The chromatin accessibility landscape in the dental pulp of mouse molars and incisors

Pujan Joshi\(^1\), Anushree Vijaykumar\(^2\), Badam Enkhmandakh\(^3\), Dong-Guk Shin\(^1\), Mina Mina\(^2\) and Dashzeveg Bayarsaihan\(^3,4,\)\(\dagger\)

\(^1\)Computer Science and Engineering Department, University of Connecticut, Storrs, CT 06269, USA; \(^2\)Department of Craniofacial Sciences, University of Connecticut Health Center, Farmington, CT 06030, USA; \(^3\)Center for Regenerative Medicine & Skeletal Development, Department of Reconstructive Sciences, University of Connecticut Health Center, Farmington, CT 06030, USA; \(^4\)Institute for System Genomics, University of Connecticut, Storrs, CT 06269, USA

The dental pulp is a promising source of progenitor cells for regenerative medicine. The natural function of dental pulp is to produce odontoblasts to generate reparative dentin. Stem cells within the pulp tissue originate from the migrating neural crest cells and possess mesenchymal stem cell properties with the ability to differentiate into multiple lineages. To elucidate the transcriptional control mechanisms underlying cell fate determination, we compared the transcriptome and chromatin accessibility in primary dental pulp tissue derived from 5–6-day-old mice. Using RNA sequencing and assay for transposase-accessible chromatin using sequencing (ATAC-seq), we correlated gene expression with chromatin accessibility. We found that the majority of ATAC-seq peaks were concentrated at genes associated with development and cell differentiation. Most of these genes were highly expressed in the mouse dental pulp. Surprisingly, we uncovered a group of genes encoding master transcription factors that were not expressed in the dental pulp but retained open chromatin states. Within this group, we identified key developmental genes important for specification of the neural crest, adipocyte, neural, myoblast, osteoblast and hepatocyte lineages. Collectively, our results uncover a complex relationship between gene expression and the chromatin accessibility landscape in the mouse dental pulp.

Acknowledgement of Financial Support: This study was supported by OVPF from the University of Connecticut, REP-401754-20144-20 to D.B., R01-DE016689 to M.M., a predoctoral fellowship provided from the Department of Computer Science and Engineering, University of Connecticut to P.J. and NIH grant no. HD098636 to D.G.

Keywords: dental pulp, ATAC-seq, chromatin; gene body, promoter

INTRODUCTION

Over the past few years, dental pulp-derived mesenchymal stem cells (MSCs) have emerged as strong candidates for tissue engineering and regenerative medicine (Masuda et al., 2021; Augustine et al., 2021; Gan et al., 2020; Bonaventura et al., 2020). MSCs are capable of differentiating into multiple lineages such as osteogenic, angiogenic, myogenic, and chondrogenic cells (Nagata et al., 2021; Luo et al., 2018; Potdar et al., 2015). However, we know very little about the molecular mechanisms that determine specification of MSCs along these lineages. The regulation of MSCs during cell fate determination is a complex process and can be delineated at the level of chromatin accessibility. A growing body of evidence indicates that the higher-order chromatin architecture is a critical component of gene regulation and cell fate specification (Harada et al., 2017; Guo et al., 2017; Hernandez-Hernandez et al., 2020). Genome-wide chromatin accessibility can be measured based on the assay for transposase-accessible chromatin using sequencing (ATAC-seq), a powerful method for profiling open chromatin regions and delineating regulatory sequences in the mammalian genome (Marinov et al., 2021).

Chromatin accessibility deduced from ATAC-seq is often used with RNA sequencing (RNA-seq) to correlate cis-acting regulatory regions such as enhancers and promoters with gene expression (Nair et al., 2021). Both approaches have become increasingly powerful for investigating distinct differences in the regulatory wirings among closely related cell types or tissues (Mao et al., 2021). For example, using ATAC-seq and RNA-seq, Arenas-Mena and others reported evidence for the existence of a gene regulatory hierarchy during sea urchin embryogenesis (Arenas-Mena et al., 2021). In another study, chromatin accessibility and transcriptome analyses provided a more refined understanding of differences in the transcriptional programs between embryonic and extraembryonic tissues (Zhang et al., 2021a).

