

Nucleosome Positioning as a Potential Biomarker for Rheumatoid Arthritis

Vaidika Goyani

University of Essex

Abstract

Rheumatoid arthritis (RA) is a chronic autoimmune disease that causes joint inflammation, pain, and potential disability if left untreated, affecting approximately 0.24% to 1% of the global population. Early diagnosis is crucial for effective treatment but is hindered by a lack of specific biomarkers. Epigenetic mechanisms, particularly nucleosome positioning, play an important role in RA pathogenesis by regulating gene expression through the control of DNA accessibility to transcription factors. This study aimed to investigate nucleosome positioning as a potential biomarker for RA by analysing chromatin accessibility using ATAC-seq data to identify differentially accessible chromatin regions between RA and healthy samples. Publicly available ATAC-seq datasets from the NCBI GEO database were analysed using bioinformatics tools including Bowtie, MACS2, BEDTools, the UCSC Genome Browser, and WebGestalt. Differentially accessible chromatin regions linked to RA were identified through peak calling and annotation, followed by functional analysis to assess their biological relevance. The comparative analysis revealed significant accessibility differences between RA and healthy samples, with peaks in RA enriched in satellite regions and reduced in CpG islands and intergenic regions. Functional annotation identified histone modifications and chromatin remodelling as important pathways in RA, with genes like SMYD3, UTY, and DNMT1 linked to histone modifications, suggesting a role in disease progression. This study provides evidence that altered nucleosome positioning in RA could serve as a potential biomarker for early diagnosis and highlights the need for further investigation into chromatin modifications for targeted therapies.

Keywords: Anti-Citrullinated Peptide Antibodies, Assay for Transposase-Accessible Chromatin sequencing, cell-free DNA, Fibroblast-Like Synoviocytes, Nucleosome Repeat Length, Osteoarthritis, Rheumatoid Arthritis, Rheumatoid Factor, Sequence Read Archive

Date of Submission: 23.08.2025

Date of Acceptance: 12.12.2025

Acknowledgements

I would like to thank my supervisor Dr. Vladimir Teif for his constant guidance and support throughout this project.

A special thanks to my friends for always keeping me motivated and encouraging me along the way.

Introduction

Rheumatoid Arthritis: Epidemiology, Diagnosis, and Epigenetics

Rheumatoid arthritis (RA) is a chronic, systemic autoimmune disease characterised by joint inflammation causing pain, stiffness, swelling, and potential joint damage (Lee *et al.*, 2017). It often begins in small joints, progressing to larger ones, and can affect extra-articular sites like the skin, eyes, and lungs, leading to deformities, bone erosion, and disability if untreated (Bullock *et al.*, 2018). RA affects 0.24% to 1% of the global population, with higher rates in the U.S., Europe, Australia (2%), and Native American populations (Pima 5.3%, Chippewa 6.8%), while being rarer in Africa and Asia (Venetsanopoulou *et al.*, 2023). Early diagnosis is crucial for effective management, yet RA diagnosis remains challenging due to the lack of highly specific biomarkers.

Studies show that rheumatoid factor (RF) and anti-citrullinated peptide antibodies (ACPA) are key biomarkers for RA (Zhao and Li, 2018), but they are not specific to RA and can occur in other conditions or healthy individuals. Additionally, early RA symptoms often overlap with diseases like osteoarthritis (OA), complicating diagnosis. Recent studies have emphasised the role of epigenetic mechanisms in RA development, which regulate chromatin structure and gene expression without changing the DNA sequence, influencing cellular behaviour and contributing to disease progression (Araki and Mimura, 2016).

