Proceedings of the International Conference on Modern Medicine and Global Health (ICMMGH 2023)
DOI: 10.54254/2753-8818/6/20230267

ATAC-seq: A powerful tool for investigating chromatin accessibility and transcription factor binding

Xilin Wang[1,2]

[1]Shenyang Agricultural University, No. 120 Dongling Road, Dongling District, Shenyang, Liaoning, LN, CHN 110866.

[2]renxing@syau.edu.cn

Abstract. Chromatin Transposase Accessibility Sequencing is a new high-throughput sequencing technique developed by Professor William Greenleaf in 2013 which uses DNA transposase to probe chromatin accessibility with Tn5 transposase. This technique, which is simpler and more sensitive than DNase-seq, MNase-seq and FAIRE-seq and requires fewer cells, has been used to study chromatin accessibility using Tn5 transposase. ATAC-seq is important for the study of epigenetic molecular mechanisms because it can map chromatin accessibility on a genome-wide scale, compare open chromatin regions in different tumour samples, compare differences in transcription factor binding between treatments, reveal nucleosome localisation information and transcription factor binding sites, and can be used to locate specific unknown transcription factors, which can be used in combination with other methods to screen for specific transcription factors of interest. It is possible to combine this approach with others to investigate specific regulatory factors. Herein, ATAC-seq is systemically profiled to present that ATAC-seq has enormous potential to drive future discoveries in the field of genomics and beyond.

Keywords: ATAC-seq, Tn5 transposase, chromatin accessibility, epigenetic molecular mechanisms, genome-wide scale.

1. Introduction

Eukaryotic chromosomes are made up of bead-like nucleosomes that are folded and coiled in a helical pattern [1], with the DNA and histones wrapped around the nucleosomes[2,3]. Within chromosomes, most of the highly ordered packaging of DNA is not open, and only open DNA can undergo replication, transcription and translation for gene expression [4]. During transcription, some of the higher-order structure of the DNA is unwound, leaving the region accessible to transcription factors and other regulatory elements that can bind to it and facilitate transcription [5]. Accessibility of chromatin is associated with spatial and temporal conditions, i.e. the state of chromatin varies from site to site [6]. This may be due to gene silencing regions where chromatin is tightly packed or active transcriptional regions where genes are loosely packed [7].

In the 1940s, the American geneticist Barbara Mc Clintock discovered that transposons, which occur in large numbers in the genomes of eukaryotic organisms and represent a DNA sequence that can alter its position in the genome, can be divided into type I transposons, i.e. RNA transposons, which can be copied and pasted, and type II transposons, i.e. DNA transposons, which can be directly

DOI: 10.54254/2753-8818/6/20230267

"cut" [8]. The type II cut-and-paste transposon, also known as a DNA transposon [9], cuts out its own DNA sequence and inserts it into a new position without increasing the number of copies. ATAC-seq [10], on the other hand, uses a DNA transposase, the Tn5 transposase [11], which, when combined with open chromatin regions, enables the transfer of specific DNA sequences across regions of the chromosome where the higher spatial structure of the DNA would otherwise prevent the process. The Tn5 transposase, a specific transposable complex carrying a known DNA sequence tag, is added to the nucleus and the Tn5 transposase inserts the sequencing junction directly into the open chromatin region, fragments the DNA, end-repairs the broken product and uses the sequence-specific tag to build a library for second-generation, high-throughput sequencing to obtain genome-wide chromatin accessibility information. The genome is interrogated in exposed active transcribed DNA regions and tightly bound heterochromatin regions.

Compared to traditional methods such as DNase-seq [12] and FAIRE-seq [13], ATAC-seq has several advantages. Firstly, ATAC-seq eliminates the need to precisely control the amount of enzyme required for DNase-seq and the need to determine the duration of formaldehyde crosslinking for FAIRE-seq [14]. Second, ATAC-seq requires only a small sample size of 500 to 50,000 cells for library preparation and takes only three hours, as opposed to the two techniques that require two to three days, making it much simpler and faster [15]. Thirdly, ATAC-seq provides higher resolution and sensitivity for the detection of open chromatin regions, allowing the identification of smaller regulatory elements such as enhancers and transcri`ption factor binding sites [16]. Fourth, ATAC-seq is more suitable for studying rare or low abundance cell populations because it is less susceptible to background noise and non-specific signals [17]. Together, these advantages make ATAC-seq a highly valuable tool for the study of gene regulation [18], chromatin accessibility [19] and epigenetic modifications [20] in a wide range of biological systems.

