Immunol *17*, 1216-1225.

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schubeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature *520*, 243-U278.

Bell, A., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482-485.

Berger, S.L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. Genes & Development *23*, 781-783.

Bernstein, B., Mikkelsen, T., Xie, X., Kamal, M., Huebert, D., 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.
Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J Mol Biol *203*, 971-983.
Bestor, T.H., Hellewell, S.B., and Ingram, V.M. (1984). Differentiation of two mouse cell lines is associated with hypomethylation of their genomes. Mol Cell Biol *4*, 1800-1806.
Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell *40*, 91-99.
Bird, A.P., and Southern, E.M. (1978). Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from Xenopus laevis. J Mol Biol *118*, 27-47.
Bird, A.P., and Taggart, M.H. (1980). Variable patterns of total DNA and rDNA methylation in animals. Nucleic Acids Res *8*, 1485-1497.
Blaschke, K., Ebata, K.T., Karimi, M.M., Zepeda-Martinez, J.A., Goyal, P., Mahapatra, S., Tam, A., Laird, D.J., Hirst, M., Rao, A.*, et al.* (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. Nature *500*, 222-226.
Bock, C., Beerman, I., Lien, W.H., Smith, Z.D., Gu, H., Boyle, P., Gnirke, A., Fuchs, E., Rossi, D.J., and Meissner, A. (2012). DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. Mol Cell *47*, 633-647.
Borgel, J., Guibert, S., Li, Y.F., Chiba, H., Schubeler, D., Sasaki, H., Forne, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. Nature Genetics *42*, 1093-U1090.
Bostick, M., Kim, J., Esteve, P., Clark, A., Pradhan, S., and Jacobsen, S. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760-1764.
Bourc'his, D., and Bestor, T. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature *431*, 96-99.
Bourc'his, D., Xu, G., Lin, C., Bollman, B., and Bestor, T. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science *294*, 2536-2539.
Boyle, A., Davis, S., Shulha, H., Meltzer, P., Margulies, E., Weng, Z., Furey, T., and Crawford, G. (2008). High-resolution mapping and characterization of open chromatin across the genome. Cell *132*, 311-322.
Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., and Cedar, H. (1994). Sp1 elements protect a CpG island from de novo methylation. Nature *371*, 435-438.
Brinkman, A.B., Gu, H., Bartels, S.J., 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, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (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.
Cardoso, M., and Leonhardt, H. (1999). DNA methyltransferase is actively retained in the cytoplasm during early development. Journal of Cell Biology *147*, 25-32.
Cedar, H., Solage, A., Glaser, G., and Razin, A. (1979). Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI. Nucleic Acids Res *6*, 2125-2132.
Chan, S., and Majeti, R. (2013). Role of DNMT3A, TET2, and IDH1/2 mutations in pre-leukemic stem cells in acute myeloid leukemia. International Journal of Hematology *98*, 648-657.

Charlet, J., Duymich, C.E., Lay, F.D., Mundbjerg, K., Dalsgaard Sorensen, K., Liang, G., and Jones, P.A. (2016). Bivalent Regions of Cytosine Methylation and H3K27 Acetylation Suggest an Active Role for DNA Methylation at Enhancers. Mol Cell *62*, 422-431.
Chen, T., Ueda, Y., Dodge, J., Wang, Z., and Li, E. (2003). Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Molecular and Cellular Biology *23*, 5594-5605.
Cimmino, L., Dolgalev, I., Wang, Y., Yoshimi, A., Martin, G.H., Wang, J., Ng, V., Xia, B., Witkowski, M.T., Mitchell-Flack, M.*, et al.* (2017). Restoration of TET2 Function Blocks Aberrant Self-Renewal and Leukemia Progression. Cell *170*, 1079-1095 e1020.
Clark, S.J., Argelaguet, R., Kapourani, C.A., Stubbs, T.M., Lee, H.J., Alda-Catalinas, C., Krueger, F., Sanguinetti, G., Kelsey, G., Marioni, J.C.*, et al.* (2018). scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. Nat Commun *9*, 781.
Cooper, D.N., and Krawczak, M. (1989). Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. Hum Genet *83*, 181-188.
Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B.*, et al.* (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods *14*, 959-962.
Core, L., Martins, A., Danko, C., Waters, C., Siepel, A., and Lis, J. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nature Genetics *46*, 1311-1320.
Core, L., Waterfall, J., and Lis, J. (2008). Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. Science *322*, 1845-1848.
Costa, Y., Ding, J., Theunissen, T., Faiola, F., Hore, T., Shliaha, P., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S.*, et al.* (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature *495*, 370-374.
Crawford, G., Davis, S., Scacheri, P., Renaud, G., Halawi, M., Erdos, M., Green, R., Meltzer, P., Wolfsberg, T., and Collins, F. (2006). DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. Nature Methods *3*, 503-509.
Creyghton, M., Cheng, A., Welstead, G., Kooistra, T., Carey, B., Steine, E., Hanna, J., Lodato, M., Frampton, G., Sharp, P.*, et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences of the United States of America *107*, 21931-21936.
Dawlaty, M., Breiling, A., Le, T., Barrasa, M., Raddatz, G., Gao, Q., Powell, B., Cheng, A., Faull, K., Lyko, F.*, et al.* (2014). Loss of Tet Enzymes Compromises Proper Differentiation of Embryonic Stem Cells. Developmental Cell *29*, 102-111.
Dawlaty, M., Breiling, A., Le, T., Raddatz, G., Barrasa, M., Cheng, A., Gao, Q., Powell, B., Li, Z., Xu, M.*, et al.* (2013). Combined Deficiency of Tet1 and Tet2 Causes Epigenetic Abnormalities but Is Compatible with Postnatal Development. Developmental Cell *24*, 310-323.