We reasoned that employing ATAC-seq could provide insights into the chromatin states and associated regulatory regions of key developmental genes in the mouse dental pulp. Using ATAC-seq and RNA-seq, we analyzed chromatin accessibility and gene expression of master regulators of cell lineage and fate determination in the mouse dental pulp tissue.

MATERIALS AND METHODS

Primary dental pulp isolation and genome-wide sequencing

Primary pulp from mouse incisors and molars was prepared from 5-6-day-old mice according to previously described procedures (Balic et al., 2010). ATAC-seq and RNA-seq services were performed by Active Motif (Carlsbad, CA). RNA isolation was performed using the RNAeasy Mini/Midi kit (Qiagen, Germantown, MD). RNA-seq was performed using an Illumina NextSeq 500
to generate 42-nt paired-end sequences. For ATAC-seq, the paired-end 42 bp sequencing reads generated by Il- luminata NextSeq 500 sequencing were mapped to the ge- nome using the Burrows-Wheeler Aligner algorithm with default settings.

Computational analyses

Alignment information for each read was stored in the BAM format. Genomic regions with high levels of transposition/tagging events were determined using the MACS2 peak-calling algorithm (Zhang et al., 2008). Since both reads (tags) from paired-end sequencing represent transposition events, both reads were used for peak-calling but were treated as single, independent reads. To identify the density of the transposition events along the genome, the genome was divided into 32 bp bins and the number of fragments in each bin was determined. For this purpose, the reads were extended to 200 bp, which is close to the average length of the sequenced library inserts. In the default analysis, the tag number of all samples was reduced (by random sampling) to the number of tags present in the smallest sample. To compare peak metrics between two or more samples, overlapping intervals were grouped into “merged regions”, which were defined by the start coordinate of the most upstream interval and the end coordinate of the most downstream interval (= union of overlapping intervals). In locations where only one sample had an interval, this interval defined the merged region. After defining the intervals and merged regions, their genomic locations along with their proximities to gene annotations and other genomic features were determined and presented in Excel spreadsheets.

To clarify the relationship between ATAC-seq enrichment and RNA expression, we used a previously published approach (Wang et al., 2016) where genes are separated into five groups (10–20%, 20–40%, 40–60%, 60–80% and 80–100%) based on their mRNA expression levels. Box plots of average ATAC-seq peak values were generated for each group of genes in promoter and gene bodies for studying both incisors and molars.

The RNA-seq fragments were mapped to the reference genome mm10 using the STAR aligner (Dobin et al., 2013). The fragment assignment step was carried out to count the number of fragments overlapping the genomic sequence. Only the read pairs that had both ends aligned at the same chromosome and the same DNA strand were considered for subsequent analysis. The feature count (FPKM assignment to genes) was performed using the Subread package (Liao et al., 2014). Gene annotations were originally from the NCBI RefSeq database and then adapted by merging the overlapping exons from the same gene to form a set of disjoint exons for each gene. After obtaining the same table containing the fragment (or read) of genes, differential analysis was performed using DESeq2 (Love et al., 2014).

RESULTS

Distribution of the ATAC-seq peaks in the mouse dental pulp genome

We evaluated the chromatin accessibility landscape in the mouse dental pulp from incisors and molars by ATAC-seq. Our findings demonstrated a high signal-to-noise ratio in the ATAC-seq signals in the pulp genome. We observed that ATAC-seq reads were elevated in the gene bodies (40.8% in the incisor and 40.7% in the mo- lar genomes) (Fig. 1A), and were also enriched in the intergenic regions (31.3% in the incisor and 30.4% in the molar genomes). Within the upstream regulatory regions, ATAC-seq peaks were enriched in the incisor and molar genomes (23.7% and 24.6%, respectively). A relatively small level of ATAC-seq reads was also mapped in the downstream regulatory regions (4.2% in the incisor and 4.3% in the molar genomes). This ATAC-seq distribution pattern is similar to the RNA-seq reads. RNA-seq peaks were enriched in exons of the incisor and molar genomes (81% and 84%, respectively). As a common feature, a strong ATAC-seq signal was centered in the vicinity of gene transcription start sites (TSSs), which is consistent with other studies (Fig. 1B).