Over the past few years, ATAC-seq has emerged as a powerful technique for the study of chromatin accessibility and gene regulation, providing insights into the mechanisms that underlie these processes. Through identification of open chromatin regions and regulatory elements, ATAC-seq can have numerous applications in basic and translational research, as well as potentially in clinics. However, as with any technique, it is limited in how far it goes. Here, ATAC-seq is reviewed, including its basic workflow, limitations and applications. We also discuss its potential to drive future discoveries in genomics and beyond, as well as its limitations and future developments.

2. Basic workflow of ATAC-seq

The first step in the performance of ATAC-seq is the preparation of cells and the extraction of nuclei [21]. This is done by centrifuging or scraping the cell to the appropriate density. Cells are then resuspended in hypotonic cold lysis buffer to lyse, swell and release nuclei for the subsequent Tn5 transposase reaction and DNA fragmentation. In summary, this step is essential for the generation of high-quality ATAC-seq data.

Having obtained the nuclei, the next step in ATAC-seq is to transpose and purify the nuclei immediately after extraction by resuspending the precipitate in a Tn5 transposase reaction system, followed by purifying the sample using the Qiagen MinElute PCR Purification Kit [22]. The Tn5 transposase cleaves the DNA, resulting in short, fragment-sized DNA fragments, which can be used for high throughput sequencing [23]. The Tn5 transposase reaction is a key step in ATAC-seq, which fragments and labels accessible chromatin regions of the genome through the identification and cleavage of chromatin in open regions of the chromatin structure. A transposase reaction buffer containing Tn5 transposase and reaction components is used to resuspend nuclei or chromatin. The reaction is stopped by adding EDTA or a similar chelating agent after a period of incubation at 37°C, usually between 30 minutes and 1 hour. To remove any residual reaction components or contaminants, the resulting DNA fragment is then purified using an adsorbent column or magnetic beads.

After purification, the DNA fragments are PCR amplified with primers targeting the sequencing aptamer. The resulting sequence can be compared with a reference genome and analysed to identify regions of open chromatin structure and regulatory elements. To reduce the effects of fragment size

DOI: 10.54254/2753-8818/6/20230267

and GC bias in PCR, qPCR is used to determine the number of subsequent PCR cycles and to terminate amplification before saturation, ensuring that the post-transposition fragment size is in the range of 40 bp to 1 kb without size selection and maintaining high library complexity.As the transposition process creates a 9bp gap, the first step of PCR requires a 5 min reaction at 72°C to fill the gap, and the PCR enzyme used is a non-hotstart, strand-switching enzyme.

High-throughput sequencing of the library using an appropriate sequencing platform and read length is the final step in ATAC-seq. To identify regions of open chromatin structure and regulatory elements, the resulting sequencing reads are analysed against a reference genome. Post-sequencing quality control and pre-processing of the resulting reads to remove low-quality reads and sequencing artefacts must be considered. The remaining reads are then aligned to the reference genome using an appropriate alignment algorithm. The resulting alignment file is then subjected to peak calling and downstream analysis. Peakcalling identifies open chromatin regions and regulatory elements based on the read distribution across the genomic space. After peak call, annotation is performed against genome databases and functional annotation tools to determine putative target genes and regulatory elements associated with accessible chromatin regions. The resulting data is then visualised to help understand how the genome gets regulated.