Chromatin Structure, Nucleosome Positioning, and their Role in RA

In eukaryotic cells, genetic information is stored and maintained in the form of chromatin, as shown in Figure 1. The fundamental units of chromatin are called nucleosomes; they consist of a histone octamer, which is a protein containing two copies of H2A, H2B, H3, and H4, wrapped around approximately 147 base pairs of DNA (Li, Peng and Panchenko, 2022). These nucleosomes are connected by short DNA segments called linker DNA, forming nucleosomal arrays that interact with neighbouring nucleosomes to form a polymer resembling "beads on a string," which can be termed as the primary structure of the chromatin. Histone N-terminal tails extend from nucleosomes into the nuclear lumen, providing sites for post-translational modifications that regulate chromatin accessibility. Additionally, histone H1 binds linker DNA, stabilising chromatin structure. Upon further condensation, it forms the secondary and tertiary higher-order chromatin structures (Luger, Dechassa and Tremethick 2012). Nucleosome positioning influences whether a region is in an open, transcriptionally active state (euchromatin) or a tightly packed, repressive state (heterochromatin) (Handy, Castro and Loscalzo, 2011). Chromatin remodelling complexes and

histone modifications dynamically alter nucleosome arrangement to control DNA accessibility for regulatory proteins and transcription factors. In general, coding regions tend to have dense nucleosome occupancy (condensed chromatin), while regulatory elements such as enhancers, promoters, and terminators are typically nucleosome-depleted, allowing for transcriptional activation (Struhl and Segal, 2013).

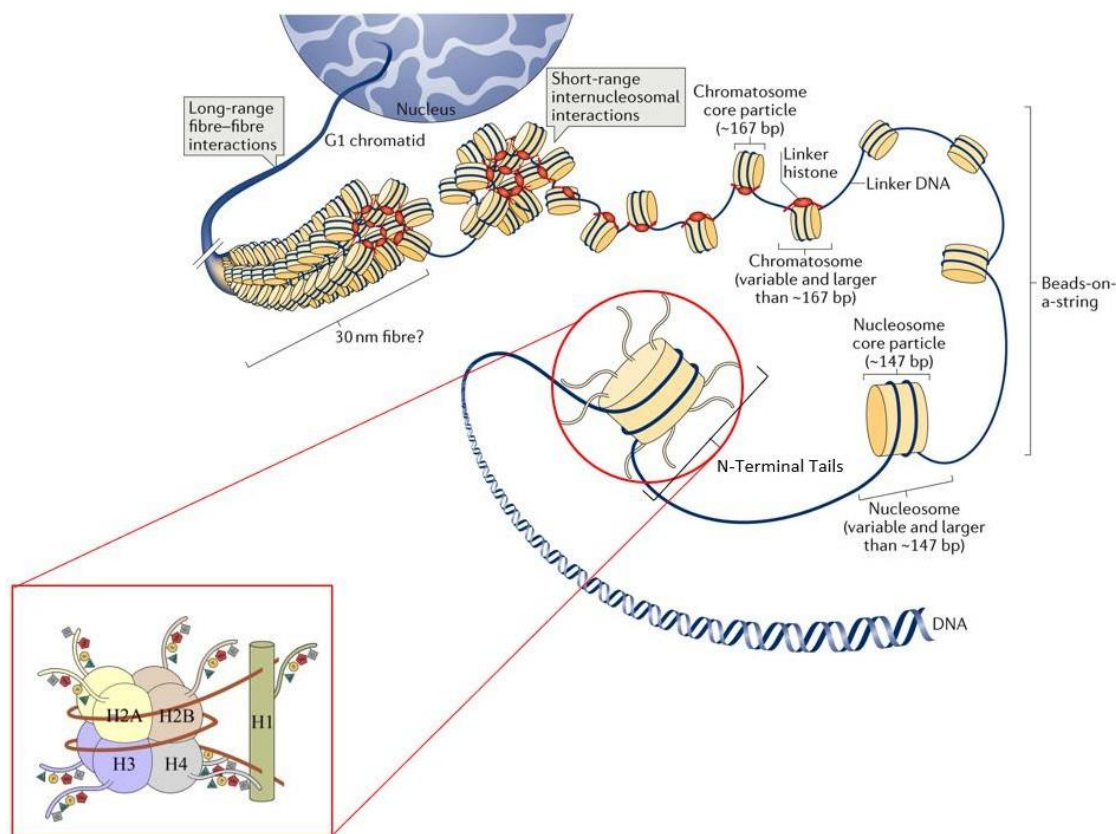


Figure 1. Chromatin Structural Organisation with a Focus on Nucleosome Structure.