Dawlaty, M., Ganz, K., Powell, B., Hu, Y., Markoulaki, S., Cheng, A., Gao, Q., Kim, J., Choi, S., Page, D.*, et al.* (2011). Tet1 Is Dispensable for Maintaining Pluripotency and Its Loss Is Compatible with Embryonic and Postnatal Development. Cell Stem Cell *9*, 166-175.
Dodge, J., Okano, M., Dick, F., Tsujimoto, N., Chen, T., Wang, S., Ueda, Y., Dyson, N., and Li, E. (2005). Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. Journal of Biological Chemistry *280*, 17986-17991.
Doege, C., Inoue, K., Yamashita, T., Rhee, D., Travis, S., Fujita, R., Guarnieri, P., Bhagat, G., Vanti, W., Shih, A.*, et al.* (2012). Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. Nature *488*, 652-655.

Domcke, S., Bardet, A., Ginno, P., Hartl, D., Burger, L., and Schubeler, D. (2015). Competition between DNA methylation and transcription factors determines binding of NRF1. Nature *528*, 575-+.
Donaghey, J., Thakurela, S., Charlton, J., Chen, J.S., Smith, Z.D., Gu, H., Pop, R., Clement, K., Stamenova, E.K., Karnik, R.*, et al.* (2018). Genetic determinants and epigenetic effects of pioneer-factor occupancy. Nat Genet *50*, 250-258.
Dunwell, T.L., and Pfeifer, G.P. (2014). Drosophila genomic methylation: new evidence and new questions. Epigenomics-Uk *6*, 459-461.
Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat Methods *9*, 215-216.
Fang, F., Hodges, E., Molaro, A., Dean, M., Hannon, G.J., and Smith, A.D. (2012). Genomic landscape of human allele-specific DNA methylation. Proceedings of the National Academy of Sciences of the United States of America *109*, 7332-7337.
Farthing, C.R., Ficz, G., Ng, R.K., Chan, C.F., Andrews, S., Dean, W., Hemberger, M., and Reik, W. (2008). Global Mapping of DNA Methylation in Mouse Promoters Reveals Epigenetic Reprogramming of Pluripotency Genes. Plos Genetics *4*.
Fatemi, M., Hermann, A., Gowher, H., and Jeltsch, A. (2002). Dnmt3a and Dnmt1 functionally cooperate during de novo methylation of DNA. European Journal of Biochemistry *269*, 4981-4984.
Fatemi, M., Hermann, A., Pradhan, S., and Jeltsch, A. (2001). The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. Journal of Molecular Biology *309*, 1189-1199.
Feinberg, A.P., and Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature *301*, 89-92.
Feldmann, A., Ivanek, R., Murr, R., Gaidatzis, D., Burger, L., and Schubeler, D. (2013). Transcription Factor Occupancy Can Mediate Active Turnover of DNA Methylation at Regulatory Regions. Plos Genetics *9*.
Floess, S., Freyer, J., Siewert, C., Baron, U., Olek, S., Polansky, J., Schlawe, K., Chang, H., Bopp, T., Schmitt, E.*, et al.* (2007). Epigenetic control of the foxp3 locus in regulatory T cells. Plos Biology *5*, 169-178.
Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., and Paul, C.L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A *89*, 1827-1831.
Galas, D.J., and Schmitz, A. (1978). DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res *5*, 3157-3170.
Gama-Sosa, M.A., Slagel, V.A., Trewyn, R.W., Oxenhandler, R., Kuo, K.C., Gehrke, C.W., and Ehrlich, M. (1983). The 5-methylcytosine content of DNA from human tumors. Nucleic Acids Res *11*, 6883-6894.
Gao, Y., Chen, J., Li, K., Wu, T., Huang, B., Liu, W., Kou, X., Zhang, Y., Huang, H., Jiang, Y.*, et al.* (2013). Replacement of Oct4 by Tet1 during iPSC Induction Reveals an Important Role of DNA Methylation and Hydroxymethylation in Reprogramming. Cell Stem Cell *12*, 453-469.
Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. J Mol Biol *196*, 261-282.
Gaudet, F., Hodgson, J., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J., Leonhardt, H., and Jaenisch, R. (2003). Induction of tumors in mice by genomic hypomethylation. Science *300*, 489-492.
Gkountela, S., Zhang, K., Shafiq, T., Liao, W., Hargan-Calvopina, J., Chen, P., and Clark, A. (2015). DNA Demethylation Dynamics in the Human Prenatal Germline. Cell *161*, 1425-1436.

Goll, M., Kirpekar, F., Maggert, K., Yoder, J., Hsieh, C., Zhang, X., Golic, K., Jacobsen, S., and Bestor, T. (2006). Methylation of tRNA(AsP) by the DNA methyltransferase homolog Dnmt2. Science *311*, 395-398.
Grant, M., Zuccotti, M., and Monk, M. (1992). Methylation of CpG sites of two X-linked genes coincides with X-inactivation in the female mouse embryo but not in the germ line. Nat Genet *2*, 161-166.
Gu, P., Le Menuet, D., Chung, A., and Cooney, A. (2009). Retraction of Differential Recruitment of Methylated CpG Binding Domains by the Orphan Receptor GCNF Initiates the Repression and Silencing of Oct4 Expression (vol 26, pg 9471, 2006). Molecular and Cellular Biology *29*, 1987-1987.
Gu, T., Guo, F., Yang, H., Wu, H., Xu, G., Liu, W., Xie, Z., Shi, L., He, X., Jin, S.*, et al.* (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature *477*, 606-U136.
