Chapter 11

Assay for Transposase Accessible Chromatin (ATAC-Seq) to Chart the Open Chromatin Landscape of Human Pancreatic Islets

Helena Raurell-Vila, Mireia Ramos-Rodríguez, and Lorenzo Pasquali

Abstract

The regulatory mechanisms that ensure an accurate control of gene transcription are central to cellular function, development and disease. Such mechanisms rely largely on noncoding regulatory sequences that allow the establishment and maintenance of cell identity and tissue-specific cellular functions.

The study of chromatin structure and nucleosome positioning allowed revealing transcription factor accessible genomic sites with regulatory potential, facilitating the comprehension of tissue-specific cis-regulatory networks. Recently a new technique coupled with high-throughput sequencing named Assay for Transposase Accessible Chromatin (ATAC-seq) emerged as an efficient method to chart open chromatin genome wide. The application of such technique to different cell types allowed unmasking tissue-specific regulatory elements and characterizing cis-regulatory networks. Herein we describe the implementation of the ATAC-seq method to human pancreatic islets, a tissue playing a central role in the control of glucose metabolism.

Key words Open chromatin, Pancreatic islets, Gene transcription, Epigenetics

1 Introduction

Transcription regulation is central to cellular function, development, and disease. Thus, the regulatory landscape controlling gene expression is highly dynamic, cell type-specific, and varies by individual genomes. Much of the transcriptional regulation of a cell is orchestrated by transcription factors that, by binding the DNA at proximal and distal regulatory sites, fine tune the amount of RNA produced.

In eukaryotic cells, DNA is packed within a nucleus through a hierarchical folding of 147 bp of DNA around a histone octamer to form nucleosomes, and their further compaction into chromatin.

Helena Raurell-Vila and Mireia Ramos-Rodríguez contributed equally to this work.

Tanya Vavouri and Miguel A. Peinado (eds.), *CpG Islands: Methods and Protocols*, Methods in Molecular Biology, vol. 1766, https://doi.org/10.1007/978-1-4939-7768-0_11,

Major insights into the epigenetic information encoded within the nucleoprotein structure of chromatin have come from high-throughput, genome-wide methods for assaying the accessibility of DNA to the machinery of gene expression also referred as chromatin "openness" [1, 2]. In fact, regulatory DNA often coincides with regions of remodeled chromatin, resulting in genomic sites open or accessible to transcription factor binding.

The use of techniques such as FAIRE and DNase I hypersensitive coupled with high-throughput sequencing, enabled the identification genome-wide of active transcription start sites, enhancers, and insulators in a wide variety of cell lines and tissue samples including the pancreatic islets [2–5]. Despite these successes, such methods employ multiple biochemical steps and substantial amounts of starting sample material. Recently, Buenrostro et al. reported on a robust and sensitive technology for profiling open chromatin by direct transposition of sequencing adapters into native chromatin [6] termed Assay of transposase Accessible Chromatin or "ATAC-seq". This novel assay takes advantage of a hyperactive Tn5 transposase that can simultaneously fragment and tag with sequencing adaptors genomic regions of accessible chromatin allowing an open chromatin library construction in a single enzymatic step.

Importantly ATAC-seq assay and library construction is achieved with considerably less starting material than conventional methods, thus being especially well-suited for the study of rare or difficult to obtain cell types, such as human pancreatic islet.

In this chapter we describe the implementation of the ATAC-seq protocol to chart open chromatin in isolated human pancreatic islets. We also provide a short technical description of the computational procedure to analyze ATAC-seq data. The experiments and analysis here described may be generally applicable to cell lines and primary tissue cellular aggregates.

2 Materials

2.1 Tissue Culture

1. Dithizone (diphenylthiocarbazone) stock solution: add 10 mg dithizone (Sigma) to 2 ml dimethylsulfoxide (DMSO) (Sigma). Store at −20 °C.
2. Islet medium: Ham's F10 medium supplemented with 10% FBS, 2 mM GlutaMAX, 50 U/ml penicillin and 50 μg/ml streptomycin (GIBCO), 6.1 mM glucose, 50 μM 3-isobutyl-1-methylxanthine, 1% BSA (Sigma).
3. Hanks' balanced salt solution (Sigma).
4. Phosphate buffered saline (PBS).

