Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

HHS Public Access
Author manuscript
Methods Cell Biol. Author manuscript; available in PMC 2020 May 29.

Published in final edited form as:
Methods Cell Biol. 2019 ; 151: 219–235. doi:10.1016/bs.mcb.2018.11.002.

Genome-wide analysis of chromatin accessibility using ATAC-seq

Tanvi Shashikant, **Charles A. Ettensohn***
Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA, United States

Abstract

Programs of gene transcription are controlled by *cis*-acting DNA elements, including enhancers, silencers, and promoters. Local accessibility of chromatin has proven to be a highly informative structural feature for identifying such regulatory elements, which tend to be relatively open due to their interactions with proteins. Recently, ATAC-seq (assay for transposase-accessible chromatin using sequencing) has emerged as one of the most powerful approaches for genome-wide chromatin accessibility profiling. This method assesses DNA accessibility using hyperactive Tn5 transposase, which simultaneously cuts DNA and inserts sequencing adaptors, preferentially in regions of open chromatin. ATAC-seq is a relatively simple procedure which can be applied to only a few thousand cells. It is well-suited to developing embryos of sea urchins and other echinoderms, which are a prominent experimental model for understanding the genomic control of animal development. In this chapter, we present a protocol for applying ATAC-seq to embryonic cells of sea urchins.

1 Introduction

The transformation of a single cell into a multicellular organism is encoded in the genome. A cardinal feature of this process is the progressive emergence of distinct programs of gene transcription in various cells of the embryo. These gene regulatory programs endow cells with their distinct identities. Transcriptional programs are intimately associated with changes in chromatin state, including epigenetic modifications to DNA, local changes in chromatin structure (including nucleosome positioning and density), changes in the three-dimensional topology of chromatin domains, and the overall organization of chromosomes within the cell nucleus (Atlasi & Stunnenberg, 2017; Perino & Veenstra, 2016).

The local accessibility of chromatin has proven to be a particularly informative structural feature. The advent of high-throughput DNA sequencing has led to the development of several technologies for assessing local patterns of chromatin accessibility on a genome-wide scale. These methods include DNase-seq (Boyle, Guinney, Crawford, & Furey, 2008; Song & Crawford, 2010), FAIRE (Giresi, Kim, McDaniell, Iyer, & Lieb, 2007), NOMe-seq (Kelly et al., 2012), NicE-Seq (Ponnaluri et al., 2017), and others. All these methods rely on the same basic principle; namely, that regions of the chromatin that are relatively open (see below) exhibit enhanced accessibility to enzymes that modify DNA (or, in the case of

*Corresponding author: ettensohn@cmu.edu.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

FAIRE, enhanced susceptibility to mechanical shearing). Local regions of open chromatin are often associated with enhancers, promoters, and other regulatory DNA elements; therefore, chromatin accessibility profiling has emerged as an important tool for the identification of such elements. Although only a handful of studies have applied more than one chromatin accessibility profiling method to identical biological samples, these have generally found reasonable agreement among different approaches (40–90% of peaks identified by one approach overlap peaks identified by a different method) (Ponnaluri et al., 2017; Shashikant, Khor, & Ettensohn, 2018b). It should be noted, however, that there are method-specific differences among the various chromatin accessibility profiling techniques that have not been systematically explored. At least some of these differences are likely due to different DNA sequence biases inherent in the various procedures (see below).

There is a growing appreciation that many protein-DNA interactions are highly dynamic. For example, recent live-cell imaging studies of transcription factor (TF)-enhancer interactions have revealed that at least some TFs display surprisingly short (<20s) residence times at their specific binding sites in vivo (reviewed by Liu & Tjian, 2018). Obviously, highly dynamic, molecular interactions might not be detected by current, genome-wide methods for profiling chromatin accessibility. These methods typically average signals over many cells and identify regions of open chromatin that are sufficiently stable to be recovered across multiple biological replicates. Thus, they provide a snapshot of relatively stable, open regions of chromatin, within which dynamic molecular interactions might be taking place.

In addition to their well-documented utility for the identification of *cis*-regulatory elements, there is considerable interest in using ATAC-seq and other chromatin accessibility profiling methods for the genome-wide identification of individual TF binding sites, through the identification of locally protected sites within larger, hyperaccessible regions of DNA. This approach, known as genomic footprinting, has proven more challenging than originally anticipated due to various factors, including the sequence bias of the enzymes used to fragment DNA (Sung, Baek, & Hager, 2016). The development of improved computational methods for genomic footprinting continues to be a very active area of work (Baek, Goldstein, & Hager, 2017; Quach & Furey, 2017).

Echinoderms provide outstanding experimental material for the analysis of developmental changes in chromatin organization. The ease with which large numbers of synchronously developing embryos can be obtained is a tremendous advantage in this regard. In addition, a variety of approaches are available for the isolation of specific embryonic cell types (see Section 6). Importantly, models of developmental gene regulatory networks (GRNs) are particularly well-developed in echinoderms (Peter & Davidson, 2015; Shashikant, Khor, & Ettensohn, 2018a), and changes in chromatin accessibility can be interpreted in the context of these dynamic regulatory networks. Chromatin accessibility profiling in sea urchins promises to be a powerful tool for the identification of regulatory DNA elements that control developmental GRNs. This approach will also shed light on the dynamic changes in chromatin accessibility that occurs during the progressive differentiation of cell lineages during embryonic development.

10x Genomics, Inc. Exhibit 2005
Parse Biosciences, Inc. v. 10x Genomics, Inc.
IPR2023-01030
Page 00002

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

To date, one published study has applied ATAC-seq to sea urchin embryonic cells. Shashikant et al. (2018b) used ATAC-seq for the high-throughput identification of *cis*-regulatory modules that are components of a transcriptional network deployed in skeletogenic cells of *Strongylocentrotus purpuratus* (Shashikant et al., 2018a). In related work, Tulin, Barsi, Bocconcelli, and Smith (2016) used a different, immunoprecipitation-based approach, termed GRIP-seq, to produce a database of putative, active enhancers from whole, 24-h *S. purpuratus* embryos.

