

2023

Assuit University

Bioinformatics department

CHROMATIN ACCESSIBILITY PROFILING BY ATAC- SEQ

CHROMATIN ACCESSIBILITY PROFILING BY ATAC-SEQ

#1 Sara Ashraf, **#2 Rana Hamed,** **#3 Rewaa Salah,** **#4 Amira Ahmed**
Bioinformatics dep. *Bioinformatics dep.* *Bioinformatics dep.* *Bioinformatics dep.*

Assuit University in Assuit city -Egypt

sara.20367681@compit.aun.edu.eg, Rana.20367670@compit.aun.edu.eg,

Rewaa.20367671@compit.aun.edu.eg, Amira.20367605@compit.aun.edu.eg

Abstract

ATAC-seq stands for assay for transposase accessible chromatin with high-throughput sequencing. This technique was first published in 2013 in the journal Nature Methods by lead researcher Jason Buenrostro in the labs of Howard Chang and William Greenleaf at Stanford University.

ATAC-seq Method gives researchers information about chromatin accessibility across the genome. This method can access DNA with the hyperactive transposase Tn5 used in tagmentation; the tagmentation process can be used to fragment genomic DNA. Tn5 is preloaded with adapters which inserts these adapters into open chromatin regions. It relies on next generation sequencing (NGS) library construction. Then the sequencing reads can be utilized to map nucleosome position, transcription factor binding areas, and regions of enhanced accessibility. With instructions for sample preparation and subsequent bioinformatics analysis, this protocol describes how to create and sequence ATAC-seq libraries.

All the previously existing techniques had several drawbacks, such as the requirement for large quantities of starting material, convoluted and time-consuming protocols, and the inability to simultaneously evaluate the three chromatin mechanisms discussed earlier together in one assay. Researchers developed an assay that was able to examine the whole epigenetic profile from a substantially smaller number of cells than the other techniques that will be covered later in an effort to get beyond the limits of the previous techniques.

Introduction

Understanding the biological system functionality is important, though researchers need to study well the cell alternations, specifically the DNA structure, which binds to the histones in a condense format to make the chromatin. In DNA there is some open regions found during the DNA replication and transcription processes, as known that the regulatory elements can bind to open regions and regulate DNA processes. In addition, epigenetic modifications of the chromatin can undergo dynamic, those modifications are: DNA methylation, chromatin remodeling, and histone modification, which by using traditional molecular techniques, they are difficult to detect completely.

Scientists need a way that can help discover through how can the chromatin structures and other elements affect the gene expression, so a method was developed to determine open chromatin regions across the genome called transposase accessible chromatin with high throughput sequencing (ATAC-seq). ATAC-seq is an innovative epigenetic technology, a method for mapping chromatin accessibility by probing Tn5 transposase to a DNA-Sequence at a genome-wide level, and then the open chromatin can be identified by using a library for sequencing the output fragments.

This survey view studies of some scientific research on the achievements of using ATAC seq and how it is linked to substantial advancements in biological research. According to recent studies, ATAC-seq has also been used to

refine map disease-associated genetic variants discovered through genome-wide association studies (GWASs), allowing the prediction of functional noncoding changes.

we define the landscape and dynamics of active regulatory DNA in cutaneous T cell lymphoma (CTCL) by ATAC-seq by identifying the transcription factor (TF) activation that drive leukemia regulates, which are We identify three dominant patterns of transcription factor (TF). By this way we can study personal regulates in human cancer and epigenetic therapy.

Assay of Transposase Accessible Chromatin with sequencing (ATAC-seq) is a recently introduced and sensitive method to map open chromatin sites, predict transcription factor binding, and determine nucleosome position from as few as 500 cell (ATAC-seq) enable us to track the epigenome state of patient. In this technique we can characterize chromatin dynamics in CTCL using ATAC-seq, and addressed the regulatory dynamics in leukemic epigenomes from CTCL patients treated with HDACi, Resulting in Landscape of DNA accessibility in normal CD4+, CTCL leukemia, and host T cells.

Age-related macular degeneration (AMD) is a significant disease cause of vision loss in the elderly, we profile chromatin accessibility using ATAC-Seq in the retina and retinal pigmented epithelium (RPE) from AMD and control patients, also a study of tissues of both male and female mice tissues helped to identify key transcription factors specific to distinct tissues using ATAC-seq. Non-small cell lung carcinoma (NSCLC) is a significant cancer type whose epigenetic alteration remains unclear. Whole-genome sequencing for each sample, performing ATAC-seq and other techniques, we need to find out how did ATAC-seq method helped wildly to analyze such big complex disease data.

Keywords

ATAC-seq, Chromatin accessibility, T cells, Age-related macular degeneration (AMD), Non-small cell lung carcinoma (NSCLC).

Background

The latest researches used to analyze such big data of open chromatin regions using the ChIP-seq Technique, this method depends on antibody-based enrichment, As it requires more input cells range(1-5 million)than ATAC-Seq that takes a small number of input cells range(50000-100000cells) so it takes less time in processing data unlike Chip-seq, which needs time for occupation /assay, ATACT-seq does not require a priori knowledge of the epigenetic marks or transcription, and can be integrated with RNA-seq, ChIP-seq, DNA methylation, and other techniques, not to mention the least cost and troubleshooting, this method shows all the active loci(promoters and enhancers) reveal all open chromatin and can detect active transcription factor binding within promoters, informing a choice of drug candidates, which make it a better method to detect open chromatin loci.

Problem definition

Given that in the Open Chromatin Landscape of Non-Small Cell Lung Carcinoma needs a better understanding of intratumor heterogeneity, So by using 3 different methods (ATAC-seq, WGS, and RNA-seq data), Applied on 50 primary NSCLC cases that have specific tumor cells, After analyzing the results from those different techniques, It should help in identifying the integration of multiple-omics and get more information about the functionality and the NSCLC mechanism, Also to get a better concept about open chromatin peaks data, That might regulate genes essential to NSCLC functions.

In the other side T cells produce a wide variety of genes encoding their T cell receptors (TCRs), allowing individual clones to recognize specific peptide _major histocompatibility complex (MHC) ligands, Each T cell expresses a TCR that recognizes antigens in the context of MHC molecules displayed on the surface of antigen-presenting or pathogen-infected cells. The major TCR species is composed of α - and β -subunits that are encoded by genes generated due to somatic V (D) J recombination. Therefore, the identification of TCR- $\alpha\beta$ -encoding sequences is critical to understanding the identity of single T cells, and methods that pair TCR- $\alpha\beta$ -encoding

sequences with cell and activation states may uncover clonal gene regulatory pathways missed by ensemble measurements.