The ATAC-seq signal at promoters and gene bodies is correlated with RNA expression

To analyze the correlation between gene expression and the ATAC-seq signals at the promoter and gene body regions, we used a previously published approach (Wang et al., 2016). Briefly, we separated genes into five groups according to their expression levels (from low to high: 0–20%, 20–40%, 40–60%, 60–80% and 80–100%)
and plotted ATAC-seq signals across the gene body regions and promoters of the corresponding genes. We measured the RNA expression profiles using transcripts per million. We found a strong correlation between the ATAC-seq peaks and RNA expression (Fig. 1C). The average ATAC-seq signal was elevated in the promoter regions and in the corresponding genes that showed the highest expression levels (80–100% and 60–80%). We observed a similar pattern of chromatin accessibility and gene expression in both the incisor and molar dental pulp genomes. In summary, our data indicated that the open chromatin landscape correlates well with RNA expression in the mouse pulp tissues.

Chromatin accessibility varies across the genome of transcription factors associated with neural crest development

The dental pulp derives from the neural crest during embryonic development (Miletich & Sharpe, 2004; Yu et al., 2018). The neural crest regulatory program is composed of a set of master transcription factors (TFs) and several signaling pathways (Simoes-Costa & Bonner, 2015). We found that many genes involved in neural crest specification and development were expressed in the mouse dental pulp. Among these well-expressed genes, Myc, Mxi1, Six1 and Snai2 retained an open chromatin structure (Fig. 2). However, a number of genes involved in neural crest specification were not expressed in the dental pulp. Among these genes were Sox10, Pax3, Zic1, Zic4 and Foxd3. ATAC-seq analysis revealed that the chromatin in the vicinity of these genes was either in a closed or a partially accessible state. The x-axis represents genomic coordinates; the y-axis represents the ATAC-seq and RNA-seq normalized read counts.

Figure 2. Chromatin accessibility varies across the core master regulators of neural crest development.

Key developmental genes involved in neural crest specification, i.e., Six1, Snai2, Myc and Mxi1, were well expressed in the mouse dental pulp and retained the open chromatin state. Sox10, Pax3, Zic1, Zic4 and Foxd3 were not expressed in the dental pulp. ATAC-seq analysis revealed that the chromatin in the vicinity of these genes was either in a closed or a partially accessible state. The x-axis represents genomic coordinates; the y-axis represents the ATAC-seq and RNA-seq normalized read counts.
Me2c and Mesp1 (Fig. 3). Of these genes, only Me2c was expressed in the dental pulp, and we found high ATAC-seq peaks along the gene body region. Gata4, Tbx5 and Mesp1 were not expressed in the dental pulp, however, we detected strong ATAC-seq signals in the promoter region of Gata4 and within the gene body regions of Mesp1. By contrast, we detected no ATAC-seq signals in the promoter or gene body regions of Tbx5, indicating a closed chromatin state. Second, we explored the chromatin accessibility and RNA expression of genes associated with adipo-cyte differentiation (i.e., Mief1, Pparg, Egr2 and Cebpδ) (Fig. 4). We found chromatin accessibility in the promoter and gene body regions of Pparg, Egr2 and Cebpδ. The RNA-seq analysis was consistent with the ATAC-seq results, revealing strong expression of all three genes in the dental pulp. For Mief1, we failed to detect transcriptional activity. We did not detect ATAC-seq peaks in the gene body region, indicating that the chromatin was closed; however, strong ATAC-seq signal was present in the promoter of Mief1 (Fig. 4). Third, we analyzed the core regulators of hepatocyte differentiation. The promoter and gene body regions of Cebpδ, Hlxεc and Gata2 displayed open chromatin features. All three genes were well expressed in the dental pulp. We did not detect significant expression of Foxa1, although the chromatin along the gene body and promoter region was modestly accessible (Fig. 4). Fourth, we interrogated genes involved in neuronal differentiation. According to RNA-seq analysis, Neurod1, Ascl1, Nkh2-1 and Olig2 were not expressed in the dental pulp (Fig. 5). However, we detected strong ATAC-seq reads at the promoter and gene body regions of Ascl1, Nkh2-1 and Olig2, indicating accessible chromatin states in the vicinity of these genes. We also found a strong ATAC-seq peak in the gene body of Neurod1, although we failed to detect ATAC-seq signals in the promoter region.