2 Principles of ATAC-seq

ATAC-seq (assay for transposase-accessible chromatin using sequencing) has rapidly emerged as one of the most powerful approaches for chromatin accessibility profiling. This method assesses DNA accessibility using hyperactive Tn5 transposase, which simultaneously cuts DNA and inserts sequencing adaptors, preferentially in regions of open chromatin. As with other approaches, DNA sequencing libraries that are enriched for hyperaccessible regions of DNA are generated and subjected to high-throughput sequencing. Reads are then aligned to an assembled genome to identify regions marked by high densities of aligned reads. Since this method was first developed (Buenrostro, Giresi, Zaba, Chang, & Greenleaf, 2013) it has been applied to hundreds of cell types, including embryonic cells from several organisms (Cusanovich et al., 2018; Koh et al., 2016; Shashikant et al., 2018b; Simon et al., 2017; Wu et al., 2016).

ATAC-seq relies on the activity of Tn5 transposase. Tn5 is a prokaryotic enzyme naturally encoded by the Tn5 transposon, which also contains specific, 19-bp flanking sequences known as “end sequences” (ESs) (Reznikoff, 2008). In its native environment, homodimeric Tn5 transposase recognizes the ESs of the transposon and, through a cut-and-paste mechanism, excises the DNA and inserts it into a new (host) position. The interaction with the host site has only limited sequence specificity. This biology has been exploited as a tool for in vitro transposition by utilizing a hyperactive derivative of the Tn5 transposase pre-bound to a DNA sequence which is to be inserted. The donor DNA is flanked on each end by a “mosaic end” (ME) sequence, a 19-bp sequence similar to endogenous ES sequences but even more active in facilitating integration. Under these conditions, ME-flanked DNA is inserted into target DNA with high efficiency. More recently, this in vitro transposition reaction has been further modified to allow the rapid production of DNA sequencing libraries using a procedure known as “tagmentation.” When hyperactive, dimeric Tn5 transposase is pre-bound to free, synthetic ME-tagged adaptor sequences (in contrast to ME-flanked DNA in which two ME sequences are linked by intervening DNA), the resulting “transposome” catalyzes fragmentation of host DNA and 5′ end-joining of the ME-tagged adaptor sequences (Fig. 1). By varying the concentration of transposome complexes relative to the target DNA, the size distribution of the DNA fragments can be controlled. To generate DNA libraries compatible with next-generation DNA sequencing, additional sequences are appended to these adaptors by PCR. The additional sequences can include sample-specific barcodes, facilitating multiplexed sequencing on a single instrument run.

The structural basis of local hypersensitivity to Tn5 transposase is not fully understood. Several studies have indicated that enhancers, promoters, and other regulatory regions are

10x Genomics, Inc. Exhibit 2005
Parse Biosciences, Inc. v. 10x Genomics, Inc.
IPR2023-01030
Page 00003

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

relatively nucleosome-free, suggesting that local depletion of nucleosomes (via nucleosome movement or loss) might underlie the formation of hyperaccessible regions (Mavrich et al., 2008; Yuan et al., 2005). More recent work, however, has shown that local increases in DNA accessibility detectable by ATAC-seq can arise without changes in nucleosome occupancy (Mueller et al., 2017). These findings suggest that other changes in chromatin structure, perhaps weakening of DNA-nucleosome or nucleosome-nucleosome interactions, might contribute to changes in DNA hyperaccessibility.

Complete, unbiased sampling of open regions of chromatin by ATAC-seq and related methods is only possible if other factors do not constrain the sites in the target DNA that are subject to fragmentation. In practice, data from all forms of high-throughput chromatin accessibility profiling, including ATAC-seq, exhibit some degree of DNA sequence bias (Wang, Quach, & Furey, 2017). In the case of ATAC-seq data, one source of this bias arises from the fact that both native and hyperactive forms of Tn5 transposase exhibit a slight preference for inserting DNA at particular target sequences (Adey et al., 2010; Goryshin, Miller, Kil, Lanzov, & Reznikoff, 1998). Recently, a new form of Tn5 has been described that shows reduced insertion bias (Kia et al., 2017). Other steps in the preparation of the DNA for sequencing also introduce some sequence bias. Most of these steps are shared by all current methods of whole-genome chromatin profiling, however, and are therefore not unique to ATAC-seq (Wang et al., 2017). In practice, the overall degree of sequence bias observed in tagmentation-based sequencing libraries is similar to that of other widely-used DNA library construction methods, and the advantages of ATAC-seq (outlined below) often make it the most desirable approach (Adey et al., 2010).

A major advantage of ATAC-seq over most other methods for chromatin accessibility profiling is its simplicity. Relatively few experimental steps are involved and the entire procedure through final library generation (i.e., prior to DNA sequencing) can be completed in a single day. A second important advantage of ATAC-seq is its sensitivity. The original ATAC-seq protocol is based on an input of 50,000 cells, but this method can also be applied to much smaller numbers of cells, albeit with some loss of signal (Buenrostro et al., 2013). Thus, ATAC-seq is well suited for studies in which the amount of starting material is limited; e.g., studies using FACS-isolated cells or cells from experimentally-modified embryos. Procedures have recently been described for ATAC-seq analysis of single cells and for deconstructing single-cell ATAC-seq profiles based on combinatorial cellular “indexing” (Buenrostro, Wu, Chang, & Greenleaf, 2015; Cusanovich et al., 2015, 2018), but these methods have not yet been applied to echinoderm embryos.

3 ATAC-seq experimental protocol

The procedure which follows is based closely on the protocol of Buenrostro, Wu, Chang, et al. (2015) and Buenrostro, Wu, Litzenburger, et al. (2015), with only minor modifications (Fig. 1). The protocol is designed for Illumina-based sequencing. Our procedure uses ~150,000 sea urchin embryo nuclei per sample. The standard ATAC-seq procedure calls for 50,000 mammalian cells per sample—we scale up in order to compensate for the smaller size of the sea urchin genome (i.e., 800 MB versus 3 GB). As noted above, ATAC-seq can also be carried out on much smaller numbers of mammalian cells (e.g., 5000) but with some

10x Genomics, Inc. Exhibit 2005
Parse Biosciences, Inc. v. 10x Genomics, Inc.
IPR2023-01030
Page 00004

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

loss of signal (Buenrostro et al., 2013). We assume that smaller numbers of sea urchin embryonic nuclei could also be used, but detection of differences in peak intensities would presumably be more challenging due to the reduced signal-to-noise ratio.

