Published in final edited form as:

Methods Cell Biol. 2016 ; 135: 387–412. doi:10.1016/bs.mcb.2016.04.019.

Chromatin immunoprecipitation and an open chromatin assay in zebrafish erythrocytes

S. Yang^{*,§,¶}, C.J. Ott[§], M.P. Rossmann^{||}, M. Superdock^{*,§,¶}, L.I. Zon^{*,§,¶,||}, and Y. Zhou^{*,§,¶,||,1}

^{*}Boston Children's Hospital, Boston, MA, United States

[§]Dana Farber Cancer Institute, Harvard Stem Cell Institute, Boston, MA, United States

[¶]Harvard Medical School and Howard Hughes Medical Institute, Boston, MA, United States

^{||}Harvard University, Harvard, Cambridge, MA, United States

Abstract

Zebrafish is an excellent genetic and developmental model for the study of vertebrate development and disease. Its ability to produce an abundance of transparent, externally developed embryos has facilitated large-scale genetic and chemical screens for the identification of critical genes and chemical factors that modulate developmental pathways. These studies can have profound implications for the diagnosis and treatment of a variety of human diseases. Recent advancements in molecular and genomic studies have provided valuable tools and resources for comprehensive and high-resolution analysis of epigenomes during cell specification and lineage differentiation throughout development. In this chapter, we describe two simple methods to evaluate protein–DNA interaction and chromatin architecture in erythrocytes from adult zebrafish. These are chromatin immunoprecipitation coupled with next-generation sequencing (ChIP-seq) and an assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq). These techniques, together with gene expression profiling, are useful for analyzing epigenomic regulation of cell specification, differentiation, and function during zebrafish development in both normal and disease models

INTRODUCTION

As a well-established model for studies concerning vertebrate developmental biology and genetics, including areas of developmental hematopoiesis (Paik & Zon, 2010) and organ formation and function (Cox & Goessling, 2015; Janasik & Sallin, 2016; Lagendijk, Yap, & Hogan, 2014; Quiros-Gonzalez & Yadav, 2014; Schlegel & Gut, 2015; Takeuchi, 2014), zebrafish has been reliably utilized to model human biology and disease and develop therapies and drugs for human disease treatment (Asnani & Peterson, 2014; Carroll & North, 2014; Goessling & Sadler, 2015; Gurevich, Siegel, & Currie, 2015; Haesemeyer & Schier, 2015; van Houcke, De Groef, Dekeyster, & Moons, 2015; Jörgens et al., 2015; McCammon & Sive, 2015; Mohseny & Hogendoorn, 2014; Mort, Jackson, & Patton, 2015; Phillips &

¹Corresponding author: yzhou@enders.tch.harvard.edu.

Conflicts of interest: L.I.Z. is a founder and stock holder of Fate Therapeutics Inc. and Scholar Rock, and a scientific advisor for Stemgent.

Westerfield, 2014; Plantié, Migocka-Patrzałek, Daczewska, & Jagla, 2015; Powles-Glover, 2014; Saralahti & Rämetsä, 2015; Wager, Mahmood, & Russell, 2014; White, 2015; Wilkinson, Jopling, & van Eeden, 2014; Wilkinson & van Eeden, 2014). Zebrafish offers a number of unique advantages for studying development in particular, including the transparency of zebrafish embryos, which facilitates the use of in vivo microscopy coupled with fluorescent labeling for the direct visualization of organ development and function. A major breakthrough that has extended the advantages of transparency to adult fish is the creation of the “casper” fish (White et al., 2008), which lacks melanocytes and iridophores. These traits cause the “casper” fish to remain transparent through adulthood. An additional benefit of zebrafish is its fecundity, which facilitates large-scale phenotype-based forward genetic and chemical genetic screens. These classic advantages of zebrafish make it ideal for developmental and genetic studies.

Many advantageous tools have been created to facilitate the use of this powerful system. For example, the introduction of a high-quality reference genome has supported many comprehensive genomic and epigenomic studies, including genome-wide transcriptome analyses and detailed studies of the effect of transcription factors and chromatin architecture on transcription regulation. These tools have accelerated in vivo molecular mechanism studies regarding stem cell specification, maintenance, and subsequent lineage differentiation throughout vertebrate development. However, many molecular and genomic tools still remain largely underutilized in zebrafish studies.

Despite the relative abundance of gene expression data, limited chromatin immunoprecipitation (ChIP) experiments have been used to examine the changes in histone marks that occur during zebrafish embryogenesis (Lindeman, Vogt-Kielland, Alestrom, & Collas, 2009; Vastenhouw et al., 2010). Using ChIP to examine histone modifications provides a better molecular understanding of the alterations to chromatin architecture that accompany gene expression changes. ChIP can also be employed to define the specific DNA binding sites of transcription factors and infer which genes they directly regulate. The assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) has been similarly underutilized in zebrafish studies. ATAC-seq is an assay that is used to identify regions of open chromatin that are more accessible to DNA-interacting factors at a genome-wide level. It functions similarly to assays of DNase I hypersensitivity coupled with next-generation sequencing (DNase-seq) but requires many fewer cells. This method supports the detection of known and novel DNA regulatory elements, such as promoters and enhancers. It also allows for the confirmation or inference of known and novel cognate transcription factors and epigenetic factors through the analysis of the underlying DNA sequence motifs in these regulatory regions. Here we describe both protocols for ChIP-seq and ATAC-seq analysis of adult zebrafish erythrocytes.

1. RATIONALE

ChIP has been used extensively for studying the chromosome landscape, including histone modifications such as methylation and acetylation as well as the various transcription and chromatin factors that bind to DNA. This method describes the isolation of the zebrafish cell type of interest, the ChIP procedure, and the subsequent analysis of sequence datasets. We

describe this method here briefly. DNA and proteins are cross-linked to maintain the chromatin architecture and fix the interaction between factors bound to chromatin and DNA. Antibodies against specific histone modifications or transcription and chromatin factors are conjugated to solid beads and used to bind and subsequently precipitate the modified histones or transcription and chromatin factors of interest, respectively. After immunoprecipitation, the beads are thoroughly washed and the cross-links are reversed. Isolated DNA fragments from a specific chromosome region of interest can then be detected by quantitative PCR. Binding to all immunoprecipitated sites across the genome or specific sites of interest can be identified in a genome-wide manner by next-generation sequencing.

For ATAC-seq, cells are permeabilized to allow the modified bacterial Tn5 transposase to effectively create double-stranded breaks and add a specific short sequencing oligonucleotide to both 5'-ends of the DNA at each break. The "tagged" DNA fragments are then purified from their interacting proteins. DNA fragments from each sample are amplified and indexed for high-throughput sequencing. These techniques allow the sequences from precipitated or "tagged" fragments to be mapped to a reference genome, at which point comprehensive bioinformatics analysis can be applied to evaluate the function of these target regions by the identification of specific enriched genomic locations, binding motifs, and associated functional genes.

2. ADULT ZEBRAFISH EXSANGUINATION

For each ChIP, a minimum of 10^7 cells should be sufficient to avoid an amplification step post ChIP. Using the following exsanguination protocol, approximately 10^6 red blood cells can be acquired from one adult fish. Therefore, at least 10 adult zebrafish should be processed per ChIP. Fish are first euthanized in an overdose of Tricaine-S (MS-222, TMS, tricaine methanesulfonate, 200–300 mg/L), and they are bled quickly using a heparinized 200- μ L tip attached to a pipette. Blood cells are collected by puncturing each fish through the gills deeply enough to penetrate the heart. To aid bleeding and prevent clotting, 200 μ L of red cell isolation buffer (0.9% PBS, 5% fetal bovine serum, and 0.3–1 U heparin/mL) should be applied near the puncture site before the initial puncture. Using the same heparinized pipette tip, the blood is extracted by repeated pipetting and isolated blood is transferred to an Eppendorf tube containing the red cell isolation buffer. Samples should be kept at room temperature, since zebrafish erythrocytes are fragile and will lyse at lower temperatures.