A study combined sequencing of the TCR-encoding genes with assay for transposase-accessible chromatin with sequencing (ATAC-seq) analysis at the single-cell level to provide information on the TCR specificity and epigenomic state of individual T cells. In patients with a leukemic form of cutaneous T cell lymphoma, T-ATAC-seq enabled identification of leukemic and nonleukemic regulatory pathways in T cells from the same individual by allowing separation of the signals that arose from the malignant clone from the background T cell noise.

Age-related macular degeneration (AMD) is a well-known pathologic condition affecting the middle part of the visual system and leaves the peripheral sideways of the vision untouched. It does not cause total blindness; however, it can make daily tasks such as reading and face recognition more difficult. There are two stages of AMD, early stage (dry) which is characterized by drusen deposits on the retina which lead to seeing a spot in the central vision, and late stage (wet) which is less common, more severe, and rapidly develop and this stage causes blindness as there is a bleeding (blood vessels) at the back of the eye, and this blood leaves a scar which led to dark vision. Several studies have shown a link between inflammatory and AMD-related variables such as smoking and obesity, it has also been correlated with high blood pressure and family history of AMD onset.

This research focuses on genome-wide chromatin accessibility studies and observes global and progressive decreases in chromatin accessibility associated with AMD progression. Assay for transposase accessible chromatin using sequencing (ATAC-Seq), used for studying the global epigenetic landscape of AMD. It detects chromatin accessibility regions.

Chromatin accessibility is globally profiled using ATAC-Seq in the retina and retinal pigmented epithelium (RPE) from AMD and control patients. This study focused on changes at the highly prevalent early stage and the less reported late, atrophic stage of AMD. Studies show changes in chromatin accessibility in AMD. These changes in chromatin accessibility are seen first in the RPE, and then later in the retina. It is also found that the changes are greater at the macular retina than the peripheral. This clarifies that dysfunction in RPE drives disease onset.

Related work

In the past decades, there were many different methods used for chromatin accessibility profiling, and over time they have been developed and to provide better studies for genome regulation in many complicated tissues such as tumors and brain tissues.

Whole genome sequencing (WGS):

All organisms like bacteria have a unique genome that is consist of millions of bases (A's , C's ,T's and G's),by knowing the sequence of the bases in an organism, you have identified its unique DNA fingerprint , this bases are ordered in special way this order is called sequencing

Whole genome sequencing is a laboratory procedure that determines the order of bases in the genome of an organism in one process.

There are four main steps for sequencing the whole genome:

1. DNA shearing:

As we mentioned before that DNA is consist of millions of bases (A's , C's ,T's and G's),In this step we begin by cutting DNA into small pieces of DNA tags ,or bar codes that should be suitable for sequencing machine to read.

2. DNA bar coding:

In this step we add small pieces of DNA tags ,or bar codes for using in identifying the piece of sheared DNA belongs to which bacteria.

3. DNA sequencing:

We put the bar-coded DNA from multiple bacteria in a DNA sequencer , then The sequencer identifies the A's, C's, T's, and G's , that make up the bacterial sequence, The sequencer uses the bar code to identify which bases belong to which bacteria.

4. **Data analysis:**

We use computer analysis tools to compare sequences from multiple bacteria and identify differences like (sequence alignment), this will help us to identify how closely related the bacteria are, and how likely it.

Whole genome sequencing provides detailed and precise data for identifying outbreaks sooner. Additionally, whole genome sequencing is used to characterize bacteria as well as track outbreaks.

MNase-seq is one of four methods (DNase-seq, MNase-seq, FAIRE-seq, and ATAC-seq) that determine the chromatin accessibility first discovered in 1956 .

FIGURE 1 :

MNase-seq	DNase-seq
<ul style="list-style-type: none">• Indirect• MNase to isolate DNA bound to nucleosomes or crosslinked transcription factors• Single-cell variant exists• Analyzes protein-bound DNA	<ul style="list-style-type: none">• Direct• DNase to isolate nucleosome-bound DNA or identify promoters/enhancers• Single-cell variant exists• Analyzes DNase I hypersensitive sites
FAIRE-seq	ATAC-seq
<ul style="list-style-type: none">• Direct• Uses crosslinking, sonication, and phenol-chloroform extraction• Minimum 100,000 cell input• Analyzes protein-free DNA	<ul style="list-style-type: none">• Direct• Uses transposase to insert markers in nucleosome-free DNA for sequencing• Single-cell variant exists• Assesses open DNA

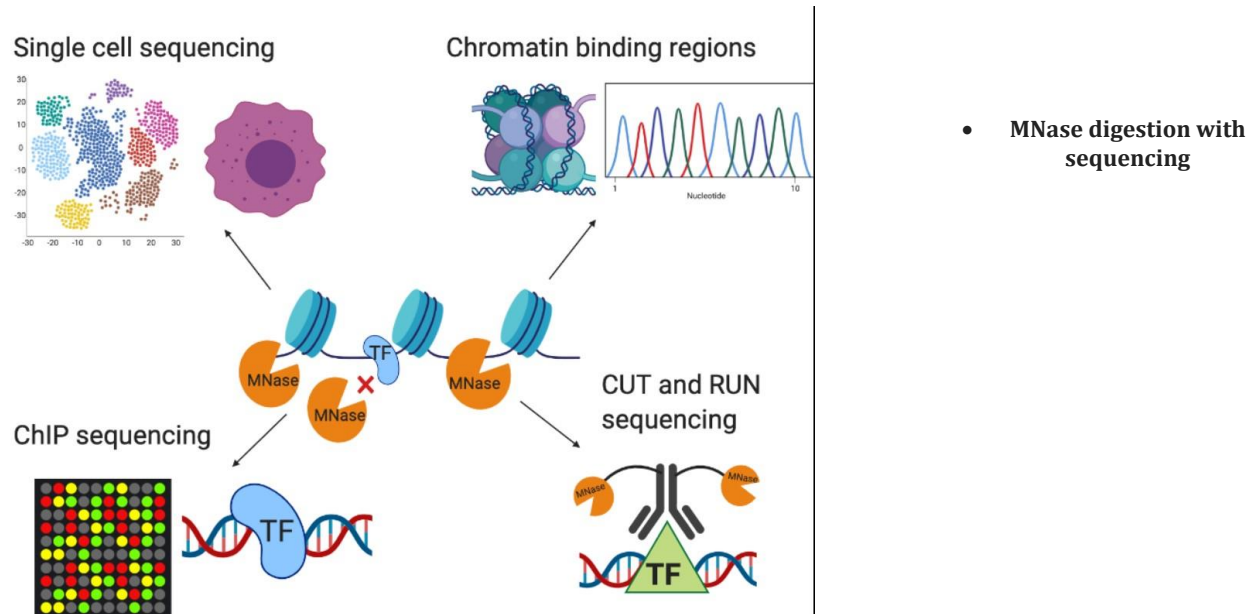
Table 1: Comparison of four methods.