2.2 ATAC-Seq

1. Lysis buffer (10 mM Tris–HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Igepal CA-630). Store up to 1 week at +4 °C.
2. Nuclear staining: Methyl Green-Pyronin Staining (#HT70116, Sigma).
3. Insulin needle (29G×1/2″).
4. Tn5 transposase and TD buffer (Nextera DNA Library Prep Kit, #15028212, Illumina).
5. Clean up buffer (900 mM NaCl, 300 mM EDTA).
6. Sodium Dodecyl Sulfate (SDS) 20% Solution (Omnipur).
7. Proteinase K (Thermo Scientific).
8. SPRI beads cleanup (Agencourt AMPure XP—5 ml, #A63880, Beckman Coulter).
9. 25 μM PCR Primer 1 (sequences provided in Buenrostro et al. [6]).
10. 25 μM Barcoded PCR Primer 2 (sequences provided in Buenrostro et al. [6]).
11. NEBNext High-Fidelity 2× PCR Master Mix (#M0541, New England Biolabs).
12. SYBR Green (Roche).
13. MinElute PCR Purification Kit (Qiagen).
14. Qubit® dsDNA HS Assay (#Q32851, Invitrogen).

3 Methods

The ATAC-seq method here described is based on that developed by Buenrostro et al. [6] and further modified by other authors [7, 8]. These modifications are aimed to improve the technique efficiency in purified primary human islets and may be used to enhance this assay in other cell lines and primary tissue cellular aggregates.

A successful ATAC-seq experiment relies on the preservation of the native chromatin architecture and the original nucleosome distribution patterns [7]. Thus, ATAC-seq is most efficient when applied to freshly isolated nuclei, in this case freshly human pancreatic endocrine islets, while fixed or frozen cells may reduce the sensitivity of the methodology [8]. Human pancreatic islets are obtained from organ donors after a laborious isolation procedure involving collagenase digestion and gradient purification to separate endocrine pancreatic islets from the exocrine tissue [9, 10]. Human pancreatic islets are then cultivated for 48 h to recover from the stress and minimize the environment induced variation. Freshly isolated pancreatic islets are tissue fragments of typically ~500–2000 cells that can be cultured in suspension.

This protocol is optimized for 50,000 cells corresponding to approximately 50 human pancreatic islets. The number of starting cells to be processed is crucial, as the transposase-to-cell ratio determines the distribution of DNA fragments generated [6]. A correct proportion between the quantity of Tn5 transposase and the number of cells is clue to a successful ATAC-seq experiment.

3.1 Tissue Culture

Prior to culture, the ratio of islets to exocrine tissue is ascertained by dithizone staining. Dithizone binds to zinc ions present in the secretory granules of the β-cells allowing to differentiate them from the acinar contaminants, thus to estimate islet size and purity.

Collect a representative volume of the sample, containing a minimum of 50 islets, to a 1.5 ml eppendorf tube. Let the islet sediment, remove the supernatant and resuspend with 100 μl of culture medium. Add 900 μl of dithizone solution and incubate for 15 min at 37 °C. Dithizone solution is obtained by mixing 200 μl dithizone stock solution (*see* Subheading 2) to 800 μl Hanks' balanced salt solution and centrifuging at maximum speed at room temperature for 5 min. After rinsing the islets with phosphate buffered saline twice, stained islets can be counted under a microscope. The sample purity is extrapolated by calculating the ratio of stained islets to the total tissue aggregates.

In a primary tissue culture hood, transfer the islets to 50 ml falcon tubes and to spin the sample for 1 min at 100 × *g* at room temperature. Discard the supernatant and resuspend, in 75 ml flasks, ~5 ml islet medium (*see* Subheading 2) every 1000 islets and, place in an incubator at 37 °C for 48 h.

3.2 Islet Preparation

Preheat a thermo-block to +37 °C, prepare 5 ml of cold PBS and a table-top centrifuge at +4 °C.

Carefully hand pick 50 healthy and acinar-free islets corresponding approximately to 50,000 cells and transfer them to a 1.5 ml eppendorf tube and keep on ice (*see* Subheading 5, **Note 1**).