3.1 Cell preparation

1. Collect embryos at the desired developmental stage. As noted above, we use sufficient numbers of embryos or cells to provide ~150,000 nuclei for each sample. For whole-embryo studies, start with sufficient numbers of embryos to provide at least twice the final number of nuclei desired, as there will be some loss of material during the lysis and washing steps. The approximate numbers of cells per embryo at various developmental stages have been determined for a few sea urchin species (Nislow & Morrill, 1988; Poccia & Hinegardner, 1975) and can be used as an initial guide. For example, in *Lytechinus variegatus* there are ~1000 cells/embryo at the early gastrula stage, and therefore <500 gastrulae are required to provide sufficient starting material. If a purified cell population is to be used, methods for cell isolation and yield will vary depending on the cell type (e.g., Barsi, Tu, & Davidson, 2014; Rafiq, Shashikant, McManus, & Ettensohn, 2014; Swartz et al., 2014; Wilt & Benson, 2004).
2. Pellet cells or embryos by low-speed centrifugation at 4°C (500 × *g*; 30s for whole embryos, 5min for cells). Remove and discard the supernatant. Wash embryos/cells once with 4°C seawater. Centrifuge again at 4°C as above.
3. Remove and discard the supernatant. Add 500 μL of cold lysis buffer (formula below) and gently pipette up and down to resuspend the embryos/cells. Repeat this procedure twice more (three total washes with cold lysis buffer). After resuspending in lysis buffer for the third time, remove a small aliquot and determine the concentration of nuclei using a hemocytometer. Nuclei can be visualized using differential interference contrast (DIC) optics or fluorescently labeled by adding an equal volume of a 1:2500 dilution of a stock solution of DAPI (ThermoFisher D3571; stock solution is 5mg/mL in H_2O). Keep the remainder of the sample on ice during this time.

Lysis buffer—10mM Tris-HCl, pH 7.4

10mM NaCl

3 mM $MgCl_2$

0.1% (v/v) Igepal CA-630

Store up to 1 week at 4°C

1. After determining the concentration of nuclei, transfer enough of the sample to provide ~150,000 nuclei into a fresh microfuge tube and centrifuge for 10min at 500 × *g*, 4°C to pellet the nuclei.

10x Genomics, Inc. Exhibit 2005
Parse Biosciences, Inc. v. 10x Genomics, Inc.
IPR2023-01030
Page 00005

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2. Discard the supernatant, and proceed immediately to the transposition reaction. Keep the nuclear pellet on ice.

Note: although the above nuclear isolation procedure works well for ATAC-seq studies of post-blastula stage sea urchin embryos, a number of alternative procedures have been described for the isolation of nuclei from echinoderm eggs and early embryos that we have not systematically explored (Coffman & Yuh, 2004; Wessel & Vacquier, 2004).

3.2 Transposition reaction

1. To make the transposition reaction mix, combine the following:

   2.5 μL TD (2 × reaction buffer from Nextera kit)

   2.5 μL TDE1 (Nextera Tn5 transposase from the Nextera kit)

   22.5 μL nuclease-free H_2O

2. Resuspend the pelleted nuclei (from step 3) in the transposition reaction mix and incubate at 37°C for 30 min. Gentle mixing may increase fragment yield.

3. Immediately following transposition, purify the transposed DNA using a Qiagen MinElute PCR purification kit. Elute the DNA in 10 μL elution buffer (Buffer EB from the MinElute kit, consisting of 10mM Tris-Cl, pH 8).

4. This is a convenient stopping point, as the purified DNA can be stored at −20°C.

3.3 Adaptor extension and initial PCR amplification

1. Amplify the purified DNA for 5 cycles using primers that include Illumina adaptors. Combine the following in a 0.2mL PCR tube:

   10 μL transposed DNA

   10 μL nuclease-free H_2O

   2.5 μL 25 μM PCR Primer 1

   2.5 μL 25 μM Barcoded PCR Primer 2

   25 μL NEBNext High-Fidelity 2 × PCR Master Mix

Note: primers and PCR conditions are optimized for amplifying high molecular weight fragments from low-input material; therefore, PCR reagents provided by Illumina are not recommended. Primers are synthesized by Integrated DNA Technologies (IDT) with no additional modifications. A complete list of primers is provided by Buenrostro et al. (2013). Samples should be barcoded appropriately for subsequent pooling and Illumina sequencing.

1. Thermal cycler conditions:

1 cycle:	5min @ 72°C
	30s @ 98°C
5 cycles:	10s @ 98°C
	30s @ 63°C

10x Genomics, Inc. Exhibit 2005
Parse Biosciences, Inc. v. 10x Genomics, Inc.
IPR2023-01030
Page 00006

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

	1min @ 72°C

Important note: The initial 5min extension at 72°C is crucial as it allows extension of both adaptors after transposition, thereby producing amplifiable fragments.

3.4 Additional PCR amplification

The tagged DNA fragments must be further amplified to provide sufficient material for sequencing. To reduce GC and size bias due to PCR, an appropriate, minimal number of additional PCR cycles (*N*) is determined using qPCR, preventing saturation of the amplification reaction.

1. To perform a pilot qPCR reaction, combine the following:

   5 μL of previously amplified DNA (from step 11)

   4,41 μL nuclease-free H_2O

   0.25 μL 25 μM Custom Nextera PCR Primer 1

   0.25 μL 25 μM Custom Nextera PCR Primer 2

   0.09 μL 100 × SYBR Green I

   5 μL NEBNext High-Fidelity 2 × PCR Master Mix

2. Carry out the following amplification using a qPCR instrument:

1 cycle:	30s @ 98°C
20 cycles:	10s @ 98°C
	30s @ 63°C
	1min @ 72°C

1. To calculate the additional numbers of cycles (*N*) required to amplify the library, plot the Rn value (fluorescent signal from SYBR Green I, corrected for background signal) versus cycle number. Determine the cycle number that corresponds to approximately *one-third of the maximum fluorescent intensity* (see Buenrostro, Wu, Chang, et al., 2015; Buenrostro, Wu, Litzenburger, et al., 2015). This operation is designed to minimize the number of PCR cycles, as most PCR bias arises from later PCR cycles when reagents become limited. In practice, *N* is typically 5–7; much larger *N* values are a concern as they suggest that library complexity may be low.