MNase-seq short for micrococcal nuclease digestion with deep sequencing. this technique uses the non-specific endo-exonuclease micrococcal nuclease, an enzyme derived from the bacteria *Staphylococcus aureus*, to bind and cleave the regions of DNA on chromatin that unbound by protein .

The regions of DNA on chromatin that bound to histones or other proteins may remain undigested. The uncut DNA is then purified from the proteins and sequenced

This technique was first applied in 2006 to measure nucleosome occupancy in the *C. elegans* genome, and then applied to the human genome in 2008.

FIGURE 2 :



ATAC-seq has several advantages over MNase-seq for assessing chromatin accessibility. ATAC-seq does not rely on variable digestion with micrococcal nucleases, or on crosslinking or phenol-chloroform extraction. It generally preserves chromatin structure, so ATAC-seq results can be used to assess chromatin accessibility directly, rather than indirectly via MNase-seq. ATAC-seq can also be completed within hours.

DNase I hypersensitive site sequencing (DNase-seq) is one of the biochemical accurate methods and one of the first methods used for mapping Open chromatin regions (OCRs), was published in 2008 by sequencing genomic DNA fragments following digestion by DNase I, and first published in 1978 after MNase-seq method:

FIGURE 3:

From: [Chromatin accessibility profiling methods](#)

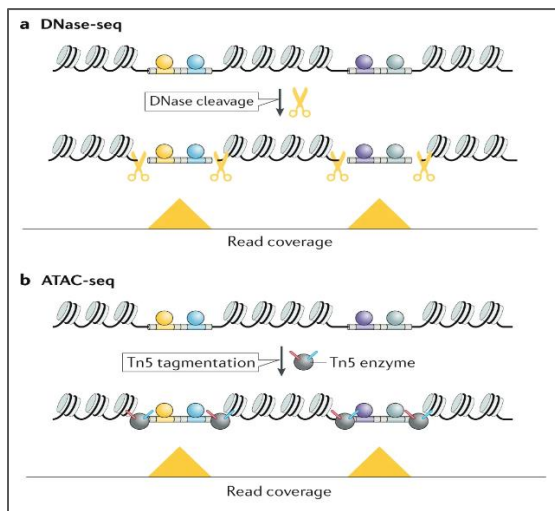
Feature	DNase-seq	ATAC-seq
Type of data produced	Accessible chromatin	Accessible chromatin
Type of input	Fresh, fixed paraffin-embedded samples or formaldehyde cross-linked samples	Fresh, slowly cooled cryopreserved, snap-frozen or formaldehyde cross-linked samples
Number of input cells ^a	1–10 million	500–50,000
Sequencing depth (for human samples)	20–50 million uniquely mapping reads (200 million for transcription factor footprinting analysis)	25 million non-mitochondrial uniquely mapping reads for standard analysis
Enzyme-specific cleavage bias?	Yes	Yes
Difficulty	Requires careful enzyme calibration and complicated sample preparations	Simple protocol, requires limited or no experimental calibration
Time	Lengthy protocol (1–3 days)	Fast protocol (<1 day)

• Table 2: Comparison of the two most commonly used chromatin accessibility profiling methods.

in the 1970s there found a research about transcription regions and how active they are when using the enzyme Deoxyribonuclease-I (DNase I), which can indicate the accessible regions in chromatin when used in the process of TF footprinting.

This method combined with ATAC-seq ,consider the most used techniques for today.

FIGURE 4:

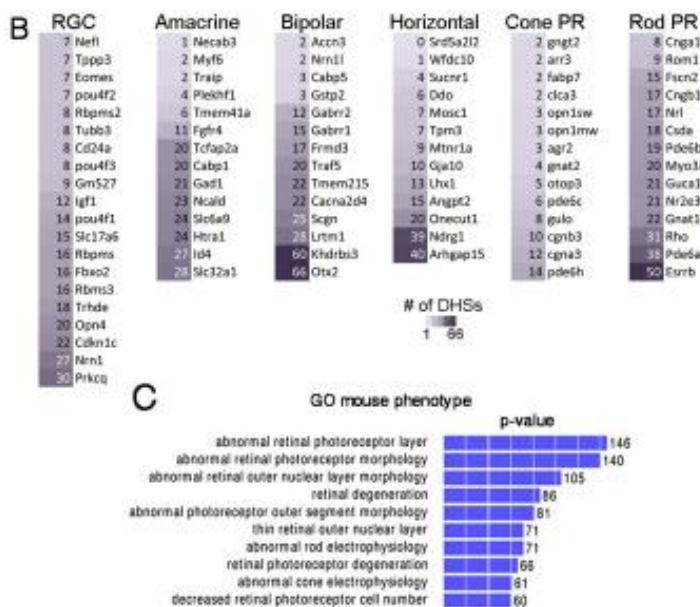


- a | In (DNase-seq) the yellow parts represent the scissors(DNase I) that is used for cleave (OCRs),to create the fragments , which later could be amplified in Next-generation sequencing (NGS),to create the libraries .
- b | in ATAC-seq, the grey circle parts represents the Tn5 transposase which inserted in (OCRs)and adding the blue and red lines in the Tn5 ,to create the fragments to be amplified and create the sequence libraries .

DNase-seq in analysis of diseases

Research paper done to a mouse's brains and retina to identify the regulatory elements in central nervous system (CNS). The results showed that DNase-seq using DNase I helped to create an extensive catalogue of possible rare cell type-specific regulatory elements in (CNS) and common DNase I-hypersensitive sites (DHSs) in related to CNS tissues.