All steps are to be done swiftly, to prevent cell stress and disturbance of chromatin.

Centrifuge the islets aliquot for 1 min at 100 × *g*, +4 °C, remove the medium and rinse with 500 μl of ice-cold PBS (1 min, 100 × *g*, +4 °C), remove all supernatant.

3.3 Nuclei Preparation

Add to the islets pellet 300 μl cold lysis buffer (*see* Subheading 2), resuspend, by gentle pipetting, and incubate 25 min on ice. In order to optimize the nuclei isolation, while incubating on ice, resuspend after 5 and 15 min the lysed islets using a syringe with a 29G needle.

During the lysis reaction, check the number and integrity of the nuclei obtained by instant nucleus staining (*see* Subheading 2) 5% of the sample aliquot.

Spin down the nuclei for 15 min at 500 × *g*, +4 °C using a swing rotor with low acceleration and brake settings. Carefully discard the supernatant (the pellet should be translucent).

Gently wash the pellet in 100 μl of lysis buffer and centrifuge 15 min at 500 × *g*, +4 °C using a swing rotor with low acceleration and brake settings and carefully to remove the supernatant.

While washing, prepare the transposase reaction mix described in the next step and keep at room temperature.

3.4 Transposition DNA Purification

Resuspend nuclei pellet in a 25 μl transposase reaction mix containing 2 μl of Tn5 transposase, 12.5 μl of TD buffer (*see* Subheading 2) and 10.5 μl DEPC treated water, per reaction and incubate at 37 °C for 1 h. Gentle mixing may increase fragment yield.

Add 5 μl of clean up buffer (*see* Subheading 2), 2 μl of 5% SDS and 2 μl of Proteinase K and incubate for 30 min at 40 °C.

3.5 DNA Purification

Right after the transposition reaction, place the sample on ice and proceed immediately with the isolation of the tagmented DNA using 2× SPRI beads cleanup (*see* Subheading 2), following kit instructions. Elute in 20 μl DEPC treated water. The isolated DNA samples can be stored at −20 °C before library amplification or used immediately for the next steps.

3.6 Library Amplification and Purification

Two sequential 9-cycle PCR are performed in order to enrich for small tagmented DNA fragments.

Prepare the PCR mix consisting of 2 μl of PCR Primer 1 (25 μM working stock), 2 μl of Barcoded PCR Primer 2 (25 μM working stock), 25 μl of NeBNext High-Fidelity 2× PCR Master Mix, 1μl of DEPC treated water and 20 μl of the eluted sample (add DEPC water to compensate in case the volume of eluted DNA is less than 20 μl).

Amplify the library in a thermocycler using the following program (leave preheated to 72 °C): 72 °C for 5 min; 98 °C for 30 s; 9 cycles of 98 °C for 10 s, 63 °C for 30 s; and 72 °C for 1 min; and at 4 °C hold. After the first PCR round, select for fragments smaller than 600 bp using SPRI cleanup beads (*see* Subheading 2).

Perform a second PCR applying the same conditions in order to obtain the final library.

Finally purify the DNA library using the MinElute PCR Purification Kit, following kit instructions and eluting 2 × 10μl with the elution buffer. The purified libraries can then be stored at −20 °C.

3.7 Library Quality Control and Quantification

Accurate assessment of library quality and concentration is critical to successful sequencing and data analysis. A sample aliquot of 1.2 μl is used to check, by TapeStation or Bioanalyzer, the library quality and the distribution of the fragment size. The library obtained from a successfully tagmented sample will display a nucleosomal pattern (Fig. 1) (*see* Subheading 5, **Note 1**).