Guibert, S., Forne, T., and Weber, M. (2012). Global profiling of DNA methylation erasure in mouse primordial germ cells. Genome Research *22*, 633-641.
Guilhamon, P., Eskandarpour, M., Halai, D., Wilson, G., Feber, A., Teschendorff, A., Gomez, V., Hergovich, A., Tirabosco, R., Amary, M.*, et al.* (2013). Meta-analysis of IDH-mutant cancers identifies EBF1 as an interaction partner for TET2. Nature Communications *4*.
Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., Hu, Y., Wang, X., Wei, Y.*, et al.* (2015). The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. Cell *161*, 1437-1452.
Hackett, J., Sengupta, R., Zylicz, J., Murakami, K., Lee, C., Down, T., and Surani, M. (2013). Germline DNA Demethylation Dynamics and Imprint Erasure Through 5-Hydroxymethylcytosine. Science *339*, 448-452.
Hah, N., Danko, C.G., Core, L., Waterfall, J.J., Siepel, A., Lis, J.T., and Kraus, L.W. (2011). A Rapid, Extensive, and Transient Transcriptional Response to Estrogen Signaling in Breast Cancer Cells. Cell *145*, 622-634.
Han, L., Ren, C., Zhang, J., Shu, W., and Wang, Q. (2019). Differential roles of Stella in the modulation of DNA methylation during oocyte and zygotic development. Cell Discovery *5*.
Handa, V., and Jeltsch, A. (2005). Profound flanking sequence preference of Dnmt3a and Dnmt3b mammalian DNA methyltransferases shape the human epigenome. Journal of Molecular Biology *348*, 1103-1112.
Hashimoto, H., Liu, Y., Upadhyay, A., Chang, Y., Howerton, S., Vertino, P., Zhang, X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Research *40*, 4841-4849.
Hata, K., Okano, M., Lei, H., and Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development *129*, 1983-1993.
He, Y., Li, B., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L.*, et al.* (2011). Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. Science *333*, 1303-1307.
Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and 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*, 576-589.
Hellman, A., and Chess, A. (2007). Gene body-specific methylation on the active X chromosome. Science *315*, 1141-1143.
Hermann, A., Goyal, R., and Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. Journal of Biological Chemistry *279*, 48350-48359.

Hesselberth, J., Chen, X., Zhang, Z., Sabo, P., Sandstrom, R., Reynolds, A., Thurman, R., Neph, S., Kuehn, M., Noble, W.*, et al.* (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nature Methods *6*, 283-289.
Hewish, D.R., and Burgoyne, L.A. (1973). Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. Biochem Biophys Res Commun *52*, 504-510.
Hodges, E., Molaro, A., Dos Santos, C., Thekkat, P., Song, Q., Uren, P., Park, J., Butler, J., Rafii, S., McCombie, W.*, et al.* (2011). Directional DNA Methylation Changes and Complex Intermediate States Accompany Lineage Specificity in the Adult Hematopoietic Compartment. Molecular Cell *44*, 17-28.
Holliday, R., and Grigg, G.W. (1993). DNA methylation and mutation. Mutat Res *285*, 61-67.
Howell, C., Bestor, T., Ding, F., Latham, K., Mertineit, C., Trasler, J., and Chaillet, J. (2001). Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell *104*, 829-838.
Hui, T., Cao, Q., Wegrzyn-Woltosz, J., O'Neill, K., Hammond, C., Knapp, D., Laks, E., Moksa, M., Aparicio, S., Eaves, C.*, et al.* (2018). High-Resolution Single-Cell DNA Methylation Measurements Reveal Epigenetically Distinct Hematopoietic Stem Cell Subpopulations. Stem Cell Reports *11*, 578-592.
Inoue, A., Shen, L., Dai, Q., He, C., and Zhang, Y. (2011). Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. Cell Research *21*, 1670-1676.
Inoue, A., and Zhang, Y. (2011). Replication-Dependent Loss of 5-Hydroxymethylcytosine in Mouse Preimplantation Embryos. Science *334*, 194-194.
Iqbal, K., Jin, S., Pfeifer, G., and Szabo, P. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proceedings of the National Academy of Sciences of the United States of America *108*, 3642-3647.
Ito, S., D'Alessio, A., Taranova, O., Hong, K., Sowers, L., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129-U1151.
Jadhav, U., Cavazza, A., Banerjee, K.K., Xie, H., O'Neill, N.K., Saenz-Vash, V., Herbert, Z., Madha, S., Orkin, S.H., Zhai, H.*, et al.* (2019). Extensive Recovery of Embryonic Enhancer and Gene Memory Stored in Hypomethylated Enhancer DNA. Mol Cell *74*, 542-554 e545.
Ji, D., Lin, K., Song, J., and Wang, Y. (2014). Effects of Tet-induced oxidation products of 5-methylcytosine on Dnmt1-and DNMT3a-mediated cytosine methylation. Molecular Biosystems *10*, 1749-1752.
Jia, D., Jurkowska, R., Zhang, X., Jeltsch, A., and Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature *449*, 248-U213.
Jin, C., Qin, T., Barton, M., Jelinek, J., and Issa, J. (2015). Minimal role of base excision repair in TET-induced global DNA demethylation in HEK293T cells. Epigenetics *10*, 1006-1013.
Jones, P., Veenstra, G., Wade, P., Vermaak, D., Kass, S., Landsberger, N., Strouboulis, J., and Wolffe, A. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nature Genetics *19*, 187-191.
Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M., and Saitou, M. (2013). Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. Embo Journal *32*, 340-353.