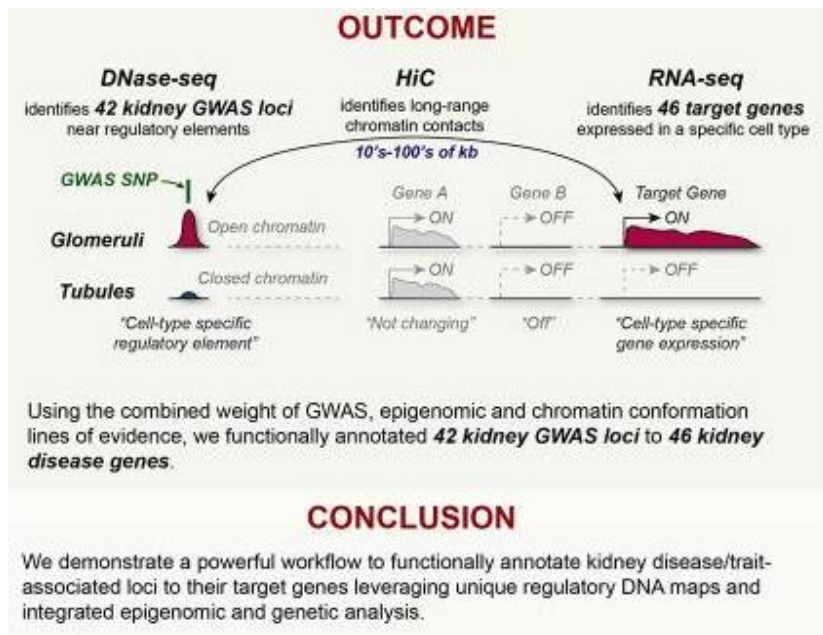
FIGURE 5:



- B | The number of DHSs near genes shows the retinal cell types in columns. Color intensity increases with DHS number.
- C | Gene Ontology represents terms for retinal, and p-values represents the number of genes associated with each term.

Another research study made to discover unique regulatory elements on cortex samples in of Kidney using (DNase-seq) and aired gene expression profiles (RNA-seq) methods , great resultus research found in figure down below .

FIGURE 6:



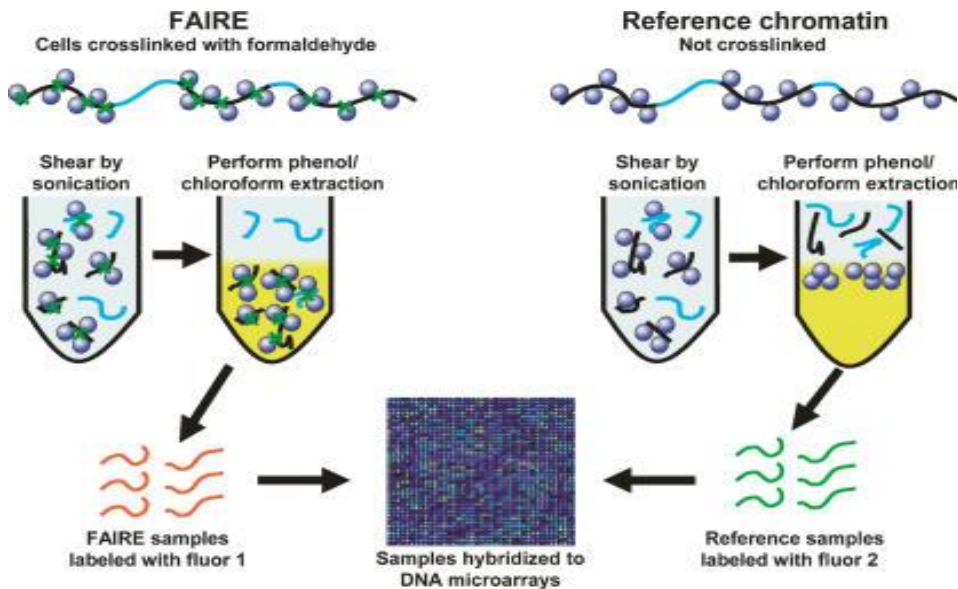
- Functionally annotated total 88 kidney disease genes (42 genes from DNase-seq, and 46 genes from RNA-seq).

FAIRE-seq is one of the easiest methods to directly investigate nucleosome-depleted regions of the genome, furthermore it differs from DNase-Seq in that it does not require permeabilization of cells or isolation of nuclei, and it can analyze all cell types. In a study of seven different human cell types, DNase-Seq and FAIRE-Seq produced a strong validation that each cell type contained 1-2% of the human genome as open chromatin. The idea of FAIRE-Seq is based on the fact that formaldehyde cross-linking is more efficient in nucleosome-bound DNA than in nucleosome-depleted regions of the genome.

Uninterrupted DNA normally found in open chromatin is then separated, which is then sequenced. The protocol consists of cross-linking, phenol extraction, and DNA sequencing in the aqueous phase. FAIRE uses biochemical properties of protein-bound DNA to separate nucleosome-depleted regions of the genome. Cells will undergo cross-linking, which ensures the stability of the interaction between nucleosomes and DNA. After sonication, using phenol-chloroform extraction the fragmented and fixed DNA is separated. This method results in two phases, an organic phase and an aqueous phase. Due to their biochemical properties, DNA fragments attached to nucleosomes will sit preferentially in the organic phase. On the other hand, regions that are nucleosome-depleted or 'open' will be found in the aqueous phase. Specifically, through aqueous phase extraction, only depleted regions of the nucleus will be purified and enriched.

Because of its speed and simplicity, FAIRE is of utility for generation of chromatin profiles of different cell types in health and disease, for isolating DNA regulatory elements en masse for further characterization, and as a screening assay for the effects of small molecules on chromatin regulation.

Figure 7:

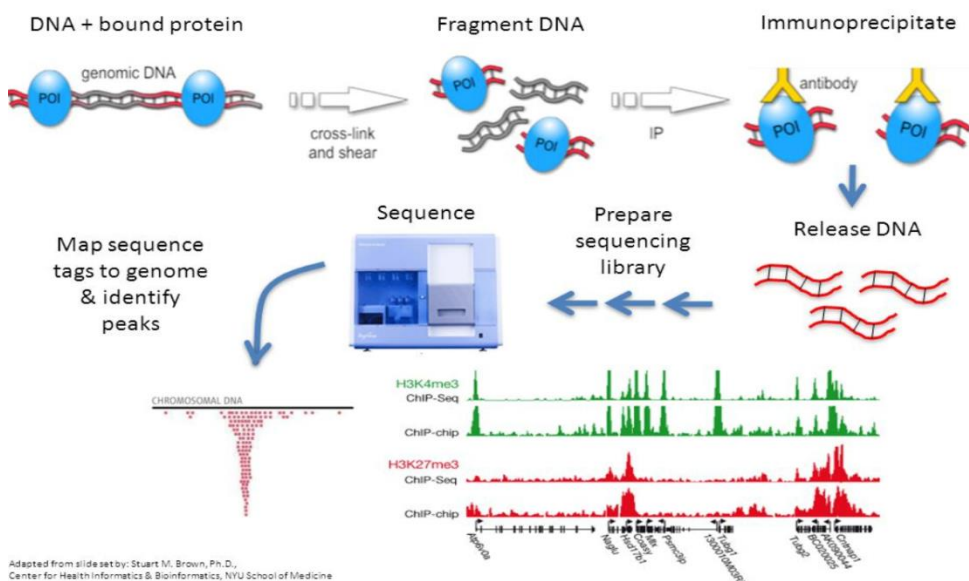


- FAIRE-seq in human cells is illustrated on the left, while preparation of the reference is illustrated on the right.