2. Once *N* has been established, amplify the remaining 45 μL of previously amplified DNA (from step 11), using the reaction mix described in step 12 but omitting SYBR Green I and carrying out the following PCR reaction:

1 cycle:	30s @ 98°C
N cycles:	10s @ 98°C

10x Genomics, Inc. Exhibit 2005
Parse Biosciences, Inc. v. 10x Genomics, Inc.
IPR2023-01030
Page 00007

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

	30s @ 63°C
	1min @ 72°C

1. The quality of the amplified library should be analyzed at this stage using polyacrylamide gel electrophoresis or a Bioanalyzer or TapeStation (Agilent) to ensure that libraries do not contain excessive numbers of long DNA fragments (see Buenrostro, Wu, Chang, et al., 2015; Buenrostro, Wu, Litzenburger, et al., 2015).

2. Purify the amplified library using the Qiagen MinElute PCR purification kit. Elute the purified library in 20 μL of elution buffer (Buffer EB from the MinElute kit, consisting of 10 mM Tris-HCl, pH 8). Ensure that the column is dry before adding elution buffer to avoid ethanol contamination in the final library.

3.5 Sequencing

The Nextera tagmentation-based library construction method is designed for Illumina-based, high-throughput DNA sequencing. For the identification of local regions of open chromatin, single-end, 50- or 75-cycle reads are sufficient. Typically, >50 million mapped reads are used/sample. Shashikant et al. (2018b) used 70 million, 76-bp mapped reads/sample, which corresponded to about 90 million total reads/sample. Note that for transcription factor footprinting using chromatin accessibility data, much larger numbers of mapped reads (>200 million) have been used (Neph et al., 2012). In addition, libraries that contain large amounts of mitochondrial DNA (mtDNA) require larger numbers of total reads, as a smaller proportion map to genomic DNA.

4 Analysis of DNA sequence data

The major steps in our ATAC-seq sequence analysis pipeline are shown in Fig. 2. Most of these steps make use of widely used, free bioinformatics tools (e.g., FastQC, Cutadapt, Bowtie, Samtools, and Bedtools). With respect to peak-callers, we use Fseq (Boyle et al., 2008), which has been shown to out-perform several other peak callers when applied to chromatin accessibility data (Koohy, Down, Spivakov, & Hubbard, 2014). Note that not all peak callers or other analysis tools are compatible with large numbers of genomic scaffolds, a feature of all current echinoderm genome assemblies. The most recent version of the *S. purpuratus* genome assembly, for example, consists of ~38,000 scaffolds (Kudtarkar & Cameron, 2017).

After mapping reads, several quality control (QC) measures are applied. A very low fraction of mapped reads is cause for concern, and may suggest that Bowtie parameters are too stringent (we use default settings) or that other problems exist. As noted above, if early embryonic stages are used, the fraction of reads that represent mtDNA will reduce the fraction of reads that map to genomic DNA. Another QC measure is to determine the fraction of reads within peaks (the FRiP score) for each sample. A low FRiP score indicates a high degree of noise in the data. We use a minimum FriP score threshold of 0.4 but values

10x Genomics, Inc. Exhibit 2005
Parse Biosciences, Inc. v. 10x Genomics, Inc.
IPR2023-01030
Page 00008

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

above 0.6 are typical. A third QC measure is based on the use of biological replicates. We typically perform at least three biological replicates of each experiment and assess the degree of agreement among samples using deepTools (Ramìrez, Dündar, Diehl, Grüning, & Manke, 2014). Replicates should be highly concordant (Pearson correlation coefficient >0.9) and discordant replicates should be discarded.

There are various options for amalgamating data across multiple biological replicates. We generate a "reference peak set" (RPS) that consists of all peaks that overlap by at least 75% in one direction (i.e., non-reciprocally). For example, for experiments with three replicates, we first merge peaks from replicates 1 and 2 that overlap by 75% in one direction, then merge peaks in replicate 3 with the merged replicate (1+2) peaks if they overlap by 75% or more in one direction. The 75% overlap criterion is enforced non-reciprocally in order to account for differences in peak sizes across replicates. For example, if a 75% or greater overlap were to be enforced reciprocally, a peak that was >25% larger in one replicate or sample would be missing from the RPS. Pairwise comparisons can then be made between different RPS. A different approach that can be taken to amalgamate data from replicates is to pool sequence reads from all highly concordant replicates into a single sample.

For identification of differentially accessible peaks (i.e., regions of chromatin that are more accessible under one condition than another), total read counts corresponding to each peak in the RPS are first determined for each individual replicate using HTseq (Anders, Pyl, & Huber, 2015). Differential peaks are identified using DESeq2 (Love, Huber, & Anders, 2014) using nominal *P*-values selected by the user (we typically use 0.1). Once peaks (or differential peaks) are identified, we use a custom Python script to analyze the position of peaks relative to annotated genes. This makes it possible to determine the overall distribution of peaks relative to gene bodies and to identify and characterize nearby genes (Fig 3).

5 Special considerations regarding early developmental stages

Relatively few chromatin accessibility studies have focused on early (pre-blastula stage) animal embryos (Blythe & Wieschaus, 2016; Lu et al., 2016; Wu et al., 2016). Several special considerations arise when applying ATAC-seq or other chromatin accessibility profiling methods to early embryos. First, because the number of nuclei per embryo is relatively low at early stages, larger numbers of embryos are required to provide sufficient starting material for standard ATAC-seq analysis. Second, at early developmental stages, animal embryos are typically enclosed in one or more extracellular coats that might affect the ATAC-seq procedure. Early echinoderm embryos are encased in a fertilization envelope and this should be removed (Sweet et al., 2004) prior to nuclear isolation. Lastly, as noted above, the vast majority of DNA in the early embryo is maternally-derived mtDNA. This DNA is packaged in a manner distinct from nuclear chromatin (Bogenhagen, 2012) and its susceptibility to Tn5 transposase is poorly defined. The low-speed centrifugation steps in the standard ATAC-seq procedure (steps A2 and A3) eliminate some fraction of oocyte mitochondria, which are typically collected at higher centrifugal forces (Wessel & Vacquier, 2004). One CRISPR-based strategy has been described for reducing levels of mtDNA in samples derived from early mouse embryos (Wu et al., 2016), but even with this approach greater sequencing depth is required in order to offset the increased number of sequences

10x Genomics, Inc. Exhibit 2005
Parse Biosciences, Inc. v. 10x Genomics, Inc.
IPR2023-01030
Page 00009

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

reads that map to mtDNA rather than the nuclear genome. Further work will be needed to assess the impact of mtDNA contamination on ATAC-seq analysis of early embryos.