3. The limitations and recent developments of ATAC-seq

The Tn5 transposase is used to insert a cut DNA fragment and to join the sequencing junction at the ends of the two cut DNA fragments [24], so that for a given DNA fragment the junction is random, resulting in a 50% probability that the junctions at the ends of the same fragment will be the same. Since the enrichment amplification and sequencing are performed on fragments with different cleavage sites, half of the fragments cannot be used; a large number of cut DNA fragments are too large for PCR enrichment; the activity of the Tn5 transposase is affected by the reaction conditions and the composition of the solution, which still need to be optimized to improve the cutting effect; difficulties encountered with ATAC-seq include the difficulty of obtaining nuclei, especially in plant cells due to the presence of cell walls. Cell walls in plant cells and contamination from organelles like mitochondria and chloroplasts in plant cells may result in high levels of invalid sequencing data, increasing experimental costs and time. Due to differences between tissues of different species, there is no universal lysis cell system. For example, the FANS-ATAC-seq (fluorescence-activated nucleus sorting, FANS) system developed by Lu et al. [25], the INTACT system developed by Roger et al. [26] and sucrose precipitation with similar results to INTACT ensure that high quality intact nuclei are used in assays with minimal contamination of organelle DNA. The general availability of ATAC-seq for difficult cell lines, rare primary cells and clinically relevant frozen tissues was improved with the invention of Omni-ATAC by Corces et al. [27] In addition, Sos et al. developed the THS-seq technique using a new Tn5 hypermutant (Tn5059) with higher activity than the conventional Tn5 transposase and optimized reaction solutions and conditions to overcome the problems of random junctions and oversized fragments after shearing. The T7 promoter plus transcription primers, designed to replace adapters 1 and 2 in the original Tn5 transposon complex, are transcribed to produce single-stranded RNA, which is sequenced using the same principles as RNA sequencing to obtain cDNA and adapter for library construction. This technique greatly improves the efficiency of the transposon and results in more complete sequencing data by avoiding the need for random ligation of junctions. ATAC-seq has become the mainstay of chromatin accessibility analysis as the technology continues to improve.

4. Wide applications of ATAC-seq

Since its introduction in 2013, ATAC technology has been the subject of much research into its potential and applications. Due to its high sensitivity and ease of use, it has been widely used in epigenomic research. It has been used in RNA-seq, ChIP-seq (chromatin immunoprecipitation followed by high-throughput sequencing) and Hi-Seq to identify open chromatin regions, locate organising genes, predict trans-acting factors, search for potential key cis-regulatory elements and

DOI: 10.54254/2753-8818/6/20230267

transcription factors, and analyse complex regulatory networks in vivo. In addition to RNA-seq, ChIP-seq (chromatin immunoprecipitation followed by high-throughput sequence analysis) and Hi-C (high-throughput chromosomal conformation capture) were developed for in vivo chromosomal analysis, and ATAC-seq to investigate chromatin accessibility at the level of individual cells. ATAC-seq (single-cell assay for transposase-accessible chromatin, scATAC-seq) can be used to detect heterogeneity within and between cells.

4.1. Annotate functional DNA regulatory elements

ATAC-seq was used to study the open state of the binding regions of various regulators such as promoters, enhancers and silencers and to localize nucleosomes. Using ATAC-seq, Tan investigated the inverse relationship between longevity and cancer frequency in the genetic machinery of the naked mole-rat [28], indicating that the promoter regions of the reprogrammed genes in the naked mole-rat are mostly closed. Treatment with SV40 LT (large tantigen) antigen opened the promoter region and showed that the promoter region was more accessible, thereby repressing the Rb oncogene and increasing the efficiency of reprogramming; revealing a more stable epigenome in naked mole rat cells will provide new ideas for cancer prevention and treatment. In addition, transcription binds DNA sequences where transcription factors compete with nucleosomes, so transcription factor levels at the nucleosome-binding site will be lower. Quillien et al. performed ATAC-seq analysis of nuclei from Tg(fli1a: egfp)y1 transgenic embryonic endothelial cells and then used the FANS technique to label the nuclei with green fluorescent protein in order to identify specific enhancers in the entire zebrafish genome [29]. Nuclei were isolated from endothelial cells using the FANS technique to obtain high-quality ATAC-seq data, and long DNA fragments, i.e. nucleosome-binding regions, and short DNA fragments without nucleosomes were analysed for nucleosome localisation. The study reveals patterns of nucleosome localisation at transcription start sites across the genome and provides information on the dynamics of the transcription regulatory network formed by genes controlling gene expression during embryogenesis.