Kaiser, S., Jurkowski, T., Kellner, S., Schneider, D., Jeltsch, A., and Helm, M. (2017). The RNA methyltransferase Dnmt2 methylates DNA in the structural context of a tRNA. Rna Biology *14*, 1241-1251.
Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature *429*, 900-903.

Kantarjian, H., Issa, J., Rosenfeld, C., Bennett, J., Albitar, M., DiPersio, J., Klimek, V., Slack, J., de Castro, C., Ravandi, F.*, et al.* (2006). Decitabine improves patient outcomes in myelodysplastic syndromes - Resuits of a Phase III randomized study. Cancer *106*, 1794-1803.
Kass, S.U., Landsberger, N., and Wolffe, A.P. (1997). DNA methylation directs a time dependent repression of transcription initiation. Current Biology *7*, 157-165.
Kelly, T.K., Liu, Y., Lay, F.D., Liang, G., Berman, B.P., and Jones, P.A. (2012). Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res *22*, 2497-2506.
Keshet, I., Lieman-Hurwitz, J., and Cedar, H. (1986). DNA methylation affects the formation of active chromatin. Cell *44*, 535-543.
Ko, M., An, J., Bandukwala, H., Chavez, L., Aijo, T., Pastor, W., Segal, M., Li, H., Koh, K., Lahdesmaki, H.*, et al.* (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. Nature *497*, 122-+.
Ko, M., Bandukwala, H., An, J., Lamperti, E., Thompson, E., Hastie, R., Tsangaratou, A., Rajewsky, K., Koralov, S., and Rao, A. (2011). Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. Proceedings of the National Academy of Sciences of the United States of America *108*, 14566-14571.
Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., Sato, S., Nakabayashi, K., Hata, K., Sotomaru, Y.*, et al.* (2012). Contribution of Intragenic DNA Methylation in Mouse Gametic DNA Methylomes to Establish Oocyte-Specific Heritable Marks. Plos Genetics *8*.
Koche, R., Smith, Z., Adli, M., Gu, H., Ku, M., Gnirke, A., Bernstein, B., and Meissner, A. (2011). Reprogramming Factor Expression Initiates Widespread Targeted Chromatin Remodeling. Cell Stem Cell *8*, 96-105.
Kodzius, R., Kojima, M., Nishiyori, H., Nakamura, M., Fukuda, S., Tagami, M., Sasaki, D., Imamura, K., Kai, C., Harbers, M.*, et al.* (2006). CAGE: cap analysis of gene expression. Nature Methods *3*, 211-222.
Kribelbauer, J., Laptenko, O., Chen, S., Martini, G., Freed-Pastor, W., Prives, C., Mann, R., and Bussemaker, H. (2017). Quantitative Analysis of the DNA Methylation Sensitivity of Transcription Factor Complexes. Cell Reports *19*, 2383-2395.
Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.*, et al.* (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317-330.
Kwasnieski, J., Fiore, C., Chaudhari, H., and Cohen, B. (2014). High-throughput functional testing of ENCODE segmentation predictions. Genome Research *24*, 1595-1602.
Laird, C., Pleasant, N., Clark, A., Sneeden, J., Hassan, K., Manley, N., Vary, J., Morgan, T., Hansen, R., and Stoger, R. (2004). Hairpin-bisulfite PCR: Assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. Proceedings of the National Academy of Sciences of the United States of America *101*, 204-209.
Langemeijer, S., Kuiper, R., Berends, M., Knops, R., Aslanyan, M., Massop, M., van Hoogen, P., van Kessel, A., Raymakers, R., Verburgh, E.*, et al.* (2009). Acquired mutations in TET2 are common in myelodysplastic syndromes. Leukemia Research *33*, S88-S89.
Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). CpG islands as gene markers in the human genome. Genomics *13*, 1095-1107.
Legrand, C., Tuorto, F., Hartmann, M., Liebers, R., Jacob, D., Helm, M., and Lyko, F. (2017). Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. Genome Research *27*, 1589-1596.
Lei, H., Oh, S., Okano, M., Juttermann, R., Goss, K., Jaenisch, R., and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development *122*, 3195-3205.

Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. Cell *157*, 13-25.
Ley, T., and Grant, G.A.P.P. (2011). DNMT3A Mutations in Acute Myeloid Leukemia. Blood *118*, 1815-1815.
Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915-926.
Li, E., and Zhang, Y. (2014). DNA Methylation in Mammals. Cold Spring Harbor Perspectives in Biology *6*.
Li, J., Wu, X., Zhou, Y., Lee, M., Guo, L., Han, W., Mo, W., Cao, W.M., Sun, D., Xie, R.*, et al.* (2018a). Decoding the dynamic DNA methylation and hydroxymethylation landscapes in endodermal lineage intermediates during pancreatic differentiation of hESC. Nucleic Acids Res *46*, 2883-2900.
Li, R., Cauchy, P., Ramamoorthy, S., Boller, S., Chavez, L., and Grosschedl, R. (2018b). Dynamic EBF1 occupancy directs sequential epigenetic and transcriptional events in B-cell programming. Genes Dev *32*, 96-111.
Li, Y., and Tollefsbol, T.O. (2011). Combined chromatin immunoprecipitation and bisulfite methylation sequencing analysis. Methods Mol Biol *791*, 239-251.
Li, Y., Zhang, Z., Chen, J., Liu, W., Lai, W., Liu, B., Li, X., Liu, L., Xu, S., Dong, Q.*, et al.* (2018c). Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. Nature *564*, 136-+.
Liang, G., Chan, M., Tomigahara, Y., Tsai, Y., Gonzales, F., Li, E., Laird, P., and Jones, P. (2002). Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. Molecular and Cellular Biology *22*, 480-491.