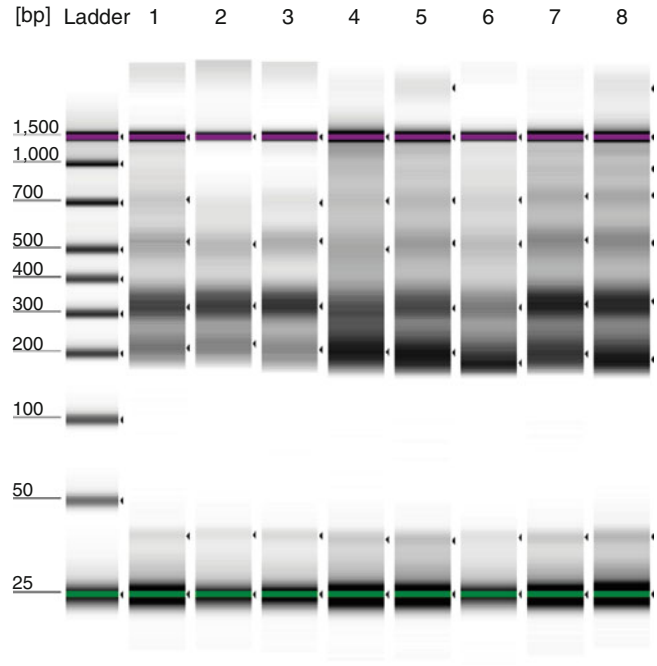

Fig. 1 Agilent TapeStation profiles showing the laddering pattern of ATAC-seq libraries. The band sizes correspond to the expected nucleosomal pattern obtained by chromatin tagmentation of human islets samples (lanes 1–8)

Fragments lower than 100 bp may represent mitochondrial fragments not properly tagged for sequencing, thus should not be included in the calculation of the molar concentration of the library. The library concentration can be measured by QuBit dsDNA HS Assay Kit following the instructions provided with the kit. As in chromatin immunoprecipitation experiments, semiquantitative PCR assays that target specific genomic sites may be carried on to estimate the efficiency of the ATAC-seq experiment in enriching for open chromatin regions. To this end we designed oligonucleotides targeting presumably open chromatin sites as well as negative control sites not expected to harbor open chromatin.

3.8 Sequencing

A minimum DNA concentration of 3.25 ng/μl in a minimum volume of 15 μl is preferred for next generation sequencing. Typically 100 million 50 bp long single or paired-end reads per library are sufficient to chart open chromatin regions genome-wide with a good enrichment over background resolution. Data yield is impacted by a fraction of mitochondrial reads that we report, for this protocol in human pancreatic islets, varying between 20% and 40%.

4 ATAC-Seq Data Analysis

The sequencing data generated is subsequently analyzed using a variety of analytical tools that allow resolving the ATAC-seq genome wide open chromatin profile. Data analysis may be carried out by applying computational procedures analogous to those applied for chromatin immunoprecipation sequencing experiments (ChIP-seq). While data analysis may require computational expertise and progressive increase of computing power and storage capacity, several ATAC-seq and ChIP-seq pipeline have been developed for researchers with limited computational experience (https://usegalaxy.org or http://cistrome.org/ap/root).

We here briefly describe the basic steps of a computational analysis addressed to resolve the ATAC-seq chromatin accessibility genome wide.

4.1 Data Quality Control and Alignment

Quality examination of the raw sequencing data is essential prior to data analysis in order to rule out sources of error and sequencing biases.

FASTQC [11] is a quality control tool which uses the raw FASTQ files to perform several simple quality control analyses. This program returns an HTML file comprehensive of a detailed analysis of the raw sequenced reads including the Phread score (per base sequencing error probability) GC content and overrepresented sequences or kmer content. Notice that for ATAC-seq experiments a "per base sequence content" warning may be due to a preference for the transposase binding [12, 13] that does not affect the quality of the experiment.

After data quality assessment sequenced reads are aligned to a reference genome. There are several mapping programs available including Maq [14], RMAP [15], Cloudburst [16], GEM [17], SHRiMP [18], BWA [19], and Bowtie [20].

Bowtie2 [21] is optimized to work with large reference genomes, such as mammalian, and reads from 50 bp up to 1 kb. This aligner software characterizes the degree of confidence of each alignment as nonnegative integer $Q = -10\ log10\ p$, where p is an estimate of the probability that the alignment does not correspond to the read's true point of origin. The "Q" value can be used, postalignment, to filter out reads with a poor mapping quality. While some reads might be mapping to several genomic locations at the same time, with default settings, bowtie2 searches for distinct valid alignments for each read and will report only the best alignment.