Another easier method for exsanguination is by decapitation. Euthanized fish are placed in a Petri dish containing several milliliters of heparin solution (0.3–1 U heparin/mL). Zebrafish are then decapitated with a sharp razor blade. This procedure needs to be approved by the Institutional Animal Care and Use Committee (IACUC) (Fan & Yazulla, 1997) at individual institutions. The blood will flow freely out of the animal into the heparin solution. To collect the cells, use a 200- μ L heparinized pipette tip and transfer cells to a 50-mL canonical tube containing the red cell isolation buffer.

3. CHROMATIN IMMUNOPRECIPITATION AND ChIP-seq LIBRARY CONSTRUCTION (LEE, JOHNSTONE, & YOUNG, 2006)

3.1 CROSS-LINKING

After isolation, the DNA and proteins should be immediately cross-linked using the following protocol:

1. Add formaldehyde solution (50 mM HEPES-KOH pH 7.5, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, and 11% formaldehyde) at 1/10 of the volume of the red cell isolation buffer. For example, 1 mL of formaldehyde solution should be added to 10^7 cells in 10 mL of the red cell isolation buffer.
2. Place samples on a shaker at room temperature for 5 min.
3. Immediately after the 5-min incubation, add 1/20 volume of 2.5 M glycine to quench the formaldehyde.
4. Incubate the cells for 5 min at room temperature.
5. Pellet the cells by centrifugation at $1100 \times g$ for 5 min at 4°C.
6. Wash the cells twice with excess amounts of PBS (for example, 50 mL). This should be done as fast as possible to prevent excessive lysis of the erythrocytes. The researcher will notice the supernatant turning red, which is an indication of free hemoglobin from the lysed cells when significant cell lysis occurs.
7. After the washes, use the cell pellets immediately or flash freeze them in liquid nitrogen. Flash frozen cell pellets can be stored for months at -80°C .

Cross-linking is crucial for an effective ChIP as inefficient cross-linking can lead to loss of interactions, whereas too much cross-linking can result in identifying false positive protein–protein and protein–DNA interactions. Cross-linking times may need to be experimentally determined for a specific antibody and/or a particular cell type.

In previously published protocols detailing ChIP analysis of mouse embryonic stem cells (ESCs), a limited amount of cells of interest were mixed with a large number of cells from other species, such as *Drosophila* cells, to minimize material loss in the process due to limited starting material (carrier ChIP or CChIP) (O'Neill, VerMilyea, & Turner, 2006). A new carrier ChIP-seq protocol was developed by Zheng et al. (2015). It employed not only bacterial cell carriers at the cross-linking and sonication steps but also DNA carriers at the DNA isolation and amplification steps. A single cell ChIP-seq analysis was also used to define subpopulations in ESCs by chromatin state (Rotem et al., 2015). It will be very interesting to determine whether or not these protocols can be adopted for ChIP-seq analysis of rare chromatin and transcription factors. However, for the protocol presented here, these modifications are unnecessary, as an adequate number of adult zebrafish erythrocytes can be isolated for performing conventional ChIP. Nevertheless, in the case that only a limited number of cells of interest can be isolated, these modified protocols could serve as useful alternatives.

3.2 SONICATION

The sonication conditions are crucial for the success of the ChIP. Sonication permits the fragmentation and solubilization of chromatin. The size of the fragments determines the precision of the identification of protein factor binding regions. Smaller fragments will allow for more precise localization, but too much sonication can lead to lower signal-to-noise ratios as a result of the loss of antibody epitopes due to protein or DNA degradation. The ideal fragment size should be around 200 bp.

Prior to sonication, fixed cells need to be lysed. This can be done using the following protocol:

1. For 10^7 erythrocytes, resuspend the cells in 10 mL of lysis buffer 1 (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, 1× protease inhibitors).
2. Incubate the cells for 10 min at 4°C.
3. Pellet the cells by centrifugation at $1350 \times g$ for 5 min at 4°C.
4. Resuspend the pellets in 10 mL of lysis buffer 2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1× protease inhibitors)
5. Incubate the cells for 10 min at room temperature.
6. Pellet the cells by centrifugation at $1350 \times g$ for 5 min at 4°C.
7. Resuspend the pellets in 1 mL of lysis buffer 3/sonication buffer (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.5% *N*-lauroylsarcosine, 1× protease inhibitors).

Once the cells have been lysed, the samples can be sonicated in a BioRuptor sonicator for 24 cycles of 30-s sonication with a 59-s resting interval after each cycle. The sonicated material should be checked for the appropriate fragment sizes before proceeding. Fragment sizes can be estimated using the following protocol:

1. Combine 10 μ L of sample with 20 μ L TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0) containing 0.1% SDS and boiled for 15 min.
2. Mix the samples with a proper amount of loading dye and run on a 1.5% agarose gel to evaluate the size of the fragments. If the mean fragment size is larger than 200 bp, perform additional cycles of sonication in increments of two to six cycles.
3. Check fragment sizes on an agarose gel as described earlier until the desired size has been achieved.

This boiling method of denaturing chromatin-bound proteins for the visualization of sonicated DNA on an agarose gel is a quick way to approximate the fragment size after sonication. To assess the sonication efficiency more accurately, the proper reversal of cross-linking is necessary (see later discussion).

After sonication, Triton X-100 is added to sonicated lysates at 1% final concentration, after which they are centrifuged for 10 min at $18,000 \times g$ to pellet cell debris. Following this centrifugation step, 50 μL of the supernatant should be set aside as a total input control for a ChIP starting with approximately 10^7 cells and kept at -20°C until step F.

As an alternative to the protocol described previously, a single buffer can be used for cell lysis and sonication so as to minimize loss of cells. Collected cell pellets should be resuspended in an SDS ChIP buffer (20 mM Tris-HCl pH 8.1, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100, 1 $\times$ protease inhibitors) (Kim, Cantor, Orkin, & Wang, 2009). Sonication conditions will be different using this buffer, so fragmented chromatin should be checked on an agarose gel as described earlier to determine the optimal number of sonication cycles.

3.3 PRETREATMENT OF BEADS AND ANTIBODIES

Prior to immunoprecipitation, the selected antibody should be attached to magnetic protein A or protein G Dynabeads. Fifty microliters of beads is used for each immunoprecipitation. To prepare the beads, they are washed twice with 1 mL blocking buffer (1 $\times$ PBS, 0.5% BSA) and then resuspended in 250 μL blocking buffer plus 5 μg of antibody. The amount of antibody can be adjusted. All washes are performed using the Dynal MPC magnet for bead/antibody collection. Beads are collected by a 30-s incubation on the magnet, the supernatant is removed, and then the next buffer is added immediately. The beads are rotated with the antibody for 6 h to overnight at 4°C . After the incubation, the beads are washed twice with 1 mL blocking buffer to remove any excess antibody. It is crucial for the beads not to get dry between washes. The beads are now ready to be used for immunoprecipitation.

3.4 CHROMATIN IMMUNOPRECIPITATION

The sonicated material from step B (after the addition of Triton X-100) can be directly added to the antibody-conjugated beads from step C. If the alternative buffer for sonication, which does not require nuclear extraction, is used, then the sonicated material can be used for immunoprecipitation directly in the same buffer. The beads and lysates are incubated together with agitation for a time ranging from 4 h to overnight at 4°C .

It is important to perform a control immunoprecipitation in parallel. Two standard controls commonly used are an IgG isotype control that is the same isotype as the antibody used for the ChIP or a control with the beads alone (Sandmann, Jakobsen, & Furlong, 2006). These controls will help eliminate false positive results caused by nonspecific binding to antibodies or beads. In addition to these two types of nonspecific binding, chromatin may also adhere nonspecifically to the surface of plastic tubes. To minimize this problem, siliconized or manufacturer-certified low-retention tubes can be used, especially when the amount of starting material is less than optimal.