Chromatin accessibility of pluripotency genes in the mouse dental pulp

The dental pulp tissue is a valuable source of MSCs that are capable of differentiating into various types of cells (Potdar et al., 2015). We analyzed the RNA expression profiles and chromatin landscapes of the core pluripotency genes Nanog, Sox2, Pou5f1 (Oct4), Maf, Salh4 and Klf4 (Fig. 6). Our analysis indicated that, of these genes, only Klf4 and Maf were strongly expressed in the dental pulp and retained an open chromatin state. For the other pluripotency genes (Nanog, Sox2 and Pou5f1), we either failed to detect RNA expression or found very weak expression (Salh4). Surprisingly, the ATAC-seq peaks were enriched in the promoter of Salh4 and were enriched along the entire gene body region of Sox2 (Fig. 6). We detected relatively weak ATAC-seq signals in the gene body of Pou5f1. At the same time, chromatin of Nanog remained in a closed state, although we noticed a strong ATAC-seq peak approximately 5kb upstream of Nanog, which could represent a repressive chromatin domain.

Overall, our findings for both mouse molars and incisors indicate that a number of master regulator genes associated with lineage specification or pluripotency are expressed in the dental pulp and maintain open chromatin states or are not expressed but still retain chromatin accessibility.

DISCUSSION

In summary, our ATAC-seq analysis demonstrated that the vast majority of key developmental genes maintain an open chromatin state in the mouse dental pulp. Cell fate commitment and differentiation involve highly coordinated events in which cell-type-specific genes are activated while others are repressed (Hernandez-Hernandez et al., 2020). Our study suggests that the differentiation of pulp-derived MSCs into different cell types most likely depends on the chromatin landscape of lineage-specific genes encoding master TFs. We found that a large number of regulatory genes encoding TFs are not expressed in the dental pulp but still retain open chromatin (Fig. 7).

Chromatin accessibility is a hallmark feature of active or repressed states of individual genes. We showed that the majority of master regulatory genes in dental pulp
maintain repressed chromatin states. Genome-wide gene silencing is initiated by the trimethylation of histone H3 at lysine 27 (H3K27me3) catalyzed by polycomb repressive complex (PRC2), whereas gene activation is dependent on trimethylation of histone H3 at lysine 4 (H3K4me3) mediated by the Set1/COMPASS complex (Piunti & Shilatifard, 2016). Hence, we believe that some of the ATAC-seq peaks could be associated with repressed chromatin enriched in H3K27me3, whereas the eviction of PRC2 from the regulatory loci may result in gene activation. Recently, Zhang et al. reported that during liver regeneration, H3K27me3 is depleted from promoters of genes involved in proliferation, thus facilitating their vigorous expression (Zhang et al., 2021b).