6 Conclusions and prospects

Because of its simplicity, speed, and reproducibility, ATAC-seq has rapidly emerged as a preferred method for chromatin accessibility profiling on a genome-wide scale. As noted above, ATAC-seq, like other chromatin accessibility profiling methods, provides a static assessment of chromatin architecture that reveals local, hyperaccessible regions. This method has proven to be valuable for the high-through identification of active *cis*-regulatory elements in a variety of cell types (Cusanovich et al., 2018; Davie et al., 2015; Koenecke, Johnston, Gaertner, Natarajan, & Zeitlinger, 2016; Quillien et al., 2017; Vrljicak et al., 2018; Wu et al., 2016), including biomineralizing cells of the early sea urchin embryo (Shashikant et al., 2018b). It is important to stress that local hyperaccessibility alone (or even differential hyperaccessibility, in the case of a comparisons between cell types) is insufficient evidence to conclude that a particular region has transcriptional regulatory function. Currently, the gold-standard for a demonstration of in vivo function is experimental analysis of putative *cis*-regulatory elements in transgenic embryos; i.e., deletion or mutation of the region in the context of the intact regulatory apparatus of a gene and/or the demonstration that the DNA element by itself has ability to drive transcription of a reporter gene in a developmentally regulated manner (Smith, 2008).

Application of ATAC-seq to other embryonic cell types from echinoderms will reveal additional, lineage-specific regulatory elements and spur the development of improved GRN models. A variety of methods are available for the isolation of specific embryonic cell types; for example, FACS can be used following the labeling of specific cell types either by selective fluorescent dyes (for example, labeling of germline progenitors with calcein) (Swartz et al., 2014) or lineage-specific expression of fluorescent reporter proteins under the control of appropriate *cis*-regulatory elements (Barsi et al., 2014; Chapter “Techniques for analyzing gene expression using BAC-based reporter constructs” by Buckley and Ettensohn, this volume).

ATAC-seq-based genomic footprinting is potentially a very powerful approach, but one that awaits further development to minimize sequence-bias issues. High-throughput analysis of transcription factor binding sites in echinoderms is also currently limited by the fact that relatively few consensus DNA target sequences have been empirically determined in these organisms, necessitating a reliance on binding site data from other experimental models. Further technical developments along these lines hold promise for the high-throughput identification of specific transcription-factor/DNA interactions, which would dramatically enhance the analysis of developmental GRNs.

References

Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, et al. (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biology, 11, R119. [PubMed: 21143862]

10x Genomics, Inc. Exhibit 2005
Parse Biosciences, Inc. v. 10x Genomics, Inc.
IPR2023-01030
Page 00010

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Anders S, Pyl PT, & Huber W (2015). HTSeq—A Python framework to work with high-throughput sequencing data. Bioinformatics, 31, 166–169. [PubMed: 25260700]

Atlasi Y, & Stunnenberg HG (2017). The interplay of epigenetic marks during stem cell differentiation and development. Nature Reviews Genetics, 18, 643–658.

Baek S, Goldstein I, & Hager GL (2017). Bivariate genomic footprinting detects changes in transcription factor activity. Cell Reports, 19, 1710–1722. [PubMed: 28538187]

Barsi JC, Tu Q, & Davidson EH (2014). General approach for in vivo recovery of cell type-specific effector gene sets. Genome Research, 24(5), 860–868. 10.1101/gr.167668.113. PubMed PMID: 24604781; PubMed Central PMCID: PMC4009615. [PubMed: 24604781]

Blythe SA, & Wieschaus EF (2016). Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis. eLife, 5, e20148. [PubMed: 27879204]

Bogenhagen DF (2012). Mitochondrial DNA nucleoid structure. Biochimica et Biophysica Acta, 1819, 914–920. [PubMed: 22142616]

Boyle AP, Guinney J, Crawford GE, & Furey TS (2008). F-Seq: A feature density estimator for high-throughput sequence tags. Bioinformatics, 24, 2537–2538. [PubMed: 18784119]

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, & Greenleaf WJ (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature Methods, 10(12), 1213–1218. 10.1038/nmeth.2688. PubMed PMID: 24097267; PubMed Central PMCID: PMC3959825. [PubMed: 24097267]

Buenrostro JD, Wu B, Chang HY, & Greenleaf WJ (2015). ATAC-seq: A method for assaying chromatin accessibility genome-wide. Current Protocols in Molecular Biology, 109, 21.29.1–9.

Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature, 523, 486–490. [PubMed: 26083756]

Coffman JA, & Yuh CH (2004). Identification of sequence-specific DNA binding proteins. Methods in Cell Biology, 74, 653–675. [PubMed: 15575625]

Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL, et al. (2015). Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science, 348, 910–914. [PubMed: 25953818]

Cusanovich DA, Reddington JP, Garfield DA, Daza RM, Aghamirzaie D, Marco-Ferreres R, et al. (2018). The *cis*-regulatory dynamics of embryonic development at single-cell resolution. Nature, 555, 538–542. [PubMed: 29539636]

Davie K, Jacobs J, Atkins M, Potier D, Christiaens V, Halder G, et al. (2015). Discovery of transcription factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open chromatin profiling. PLoS Genetics, 11, e1004994. [PubMed: 25679813]

Ettensohn CA, & Dey D (2017). KirrelL, a member of the Ig-domain superfamily of adhesion proteins, is essential for fusion of primary mesenchyme cells in the sea urchin embryo. Developmental Biology, 421, 258–270. [PubMed: 27866905]

Giresi PG, Kim J, McDaniell RM, Iyer VR, & Lieb JD (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Research, 17(6), 877–885. PubMed PMID: 17179217; PubMed Central PMCID: PMC1891346. [PubMed: 17179217]

Goryshin IY, Miller JA, Kil YV, Lanzov VA, & Reznikoff WS (1998). Tn5/IS50 target recognition. Proceedings of the National Academy of Sciences of the United States of America, 95, 10716–10721. [PubMed: 9724770]

Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, & Jones PA (2012). Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Research, 22, 2497–2506. [PubMed: 22960375]

Kia A, Gloeckner C, Osothprarop T, Gormley N, Bomati E, Stephenson M, et al. (2017). Improved genome sequencing using an engineered transposase. BMC Biotechnology, 17, 6–15. [PubMed: 28095828]