Lieberman-Aiden, E., van Berkum, N., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B., Sabo, P., Dorschner, M.*, et al.* (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science *326*, 289-293.
Lin, I., Han, L., Taghva, A., O'Brien, L., and Hsieh, C. (2002). Murine de novo methyltransferase Dnmt3a demonstrates strand asymmetry and site preference in the methylation of DNA in vitro. Molecular and Cellular Biology *22*, 704-723.
Lio, C.W., Zhang, J., Gonzalez-Avalos, E., Hogan, P.G., Chang, X., and Rao, A. (2016). Tet2 and Tet3 cooperate with B-lineage transcription factors to regulate DNA modification and chromatin accessibility. Elife *5*.
Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M.*, et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315-322.
Liu, X., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R., and Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. Cell *167*, 233-+.
Liu, Y., Siejka-Zielinska, P., Velikova, G., Bi, Y., Yuan, F., Tomkova, M., Bai, C., Chen, L., Schuster-Bockler, B., and Song, C. (2019). Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. Nature Biotechnology *37*, 424-+.
Lock, L.F., Takagi, N., and Martin, G.R. (1987). Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. Cell *48*, 39-46.
Lorincz, M., Schubeler, D., Hutchinson, S., Dickerson, D., and Groudine, M. (2002). DNA methylation density influences the stability of an epigenetic imprint and Dnmt3a/b - Independent de novo methylation. Molecular and Cellular Biology *22*, 7572-7580.
Lu, H.L., Yuan, Z.M., Tan, T., Wang, J.W., Zhang, J.Y., Luo, H.J., Xia, Y.D., Ji, W.Z., and Gao, F. (2015). Improved tagmentation-based whole-genome bisulfite sequencing for input DNA from less than 100 mammalian cells. Epigenomics-Uk *7*, 47-56.
Maegawa, S., Hinkal, G., Kim, H., Shen, L., Zhang, L., Zhang, J., Zhang, N., Liang, S., Donehower, L., and Issa, J. (2010). Widespread and tissue specific age-related DNA methylation changes in mice. Genome Research *20*, 332-340.

Mahat, D., Kwak, H., Booth, G., Jonkers, I., Danko, C., Patel, R., Waters, C., Munson, K., Core, L., and Lis, J. (2016). Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). Nature Protocols *11*, 1455-1476.
Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J Biol Chem *286*, 35334-35338.
Martos, S.N., Li, T., Ramos, R.B., Lou, D., Dai, H., Xu, J.C., Gao, G., Gao, Y., Wang, Q., An, C., *et al.* (2017). Two approaches reveal a new paradigm of 'switchable or genetics-influenced allele-specific DNA methylation' with potential in human disease. Cell Discov *3*, 17038.
Mayran, A., Khetchoumian, K., Hariri, F., Pastinen, T., Gauthier, Y., Balsalobre, A., and Drouin, J. (2018). Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. Nat Genet *50*, 259-269.
Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res *33*, 5868-5877.
Meissner, A., Mikkelsen, T.S., Gu, H.C., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X.L., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., *et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-U791.
Metivier, R., Gallais, R., Tiffoche, C., Le Peron, C., Jurkowska, R., Carmouche, R., Ibberson, D., Barath, P., Demay, F., Reid, G., *et al.* (2008). Cyclical DNA methylation of a transcriptionally active promoter. Nature *452*, 45-U42.
Mikkelsen, T., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B., Jaenisch, R., Lander, E., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49-U41.
Molaro, A., Hodges, E., Fang, F., Song, Q., McCombie, W., Hannon, G., and Smith, A. (2011). Sperm Methylation Profiles Reveal Features of Epigenetic Inheritance and Evolution in Primates. Cell *146*, 1028-1040.
Montefiori, L., Hernandez, L., Zhang, Z., Gilad, Y., Ober, C., Crawford, G., Nobrega, M., and Jo Sakabe, N. (2017). Reducing mitochondrial reads in ATAC-seq using CRISPR/Cas9. Sci Rep *7*, 2451.
Morales-Ruiz, T., Ortega-Galisteo, A.P., Ponferrada-Marin, M.I., Martinez-Macias, M.I., Ariza, R.R., and Roldan-Arjona, T. (2006). DEMETER and REPRESSOR OFSILENCING 1 encode 5-methylcytosine DNA glycosylases. Proceedings of the National Academy of Sciences of the United States of America *103*, 6853-6858.
Morselli, M., Pastor, W.A., Montanini, B., Nee, K., Ferrari, R., Fu, K., Bonora, G., Rubbi, L., Clark, A.T., Ottonello, S., *et al.* (2015). In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. Elife *4*.
Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M., *et al.* (2007). PGC7/Stella protects against DNA demethylation in early embryogenesis. Nature Cell Biology *9*, 64-U81.
Nakamura, T., Liu, Y., Nakashima, H., Umehara, H., Inoue, K., Matoba, S., Tachibana, M., Ogura, A., Shinkai, Y., and Nakano, T. (2012). PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. Nature *486*, 415-+.
Nan, X., Campoy, F., and Bird, A. (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell *88*, 471-481.
Nan, X.S., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386-389.
Nekrasov, M., Amrichova, J., Parker, B., Soboleva, T., Jack, C., Williams, R., Huttley, G., and Tremethick, D. (2012). Histone H2A.Z inheritance during the cell cycle and its impact on promoter organization and dynamics. Nature Structural & Molecular Biology *19*, 1076-+.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. Nature *543*, 72-+.
Ng, H., Jeppesen, P., and Bird, A. (2000). Active repression of methylated genes by the chromosomal protein MBD1. Molecular and Cellular Biology *20*, 1394-1406.