3.5 CHROMATIN IMMUNOPRECIPITATION WASHES

Following the immunoprecipitation, several washes are employed to remove the nonbound chromatin from the bead/antibody mixture. Beads are washed once with wash buffer 1 (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), once

with wash buffer 2 (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), once with wash buffer 3 (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 2 mM EDTA, 1% NP40), and once with TE buffer. All washes are performed at 4°C and using the Dynal magnet as described in part C. For each wash, the supernatant is removed using a pipette and 1 mL of the next solution is added. Tubes are shaken vigorously until all the beads are resuspended and then placed on a rocker for 3-min incubations at 4°C. If siliconized or low-retention tubes are not used, the beads are transferred to new tubes after each wash to minimize nonspecific binding of chromatin fragments to plastic tubes. After the last wash with TE buffer, the supernatant is removed, and the tubes are centrifuged at $960 \times g$ for 3 min at 4°C. Residual TE is then removed.

3.6 ELUTION AND REVERSAL OF CROSS-LINKING

The beads from step E are resuspended in 200 μ L elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, and 1% SDS) by heating for 30 min at 65°C in a shaking heat block. The supernatant is then transferred to a new tube for cross-link reversal that is done for 6 h to overnight at 65°C. Longer incubations may result in noisier data. It is important to also reverse the cross-links of the total input controls saved in step B. For this, 150 μ L elution buffer is added to each 50 μ L of the total input controls. They are then incubated at 65°C along with the ChIP samples.

3.7 DNA RECOVERY AND QUALITY CONTROL

After the reversal of cross-linking, the samples are treated with RNase for 2 h at 37°C, followed by proteinase K for 2 h to overnight at 55°C. Then the DNA is extracted using phenol/chloroform/isoamyl alcohol followed by an ethanol precipitation. The ChIP DNA is resuspended in 30 μ L TE buffer and the total input controls in 150 μ L TE buffer. The concentration of the ChIP DNA samples should be around 1 ng/ μ L.

To validate the ChIP, quantitative PCR is used. The primers are designed based on known information about the binding sites of the transcription factor. Standard qPCR conditions are utilized. If available, positive and negative primer sets should be used. If a 384-well qPCR platform is available, less than 0.5 μ L of the ChIP sample can be used in a 5- μ L total reaction volume. Primer sets for a housekeeping gene are included to normalize different ChIP samples since there is always residual nonspecific binding. The enrichment is quantified in comparison to the respective total input control samples and the negative controls ChIP (IgG isotype or beads alone) using the C_t method. Alternatively, equal amounts of the total input control and ChIP material can be used, and the enrichment can be measured in comparison to the respective total input control.

3.8 ChIP-seq LIBRARY PREPARATION

In general, the outlined ChIP-seq library preparation protocol is based on what has been established by Guenther et al. (2008). The sequencing libraries can be prepared with the NEBNext ChIP-Seq Library Prep Master Mix Set for Illumina (NEB, E6240) according to the manufacturer's instructions. 200–1000 ng of input DNA and as little as 5 ng of ChIP DNA have been used successfully to make high-quality libraries. DNA overhangs are turned into phosphorylated blunt ends and the samples are purified using Agencourt AMPure XP

beads (Beckman Coulter, A63880; 1.8× of the sample volume). Library preparation continues by the addition of a single dA to the 3' end to allow for directional adaptor ligation. Samples are purified again as before using the AMPure XP beads at 1.8× and Illumina adapter oligonucleotides (1/10 dilution) are added to the samples. The adapter-ligated DNA is cleaned up by bead purification of the samples using the AMPure XP beads at 1.8× followed by a size selection step using AMPure XP beads (0.9× followed by 0.2× of the sample volume), which selects for library fragment sizes of around 150 bp. The samples are amplified by PCR (limited amplification of up to 18 cycles) that adds additional linker sequence to the fragments to prepare them for annealing to binding sites in a proper Illumina sequencing flow cell. Lastly, the amplified libraries are again purified with AMPure XP beads (1× of the sample volume). The purified libraries can then be stored at −20°C. Libraries are quantified using the QuBit dsDNA HS assay kit (Thermo Fisher Scientific), and their quality is assessed by running 1 µL of appropriately diluted sample on a 2100 Bioanalyzer (Agilent Technologies), a TapeStation, or an AATI Fragment Analyzer. Using the protocol with size selection described previously, the peak fragment size usually runs at 250–300 bp. In our experience, size selection using beads is sufficient to generate libraries with this fragment range. If that is not the case, one can add another size selection step after library amplification (instead of the final bead purification), in which the libraries are separated on a 2% agarose gel and products between 150 and 350 bp are selected and extracted from the gel (the products include fragments of 50–250 bp plus approximately 100 bp of primer sequence). It is worth noting that gel purification leads to a greater loss of material than bead purification.

As an alternative to using a QuBit for quantifying the ChIP-seq libraries, library concentration can be determined using the KAPA Library Quantification Kit (Kapa Biosystems, kit varies depending on PCR machine used). However, very high dilutions (1:2,500,000 and higher) are required to accurately measure the library concentration by qPCR, which can lead to error. In our experience, the library concentration measured by a QuBit assay in combination with the fragment size, as determined by the Bioanalyzer or AATI runs, has yielded sufficiently accurate information for the calculation of molarity for pooling 12 libraries into one Illumina Rapid SE flow cell run on a HiSeq 2500 system.

All protocols for Illumina/Solexa sequence preparation, sequencing, and quality control are also available on the Illumina website (<http://www.illumina.com/pages.ilmn?ID=203>). Different platforms can also be used such as the Ion Torrent, or the Roche 454 platform (Harismendy et al., 2009; Kircher & Kelso, 2010). Distinct experiments may require diverse sequencing platforms and thus appropriate sample preparations.

4. ATAC-seq ANALYSIS OF OPEN CHROMATIN REGIONS

A modified bacterial Tn5 transposase (Adey et al., 2010) has been engineered to effectively induce random double-stranded breaks into genomic DNA both in vitro and in vivo, followed by the ligation of specific short nucleotide sequences to the 5' ends of either side of the breaks. The assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) (Buenrostro, Giresi, Zaba, Chang, and Greenleaf (2013)) utilizes the ability of this modified Tn5 transposase for “tagmentation” of chromatin in permeabilized

cells, which tags DNA ends within open chromatin regions with unique Illumina library adaptors. The assumption is that there are higher chances of breaking and tagging chromatin regions that are more open and accessible, which are probable sites of genomic regulatory elements. Coupled with gene expression profiling and ChIP-seq analysis of histone modifications, chromatin factors, and transcription factors, this assay can offer a confirmatory and complementary view of the epigenomic landscape of particular cell types. Moreover, ATAC-seq works robustly on samples with limited cell inputs (500–50,000 cells), making it a useful tool for epigenomic analysis of rare cell populations.

As the epigenomic landscape of a given cell type is unique, it is important to isolate a pure population of cells for ATAC-seq analysis to achieve an adequate signal-to-noise ratio. Given the relatively low number of cells needed for ATAC-seq, one can afford to sacrifice cell quantity for purity during cell isolation procedures from zebrafish embryos or organs. Here we outline an example protocol for ATAC-seq analysis of zebrafish peripheral red blood cells. This is an adaptation of the original protocol published in *Nature Methods* by Buenrostro et al. (2013) with slight modifications that we have found to work best for analyzing freshly harvested zebrafish red blood cells. This assay can also work with viable frozen cells. Cells after flash freezing treatment also have worked for some cell lines and primary human T cells purified from frozen peripheral blood.