Koenecke N, Johnston J, Gaertner B, Natarajan M, & Zeitlinger J (2016). Genome-wide identification of Drosophila dorso-ventral enhancers by differential histone acetylation analysis. Genome Biology, 17, 196–214. [PubMed: 27678375]

10x Genomics, Inc. Exhibit 2005
Parse Biosciences, Inc. v. 10x Genomics, Inc.
IPR2023-01030
Page 00011

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Koh PW, Sinha R, Barkal AA, Morganti RM, Chen A, Weissman IL, et al. (2016). An atlas of transcriptional, chromatin accessibility, and surface marker changes in human mesoderm development. Scientific Data, 3, 160109. [PubMed: 27996962]

Koohy H, Down TA, Spivakov M, & Hubbard T (2014). A comparison of peak callers used for DNase-Seq data. PLoS One, 9, e96303. [PubMed: 24810143]

Kudtarkar P, & Cameron RA (2017). Echinobase: An expanding resource for echinoderm genomic information. Database: The Journal of Biological Databases and Curation, 2017, 1–9.

Liu Z, & Tjian R (2018). Visualizing transcription factor dynamics in living cells. The Journal of Cell Biology, 217(4), 1181–1191. 10.1083/jcb.201710038. Review. PubMed PMID: 29378780; PubMed Central PMCID: PMC5881510. [PubMed: 29378780]

Love MI, Huber W, & Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15, 550–570. [PubMed: 25516281]

Lu F, Liu Y, Inoue A, Suzuki T, Zhao K, & Zhang Y (2016). Establishing chromatin regulatory landscape during mouse preimplantation development. Cell, 165, 1375–1388. [PubMed: 27259149]

Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, Zanton SJ, et al. (2008). Nucleosome organization in the Drosophila genome. Nature, 453, 358–362. [PubMed: 18408708]

Mueller B, Mieczkowski J, Kundu S, Wang P, Sadreyev R, Tolstorukov MY, et al. (2017). Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction. Genes and Development, 31(5), 451–462. 10.1101/gad.293118.116. PubMed PMID: 28356342; PubMed Central PMCID: PMC5393060. [PubMed: 28356342]

Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, et al. (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. Nature, 489(7414), 83–90. 10.1038/nature11212. PubMed PMID: 22955618; PubMed Central PMCID: PMC3736582. [PubMed: 22955618]

Nislow C, & Morrill JB (1988). Regionalized cell division during sea urchin gastrulation contributes to archenteron formation and is correlated with the establishment of larval symmetry. Development, Growth & Differentiation, 30, 483–499.

Perino M, & Veenstra GJ (2016). Chromatin control of developmental dynamics and plasticity. Developmental Cell, 38, 610–620. [PubMed: 27676434]

Peter IS, & Davidson EH (2015). Genomic control process: Development and evolution. Academic Press.

Poccia DL, & Hinegardner RT (1975). Developmental changes in chromatin proteins of the sea urchin from blastula to mature larva. Developmental Biology, 45, 81–89. [PubMed: 1181224]

Ponnaluri VKC, Zhang G, Estève PO, Spracklin G, Sian S, Xu SY, et al. (2017). NicE-seq: High resolution open chromatin profiling. Genome Biology, 18, 122. [PubMed: 28655330]

Quach B, & Furey TS (2017). DeFCoM: Analysis and modeling of transcription factor binding sites using a motif-centric genomic footprinter. Bioinformatics, 33(7), 956–963. 10.1093/bioinformatics/btw740. PubMed PMID: 27993786; PubMed Central PMCID: PMC6075477. [PubMed: 27993786]

Quillien A, Abdalla M, Yu J, Ou J, Zhu LJ, & Lawson ND (2017). Robust identification of developmentally active endothelial enhancers in zebrafish using FANS-assisted ATAC-seq. Cell Reports, 20, 709–720. [PubMed: 28723572]

Rafiq K, Shashikant T, McManus CJ, & Ettensohn CA (2014). Genome-wide analysis of the skeletogenic gene regulatory network of sea urchins. Development, 141(4), 950–961. 10.1242/dev.105585. Erratum in: Development. 2014;141(12):2542. PubMed PMID: 24496631. [PubMed: 24496631]

Ramìrez F, Dündar F, Diehl S, Grüning BA, & Manke T (2014). deepTools: A flexible platform for exploring deep-sequencing data. Nucleic Acids Research, 42, W187–W191. [PubMed: 24799436]

Reznikoff WS (2008). Transposon Tn5. Annual Review of Genetics, 42, 269–286.

Shashikant T, Khor JM, & Ettensohn CA (2018a). From genome to anatomy: The architecture and evolution of the skeletogenic gene regulatory network of sea urchins and other echinoderms. Genesis. 10.1002/dvg.23253 [Epub ahead of print] Review. PubMed PMID: 30264451.

10x Genomics, Inc. Exhibit 2005
Parse Biosciences, Inc. v. 10x Genomics, Inc.
IPR2023-01030
Page 00012

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Shashikant T, Khor JM, & Ettensohn CA (2018b). Global analysis of primary mesenchyme cell *cis*-regulatory modules by chromatin accessibility profiling. BMC Genomics, 19, 206–223. [PubMed: 29558892]

Simon CS, Downes DJ, Gosden ME, Telenius J, Higgs DR, Hughes JR, et al. (2017). Functional characterization of *cis*-regulatory elements governing dynamic Eomes expression in the early mouse embryo. Development, 144, 1249–1260. [PubMed: 28174238]

Smith J (2008). A protocol describing the principles of *cis*-regulatory analysis in the sea urchin. Nature Protocols, 3, 710–718. [PubMed: 18388954]

Song L, & Crawford GE (2010). DNase-seq: A high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harbor Protocols 2010 (2), pdb.prot5384, 10.1101/pdb.prot5384, PubMed PMID: 20150147; PubMed Central PMCID: PMC3627383.