Ng, H., Zhang, Y., Hendrich, B., Johnson, C., Turner, B., Erdjument-Bromage, H., Tempst, P., Reinberg, D., and Bird, A. (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nature Genetics *23*, 58-61.
Okano, M., Bell, D., Haber, D., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.
Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nature Genetics *19*, 219-220.
Otani, J., Kimura, H., Sharif, J., Endo, T.A., Mishima, Y., Kawakami, T., Koseki, H., Shirakawa, M., Suetake, I., and Tajima, S. (2013). Cell cycle-dependent turnover of 5-hydroxymethyl cytosine in mouse embryonic stem cells. PLoS One *8*, e82961.
Pacis, A., Tailleux, L., Morin, A.M., Lambourne, J., MacIsaac, J.L., Yotova, V., Dumaine, A., Danckaert, A., Luca, F., Grenier, J.C.*, et al.* (2015). Bacterial infection remodels the DNA methylation landscape of human dendritic cells. Genome Res *25*, 1801-1811.
Palii, S.S., Van Emburgh, B.O., Sankpal, U.T., Brown, K.D., and Robertson, K.D. (2008). DNA methylation inhibitor 5-aza-2 '-deoxycytidine induces reversible genome-wide DNA damage that is distinctly influenced by DNA methyltransferases 1 and 3B. Molecular and Cellular Biology *28*, 752-771.
Paz, M., Fraga, M., Avila, S., Guo, M., Pollan, M., Herman, J., and Esteller, M. (2003). A systematic profile of DNA methylation in human cancer cell lines. Cancer Research *63*, 1114-1121.
Phanstiel, D.H., Van Bortle, K., Spacek, D., Hess, G.T., Shamim, M.S., Machol, I., Love, M.I., Aiden, E.L., Bassik, M.C., and Snyder, M.P. (2017). Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs during Macrophage Development. Mol Cell *67*, 1037-1048 e1036.
Picelli, S., Bjorklund, A.K., Reinius, B., Sagasser, S., Winberg, G., and Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res *24*, 2033-2040.
Piper, J., Elze, M.C., Cauchy, P., Cockerill, P.N., Bonifer, C., and Ott, S. (2013). Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. Nucleic Acids Res *41*, e201.
Polo, J., Anderssen, E., Walsh, R., Schwarz, B., Nefzger, C., Lim, S., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J.*, et al.* (2012). A Molecular Roadmap of Reprogramming Somatic Cells into iPS Cells. Cell *151*, 1617-1632.
Pombo, A., and Dillon, N. (2015). Three-dimensional genome architecture: players and mechanisms. Nature Reviews Molecular Cell Biology *16*, 245-257.
Rada-Iglesias, A., Bajpai, R., Prescott, S., Brugmann, S.A., Swigut, T., and Wysocka, J. (2012). Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. Cell Stem Cell *11*, 633-648.
Raddatz, G., Hagemann, S., Aran, D., Sohle, J., Kulkarni, P., Kaderali, L., Hellman, A., Winnefeld, M., and Lyko, F. (2013). Aging is associated with highly defined epigenetic changes in the human epidermis. Epigenetics & Chromatin *6*.
Ragunathan, K., Jih, G., and Moazed, D. (2015). Epigenetic inheritance uncoupled from sequence-specific recruitment. Science *348*.
Riising, E.M., Comet, I., Leblanc, B., Wu, X.D., Johansen, J.V., and Helin, K. (2014). Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide. Molecular Cell *55*, 347-360.

Rossi, G., Manfrin, A., and Lutolf, M. (2018). Progress and potential in organoid research. Nature Reviews Genetics *19*, 671-687.
Russler-Germain, D., Spencer, D., Young, M., Lamprecht, T., Miller, C., Fulton, R., Meyer, M., Erdmann-Gilmore, P., Townsend, R., Wilson, R.*, et al.* (2014). The R882H DNMT3A Mutation Associated with AML Dominantly Inhibits Wild-Type DNMT3A by Blocking Its Ability to Form Active Tetramers. Cancer Cell *25*, 442-454.
Russo, V.E.A.a.R.A.D.a.M.R.A. (1996). Epigenetic mechanisms of gene regulation.
Sabo, P., Kuehn, M., Thurman, R., Johnson, B., Johnson, E., Hua, C., Man, Y., Rosenzweig, E., Goldy, J., Haydock, A.*, et al.* (2006). Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. Nature Methods *3*, 511-518.
Sanchez-Romero, M.A., Cota, I., and Casadesus, J. (2015). DNA methylation in bacteria: from the methyl group to the methylome. Current Opinion in Microbiology *25*, 9-16.
Sardina, J., Collombet, S., Tian, T., Gomez, A., Di Stefano, B., Berenguer, C., Brumbaugh, J., Stadhouders, R., Segura-Morales, C., Gut, M.*, et al.* (2018). Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate. Cell Stem Cell *23*, 727-+.
Schermelleh, L., Spada, F., Easwaran, H., Zolghadr, K., Margot, J., Cardoso, M., and Leonhardt, H. (2005). Trapped in action: direct visualization of DNA methyltransferase activity in living cells. Nature Methods *2*, 751-756.
Schlesinger, F., Smith, A., Gingeras, T., Hannon, G., and Hodges, E. (2013). De novo DNA demethylation and noncoding transcription define active intergenic regulatory elements. Genome Research *23*, 1601-1614.
Schmitt, A., Hu, M., and Ren, B. (2016). Genome-wide mapping and analysis of chromosome architecture. Nature Reviews Molecular Cell Biology *17*, 743-755.
Schones, D., Cui, K., Cuddapah, S., Roh, T., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887-898.
Schubeler, D. (2015). Function and information content of DNA methylation. Nature *517*, 321-326.
Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. Molecular Cell *48*, 849-862.
Serandour, A.A., Avner, S., Oger, F., Bizot, M., Percevault, F., Lucchetti-Miganeh, C., Palierne, G., Gheeraert, C., Barloy-Hubler, F., Peron, C.L.*, et al.* (2012). Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. Nucleic Acids Res *40*, 8255-8265.
Sheaffer, K., Kim, R., Aoki, R., Elliott, E., Schug, J., Burger, L., Schubeler, D., and Kaestner, K. (2014). DNA methylation is required for the control of stem cell differentiation in the small intestine. Genes & Development *28*, 652-664.
Shibata, H., Yoshino, K., Sunahara, S., Gondo, Y., Katsuki, M., Ueda, T., Kamiya, M., Muramatsu, M., Murakami, Y., Kalcheva, I.*, et al.* (1996). Inactive allele-specific methylation and chromatin structure of the imprinted gene U2af1-rs1 on mouse chromosome 11. Genomics *35*, 248-252.
Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature *479*, 74-U99.
Silverman, L., Demakos, E., Peterson, B., Kornblith, A., Holland, J., Odchimar-Reissig, R., Stone, R., Nelson, D., Powell, B., DeCastro, C.*, et al.* (2002). Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: A study of the cancer and leukemia group B. Journal of Clinical Oncology *20*, 2429-2440.
Singer-Sam, J., Grant, M., LeBon, J.M., Okuyama, K., Chapman, V., Monk, M., and Riggs, A.D. (1990). Use of a HpaII-polymerase chain reaction assay to study DNA methylation in the Pgk-1

CpG island of mouse embryos at the time of X-chromosome inactivation. Mol Cell Biol *10*, 4987-4989.
Smallwood, S., Lee, H., Angermueller, C., Krueger, F., Saadeh, H., Peet, J., Andrews, S., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nature Methods *11*, 817-820.
Smallwood, S., Tomizawa, S., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S., and Kelsey, G. (2011). Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nature Genetics *43*, 811-U126.
Snir, M., Kehat, I., Gepstein, A., Coleman, R., Itskovitz-Eldor, J., Livne, E., and Gepstein, L. (2003). Assessment of the ultrastructural and proliferative properties of human embryonic stem cell-derived cardiomyocytes. American Journal of Physiology-Heart and Circulatory Physiology *285*, H2355-H2363.
Song, Q., Decato, B., Hong, E., Zhou, M., Fang, F., Qu, J., Garvin, T., Kessler, M., Zhou, J., and Smith, A. (2013). A Reference Methylome Database and Analysis Pipeline to Facilitate Integrative and Comparative Epigenomics. Plos One *8*.
Sorm, F., Píškala, A., Cihák, A., and Veselý, J. (1964). 5-Azacytidine, a new, highly effective cancerostatic. Experientia *20*, 202-203.
Spektor, R., Tippens, N.D., Mimoso, C.A., and Soloway, P.D. (2019). methyl-ATAC-seq measures DNA methylation at accessible chromatin. Genome Research *29*, 969-977.
Sproul, D., Nestor, C., Culley, J., Dickson, J., Dixon, J., Harrison, D., Meehan, R., Sims, A., and Ramsahoye, B. (2011). Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. Proceedings of the National Academy of Sciences of the United States of America *108*, 4364-4369.
Stadler, M., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., Wirbelauer, C., Oakeley, E., Gaidatzis, D., Tiwari, V.*, et al.* (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490-495.
Stein, R., Razin, A., and Cedar, H. (1982). In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells. Proceedings of the National Academy of Sciences of the United States of America *79*, 3418-3422.
Stern, J.L., Paucek, R.D., Huang, F.W., Ghandi, M., Nwumeh, R., Costello, J.C., and Cech, T.R. (2017). Allele-Specific DNA Methylation and Its Interplay with Repressive Histone Marks at Promoter-Mutant TERT Genes. Cell Rep *21*, 3700-3707.
Stresemann, C., and Lyko, F. (2008). Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. International Journal of Cancer *123*, 8-13.
Sun, D., Luo, M., Jeong, M., Rodriguez, B., Xia, Z., Hannah, R., Wang, H., Le, T., Faull, K., Chen, R.*, et al.* (2014). Epigenomic Profiling of Young and Aged HSCs Reveals Concerted Changes during Aging that Reinforce Self-Renewal. Cell Stem Cell *14*, 673-688.
Suzuki, M., Liao, W., Wos, F., Johnston, A.D., DeGrazia, J., Ishii, J., Bloom, T., Zody, M.C., Germer, S., and Greally, J.M. (2018). Whole-genome bisulfite sequencing with improved accuracy and cost. Genome Research *28*, 1364-1371.
Tahiliani, M., Koh, K., Shen, Y., Pastor, W., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L., Liu, D., Aravind, L.*, et al.* (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. Science *324*, 930-935.
Takashiba, S., Van Dyke, T.E., Amar, S., Murayama, Y., Soskolne, A.W., and Shapira, L. (1999). Differentiation of monocytes to macrophages primes cells for lipopolysaccharide stimulation via accumulation of cytoplasmic nuclear factor kappaB. Infection and immunity *67*, 5573-5578.
Teissandier, A., and Bourc'his, D. (2017). Gene body DNA methylation conspires with H3K36me3 to preclude aberrant transcription. Embo Journal *36*, 1471-1473.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B.*, et al.* (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75-82.
Traore, K., Trush, M.A., George, M., Jr., Spannhake, E.W., Anderson, W., and Asseffa, A. (2005). Signal transduction of phorbol 12-myristate 13-acetate (PMA)-induced growth inhibition of human monocytic leukemia THP-1 cells is reactive oxygen dependent. Leuk Res *29*, 863-879.