Zebrafish adult peripheral red cells are harvested by heart puncture or decapitation as described in Section 1. As mentioned earlier, approval from institutional IACUC is required before carrying out any of the described handling of zebrafish adults. Since the ATAC-seq analysis requires only small numbers of cells, from 500 to 50,000, a small fraction of about 1 million peripheral red cells collected from one adult zebrafish is sufficient for analysis by ATAC-seq. Other cells can be processed similarly by following the protocol outlined later. However, it is important to optimize the transposition step by optimizing the amount of enzyme mix, number of cells, and incubation time (30–90 min) used in each transposition reaction.

4.1 CELL TREATMENT AND NUCLEUS ISOLATION

Collected cells are spun down in the cell collection medium at $500 \times g$ for 5 min at 4°C. Excess cold PBS or cell collection medium can be used to achieve a balance between different sample volumes. Extreme care should be taken not to disturb (almost) invisible cell pellets. Less than 10,000 cells can be used to produce meaningful results, but they require a careful titration of the amount of transposase, incubation time, and reaction volume that can vary significantly between different types and numbers of cells. Additionally, no purification after a transposition reaction is recommended for cell numbers between 500 and 5000. Samples with low cell numbers require immediate library PCR amplification and indexing after the transposition reaction (see later discussion).

Cell pellets are gently washed each with 50 μ L of cold $1\times$ PBS and spun down at $500 \times g$ for 5 min at 4°C. After removal of PBS, 50 μ L of cold cell lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% IGEPAL CA-360) is gently added, and cells are resuspended by gentle pipetting followed by an immediate spin at $500 \times g$ for 5 min at 4°C. The transposition reaction mix can be set up during the spin. For each sample, the following

mix is prepared: 25 μ L of 2 $\times$ TD Buffer (in the Nextera DNA Library Preparation Kit, Illumina, FC-121-1031), 2.5 μ L of transposase enzyme (in the same kit), and 22.5 μ L of nuclease-free water for a total of 50 μ L.

4.2 TRANSPOSITION OF OPEN CHROMATIN REGIONS

The supernatant of the cell lysis reaction buffer is carefully pipetted off from all samples, and the cell pellets are left on ice. Immediately the transposition reaction mix is carefully added, and the cell pellet is resuspended by pipetting up and down very gently. Since the incubation time, amount of transposition enzyme and transposition reaction volume can vary significantly between different cell types, it is a good practice to run an optimization experiment by varying the time of incubation and the amount of transposase. Normally, an incubation time of 30–90 min at 37°C can work well for a 50,000-cell sample.

4.3 DNA ISOLATION AND LIBRARY PREPARATION

Total DNA from each transposition reaction is isolated using the Qiagen MinElute PCR purification kit and eluted into 10 μ L of the elution buffer provided in the kit. The isolated DNA samples can be stored at –20°C before library amplification and indexing or used immediately for the next step. If 50,000 cells are used, it is also possible to run a fraction (10%) of the DNA samples (in a nanogram range) on a TapeStation, Bioanalyzer DNA chip, or AATI fragment analyzer before library amplification. This can help with a quick optimization of the transposition reactions for new cell types. This is not recommended for samples with a fewer number of cells since the amount of purified DNA will be below the detection limits of the aforementioned DNA analysis systems.

4.4 LIBRARY AMPLIFICATION AND INDEXING

Any high-fidelity PCR master mixes can be used for the outlined initial library amplification. Here we use the high-fidelity PCR master mix (New England Biolabs, M0541) as an example. This PCR reaction can be carried out according to the following protocol:

1. Make a PCR reaction mix consisting of 7 μ L nuclease-free water (adjustable), 2.5 μ L Nextera primer Ad1 (at 25 μ M), 2.5 μ L Nextera primer Ad2.[x] (at 25 μ M), 3 μ L 10 $\times$ SYBR Green I (a 1:1000 dilution of Invitrogen 10,000 $\times$ SYBR Green I, S-7563, in 10 mM Tris–HCl), and 25 μ L NEBNext High-Fidelity 2 $\times$ PCR master mix (NEB, M0541). Be sure to prepare some extra mix to account for loss in aliquots in each sample.
2. Add 40 μ L of the above mix to each of the 10 μ L transposed DNA samples and mix thoroughly on ice.
3. Perform an initial PCR amplification of the libraries:
 - a. 72°C, 5 min
 - b. 98°C, 30 s
 - c. 98°C, 10 s
 - d. 63°C, 30 s

- e. 72°C, 1 min
- f. Repeat steps 3–5, 4 times
- g. Hold at 4°C

It is recommended to synthesize a set of custom Greenleaf ATAC primers (<http://seq.liai.org/204-2/> and Table 1) that includes one common primer ATAC_Ad1_NoMX and 24 index primers ATAC_Ad2.1-Ad2.24 (from Buenrostro et al., 2013). We have not tried to use the primers from the Nextera Index Kits (24 or 96 indexes) for amplification and indexing of ATAC-seq libraries.

A side qPCR reaction is performed to determine how many additional PCR cycles described previously are needed to obtain enough library material for sequencing without overamplification, which can lead to GC-rich sequence and size bias during the library amplification and indexing. While the initial library amplification is being performed, a qPCR side reaction is prepared in a 15-μL total volume. This can be done using the following procedure:

1. Make a qPCR reaction mix consisting of 3.9 μL nuclease-free water (adjustable), 0.25 μL Nextera primer Ad1 (at 25 uM), 0.25 μL Nextera primer Ad2.[x] (at 25 uM), 0.6 μL 10× SYBR Green I, and 5 μL 2× High-Fidelity NEB PCR Master Mix. Be sure to prepare some extra mix to account for loss in aliquots in each sample.
2. Combine 10 μL of the qPCR reaction mix with 5 μL of each of the “5-cycle” initial library amplification and indexing reactions. Mix thoroughly by gentle pipetting.
3. Store the remaining 45 μL of the initial library amplification and indexing reactions at 4°C.
4. Perform the side qPCR amplification with the following cycles:
 - a. 98°C, 30 s
 - b. 98°C, 10 s
 - c. 63°C, 30 s
 - d. 72°C, 1 min
 - e. repeat steps 2–4, 19 times
 - f. hold at 4°C

As shown in Fig. 1, fluorescent signals from the SYBR Green that are bound to the amplified products are plotted against cycle numbers. After determining the maxima of SYBR fluorescence for the plateau of each sample, plot and calculate the cycle number on the x-axis that corresponds to a quarter of the maxima. This quarter-maxima cycle number is used to perform additional library amplification and indexing for a particular sample. If the quarter-maxima cycle number falls between two cycle numbers, it is recommended to take the smaller number for additional PCR cycles.

Additional PCR cycles should be run with identical PCR settings as described previously across all samples, except for the number of cycles, which should vary based on the specific quarter-maxima value for each sample. It is more convenient to set the PCR machine to perform the highest number of cycles for a group of samples and to remove individual samples quickly once they have completed the appropriate number of cycles. The final amplified libraries are purified again using the Qiagen MinElute PCR purification kit and eluted into 20 μ L of the elution buffer. The purified libraries can then be stored at -20°C .

4.5 LIBRARY QUALITY CONTROL AND QUANTIFICATION

Accurate measurement of library concentration and quality is critical for successful sequencing and subsequent data analysis. Library concentration can be measured using the QuBit dsDNA HS Assay Kit. Alternatively, a qPCR library quantification kit (eg, from Kapa Biosystems, but see comments in Section 3, step H) can be used. A proper sensitivity chip and/or run setup of TapeStation, Bioanalyzer, or AATI Fragment Analysis is used to check for library quality and average fragment size information (Fig. 2). Successfully tagged samples will display a characteristic nucleosomal “laddering” pattern. The molar concentration of a library is calculated for sequencing flow cell cluster generation and subsequent sequencing. The dark smear peak below 100 bp is not included in the calculation of molar concentration of libraries and their average fragment sizes. This smear is believed to be DNA fragments from mitochondrial DNA without proper tagging for clustering and sequencing on an Illumina next-generation sequencing platform.