Sung MH, Baek S, & Hager GL (2016). Genome-wide footprinting: Ready for prime time? Nature Methods, 13, 222–228. [PubMed: 26914206]

Swartz SZ, Reich AM, Oulhen N, Raz T, Milos PM, Campanale JP, et al. (2014). Deadenylase depletion protects inherited mRNAs in primordial germ cells. Development, 141, 3134–3142. [PubMed: 25100654]

Sweet H, Amemiya S, Ransick A, Minokawa T, McClay DR, Wikramanayake A, et al. (2004). Blastomere isolation and transplantation. Methods in Cell Biology, 74, 243–271. [PubMed: 15575610]

Tulin S, Barsi JC, Bocconcelli C, & Smith J (2016). Genome-wide identification of enhancer elements. The International Journal of Developmental Biology, 60, 141–150. [PubMed: 27389984]

Vrljicak P, Lucas ES, Lansdowne L, Lucciola R, Muter J, Dyer NP, et al. (2018). Analysis of chromatin accessibility in decidualizing human endometrial stromal cells. The FASEB Journal, 32, 2467–2477. [PubMed: 29259032]

Wessel GM, & Vacquier VD (2004). Isolation of organelles and components from sea urchin eggs and embryos. Methods in Cell Biology, 74, 491–522. [PubMed: 15575619]

Wang JR, Quach B, & Furey TS (2017). Correcting nucleotide-specific biases in high-throughput sequencing data. BMC Bioinformatics, 18(1), 357 10.1186/s12859-017-1766-x. PubMed PMID: 28764645; PubMed Central PMCID: PMC5540620. [PubMed: 28764645]

Wilt FH, & Benson SC (2004). Isolation and culture of micromeres and primary mesenchyme cells. Methods in Cell Biology, 74, 273–285. Review. PubMed PMID: 15575611. [PubMed: 15575611]

Wu J, Huang B, Chen H, Yin Q, Liu Y, Xiang Y, et al. (2016). The landscape of accessible chromatin in mammalian preimplantation embryos. Nature, 534, 652–657. [PubMed: 27309802]

Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, et al. (2005). Genome-scale identification of nucleosome positions in *S. cerevisiae*. Science, 309, 626–630. [PubMed: 15961632]

Crawford GE, Holt IE, Whittle J, Webb BD, Tai D, Davis S, et al. (2006). Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome Research, 16, 123–131. [PubMed: 16344561]

John S, Sabo PJ, Canfield TK, Lee K, Vong S, Weaver M, et al. (2013). Genome-scale mapping of DNase I hypersensitivity Current Protocols in Molecular Biology, Ch. 27, Unit 21.27. 10.1002/0471142727.mb2127s103. PubMed PMID: 23821440; PubMed Central PMCID: PMC4405172.

10x Genomics, Inc. Exhibit 2005
Parse Biosciences, Inc. v. 10x Genomics, Inc.
IPR2023-01030
Page 00013

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

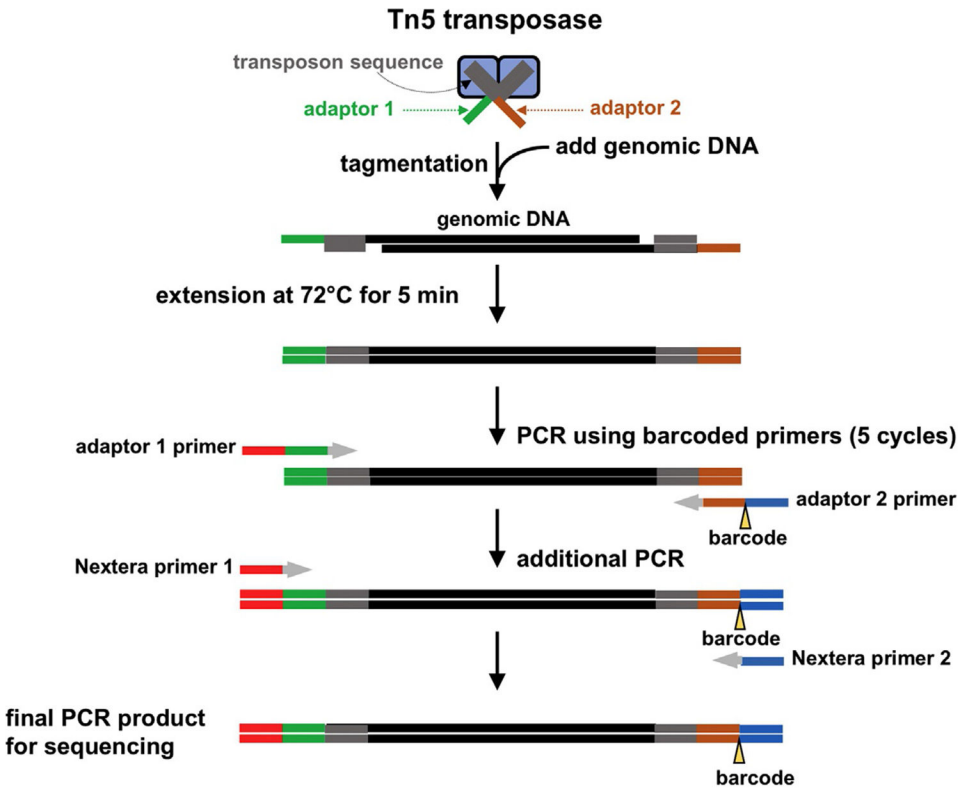

FIG. 1.

Summary of the ATAC-seq procedure (after preparation of nuclei). The Nextera kit provides a hyperactive form of Tn5 transposase loaded with adaptors, creating an active, dimeric transposome complex. The transposome cleaves the target (genomic) DNA and adds the adaptors to the 5′ ends of the DNA fragments. After filling in single-stranded gaps (by extension at 72°C for 5min) the adaptors are used for limited PCR amplification (5 cycles), usually including barcodes that allow multiplexing of samples during subsequent sequencing. The library of tagged DNA fragments is subjected to limited, additional amplification using flanking Nextera primers (“additional PCR” step).