4.2. Combined multi-omics analysis of ATAC-seq with ChIP-seq, RNA-seq, etc.

It has been shown that the data generated by ATAC-seq can be cleverly combined with other epigenetic information to provide a live picture of the enhancers and to explain their role. Ackermann et al. first analysed open chromatin regions in human islets using ATAC-seq and RNA-seq data. The screen identified binding sites for known islet cell transcription factors and a single nucleotide polymorphism (SNP) for the identified susceptibility locus for type II diabetes in both types of cells. More importantly, the discovery that the 'group-specific protein', vitamin D binding protein, is only present in α-cells and that immunoreactivity for chondrolectin is only present in β-cells, is a novel signature gene for both cell types. In 2018, the mechanism of action of the anti-inflammatory factor IL-10 was revealed by Rajbhandari et al. using ATAC-seq, ChIP-seq and RNA-seq. Using ATAC-seq and RNA-seq, they investigated the potential mechanisms by which IL-10 inhibited adipocyte thermogenesis and energy output, and verified that IL-10 altered the chromatin opening state of adipocytes and reduced the expression of UCP1, a gene involved in thermogenesis. Finally, ChIP-seq was used to elucidate the mechanism by which IL-10 inhibited the recruitment of ATF and C/EBPβ to the enhancer region of thermogenic transcription factors, resulting in anti-inflammatory activity.

4.3. Sc ATAC-seq sequencing technology

Sc ATAC-seq can be used for the study of cellular heterogeneity, for the analysis of gene regulatory networks in subpopulations of cells, and for the epigenomics of single cells. Efficient access to the whole genome of a single cell is now achieved by microfluidics and ATAC-seq (single-cell combinatorial indexed ATAC-seq, sci ATAC-seq). Microfluidics is a low-cost, high-capacity, high-capture technique based on the principle of wrapping transposase-treated nuclear DNA with oil droplets containing specific barcode sequences to form GEMs (gel beads-in emulsion). sci ATAC-seq allows individual cells to be uniquely labelled by double-dilution labelling mixing, thus providing

DOI: 10.54254/2753-8818/6/20230267

chromatin opening information for a large number of single cells. The combination of sci ATAC-seq and RNA-seq allows the identification of cis-acting elements and transcription factors in various cells, the investigation of mechanisms and networks of gene expression regulation, and the analysis of tumour and haematopoietic heterogeneities [30]. In conclusion, we believe that the combination of Sc-seq with other techniques has a very promising future.

5. Conclusion

Identifying and pinpointing open chromatin regions is critical to epigenetic research. With the development of high-throughput sequencing technologies, ATAC-seq, the chromatin-open-region method, enables the systematic study of the binding sites of key regulatory elements, including promoters, enhancers, insulators and transcription factors, across the entire genome. Beyond tissue-specific gene discovery and nucleosome localisation, ATAC-seq chromatin opening information can be used to further integrate genomic, transcriptomic and methylation data for a more three-dimensional and intuitive understanding of complex gene interactions and phenotypic effects. Currently, ATAC-seq has shown unparalleled advantages and potential for capturing open regions of chromatin, and has become an important tool for the study of epigenetic regulation due to its simpler procedure and easier-to-use materials. It is undeniably a breakthrough technique for epigenetic research, although the tools for data analysis are not yet mature. With further improvements in experimental methods, ATAC-seq can be expected to become one of the tools for the genetic analysis of complex traits, thus contributing to the continuous development of life sciences in man, mouse and other animals and plants. ATAC-seq also has the potential for clinical applications, such as the identification of biomarkers for disease diagnosis or the prediction of therapeutic response.

References

[1] Kahn, T. G., Savitsky, M. and Schwartz, Y. B., “Topological screen identifies hundreds of Cp190- and CTCF-dependent Drosophila chromatin insulator elements,” Sci Adv. Papers 9(5), eade0090 (2023).