In order to align ATAC-seq single end raw reads bowtie2 may be run via command line. Default options can be used and the following useful flags can be specified: "*-t*" for printing to the "stderr", the time taken to load the indexes and align the reads

("*2>*" can be used to keep the statistics of the alignment for posterior usage); "*-S*" to return a SAM format file output; "*-p*" to use a specified number of cores.

4.2 Postalignment Processing

Prior to downstream data analysis several steps are needed to convert file formats and reduce the impact of potential artifact reads.

The typical output of an alignment software is a SAM or a BAM format file, the latter being a binary compressed format for storing sequence data. Conversion between these two file formats and sorting by genomic coordinates, necessary for downstream analyses, can be easily performed by using the SAMtools [22] programs suite.

Reads aligning to mitochondrial genome are discarded as unrelated to the scope of the experiment. Reads must also be filtered to remove overrepresented areas of the genome due to technical bias. A collection of signal artifact blacklist regions in the human genome is provided by the ENCODE project (ENCODE blacklist: https://personal.broadinstitute.org/anshul/projects/encode/rawdata/blacklists/) and can be used for this purpose. Reads filtering can be performed with SAMtools or Picard tools (http://broadinstitute.github.io/picard). Additionally, tools such as Picard allow easily tagging and eventually removing duplicates reads that may represent PCR amplification artifacts (*see* Subheading 5, **Note 2** and **3**).

A number of publicly available genome browser tools, can be used to visualize ATAC-seq profiles relative to available annotation tracks. Such tools include Artemis [23], EagleView [24], MapView [25], Apollo [26], and the Islet Regulome Browser [27]. The University of California Santa Cruz (UCSC) [28] and the Integrative Genomics Viewer (IGV) [29] are the most widely used genome browser tools. Density profile tracks can be loaded to the UCSC browser (Fig. 2) in BigWig format, indexed binary files which associate a genomic location with number of aligning reads. Bedtools utilities [30] and UCSC provide a series of tools to convert different file formats to bedgraph and BigWig.

Finally, identification of those regions that are significantly enriched of mapped reads compared to the background allows charting genome wide the predicted open chromatin sites. Several peak calling algorithms are available such as ZINBA [31], F-seq [32], HOMER [33], or MACS [34]. Differently from ChIP-seq experiments, ATAC-seq cannot rely on input DNA or mock-IP as control background to identify the reads enriched regions. The peak caller MACS2 [34] allows to fine tune the algorithm to the ATAC-seq experiment, typical parameter added to the default arguments are (*–no model --shift – 100 --exsize-200)* that allow to center the peaks to the Tn5 cutting sites.