10x Genomics, Inc. Exhibit 2005
Parse Biosciences, Inc. v. 10x Genomics, Inc.
IPR2023-01030
Page 00014

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

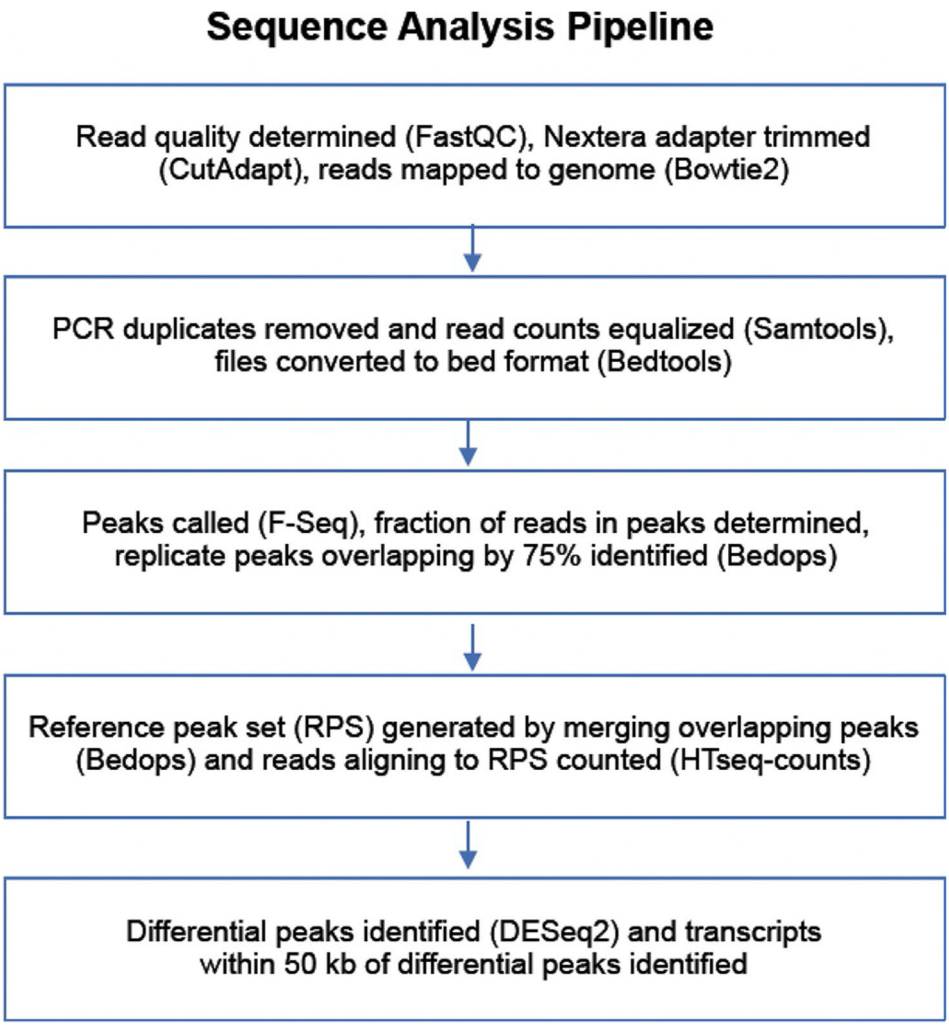

FIG. 2.
Sequence analysis pipeline.

10x Genomics, Inc. Exhibit 2005
Parse Biosciences, Inc. v. 10x Genomics, Inc.
IPR2023-01030
Page 00015

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

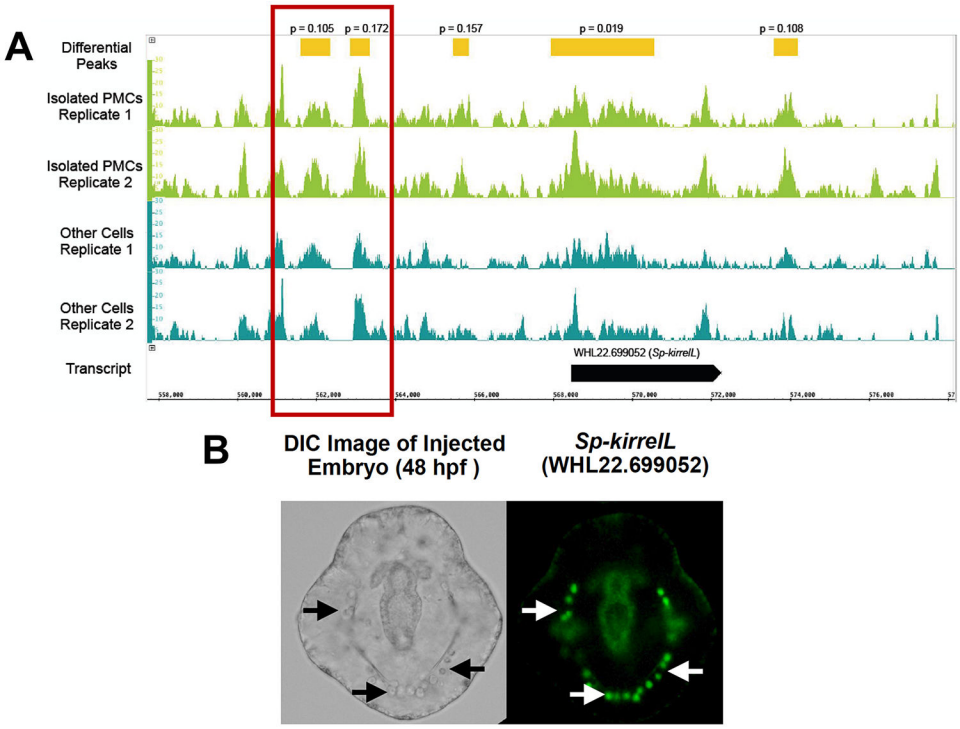

FIG. 3.

(A) Examples of differential ATAC-seq peaks (Shashikant et al., 2018b). *S. purpuratus* embryos were cultured for 24h at 15°C and skeletogenic cells (PMCs) were separated from all other cells. ATAC-seq libraries were generated, sequenced, and the sequence reads analyzed as described above. The aligned reads for two replicates are shown here, and differences in peak magnitudes can be seen when comparing differential peaks in the PMC replicates (light green peak trace) to the "other cell" replicates (dark green trace). Yellow rectangles indicate peaks located near the *Sp-kirrelL* gene that are differentially accessible in PMCs relative to other cells, with nominal *P*-values indicated. *Sp-kirrelL* is normally expressed only by PMCs and encodes a protein required for cell-cell fusion (Ettensohn & Dey, 2017). (B) *cis*-regulatory activity of a differential ATAC-seq peak. The region shown in the red box was cloned into a reporter plasmid that contained a basal sea urchin promoter upstream of the GFP coding sequence. Linearized plasmids were injected into fertilized *S. purpuratus* eggs and GFP expression was analyzed by epifluorescence microscopy. Under the control of the *cis*-regulatory element, the reporter gene is expressed selectively by PMCs (arrows).

10x Genomics, Inc. Exhibit 2005
Parse Biosciences, Inc. v. 10x Genomics, Inc.
IPR2023-01030
Page 00016