4.6 SEQUENCING CONSIDERATIONS

Because a low number of cells are sufficient for ATAC-seq analysis, the sequencing of libraries with relatively higher amplification normally affords a lower mapping rate (50–70%) of the resulting sequence reads. We typically aim to obtain 40–50 million reads from each library for a good representation of genomic complexity. We typically pool 12 libraries into one Rapid PE flow cell for a 75- or 100-bp paired-end sequencing run on a HiSeq 2500 platform. Longer reads and paired-end sequence reads typically result in more accurate mapping and higher mapping rates.

5. ChIP-seq and ATAC-seq DATA ANALYSIS

As a method to detect DNA-protein binding in vivo, ChIP-seq has become the most widely used procedure for investigation of genome-wide transcriptional regulation and epigenetic modifications (Bailey et al., 2013). The typical ChIP-seq raw reads from a sequencing machine are short DNA fragments from the DNA-protein binding regions. ATAC-seq produces a library of DNA read tags that map to genomic regions that are relatively more accessible to transposase activity due to a looser organization of nucleosomes. Since both approaches aim to find enriched genomic regions from short DNA fragments, the computational analysis procedures of ChIP-seq and ATAC-seq are similar. Most of the tools discussed here are command-line tools. However, user-friendly integrated online ChIP-seq and ATAC-seq pipelines have been developed for life science researchers who do not have extensive computational experience. Galaxy (<https://usegalaxy.org>) and GenePattern ([*Methods Cell Biol.* Author manuscript; available in PMC 2017 July 20.](http://</p></div><div data-bbox=)

www.broadinstitute.org/cancer/software/genepattern/) are alternative tools that are commonly used.

Sufficient sequencing depth in ChIP-seq and ATAC-seq is required for effective data analysis. The necessary sequencing depth depends on the genome size and the size of the genomic regions targeted for investigation. For typical zebrafish transcription factor binding sites and ATAC-seq regions, a total of 20 million mappable reads per sample are sufficient. For most histone marks and for proteins with many binding sites, more sequence reads (40–50 million) can help identify binding sites with greater confidence. An abundance of reads can also help to reveal weaker binding sites that may be harder to detect. Similarly, for ATAC-seq, open chromatin regions can be located more accurately and with greater certainty when more reads are available from a library.

Preprocessing and quality examination of the raw sequencing data are necessary for assessing the quality of the raw reads and identifying sources of error or bias. FastQC is used to provide an efficient and comprehensive evaluation of data quality. This tool can be incorporated into the whole data processing pipeline to ensure the reliability of the downstream analysis. Key parameters to examine include the Phred score (http://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf), which measures the error probability at each base (eg, a Q30 value corresponds to a 1-in-1000 base error rate), the overrepresented sequences that are possibly sequencing adapters or other contaminants, and the overall GC content. Cutadapt (<http://cutadapt.readthedocs.org/en/stable/index.html>) can be used to trim the raw reads based on quality scores, adapter contamination, or other criteria, such as read length to remove short reads.

After quality filtration, the sequence reads are mapped to the zebrafish reference genome using available mapping programs, such as Bowtie (Langmead, Trapnell, Pop, & Salzberg, 2009) and BWA (Li & Durbin, 2009). Sequence variations are more common in the zebrafish genome than they are in the human genome, so a higher mismatch rate (2 in 100 bp) is allowed for individual sequence alignment. These mapping programs output all mapped reads in a SAM or BAM format. After these files are sorted by genomic locations and indexes—procedures which can both be done using either SAMtools by Li et al. (2009) and Li (2011). (<http://samtools.sourceforge.net>) or Picard Tools by the Broad Institute (<http://broadinstitute.github.io/picard/>)—all mapped reads can be visualized in IGV by the Broad Institute (<https://www.broadinstitute.org/igv/>) and/or UCSC genome browser by Kent et al. (2002) (<https://genome.ucsc.edu>).

The next step is peak calling, the prediction of genomic regions characterized by an enrichment of mapped reads. Several peak-calling algorithms and tools are available, such as MACS (Zhang et al., 2008), SPP (Kharchenko, Tolstorukov, & Park, 2008), and ZINBA (Rashid, Giresi, Ibrahim, Sun, & Lieb, 2011). Peak-calling steps are generally similar between ChIP-seq and ATAC-seq, but there are a few differences: 1. ChIP-seq of input DNA or mock-IP are used as controls in ChIP-seq analysis, but not in ATAC-seq; 2. In general, ChIP-seq uses single-end reads and ATAC-seq uses paired-end reads; 3. When calling ATAC-seq peaks using MACS2, three parameters (–nomodel –shift –100 (or 75) –extsize 200 (or 150)) are added to the default MACS2 arguments, which are initially designed for

ChIP-seq analysis. These different parameters are used so that called ATAC-seq peaks are centered on the cutting sites created by the modified transposase, Tn5, rather than on the locations of the reads themselves. Common enrichment metrics such as p-value or q-value (FDR) are used as statistical threshold for calling peaks in both ChIP-seq and ATAC-seq analyses. It is worth noting that the number of detected peaks can vary significantly between samples from ChIP-seq and ATAC-seq analyses. Peak numbers can also vary significantly between ChIPs when different antibodies are used to analyze the same cells, even when the different antibodies are against the same protein in the same cells.

It is of great interest to identify similar and different genomic regions defined by ChIP-seq and ATAC-seq in different but related cell types. It is also illuminating to compare the peaks either within ChIP-seq and ATAC-seq results or between ChIP-seq and ATAC-seq results for the same cells under different treatment conditions or for entirely different cells. Comparison of protein-bound or open chromatin regions across different conditions or cell types can be simply performed by peak overlap using BEDTools (Quinlan & Hall, 2010) and can provide useful information about similar and different mechanisms that are used to regulate cell function and define cell identity. However, reliable quantitative and differential binding analyses rely on an adequate number of reads at the relevant peaks, which can be determined using MANorm (Shao, Zhang, Yuan, Orkin, & Waxman, 2012) or DBChIP (Liang & Keles, 2012). It may be necessary to slightly modify parameters for each individual analysis to evaluate more accurately the peaks of interest in different experiments. Experimental and functional verifications of called peaks are also invaluable whenever possible.

The called peaks for both ChIP-seq and ATAC-seq experiments are presented in a BED file, which highlights the genomic positions of the peaks, and a bigWig file, which displays the genomic profile of the mapped reads, which can both be visualized with the IGV or UCSC genome browser tools. Using these visualization tools, the genomic annotation of peaks, such as whether they occupy a promoter region, gene body, etc., is made readily visible. A genome-wide peak annotation can be achieved by GREAT (McLean et al., 2010) or other methods that associate genome regions with annotated genes and their structures. Then functional analysis of these annotated peaks, made possible through tools such as DAVID (Huang da, Sherman, & Lempicki, 2009), GREAT, and GSEA (Subramanian et al., 2005), can provide information about gene ontology enrichment or gene set enrichment of the genes associated with called peaks.

DNA-protein binding motifs enriched in ChIP-seq peaks can be found using motif discovery algorithms that are designed for analyzing ChIP-seq data such as MEME-ChIP (Machanick & Bailey, 2011). Motif analysis can be used to validate the ChIP-seq experiment if the motif is known or for the discovery of novel motifs and binding partners that are associated with ChIPed proteins or histone modifications. Likewise, motifs underneath and around ATAC-seq peaks provide useful information about protein factors that interact with these DNA segments. All of these analyses will facilitate the studies of molecular mechanisms that modulate transcriptional control of cell fate and function.