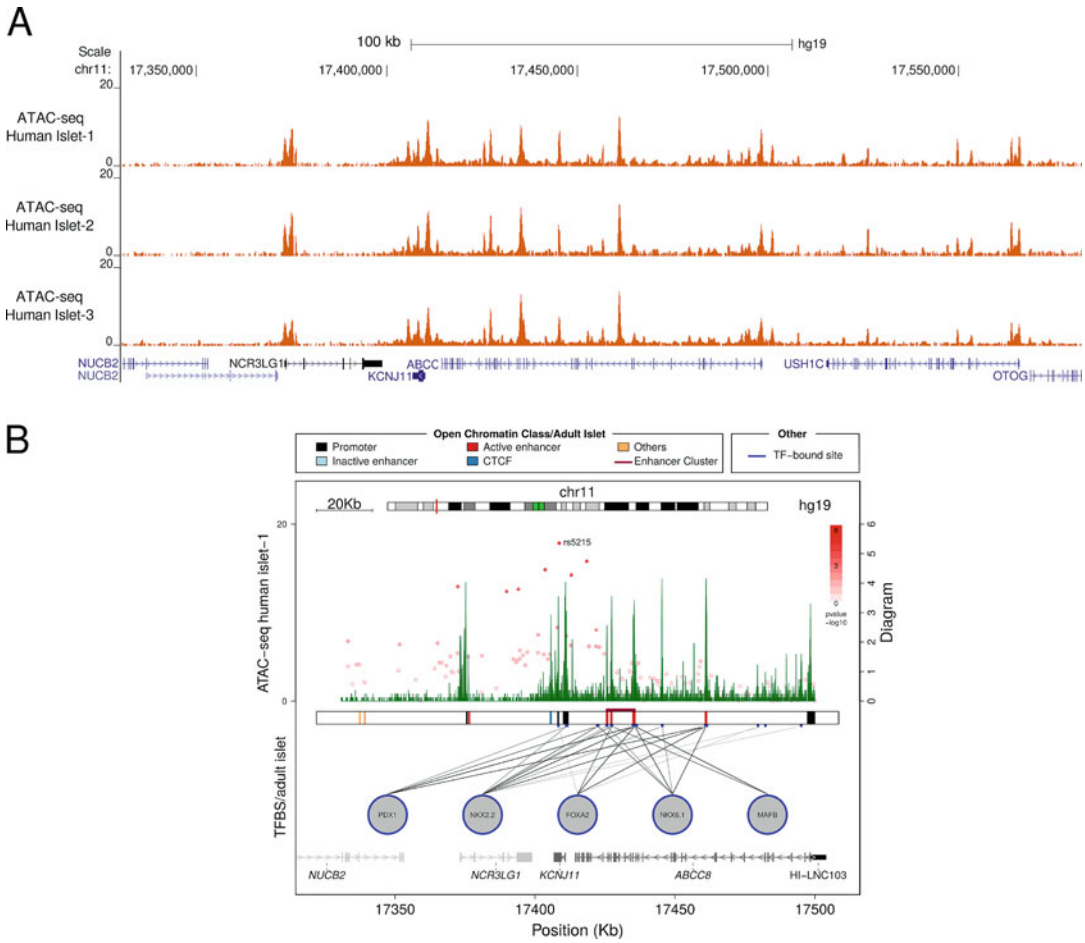

Fig. 2 Genome browser views of ATAC-seq signal in the proximity of the ATP-sensitive potassium ion channels genes, *ABCC8* and *KCNJ11*. (**a**) UCSC genome browser view of 3 human islets ATAC-seq libraries. ATAC-seq tracks display high concordance and a strong resolution over background of the signal. Notice that each ATAC-seq library was prepared by processing only ~50 human pancreatic islets. (**b**) Islet Regulome Browser [27] integrative view (www.isletregulome.com) showing that ATAC-seq open chromatin enrichment are often found to coincide with regulatory elements, promoters and transcription factor binding sites previously identified in human pancreatic islets. The plot also depicts the signal of Type 2 Diabetes Genome Wide Association Studies (GWAS), each red dot represents a genomic variant, being the color intensity of the dot proportional to -Log *p*-value of association, as indicated on the side of the plot

5 Notes

1. Variations in the number of cells processed may result in "under or over transposition" of the DNA. Such variations will be reflected in the library fragment size distribution measured by TapeStation or Bioalnalyzer. An overrepresentation of large fragments may mirror an "under tagmentation" of the DNA due to an excess of cells relative to the transposase. For this reason we recommend hand picking approximately

50 fragments of the islet prep, taking in account the purity of the sample being processed.

2. Identical reads can arise from independent transposition events or PCR amplification of a single fragment. In order to reduce experimental noise we thus recommend removing identical reads. Alternatively a solution proposed for ChIP-seq experiments is that of retaining a fixed number of tags per genomic location according to the sequencing depth [35].
3. In case of using ATAC-seq data to infer the transcription factor binding by footprint analysis [36], we recommend to adjust the read start sites to the center of the transposon's binding event. This can be done by offsetting the reads 4 bp on the + strand and 5 bp on the – strand since Tn5 transposase was shown to bind as a dimer inserting two adaptors separated by 9 bp [12].

Acknowledgment

This work was supported by a grant from the Spanish Ministry of Economy and Competiveness (BFU2014-58150-R), the Spanish Diabetes Society and Fundació La Marató de TV3. LP is a recipient of a Ramon y Cajal contract from the Spanish Ministry of Economy and Competitiveness (RYC 2013-12864). Helena Raurell-Vila and Mireia Ramos-Rodríguez contributed equally to this work.