During binding of sequence-specific regulatory factors and their chromatin remodeling activities, nucleosomes are expelled from chromatin in eukaryotic cells. Typically, these active sites have been identified experimentally through their sensitivity to nucleases. Here we present a simple protocol to perform genome-wide isolation of nucleosome-depleted DNA from chromatin human. Is called FAIRE(Formaldehyde Assisted Isolation of Regulatory Elements).

CHIP_{seq}, short for Chromatin immunoprecipitation is a robust technique for locating the genome-wide DNA binding sites of transcription factors and other proteins by combining chromatin immunoprecipitation with real-time next generation sequencing. Examining transcription factor-DNA interactions using chromatin immunoprecipitation (ChIP) experiments is essential for developing research on the regulation of gene expression and epigenetic changes. In vivo at a single locus or numerous loci, ChIP can identify and roughly quantify particular protein-DNA and protein-protein interactions.

FIGURE 8



- This figure shows the steps of Chip seq

Antibodies are used to select specific proteins or nucleosomes, which enriches for DNA fragments that are bound to these proteins or nucleosomes. Following that, the bound DNA is precipitated, cleaned, and sequenced. Formaldehyde is used to crosslink proteins and related chromatin in living cells or tissues. The cross-linked DNA-protein complexes (chromatin-protein) are subsequently physically or enzymatically digested into 500 bp DNA fragments by sonication. A suitable protein-specific antibody is then used to immunoprecipitate the DNA-protein complexes. The related DNA fragments are eluted when the cross-links are reversed, and the cross-linked complexes are then immunoprecipitated. Next-generation sequencing, microarrays (ChIP-chip) are then used to analyze the resultant DNA (ChIP-seq).

ChIP-seq has become a crucial tool for studying gene regulation and epigenetic processes. It offers greater coverage, less noise, and better resolution and therefore provides substantially improved data.

What has been achieved by applying ChIP-seq?

A study that used ChIP-seq to characterize transcription factors in embryonic stem cells and shown that the regulatory elements are organized into "enhanceosomes" allowed researchers to better understand the roles of transcription factors during embryonic stem cell differentiation. This made it possible for researchers to comprehend the functions of transcription factors during the development of embryonic stem cells.

Another study shows that ChIP-Seq has made some significant advances, the identification of additional potential cardiac enhancers across the genome. The greatest cause of death in both newborns and adults is heart disease. Although protein-coding areas have been extensively screened, the genetic cause of many heart abnormalities is still unknown. Among the various kinds of regulatory sequences, transcriptional enhancers are particularly difficult to discover since they might be found far away from the genes they control in the genome. Chromatin immunoprecipitation with p300 directly from animal tissues coupled with massively parallel sequencing (ChIP-Seq) can accurately predict the genomic location and tissue specificity of active developmental.

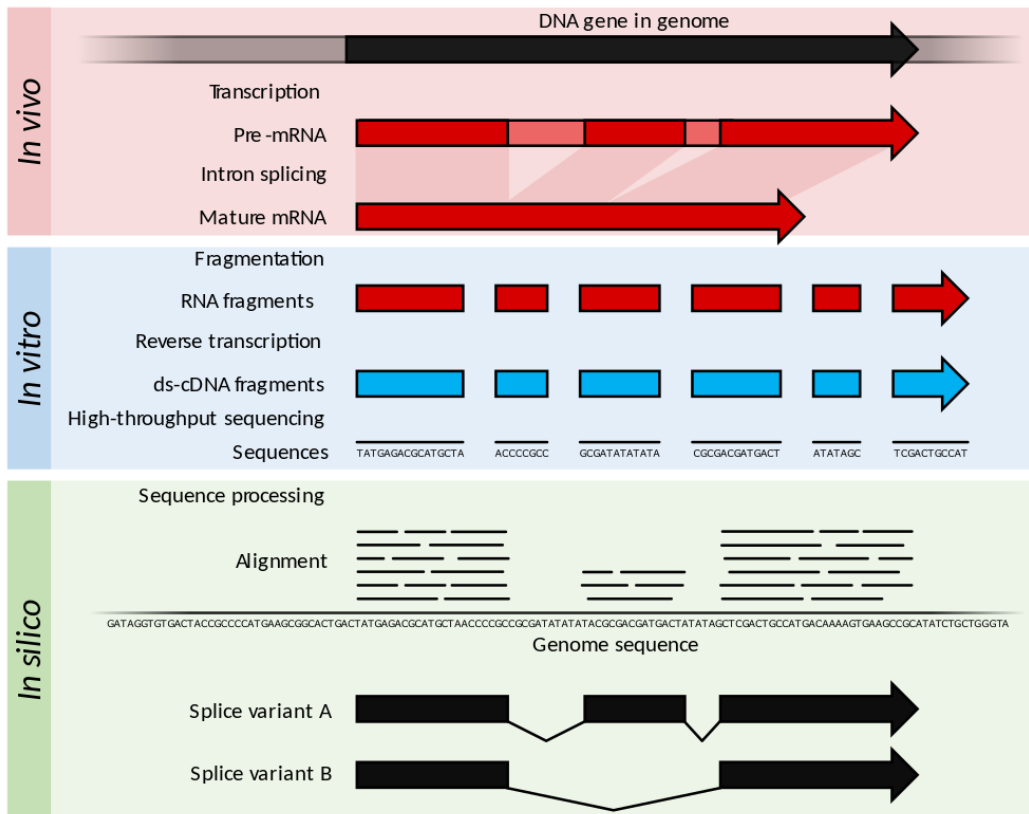
RNA-Seq

RNA-Seq(RNA sequencing) first developed in 2000s, appeared in 2008,this method uses the next-generation sequencing (NGS) technology which also called second-generation sequencing ,it was used in 2006 even before it has been given its name in a research about prostate cancer cell ,the RNA-Seq is helpful technique to determine or define the exons(the exons regions consider as small part of the gene that will be later a part of RNA and then translated to proteins) and intron(introns are parts which are going to be removed as it won't be expressed as RNA) regions boundaries ,the RNA-Seq method could also be used in three different

ways such as:

- in silico
- in vitro
- in vivo

FIGURE 9



- Summary of RNA-Seq. Within the organism

cancer and detection:

as RNA-seq is unbiased in the analysis way it uses , it could be very effective to define common events in some diseases like cancer.