6. RESULTS

We applied ChIP-seq and ATAC-seq to analyze the molecular mechanisms that modulate transcriptional regulation and to reveal the corresponding epigenetic landscape in the major globin locus. Globin switch is a biological process that is conserved from zebrafish to mammals. The current understanding of the mechanism that governs the globin switch has important implications in the diagnosis and treatment of anemia, especially thalassemia, which results from an imbalance in the production of alpha- and beta-globin chains. Zebrafish has been shown to be an excellent model for the understanding of vertebrate globin switch and transcriptional control of globin gene expression (Brownie et al., 2003; Ganis et al., 2012; Song et al., 2004). There is little or no embryonic *globin* mRNA expression in adult erythroid cells, even though both embryonic and adult *globin* genes are located near each other at the major globin locus and are under the control of the same locus control region (LCR) that is located in the intron 4 of its neighboring gene *npr13*. ChIP-seq analyses of the H3K4me3 mark, an indicator of active transcription, and *gata1*, the erythroid master transcription regulator, reveal specific binding at the promoter sites of adult *globin* genes, while there is no binding at the embryonic *globin* promoter sites (Fig. 3). These evaluations at a molecular and genomic level exhibit cell-specific histone modifications and molecular interactions of the cell-specific transcription factor. They are consistent with cell-specific transcription in adult erythroid cells revealed by qPCR and gene expression profiling.

It is also interesting that *gata1* interacts only with promoters of adult globins and the LCR. We examined the chromatin landscape in adult erythroid cells by ATAC-seq and DNase-seq to see if the adult globin cluster and LCR have a different epigenetic organization in adult erythrocytes. As expected, the adult globin cluster is open while the embryonic globin cluster stays closed, as evidenced by the peaks from ATAC-seq and DNase-seq analyses. Together, these comprehensive analyses demonstrate that there is an adult erythroid epigenetic organization that supports adult erythroid specific gene transcription and therefore cell identity and function.

SUMMARY AND CONCLUSION

This chapter describes effective methods for the identification of transcription factor binding sites and open chromatin regions of the zebrafish epigenome in a genome-wide manner with the use of ChIP-seq and ATAC-seq technologies. These methods can be easily extended to other abundant and relatively pure cell types besides adult erythrocytes. However, as with many other molecular and biochemical assays, the cell populations analyzed need to be as homogenous as possible to achieve consistent and easily interpretable results. It should also be noted that the choice of antibody for ChIP-seq studies is crucial and that not all antibodies can be used successfully in ChIP, even though they may work well in other methods such as Western blotting or immunohistochemistry. The conditions of the experiment may need to be optimized for the number of cells, the abundance of the protein factors of interest, and the available antibodies. The number of cells required for open chromatin detection by ATAC-seq is much smaller. Furthermore, we recommend trying to optimize the methods presented here to achieve the best outcomes, especially when they are

applied to different cell types. In conclusion, these epigenomic analyses can reveal the landscape of transcription and chromatin factors as well as the chromatin states of a variety of zebrafish cell types. The in-depth and comprehensive bioinformatics analyses described here, especially in conjunction with gene expression profiling information, are essential for our understanding of the molecular mechanisms that regulate zebrafish cell function and fate.

Acknowledgments

This work was supported by NIH #5PO1HL32262-34 (to S.H.O), 3P30DK049216-21S1 (to S. H. O), #5R01HL048801-23 (to L.I.Z), and NIH #5K99CA190861-01 (to C.J.O).

References

- Adey A, Morrison HG, Asan Xun X, Kitzman JO, Turner EH, ... Shendure J. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biology*. 2010; 11(12):R119. <http://dx.doi.org/10.1186/gb-2010-11-12-r119>. Epub 2010 December 8. [PubMed: 21143862]
- Asnani A, Peterson RT. The zebrafish as a tool to identify novel therapies for human cardiovascular disease. *Disease Models & Mechanisms*. Jul; 2014 7(7):763–767. <http://dx.doi.org/10.1242/dmm.016170>. Review. [PubMed: 24973746]
- Bailey T, Krajewski P, Ladunga I, Lefebvre C, Li Q, Liu T, ... Zhang J. Practical guidelines for the comprehensive analysis of chip-seq data. *PLoS Computational Biology*. 2013; 9(11):e1003326. <http://dx.doi.org/10.1371/journal.pcbi.1003326>. [PubMed: 24244136]
- Brownlie A, Hersey C, Oates AC, Paw BH, Falick AM, Witkowska HE, ... Zon L. Characterization of embryonic globin genes of the zebrafish. *Developmental Biology*. Mar 1; 2003 255(1):48–61. [PubMed: 12618133]
- Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*. Oct 6; 2013 10(12):1213–1218. <http://dx.doi.org/10.1038/nmeth.2688>. [PubMed: 24097267]
- Carroll KJ, North TE. Oceans of opportunity: exploring vertebrate hematopoiesis in zebrafish. *Experimental Hematology*. Aug; 2014 42(8):684–696. <http://dx.doi.org/10.1016/j.exphem.2014.05.002>. Epub 2014 May 9. Review. [PubMed: 24816275]
- Cox AG, Goessling W. The lure of zebrafish in liver research: regulation of hepatic growth in development and regeneration. *Current Opinion in Genetics & Development*. Jun.2015 32:153–161. <http://dx.doi.org/10.1016/j.gde.2015.03.002>. Epub 2015 April 6. Review. [PubMed: 25863341]
- Fan SF, Yazulla S. Electrogenic hyperpolarization-elicited chloride transporter current in blue cones of zebrafish retinal slices. *Journal of Neurophysiology*. 1997; 77:1447–1459. [PubMed: 9084610]
- Ganis JJ, Hsia N, Trompouki E, de Jong JL, DiBiase A, Lambert JS, ... Zon LI. Zebrafish globin switching occurs in two developmental stages and is controlled by the LCR. *Developmental Biology*. Jun 15; 2012 366(2):185–194. <http://dx.doi.org/10.1016/j.ydbio.2012.03.021>. Epub 2012 April 19. [PubMed: 22537494]
- Goessling W, Sadler KC. Zebrafish: an important tool for liver disease research. *Gastroenterology*. Nov; 2015 149(6):1361–1377. <http://dx.doi.org/10.1053/j.gastro.2015.08.034>. Epub 2015 August 28. Review. [PubMed: 26319012]
- Guenther MG, Lawton LN, Rozovskaia T, Frampton GM, Levine SS, Volkert TL, ... Young RA. Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. *Genes Development*. 2008; 22:3403–3408. [PubMed: 19141473]
- Gurevich D, Siegel A, Currie PD. Skeletal myogenesis in the zebrafish and its implications for muscle disease modelling. *Results and Problems in Cell Differentiation*. 2015; 56:49–76. http://dx.doi.org/10.1007/978-3-662-44608-9_3. Review. [PubMed: 25344666]