References

1. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutyavin T, Lajoie B, Lee BK, Lee K, London D, Lotakis D, Neph S, Neri F, Nguyen ED, Qu H, Reynolds AP, Roach V, Safi A, Sanchez ME, Sanyal A, Shafer A, Simon JM, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Lenhard B, Tewari M, Dorschner MO, Hansen RS, Navas PA, Stamatoyannopoulos G, Iyer VR, Lieb JD, Sunyaev SR, Akey JM, Sabo PJ, Kaul R, Furey TS, Dekker J, Crawford GE, Stamatoyannopoulos JA (2012) The accessible chromatin landscape of the human genome. Nature 489(7414):75–82. https://doi.org/10.1038/nature11232
2. Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, Panhuis TM, Mieczkowski P, Secchi A, Bosco D, Berney T, Montanya E, Mohlke KL, Lieb JD, Ferrer J (2010) A map of open chromatin in human pancreatic islets. Nat Genet 42(3):255–259. https://doi.org/10.1038/ng.530
3. Pasquali L, Gaulton KJ, Rodriguez-Segui S, Mularoni L, Miguel-Escalada I, Akerman I, Tena JJ, Moran I, Gómez-Marín C, van de Bunt M, Ponsa-Cobas J, Castro N, Nammo T, Cebola I, Garcia-Hurtado J, Maestro MA, Pattou F, Piemonti L, Berney T, Gloyn AL, Ravassard P, Muller F, McCarthy MI, Ferrer J (2014) Pancreatic islet enhancer clusters enriched in type 2 diabetes risk–associated variants. Nat Genet 46(2):136–143. https://doi.org/10.1038/ng.2870
4. Stitzel ML, Sethupathy P, Pearson DS, Chines PS, Song L, Erdos MR, Welch R, Parker SC, Boyle AP, Scott LJ, Margulies EH, Boehnke M, Furey TS, Crawford GE, Collins FS (2010) Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. Cell Metab 12(5):443–455. https://doi.org/10.1016/j.cmet.2010.09.012
5. Nammo T, Rodriguez-Segui SA, Ferrer J (2011) Mapping open chromatin with

formaldehyde-assisted isolation of regulatory elements. Methods Mol Biol 791:287–296. https://doi.org/10.1007/978-1-61779-316-5_21

6. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 10(12):1213–1218. https://doi.org/10.1038/nmeth.2688
7. Lavin Y, Winter D, Blecher-Gonen R, David E, Keren-Shaul H, Merad M, Jung S, Amit I (2014) Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. Cell 159(6):1312–1326. https://doi.org/10.1016/j.cell.2014.11.018
8. Milani P, Escalante-Chong R, Shelley BC, Patel-Murray NL, Xin X, Adam M, Mandefro B, Sareen D, Svendsen CN, Fraenkel E (2016) Cell freezing protocol suitable for ATAC-Seq on motor neurons derived from human induced pluripotent stem cells. Sci Rep 6:25474. https://doi.org/10.1038/srep25474
9. Shapiro AM, Pokrywczynska M, Ricordi C (2016) Clinical pancreatic islet transplantation. Nat Rev Endocrinol. https://doi.org/10.1038/nrendo.2016.178
10. Piemonti L, Pileggi A (2013) A 25 years of the Ricordi automated method for islet isolation. CellR 4(1):e128
11. Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc
12. Adey A, Morrison HG, Asan XX, Kitzman JO, Turner EH, Stackhouse B, MacKenzie AP, Caruccio NC, Zhang X, Shendure J (2010) Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol 11(12):R119. https://doi.org/10.1186/gb-2010-11-12-r119
13. Goryshin IY, Miller JA, Kil YV, Lanzov VA, Reznikoff WS (1998) Tn5/IS50 target recognition. Proc Natl Acad Sci U S A 95 (18):10716–10721
14. Li H, Ruan J, Durbin R (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res 18 (11):1851–1858. https://doi.org/10.1101/gr.078212.108
15. Smith AD, Chung WY, Hodges E, Kendall J, Hannon G, Hicks J, Xuan Z, Zhang MQ (2009) Updates to the RMAP short-read mapping software. Bioinformatics 25 (21):2841–2842. https://doi.org/10.1093/bioinformatics/btp533
16. Schatz MC (2009) CloudBurst: highly sensitive read mapping with MapReduce. Bioinformatics 25(11):1363–1369. https://doi.org/10.1093/bioinformatics/btp236
17. Marco-Sola S, Sammeth M, Guigo R, Ribeca P (2012) The GEM mapper: fast, accurate and versatile alignment by filtration. Nat Methods 9(12):1185–1188. https://doi.org/10.1038/nmeth.2221
18. Rumble SM, Lacroute P, Dalca AV, Fiume M, Sidow A, Brudno M (2009) SHRiMP: accurate mapping of short color-space reads. PLoS Comput Biol 5(5):e1000386. https://doi.org/10.1371/journal.pcbi.1000386
19. Li H, Durbin R (2009) Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics 25(14):1754–1760. https://doi.org/10.1093/bioinformatics/btp324
20. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10(3):R25. https://doi.org/10.1186/gb-2009-10-3-r25
21. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with bowtie 2. Nat Methods 9 (4):357–359. https://doi.org/10.1038/nmeth.1923
22. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25 (16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352
23. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA (2012) Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 28(4):464–469. https://doi.org/10.1093/bioinformatics/btr703
24. Huang W, Marth G (2008) EagleView: a genome assembly viewer for next-generation sequencing technologies. Genome Res 18 (9):1538–1543. https://doi.org/10.1101/gr.076067.108
25. Wolfsberg TG (2011) Using the NCBI Map Viewer to browse genomic sequence data. Curr Protoc Hum Genet Chapter 18:Unit18 15. doi:https://doi.org/10.1002/0471142905.hg1805s69
26. Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes IH, Elsik CG, Lewis SE (2013) Web Apollo: a web-based genomic annotation editing