In this kind of era, it is an urgent need develop such techniques in the sequencing of the genomic; as a result RNA-Seq is now able to be applied on the analysis of the microbe and its cells.

The combination of ATAC-Seq and the RNA-Seq is also used in a very complex problem which is gene regulatory networks(GRN).

RNA-Seq

To assess similarities and transcriptional differences within a group of cells, scRNA-seq is considered as one of the primary uses to compare transcriptomes of individual cells. Early reports show previously unappreciated levels of heterogeneity. Thus, the analysis of heterogeneity remains a primary reason for initiating scRNA-seq studies.

scRNA-seq is ideal as it is used to identify rare cell populations that may not be detected in pooled cell analyses. Assessments of transcriptional differences between individual cells, such as malignant tumor cells within a tumor mass, or highly responsive immune cells within an apparently homogenous population, have been used in addition to the assay. Single cells where each cell is essentially unique, such as single T

lymphocytes expressing highly diverse T-cell receptors, neurons within the brain or cells within the early embryo. SRNA sequencing is also being increasingly used to trace lineage and developmental relationships between heterogeneous, but related, cellular states in scenarios such as embryonic development, cancer, myoblast differentiation, lung and lymphocyte fate diversity.

The advantage of scRNA-seq is that it can provide important information about essential properties of gene expression, as well as resolving cellular heterogeneity. This includes studying mono-parallel gene expression, splicing patterns, as well as noise during transcriptional responses. Importantly, studying gene co-expression patterns at the single-cell level may allow identification of co-regulated gene modules and even inference of gene regulatory networks that underlie functional heterogeneity. and cell type specification.

What are the basic steps in conducting scRNA-seq?

FIGURE 10

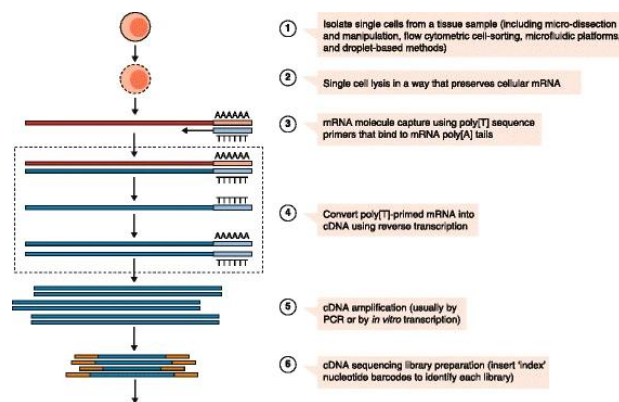
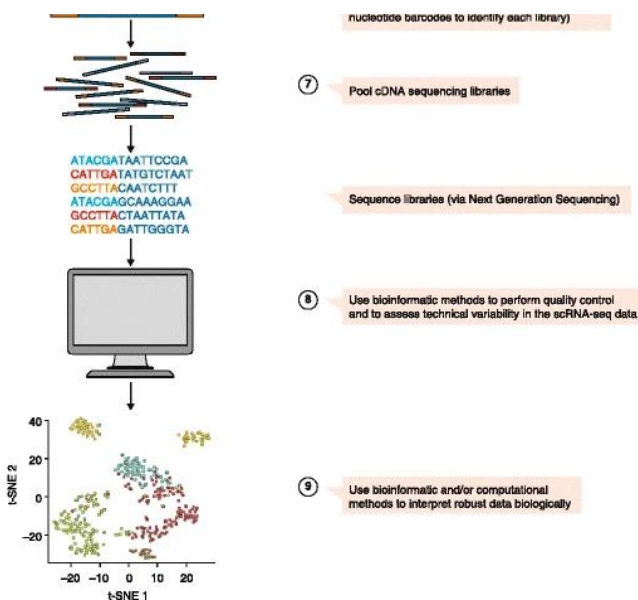


FIGURE 11



single-cell RNA-sequencing (scRNA-seq) experiments workflow include most of the following steps:

- 1) isolation of single cells.
- 2) cell lysis while preserving mRNA.
- 3) mRNA capture.
- 4) reverse transcription of primed RNA into complementary DNA (cDNA).
- 5) cDNA amplification.
- 6) preparation of cDNA sequencing library.
- 7) pooling of sequence libraries.
- 8) use of bio-informatic tools to assess quality and variability.
- 9) use of specialized tools to analyze and present the data.

Which protocol should be employed?

In order to decide which scRNA-seq protocol and platform to use, a research question has to be asked. Whereas, future studies of poorly characterized heterogeneous tissues versus characterization of transcriptional responses within a given cell population may be presented optimally through different experimental approaches. About 20 different scRNA-seq protocols have been published to date. One of the main differences between these methods is that some provide full text data, while others specifically calculate only the three endings of texts.

FIGURE 12

Protocol example	C1 (SMARTer)	Smart-seq2	MATQ-seq	MARS-seq	CEL-seq	Drop-seq	InDrop	Chromium	SEQ-well	SPLIT-seq
Transcript data	Full length	Full length	Full length	3'-end counting	3'-end counting	3'-end counting	3'-end counting	3'-end counting	3'-end counting	3'-end counting
Platform	Microfluidics	Plate-based	Plate-based	Plate-based	Plate-based	Droplet	Droplet	Droplet	Nanowell array	Plate-based
Throughput (number of cells)	10^2 – 10^3	10^2 – 10^3	10^2 – 10^3	10^2 – 10^3	10^2 – 10^3	10^3 – 10^4	10^3 – 10^4	10^3 – 10^4	10^3 – 10^4	10^3 – 10^5
Typical read depth (per cell)	10^6	10^6	10^6	10^4 – 10^5	10^4 – 10^5	10^4 – 10^5	10^4 – 10^5	10^4 – 10^5	10^4 – 10^5	10^4
Reaction volume	Nanoliter	Microliter	Microliter	Microliter	Nanoliter	Nanoliter	Nanoliter	Nanoliter	Nanoliter	Microliter
Reference	[63]	[57]	[39]	[10]	[64]	[45]	[46]	[47]	[101]	[38]

Experimental results and discussion

Result

We will talk here in this section about the expected results of ATAC-seq analysis and what we have reached through our readings of research papers and websites:

1. Overview of the ATAC-seq analysis:

To perform an ATAC-seq analysis, 3 basic steps must be followed:

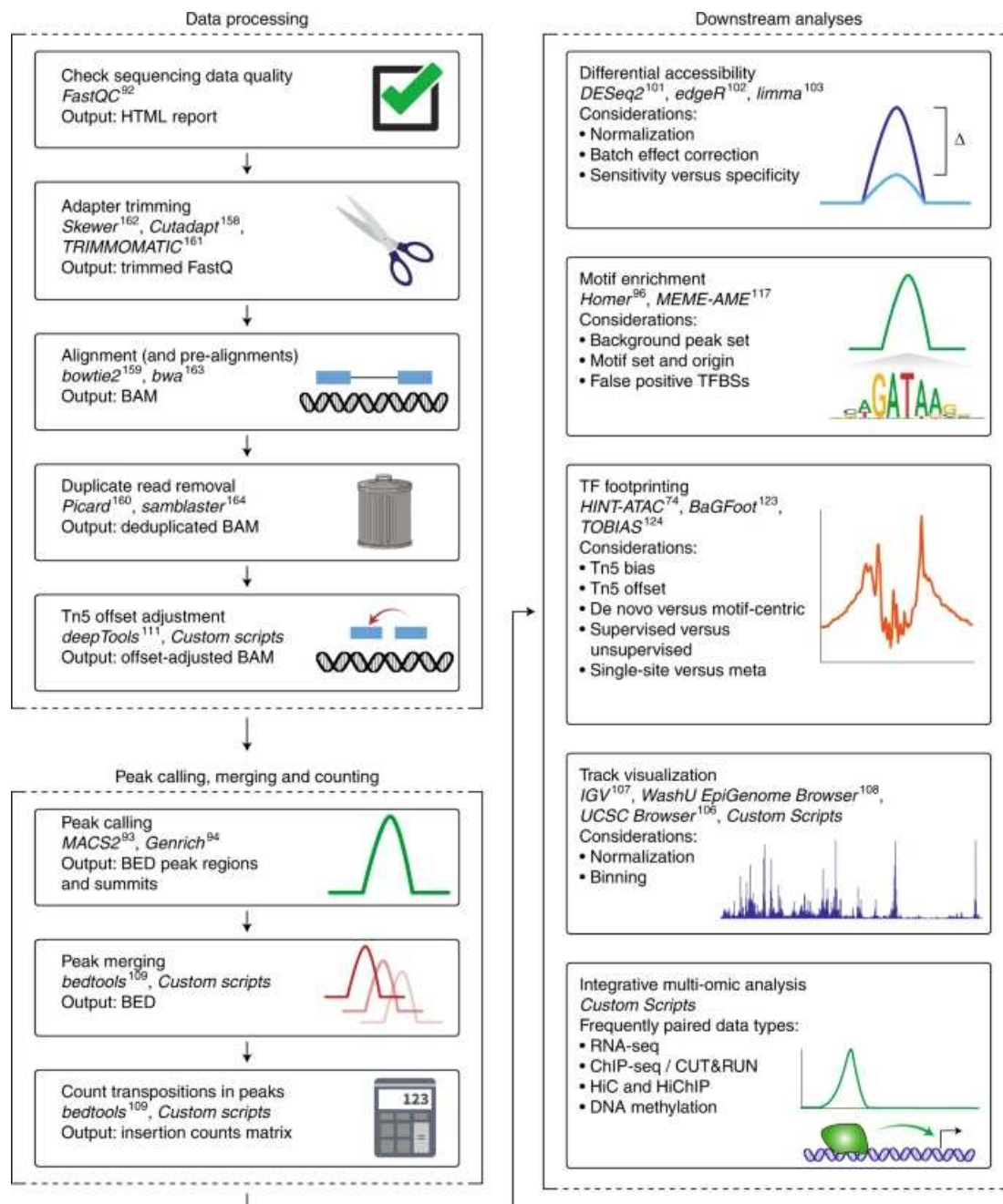
(1) data processing,

(2) peak Calling, merge and insertion counting

(3) downstream analysis.

As shown in the following figure :

FIGURE 13



Data processing describes the steps necessary to take the raw Fastq, prepare it for alignment with the genome, and adjust for the Tn5 offset. This is then followed by the use of clean and normalized readings for peak calling on a sample-by-sample basis. These peaks are then combined for all samples being compared, forming a union peak set, and the number of insertion events is calculated for each peak of the union peak set for all samples. This matrix preparing matrix and union vertex set can be used for a variety of downstream analyses, including differential accessibility analysis to identify vertices that are accessed more frequently or less frequently and furthermore enrichment of motifs on related groups of vertices.

2. What we have reached through our search for this assay, which is the focus of our interest:

This assay can determine the accessibility of chromatin, which in turn helps in understanding transcriptional regulation and locating regulatory elements of the genome. Whereas, as we mentioned before, there are several different methods that enable us to access the open chromatin, but the ATAC-seq assay has many advantages, including:

- It is widely adopted due to its low input material requirements (<50,000 cells) and short processing time, which facilitates data generation for large numbers of accessions.
- Generates genome-wide regulatory maps that closely resemble those derived from DNase-seq and MNase-seq, while reducing library preparation complexity and hands-on time.
- It requires a Simple protocol, requires limited or no experimental calibration.
- It used a Fast protocol, that takes about less than one day, unlike its variants.
- ATAC-seq does not rely on variable digestion with micrococcal nucleases, unlike MNase-seq.
- ATAC-seq results can be used to assess chromatin accessibility directly because it does not rely on crosslinking or phenol-chloroform extraction.
- ATAC-seq does not require a Prior knowledge of the regulatory elements, this method shows all the active loci (promoters and enhancers)

Thus, the ATAC-seq assay is a valuable tool for understanding how cells control gene expression, by mapping and locating putative gene regulatory elements.

In the previous figure : A schematic overview of the necessary steps in the ATAC-seq data analysis workflow, starting with the raw FASTQ files generated by the sequencer.