[2] Flury, V., Reverón-Gómez, N. and Groth, A., “Recycling of modified H2A-H2B provides short-term memory of chromatin states,” Cell Paper 186(5), 1050-1065.e19 (2023).

[3] Bruhn, C., Bastianello, G. and Foiani, M., “Cancer cell histone density links global histone acetylation, mitochondrial proteome and histone acetylase inhibitor sensitivity,” Commun Biol. Papers 5(1), 882 (2022).

[4] Bhushan, S., Singh, A. K., Thakur, Y. and Baskar, R., “Persistence of parental age effect on somatic mutation rates across generations in Arabidopsis,” BMC Plant Biol. Papers 23(1), 152 (2023).

[5] Kyung, J., Jeon, M. and Lee, I., “Recent advances in the chromatin-based mechanism of FLOWERING LOCUS C repression through autonomous pathway genes,” Front Plant Sci. Papers 13, 964931 (2022).

[6] Booth, G. T., Daza, R. M., Srivatsan, S. R., McFaline-Figueroa, J. L., Gladden, R. G., Furlan, S. N., Shendure, J. and Trapnell, C., “High-Capacity Sample Multiplexing for Single Cell Chromatin Accessibility Profiling,” bioRxiv, Papers 531201 (2023).

[7] Tang, Y., “Plant nuclear envelope as a hub connecting genome organization with regulation of gene expression,” Nucleus Papers 14(1), 2178201 (2023).

[8] Wright, S. E. and Todd, P. K., “Native functions of short tandem repeats,” Elife Papers 12, e84043 (2023).

[9] Gasparotto, E., Burattin, F. V. and Bodega, B., “Transposable Elements Co-Option in Genome Evolution and Gene Regulation,” Int J Mol Sci. Papers 24(3), 2610 (2023).

[10] Agoro, R., Nookaew, I., and White, K. E., “Single cell cortical bone transcriptomics define novel osteolineage gene sets altered in chronic kidney disease,” Front Endocrinol. Papers 14, 1063083 (2023).