H3K4me3 and H3K27me3 define bivalent domains, which are enriched in promoters of key developmental genes in different stem/progenitor cells (Klein & Hainer, 2020). Specifically, bivalent chromatin domains are formed at developmental genes that are poised for rapid activation. MSCs of dental pulp originate from the neural crest cells, a transient multipotent cell population derived from the dorsal neural tube (Masuda et al., 2021). Distinct TFs control neural crest differentiation during craniofacial and dental development (Prasad et al., 2019; Balic & Thesleff, 2015). Our findings suggest that the dental pulp genome retains memories of past events. Therefore, open chromatin was likely to have been engaged in a regulatory role during previous developmental stages. It is also possible that some of the differentiating MSCs possess distinct chromatin accessibility states and genes in their vicinity are poised for later activation. Therefore, although the majority of genes encoding pluripotency TFs and regulators of neural crest differentiation are not expressed in dental pulp, chromatin accessibility at the genome-wide level may be explained by epigenetic memory mediated by PRC2 and Set1/
Figure 5. Chromatin accessibility of neurogenic genes in the mouse dental pulp.
RNA-seq analysis demonstrated that Neurod1, Ascl1, Nkh2-1 and Olig2, which are involved in neuronal differentiation, were not expressed in the dental pulp. However, we detected strong ATAC-seq reads at the promoter and gene body regions of Ascl1, Nkh2-1 and Olig2, indicating accessible chromatin states in the vicinity of these genes. We also found a strong ATAC-seq peak in the gene body region of Neurod1, although we failed to detect ATAC-seq signals in the promoter region.

Figure 6. RNA expression and chromatin accessibility analysis of core pluripotency genes revealed that only Maf and Klf4 were strongly expressed and retained open chromatin states in the dental pulp. Although other pluripotency genes were not expressed in the dental pulp, ATAC-seq peaks were enriched in the gene body or promoter regions of Pou5f1 and Sall4 or were enriched along the entire gene body region of Sox2. Chromatin of only Nanog remained in a closed state.
COMPASS complexes. For example, the inheritance of repressed chromatin domains for DNA replication ensures the stability of H3K27me3-mediated gene silencing across cell generations (Hugues et al., 2020). This epigenetic mechanism has a crucial role in chromatin-based safeguarding of cell identity.

The mouse dental pulp is highly heterogenic and composed of several different cell types including ameloblasts, odontoblasts, pericytes, immune cells and blood cells (Nikoloudaki et al., 2020; Yianni & Sharpe, 2019; Farges et al., 2015). Developmental master TFs are selectively expressed among the various cell types; therefore, our bulk ATAC-seq approach representing the total number of reads in the tissue is not sufficiently sensitive to detect cell-type-specific peaks. In the future, using a single-cell ATAC-seq strategy, we will be able to distinguish the chromatin accessibility patterns within distinct cell populations.

Our study supports a nonlinear relationship between gene expression and chromatin accessibility in mouse dental pulp. Moreover, our work suggests a multi-level control mechanism for MSCs in the pulp tissue, which likely depends on the intricate combinatorial genome-wide principles of gene expression. Collectively, we provided the first glimpse of the chromatin accessibility landscape in the mouse dental pulp, laying the foundation for future research on the genomic mechanisms orchestrating lineage specification of the pulp-derived MSCs.

Author Contribution

D.B. conceived and designed the project and prepared the manuscript. A.V. isolated the mouse dental pulp. D.B., B.E. and P.J. performed the formal analysis. M.M. and D.G. were involved in writing and editing. All authors read and approved the manuscript.

Competing interests

The authors declare that they have no competing financial interests.

Availability of data

The data that support the findings of this study are available on request from the corresponding author.

REFERENCES


Figure 7. A large number of regulatory genes encoding TFs are not expressed in the dental pulp but still retain open chromatin. Among these genes are those encoding pluripotency TFs as well as master regulators of neural crest, myogenic, adipocyte, hepatocyte and neuronal differentiation. A relatively small number of genes such as Pax3 (neural crest), Nanog (pluripotency) and Tbx5 (myogenic) are not expressed and maintain open chromatin. Among genes that are expressed and maintain open chromatin states are Six1, Snai2, Myc and Mef2c (neural crest), Klf4, Maf and Salid4 (pluripotency), Mef2c (myogenic), Ppara, Egr2 and Cebpβ (adipocytes), and Cebpβ, Hhex and Gata2 (hepatocytes). Among genes that are not expressed but still maintain open chromatin states are Sox10, Zic1, Zic4 and Foxd3 (neural crest), Pou5f1 and Sox2 (pluripotency), Gata4 and Mespi1 (myogenic), Mief1 (adipocytes), Foxa1 (hepatocytes) and Ascl1, Nkh2-1, Olig2 and Neurod1 (neuronal).