- Haesemeyer M, Schier AF. The study of psychiatric disease genes and drugs in zebrafish. *Current Opinion in Neurobiology*. Feb.2015 30:122–130. <http://dx.doi.org/10.1016/j.conb.2014.12.002>. Epub 2014 December 16. Review. [PubMed: 25523356]
- Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, ... Frazer KA. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biology*. 2009; 10:R32. [PubMed: 19327155]
- van Houcke J, De Groef L, Dekeyser E, Moons L. The zebrafish as a gerontology model in nervous system aging, disease, and repair. *Ageing Research Reviews*. Nov; 2015 24(Pt B):358–368. <http://dx.doi.org/10.1016/j.arr.2015.10.004>. Epub 2015 October 29. Review. [PubMed: 26538520]
- Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols Other Titles: Protocols*. 2009; 4:44–57. <http://dx.doi.org/10.1038/nprot.2008.211>.
- Ja wi ska A, Sallin P. Regeneration versus scarring in vertebrate appendages and heart. *The Journal of Pathology*. Jan; 2016 238(2):233–246. <http://dx.doi.org/10.1002/path.4644>. Epub 2015 October 27. Review. [PubMed: 26414617]
- Jörgens K, Stoll SJ, Pohl J, Fleming TH, Sticht C, Nawroth PP, Hammes HP, Kroll J. High tissue glucose alters intersomitic blood vessels in zebrafish via methylglyoxal targeting the VEGF receptor signaling cascade. *Diabetes*. Jan; 2015 64(1):213–225. <http://dx.doi.org/10.2337/db14-0352>. Epub 2014 Aug 4. [PubMed: 25092676]
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. *Genome Research*. Jun; 2002 12(6):996–1006. [PubMed: 12045153]
- Kharchenko PV, Tolstorukov MY, Park PJ. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature Biotechnology*. 2008; 26:1351–1359. <http://dx.doi.org/10.1038/nbt.1508>.
- Kim J, Cantor AB, Orkin SH, Wang J. Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. *Nature Protocols Other Titles: Protocols*. 2009; 4:506–517.
- Kircher M, Kelso J. High-throughput DNA sequencing—concepts and limitations. *BioEssays*. 2010; 32:524–536. [PubMed: 20486139]
- Legendijk AK, Yap AS, Hogan BM. Endothelial cell-cell adhesion during zebrafish vascular development. *Cell Adhesion & Migration*. 2014; 8(2):136–145. Review. [PubMed: 24621476]
- Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*. 2009; 10:R25. <http://dx.doi.org/10.1186/gb-2009-10-3-r25>. [PubMed: 19261174]
- Lee TI, Johnstone SE, Young RA. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nature Protocols Other Titles: Protocols*. 2006; 1:729–748.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25:1754–1760. <http://dx.doi.org/10.1093/bioinformatics/btp324>. [PubMed: 19451168]
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N. ... 1000 Genome Project Data Processing Subgroup. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*. 2009; 25:2078–2079. [PubMed: 19505943]
- Li HA. Statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. Nov 1; 2011 27(21):2987–2993. Epub 2011 September 8. [PubMed: 21903627]
- Liang K, Keles S. Detecting differential binding of transcription factors with ChIP-seq. *Bioinformatics*. 2012; 28:121–122. <http://dx.doi.org/10.1093/bioinformatics/btr605>. [PubMed: 22057161]
- Lindeman LC, Vogt-Kielland LT, Alestrom P, Collas P. Fish'n ChIPs: chromatin immunoprecipitation in the zebrafish embryo. *Methods in Molecular Biology*. 2009; 567:75–86. [PubMed: 19588086]
- Machanic P, Bailey TL. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*. 2011; 27:1696–1697. <http://dx.doi.org/10.1093/bioinformatics/btr189>. [PubMed: 21486936]
- McCammon JM, Sive H. Challenges in understanding psychiatric disorders and developing therapeutics: a role for zebrafish. *Disease Models & Mechanisms*. Jul 1; 2015 8(7):647–656. <http://dx.doi.org/10.1242/dmm.019620>. Review. [PubMed: 26092527]

- McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, ... Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology*. 2010; 28:495–501. <http://dx.doi.org/10.1038/nbt.1630>.
- Mohseny AB, Hogendoorn PC. Zebrafish as a model for human osteosarcoma. *Advances in Experimental Medicine and Biology*. 2014; 804:221–236. http://dx.doi.org/10.1007/978-3-319-04843-7_12. Review. [PubMed: 24924177]
- Mort RL, Jackson IJ, Patton EE. The melanocyte lineage in development and disease. *Development*. Feb 15; 2015 142(4):620–632. <http://dx.doi.org/10.1242/dev.106567>. Review. Erratum in: *Development*. 2015 April 1;142(7):1387. [PubMed: 25670789]
- O'Neill LP, VerMilyea MD, Turner BM. Epigenetic characterization of the early embryo with a chromatin immunoprecipitation protocol applicable to small cell populations. *Nature Genetics*. 2006; 38:835–841. [PubMed: 16767102]
- Paik EJ, Zon LI. Hematopoietic development in the zebrafish. *International Journal of Developmental Biology*. 2010; 54:1127–1137. [PubMed: 20711990]
- Phillips JB, Westerfield M. Zebrafish models in translational research: tipping the scales toward advancements in human health. *Disease Models & Mechanisms*. Jul; 2014 7(7):739–743. <http://dx.doi.org/10.1242/dmm.015545>. Review. [PubMed: 24973743]
- Plantié E, Migocka-Patrzałek M, Daczewska M, Jagla K. Model organisms in the fight against muscular dystrophy: lessons from drosophila and zebrafish. *Molecules: a Journal of Synthetic Chemistry and Natural Product Chemistry*. Apr 9; 2015 20(4):6237–6253. <http://dx.doi.org/10.3390/molecules20046237>. Review.
- Powles-Glover N. Cilia and ciliopathies: classic examples linking phenotype and genotype-an overview. *Reproductive Toxicology*. Sep.2014 48:98–105. <http://dx.doi.org/10.1016/j.reprotox.2014.05.005>. Epub 2014 May 23. Review. [PubMed: 24859270]
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26:841–842. <http://dx.doi.org/10.1093/bioinformatics/btq033>. [PubMed: 20110278]
- Quiros-Gonzalez I, Yadav VK. Central genes, pathways and modules that regulate bone mass. *Archives of Biochemistry and Biophysics*. Nov 1.2014 561:130–136. <http://dx.doi.org/10.1016/j.abb.2014.06.005>. Epub 2014 June 17. Review. [PubMed: 24950023]
- Rashid NU, Giresi PG, Ibrahim JG, Sun W, Lieb JD. ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions. *Genome Biology*. Jul 25.2011 12(7):R67. <http://dx.doi.org/10.1186/gb-2011-12-7-r67>. [PubMed: 21787385]
- Rotem A, Ram O, Shores N, Sperling RA, Goren A, Weitz DA, Bernstein BE. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nature Biotechnology*. Nov; 2015 33(11):1165–1172. <http://dx.doi.org/10.1038/nbt.3383>. Epub 2015 October 12.
- Sandmann T, Jakobsen JS, Furlong EE. ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in *Drosophila melanogaster* embryos. *Nature Protocols Other Titles: Protocols*. 2006; 1:2839–2855.
- Saralahti A, Rämert M. Zebrafish and streptococcal infections. *Scandinavian Journal of Immunology*. Sep; 2015 82(3):174–183. <http://dx.doi.org/10.1111/sji.12320>. Review. [PubMed: 26095827]
- Schlegel A, Gut P. Metabolic insights from zebrafish genetics, physiology, and chemical biology. *Cellular and Molecular Life Sciences: CMLS*. Jun; 2015 72(12):2249–2260. <http://dx.doi.org/10.1007/s00018-014-1816-8>. Epub 2015 January 4. Review. [PubMed: 25556679]
- Shao Z, Zhang Y, Yuan GC, Orkin SH, Waxman DJ. MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. *Genome Biology*. 2012; 13:R16. <http://dx.doi.org/10.1186/gb-2012-13-3-r16>. [PubMed: 22424423]
- Song HD, Sun XJ, Deng M, Zhang GW, Zhou Y, Wu XY, ... Chen Z. Hematopoietic gene expression profile in zebrafish kidney marrow. *Proceedings of the National Academy of Sciences of the United States of America*. Nov 16; 2004 101(46):16240–16245. Epub 2004 November 1. [PubMed: 15520368]
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, ... Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.

- Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:15545–15550. <http://dx.doi.org/10.1073/pnas.0506580102>. [PubMed: 16199517]
- Takeuchi T. Regulation of cardiomyocyte proliferation during development and regeneration. *Development, Growth & Differentiation*. Jun; 2014 56(5):402–409. <http://dx.doi.org/10.1111/dgd.12134>. Epub 2014 April 16. Review.
- Vastenhouw NL, Zhang Y, Woods IG, Imam F, Regev A, Liu XS, ... Schier AF. Chromatin signature of embryonic pluripotency is established during genome activation. *Nature*. 2010; 464:922–926. [PubMed: 20336069]
- Wager K, Mahmood F, Russell C. Modelling inborn errors of metabolism in zebrafish. *Journal of Inherited Metabolic Disease*. Jul; 2014 37(4):483–495. <http://dx.doi.org/10.1007/s10545-014-9696-5>. Epub 2014 May 6. Review. [PubMed: 24797558]
- White RM. Cross-species oncogenomics using zebrafish models of cancer. *Current Opinion in Genetics & Development*. Feb.2015 30:73–79. <http://dx.doi.org/10.1016/j.gde.2015.04.006>. Epub 2015 June 9. Review. [PubMed: 26070506]
- White RM, Sessa A, Burke C, Bowman T, LeBlanc J, Ceol C, ... Zon LI. Transparent adult zebrafish as a tool for in vivo transplantation analysis. *Cell Stem Cell*. 2008; 2:183–189. [PubMed: 18371439]
- Wilkinson RN, van Eeden FJ. The zebrafish as a model of vascular development and disease. *Progress in Molecular Biology and Translational Science*. 2014; 124:93–122. <http://dx.doi.org/10.1016/B978-0-12-386930-2.00005-7>. Review. [PubMed: 24751428]
- Wilkinson RN, Jopling C, van Eeden FJ. Zebrafish as a model of cardiac disease. *Progress in Molecular Biology and Translational Science*. 2014; 124:65–91. <http://dx.doi.org/10.1016/B978-0-12-386930-2.00004-5>. Review. [PubMed: 24751427]
- Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, ... Liu XS. Model-based analysis of chip-seq (MACS). *Genome Biology*. 2008; 9:R137. [PubMed: 18798982]
- Zheng X, Yue S, Chen H, Weber B, Jia J, Zheng Y. Low-cell-number epigenome profiling aids the study of lens aging and hematopoiesis. *Cell Reports*. Nov 17; 2015 13(7):1505–1518. <http://dx.doi.org/10.1016/j.celrep.2015.10.004>. Epub 2015 November 5. [PubMed: 26549448]

FIGURE 1.
Determination of additional number of PCR cycles needed for library amplification and indexing.

FIGURE 2.
Example results from Agilent TapeStation and AATI Fragment Analysis of ATAC-seq libraries. (A) Agilent TapeStation results showing the size and laddering pattern of a set of ATAC-seq libraries. (B) AATI Fragment Analyzer results showing the size and laddering pattern of a different set of ATAC-seq libraries.

FIGURE 3.

A screen shot of H3K4me3 ChIP-seq, gata1 ChIP-seq, ATAC-seq, and DNase-seq tracks at the zebrafish major *globin* locus. The *red line* (light gray in print versions) marks the adult globin cluster, the *blue line* (black in print versions) indicates the embryonic globin cluster, the *short purple line* (gray in print versions) dictates the LCR, and the *red bar* (light gray in print versions) on the DNase I track shows the location of the LCR core site. These tracks demonstrate the coexistence of binding by the active chromatin mark (H3K4me3) and erythroid master transcription factor (*gata1*) in the adult globin gene cluster and the LCR. They also highlight an open chromatin conformation in the corresponding genomic regions, as evidenced by ATAC-seq and DNase-seq peaks.

Table 1

ATAC-seq Index Primers

Greenleaf ATAC Primers		
Primer name	Index	Primer Sequence to Order
ATAC_Ad1_noMX:	N/A	AATGATACGGCGACCACCGAGATCTACACTCTCGTCGGCAGCGTCAGATGTG
ATAC_Ad2.1_TAAGCGA	TAAAGCGA	CAAGCAGAAAGACGGCATAACGAGATTCGCCTTAGTCTCTGTGGCTCGGAGATGT
ATAC_Ad2.2_CGTACTAG	CGTACTAG	CAAGCAGAAGACGGCATAACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
ATAC_Ad2.3_AGGCAGAA	AGGCAGAA	CAAGCAGAAGACGGCATAACGAGATTTCCTGCTGTCTCGTGGGCTCGGAGATGT
ATAC_Ad2.4_TCCTGAGC	TCCTGAGC	CAAGCAGAAGACGGCATAACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
ATAC_Ad2.5_GGACTCCT	GGACTCCT	CAAGCAGAAGACGGCATAACGAGATAGGAGTCCGCTCTCGTGGGCTCGGAGATGT
ATAC_Ad2.6_TAGGCATG	TAGGCATG	CAAGCAGAAGACGGCATAACGAGATCATGCTAGTCTCTGTGGGCTCGGAGATGT
ATAC_Ad2.7_CTCTCTAC	CTCTCTAC	CAAGCAGAAGACGGCATAACGAGATGTAGAGAGGTCTCTGTGGGCTCGGAGATGT
ATAC_Ad2.8_CAGAGAGG	CAGAGAGG	CAAGCAGAAGACGGCATAACGAGATCTCTGTGGTCTCGTGGGCTCGGAGATGT
ATAC_Ad2.9_GCTACGCT	GCTACGCT	CAAGCAGAAGACGGCATAACGAGATAGGATAGCGTCTCTGTGGGCTCGGAGATGT
ATAC_Ad2.10_CGAGGCTG	CGAGGCTG	CAAGCAGAAGACGGCATAACGAGATCAGCCTCGGTCTCTGTGGGCTCGGAGATGT
ATAC_Ad2.11_AAGAGGCA	AAGAGGCA	CAAGCAGAAGACGGCATAACGAGATTGCCCTTCTCTCGTGGGCTCGGAGATGT
ATAC_Ad2.12_GTAGAGGA	GTAGAGGA	CAAGCAGAAGACGGCATAACGAGATTCTCTCTACGTCTCTGTGGGCTCGGAGATGT
ATAC_Ad2.13_GTCGTGAT	GTCGTGAT	CAAGCAGAAGACGGCATAACGAGATATCACGACGTCTCTGTGGGCTCGGAGATGT
ATAC_Ad2.14_ACCACTGT	ACCACTGT	CAAGCAGAAGACGGCATAACGAGATACAGTGGTGTCTCTGTGGGCTCGGAGATGT
ATAC_Ad2.15_TGGATCTG	TGGATCTG	CAAGCAGAAGACGGCATAACGAGATCAGATCCAGTCTCTGTGGGCTCGGAGATGT
ATAC_Ad2.16_CCGTTTGT	CCGTTTGT	CAAGCAGAAGACGGCATAACGAGATACAAACGGGTCTCTGTGGGCTCGGAGATGT
ATAC_Ad2.17_TGCTGGGT	TGCTGGGT	CAAGCAGAAGACGGCATAACGAGATACCCAGCAGTCTCTGTGGGCTCGGAGATGT
ATAC_Ad2.18_GAGGGGTT	GAGGGGTT	CAAGCAGAAGACGGCATAACGAGATAACCCCTCGTCTCTGTGGGCTCGGAGATGT
ATAC_Ad2.19_AGGTTGGG	AGGTTGGG	CAAGCAGAAGACGGCATAACGAGATCCCAACCTGTCTCTGTGGGCTCGGAGATGT
ATAC_Ad2.20_GTGTTGGTG	GTGTTGGTG	CAAGCAGAAGACGGCATAACGAGATCACACCTGTCTCTGTGGGCTCGGAGATGT
ATAC_Ad2.21_TGGGTTTC	TGGGTTTC	CAAGCAGAAGACGGCATAACGAGATGAACCCAGTCTCTGTGGGCTCGGAGATGT
ATAC_Ad2.22_TGGTCACA	TGGTCACA	CAAGCAGAAGACGGCATAACGAGATTGTGACCAAGTCTCTGTGGGCTCGGAGATGT
ATAC_Ad2.23_TTGACCCCT	TTGACCCCT	CAAGCAGAAGACGGCATAACGAGATAGGGTCAAGTCTCTGTGGGCTCGGAGATGT
ATAC_Ad2.24_CCACCTCCT	CCACCTCCT	CAAGCAGAAGACGGCATAACGAGATAGGAGTGGGTCTCTGTGGGCTCGGAGATGT