Figure 14

Methods	MNase-seq	DNase-seq	FAIRE-seq	ATAC-seq
Cell status	Any state of cells	Any state of cells	Any state of cells	Fresh cells or slowly cooled cryopreserved cells
Principle	MNase digests DNA which is not protected by protein or nucleosome on chromatin.	DNAase I preferentially excises DNA sequence without nucleosomes.	Separation of naked DNA based on formaldehyde fixation and phenol-chloroform extraction	Tn5 transposase inserts the DNA sequence without protein or nucleosome protection and excises it.
Target regions	Focus on nucleosome localization	Accessible chromatin regions, focusing on transcription factor binding sites	Accessible chromatin regions	Accessible chromatin regions in genome-wide, including transcription factors, histone modifications.
Specific features	1. A large number of cells as input materials; 2. The quantity of enzyme needs to be accurate; 3. Localization of the entire nucleosome and inactive regulatory region; 4. Detecting inactive regions by degrading active regions; 5. Standard analysis requires 150-200 M reads.	1. A large number of cells as input materials; 2. The process of sample preparation is complicated; 3. The quantity of enzyme needs to be accurate; 4. Standard analysis requires 20-50 M reads.	1. Low signal-to-noise ratio makes data analysis difficult; 2. Results depend heavily on formaldehyde fixation; 3. Standard analysis requires 20-50 M reads.	1. A lower number of cells as input materials; 2. Standard analysis requires 20-50 M reads through reducing sequencing depth; 3. Conveniently obtain accessible chromatin regions in genome-wide; 4. Mitochondrial data has an effect on the accuracy of the results.
Time	2-3 days	2-3 days	3-4 days	2-3 h

Discussion

In this work, we provide an overview of a number of studies on ATAC-seq chromatin accessibility profiling, a method for determining the distinctive chromatin landscape associated with a specific cell type and how it may be impacted by perturbation or disease. Our study reveals that the ATAC seq approach can give important insights into the regulatory structure of the genome. A rapid and scalable tool for identifying the regions of the genome that are bound by TFs and studying how these landscapes change in response to various settings or perturbations is provided by the use of ATAC-seq. By transposing sequencing adapters onto native chromatin in vitro, this is accomplished.

We discover that by applying three different techniques (ATAC-seq, WGS, and RNA-seq data) to NSCLC cases with specific tumor cells and analyzing the results, It should help in identifying the integration of multiple-omics and get more information about the functionality and the NSCLC mechanism. It should also help to get a better understanding of open chromatin peaks data, which might regulate genes essential to NSCLC functions.

At the single-cell level, the TCR-encoding genes are sequenced along with an assay for transposase-accessible chromatin with sequencing (ATAC-seq) analysis to reveal the TCR specificity and epigenome status of each T cell.

Using ATAC-Seq, the retina and retinal pigmented epithelium (RPE) of AMD and control patients were extensively analyzed for chromatin accessibility. This study concentrated on alterations in AMD's early, highly prevalent stage and its later, atrophic stage, which receives less media attention. Studies suggest that AMD causes changes in chromatin accessibility. The RPE first exhibits these alterations in chromatin accessibility, followed subsequently by the retina. Additionally, it is discovered that the macular retina experiences more modifications than the peripheral retina. This makes it clear that RPE dysfunction causes the start of illness.

Conclusion and future work

In this paper we presented ATAC-seq method and how it can access the open regions in the chromatin,we focuses on some studeis which have been done to diffrent kinds of diseases,to see the diffrent types of results using this technique,we also did a simple comparison of this technique using some studies, also we have mentioned the other thechnique that can contripute in diffrent ways to help improvemnet of the final result and how it does help researchers and scientists to define their problems and analyze such a big data as genomics is complex to be analyzed ,as future work there is upcoming technique is still under tests and studies is on progress which is single-cell Assay for Transposase-Accessible Chromatin using sequencing (scATAC-Seq),the usage of this protocol should evaluate the final result to become more accurate, as it uses single cell this may help in the noise reduction compared using a bulk of cells ,while the process is undergo the using of ATAC-seq known to have a large scale of output reads that sometimes when represented as peaks for some samples or species,some peaks considered as unimportant in the ATAC-seq so it will be ignored, which is considered a waste of sequencing resurces.

In the future a diffrent techniques such as Omni-ATAC method which is the improve of ATAC-seq that uses diffrent ways during its process to remove those samples from the beginning could help to reduce the possibility of consumption .another method such as Whole Genome Sequencing (WGS) evolves and now try to response and detect more about diseases researchers are looking in for how to make a good use of this method with many other ways, as it helps to get more desriptive details about samples or some species ,it could be said that ATAC-Seq has helped very well in solving diseases complex analysis combined with different methods.finally the reader will find refrences in the last section so it could help the user to understand and read more.

References

1. Fiorella C. Grandi, Hailey Modi, Lucas Kampman & M. Ryan Corces. (2022). Chromatin accessibility profiling by ATAC-seq. *Nature Protocols*. **17**, 1518–1552.
2. Zhoufeng Wang; Kailing Tu; Lin Xia; Kai Luo; Wenxin Luo; Jie Tang; Keying Lu; Xinlei Hu; Yijing He; Wenliang Qiao; Yongzhao Zhou; Jun Zhang; Feng Cao; Shuiping Dai; Panwen Tian; Ye Wang; Lunxu Liu; Guowei Che; Qinghua Zhou; Dan Xie; Weimin Li. (2019). The Open Chromatin Landscape of Non–Small Cell Lung Carcinoma. *AACR Journal*. **79 (19)**: 4840–4854.
3. Jie Wang, Cristina Zibetti, Peng Shang, Srinivasa R. Sripathi, Pingwu Zhang, Marisol Cano, Thanh Hoang, Shuli Xia, Hongkai Ji, Shannath L. Merbs, Donald J. Zack, James T. Handa, Debasish Sinha, Seth Blackshaw & Jiang Qian. (2018). ATAC-Seq analysis reveals a widespread decrease of chromatin accessibility in age-related macular degeneration. *Nature Communications*. **9**, Article number: 1364.
4. Carsten Carlberg & Ferdinand Molnár. (2018). *Methods and Applications of Epigenomics*. Springer Link. pp 19–38.
5. Jason D. Buenrostro, Paul G. Giresi, Lisa C. Zaba, Howard Y. Chang, and William J. Greenleaf. (2013). Transposition of native chromatin for multimodal regulatory analysis and personal epigenomics. *PubMed Central*. 10(12): 1213-1218.
6. Yuanyuan Sun, Nan Miao & Tao Sun. (2019). Detect accessible chromatin using ATAC-sequencing, from principle to applications. *Hereditas*. 156, 29.
7. Ansuman T. Satpathy, Naresha Saligrama, Jason D. Buenrostro, Yuning Wei, Beijing Wu, Adam J. Rubin, Jeffrey M. Granja, Caleb A. Lareau, Rui Li, Yanyan Qi, Kevin R. Parker, Maxwell R. Mumbach, William S. Serratelli, David G. Gennert, Alicia N. Schep, M. Ryan Corces, Michael S. Khodadoust, Youn H. Kim, Paul A. Khavari, William J. Greenleaf, Mark M. Davis & Howard Y. Chang. (2018). Transcript-indexed ATAC-seq for precision immune profiling. *Nature Medicine*. 24, pages580–590.
8. Johnson, D. S. et al. Genome-wide mapping of *in vivo* protein–DNA interactions. *Science* **316**, 1497–1502 (2007).