platform. Genome Biol 14(8):R93. https://doi.org/10.1186/gb-2013-14-8-r93

27. Mularoni L, Ramos-Rodríguez M, Pasquali L (2017) The pancreatic islet regulome browser. Front Genet 8(13). https://doi.org/10.3389/fgene.2017.00013
28. Speir ML, Zweig AS, Rosenbloom KR, Raney BJ, Paten B, Nejad P, Lee BT, Learned K, Karolchik D, Hinrichs AS, Heitner S, Harte RA, Haeussler M, Guruvadoo L, Fujita PA, Eisenhart C, Diekhans M, Clawson H, Casper J, Barber GP, Haussler D, Kuhn RM, Kent WJ (2016) The UCSC genome browser database: 2016 update. Nucleic Acids Res 44 (D1):D717–D725. https://doi.org/10.1093/nar/gkv1275
29. Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14 (2):178–192. https://doi.org/10.1093/bib/bbs017
30. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(6):841–842. https://doi.org/10.1093/bioinformatics/btq033
31. Rashid NU, Giresi PG, Ibrahim JG, Sun W, Lieb JD (2011) ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions. Genome Biol 12 (7):R67. https://doi.org/10.1186/gb-2011-12-7-r67
32. Boyle AP, Guinney J, Crawford GE, Furey TS (2008) F-Seq: a feature density estimator for high-throughput sequence tags. Bioinformatics 24(21):2537–2538. https://doi.org/10.1093/bioinformatics/btn480
33. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38 (4):576–589. https://doi.org/10.1016/j.molcel.2010.05.004
34. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9(9):R137. https://doi.org/10.1186/gb-2008-9-9-r137
35. Chen Y, Negre N, Li Q, Mieczkowska JO, Slattery M, Liu T, Zhang Y, Kim TK, He HH, Zieba J, Ruan Y, Bickel PJ, Myers RM, Wold BJ, White KP, Lieb JD, Liu XS (2012) Systematic evaluation of factors influencing ChIP-seq fidelity. Nat Methods 9 (6):609–614. https://doi.org/10.1038/nmeth.1985
36. Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, John S, Sandstrom R, Johnson AK, Maurano MT, Humbert R, Rynes E, Wang H, Vong S, Lee K, Bates D, Diegel M, Roach V, Dunn D, Neri J, Schafer A, Hansen RS, Kutyavin T, Giste E, Weaver M, Canfield T, Sabo P, Zhang M, Balasundaram G, Byron R, MacCoss MJ, Akey JM, Bender MA, Groudine M, Kaul R, Stamatoyannopoulos JA (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. Nature 489 (7414):83–90. https://doi.org/10.1038/nature11212