DOI: 10.54254/2753-8818/6/20230267

[11] Penkov, D., Zubkova, E. and Parfyonova, Y., "Tn5 DNA Transposase in Multi-Omics Research," Methods Protoc. Papers 6(2), 24 (2023).
[12] Ishihara, T., Fenelon, J. C., Griffith, O. W., Ishiguro, K.-I. and Renfree, M. B., "Conserved H3K27me3-associated chromatin remodelling allows STRA8 but not MEIOSIN expression in mammalian germ cells," Reproduction Papers 165(5), 507–520 (2023).
[13] Lima, A. R. J., Elias, M. C. and da Cunha, J. P. C., "Open chromatin analysis in Trypanosoma cruzi life forms highlights critical differences in genomic compartments and developmental regulation at tDNA loci," Epigenetics Chromatin Papers 15(1), 22 (2022).
[14] Estève, P.-O., Vishnu, U. S., Chin, H. G. and Pradhan, S., "NicE-viewSeq: An Integrative Visualization and Genomics Method to Detect Accessible Chromatin in Fixed Cells," Methods Mol Biol. Papers 2611, 293–302 (2023).
[15] Qin, W., Xie, Y., Ren, Z., Xu, C., Sun, M.-A., Yin, Z. and Bao, W., "Integrative ATAC-seq and RNA-seq analyses of IPEC-J2 cells reveals porcine transcription and chromatin accessibility changes associated with Escherichia coli F18ac inhibited by Lactobacillus reuteri," Front Microbiol. Papers 14, 1101111 (2023).
[16] Henikoff, S., Henikoff, J. G., Kaya-Okur, H. S. and Ahmad, K., "Efficient chromatin accessibility mapping in situ by nucleosome-tethered tagmentation," Elife Papers 9, e63274 (2020).
[17] Martini, L., Bardini, R., Savino, A. and Di Carlo, S., "GAGAM v1.2: An Improvement on Peak Labeling and Genomic Annotated Gene Activity Matrix Construction," Genes Papers 14(1), 115 (2022).
[18] Lu, Z.-J., Ye, J.-G., Wang, D.-L., Li, M.-K., Zhang, Q.-K., Liu, Z., Huang, Y.-J., Pan, C.-N., Lin, Y.-H., Shi, Z.-X. and Zheng, Y.-F., "Integrative Single-Cell RNA-Seq and ATAC-Seq Analysis of Mouse Corneal Epithelial Cells," Invest Ophthalmol Vis Sci 64(3), 30 (2023).
[19] Jiao, Y., Long, Y., Xu, K., Zhao, F., Zhao, J., Li, S., Geng, S., Gao, W., Sun, P., Deng, X., Chen, Q., Li, C. and Qu, Y., "Weighted Gene Co-Expression Network Analysis Reveals Hub Genes for Fuzz Development in Gossypium hirsutum," Genes Papers 14(1), 208 (2023).
[20] Li, W., Tim Wu, U., Cheng, Y., Huang, Y., Mao, L., Sun, M., Qiu, C., Zhou, L. and Gao, L., "Epigenetic Application of ATAC-Seq Based on Tn5 Transposase Purification Technology," Genet Res. Papers 2022, 8429207 (2022).
[21] de Vos, J., Crooijmans, R. P. M. A., Derks, M. F. L., Kloet, S. L., Dibbits, B., Groenen, M. A. M. and Madsen, O., "Detailed molecular and epigenetic characterization of the pig IPEC-J2 and chicken SL-29 cell lines," iScience Papers 26(3), 106252 (2023).
[22] Dang, Y., Yadav, R. P. and Chen, X., "ATAC-See: A Tn5 Transposase-Mediated Assay for Detection of Chromatin Accessibility with Imaging," Methods Mol Biol. Papers 2611, 285–291 (2023).
[23] Crofts, T. S., McFarland, A. G. and Hartmann, E. M., "Mosaic Ends Tagmentation (METa) assembly for extremely efficient construction of functional metagenomic libraries," bioRxiv. Papers 429292 (2021).
[24] Zhang, Y., Tang, Y., Sun, Z., Jia, J., Fang, Y., Wan, X. and Fang, D., "Tn5 tagments and transposes oligos to single-stranded DNA for strand-specific RNA sequencing," Genome Res. Papers (2023).
[25] Marinov, G. K., Shipony, Z., Kundaje, A. and Greenleaf, W. J., "Genome-Wide Mapping of Active Regulatory Elements Using ATAC-seq," Methods Mol Biol. Papers 2611, 3–19 (2023).
[26] Raus, A. M., Fuller, T. D., Nelson, N. E., Valientes, D. A., Bayat, A. and Ivy, A. S., "Early-life exercise primes the murine neural epigenome to facilitate gene expression and hippocampal memory consolidation," Commun Biol. Papers 6(1), 18 (2023).
[27] Levinsky, A. J., McEdwards, G., Sethna, N. and Currie, M. A., "Targets of histone H3 lysine 9 methyltransferases," Front Cell Dev Biol. Papers 10, 1026406 (2022).

DOI: 10.54254/2753-8818/6/20230267

[28] Schraverus, H., Larondelle, Y. and Page, M. M., "Beyond the Lab: What We Can Learn about Cancer from Wild and Domestic Animals," Cancers Papers 14(24), 6177 (2022).

[29] Quillien, A., Abdalla, M., Yu, J., Ou, J., Zhu, L. J. and Lawson, N. D., "Robust Identification of Developmentally Active Endothelial Enhancers in Zebrafish Using FANS-Assisted ATAC-Seq," Cell Rep. Papers 20(3), 709–720 (2017).

[30] Lin, Y., Wu, T.-Y., Wan, S., Yang, J. Y. H., Wong, W. H. and Wang, Y. X. R., "scJoint integrates atlas-scale single-cell RNA-seq and ATAC-seq data with transfer learning," Nat Biotechnol. Papers 40(5), 703–710 (2022).