Tsuchiya, S., Kobayashi, Y., Goto, Y., Okumura, H., Nakae, S., Konno, T., and Tada, K. (1982). Induction of maturation in cultured human monocytic leukemia cells by a phorbol diester. Cancer research *42*, 1530-1536.
Vaisvila, R., Ponnaluri, V.K.C., Sun, Z., Langhorst, B.W., Saleh, L., Guan, S., Dai, N., Campbell, M.A., Sexton, B., Marks, K.*, et al.* (2019). EM-seq: Detection of DNA Methylation at Single Base Resolution from Picograms of DNA. bioRxiv, 2019.2012.2020.884692.
Vardimon, L., Kressmann, A., Cedar, H., Maechler, M., and Doerfler, W. (1982). Expression of a Cloned Adenovirus Gene Is Inhibited by Invitro Methylation. P Natl Acad Sci-Biol *79*, 1073-1077.
Vilkaitis, G., Suetake, I., Klimasauskas, S., and Tajima, S. (2005). Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. Journal of Biological Chemistry *280*, 64-72.
Waddington, C.H. (1953). EPIGENETICS AND EVOLUTION. Symposia of the Society for Experimental Biology *7*, 186-199.
Waddington, C.H. (2012). The Epigenotype. International Journal of Epidemiology *41*, 10-13.
Walsh, C., Chaillet, J., and Bestor, T. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nature Genetics *20*, 116-117.
Wang, H., Maurano, M., Qu, H., Varley, K., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R.*, et al.* (2012a). Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Research *22*, 1680-1688.
Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T., Greven, M., Pierce, B., Dong, X., Kundaje, A., Cheng, Y.*, et al.* (2012b). Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome Research *22*, 1798-1812.
Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Li, G., Ci, W., Li, W.*, et al.* (2014). Programming and Inheritance of Parental DNA Methylomes in Mammals. Cell *157*, 979-991.
Wang, Q., Gu, L., Adey, A., Radlwimmer, B., Wang, W., Hovestadt, V., Bahr, M., Wolf, S., Shendure, J., Eils, R.*, et al.* (2013). Tagmentation-based whole-genome bisulfite sequencing. Nat Protoc *8*, 2022-2032.
Weber, A., Krawczyk, C., Robertson, A., Kusnierczyk, A., Vagbo, C., Schuermann, D., Klungland, A., and Schar, P. (2016). Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. Nature Communications *7*.
Weber, M., Davies, J., Wittig, D., Oakeley, E., Haase, M., Lam, W., and Schubeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nature Genetics *37*, 853-862.
Wiench, M., John, S., Baek, S., Johnson, T., Sung, M., Escobar, T., Simmons, C., Pearce, K., Biddie, S., Sabo, P.*, et al.* (2011). DNA methylation status predicts cell type-specific enhancer activity. Embo Journal *30*, 3028-3039.
Williams, K., Christensen, J., Pedersen, M., Johansen, J., Cloos, P., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343-U472.
Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., and Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nature Communications *2*.

Wu, H., D'Alessio, A., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y., and Zhang, Y. (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature *473*, 389-U578.
Wu, M., and Gu, L. (2019). TCseq: Time course sequencing data analysis. . R package version 1.6.1.
Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S., Ito, S., Yang, C., Xiao, M., Liu, L.*, et al.* (2011). Oncometabolite 2-Hydroxyglutarate Is a Competitive Inhibitor of alpha-Ketoglutarate-Dependent Dioxygenases. Cancer Cell *19*, 17-30.
Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J., Bian, C., Hu, Y., Hu, D., Zhao, B., Cerovina, T.*, et al.* (2012). Tet3 CXXC Domain and Dioxygenase Activity Cooperatively Regulate Key Genes for Xenopus Eye and Neural Development. Cell *151*, 1200-1213.
Yamaguchi, S., Hong, K., Liu, R., Inoue, A., Shen, L., Zhang, K., and Zhang, Y. (2013). Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. Cell Research *23*, 329-339.
Yan, X., Xu, J., Gu, Z., Pan, C., Lu, G., Shen, Y., Shi, J., Zhu, Y., Tang, L., Zhang, X.*, et al.* (2011). Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. Nature Genetics *43*, 309-U351.
Yang, X.J., Han, H., De Carvalho, D.D., Lay, F.D., Jones, P.A., and Liang, G.N. (2014). Gene Body Methylation Can Alter Gene Expression and Is a Therapeutic Target in Cancer. Cancer Cell *26*, 577-590.
Yao, L., Berman, B., and Farnham, P. (2015). Demystifying the secret mission of enhancers: linking distal regulatory elements to target genes. Critical Reviews in Biochemistry and Molecular Biology *50*, 550-573.
Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P., Kivioja, T., Dave, K., Zhong, F.*, et al.* (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science *356*.
Yue, X., Trifari, S., Aijo, T., Tsagaratou, A., Pastor, W., Zepeda-Martinez, J., Lio, C., Li, X., Huang, Y., Vijayanand, P.*, et al.* (2016). Control of Foxp3 stability through modulation of TET activity. Journal of Experimental Medicine *213*, 377-397.
Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. Science *328*, 916-919.
Zentner, G.E., Tesar, P.J., and Scacheri, P.C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res *21*, 1273-1283.
Zhang, H., Zhang, X., Clark, E., Mulcahey, M., Huang, S., and Shi, Y. (2010). TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. Cell Research *20*, 1390-1393.
Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.
Ziller, M., Hansen, K., Meissner, A., and Aryee, M. (2015). Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. Nature Methods *12*, 230-+.