This figure provides a labelled schematic of chromatin structure, with a detailed zoom-in on the nucleosome, highlighted in red.

Combined figure adapted from Li, Ding and Zheng (2014) and Fyodorov *et al.* (2018).

Epigenetic modifications, including DNA methylation and histone modifications, are key regulators of chromatin structure and gene expression. DNA methylation involves the addition of a methyl group to the C5 position of cytosine, forming 5-methylcytosine, which represses gene expression by recruiting silencing proteins or preventing the binding of transcription factors (Moore, Le and Fan, 2013). Post-translational modifications of histones, like methylation, acetylation, phosphorylation, and ubiquitination, regulate chromatin structure and gene expression. Histone methylation provides long-term epigenetic memory, controlling gene activation or silencing, while acetylation loosens chromatin structure to promote transcription (Lui *et al.*, 2023). Ubiquitination targets histones for

degradation or modulates transcriptional activity, while phosphorylation allows for rapid chromatin changes in response to cellular signals like DNA damage (Lui *et al.*, 2023). Together, these modifications dynamically influence chromatin organisation, ensuring both stable gene regulation and adaptability to environmental signals.

The positioning of nucleosomes is another important component of gene regulation, as the presence and absence of nucleosomes around promoter regions and transcription start sites blocks the retrieval of the regulatory elements. Several factors influence nucleosome positioning and chromatin structure, as highlighted by Teif and Clarkson (2019). The dyad position, marking the nucleosome centre, determines whether regulatory DNA elements like promoters or enhancers are accessible or occluded. Nucleosome occupancy—the likelihood of a DNA site being bound by a nucleosome—regulates transcription, with high nucleosome occupancy repressing gene expression and low nucleosome occupancy facilitating it. The stability, accessibility, and fuzziness of nucleosomes also affect DNA exposure, impacting transcription factor binding and polymerase activity. The nucleosome repeat length (NRL) further shapes chromatin organisation, as shorter NRLs promote tighter packing, while longer NRLs create a more open and transcriptionally active chromatin state (Kornberg and Lorch, 1999; Struhl and Segal, 2013).

Nucleosomes undergo dynamic changes, including histone eviction, reassembly, and the incorporation of specialised histone variants, which influence gene regulation and DNA repair. Covalent histone modifications further impact chromatin structure, altering histone–DNA interactions and regulatory factor engagement. This interplay shapes a chromatin landscape that can either promote or repress transcription, depending on cellular needs (Bai and Morozov, 2010). Altered nucleosome positioning in monocytes may drive RA-related chromatin dysregulation in peripheral blood, contributing to inflammation and osteoclast differentiation. The extent of this chromatin remodelling correlates with serum C-reactive protein levels, linking nucleosome positioning to RA-associated inflammatory responses (Zong *et al.*, 2021).

ATAC-seq and its Role in Studying Chromatin Accessibility

To study the epigenetic landscape, several high-throughput sequencing technologies have been developed, with Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq) being one of the most efficient methods for identifying chromatin accessibility (Yan *et al.*, 2020). Figure 2 shows the ATAC-seq mechanism and peak calling process. ATAC-seq efficiently maps open chromatin regions where transcription factors bind, enabling the comparison of chromatin accessibility across different conditions (Grandi *et al.*, 2022). This technique uses a hyperactive Tn5 transposase, preloaded with sequencing adapters, which targets accessible chromatin regions. It inserts these adapters into nucleosome-depleted areas while creating 9-bp staggered nicks in the chromatin. These nicks are repaired, resulting in a 9-bp duplication at the insertion sites, which serves as a

marker to accurately map open chromatin regions. Following fragmentation, paired-end sequencing is performed, and the raw reads undergo quality control to assess base quality, GC content, and adapter contamination. Overrepresented adapters are trimmed, and the cleaned reads are aligned to a reference genome. Peak calling then identifies accessible chromatin regions, such as promoters and enhancers, which are crucial for understanding gene regulation and chromatin accessibility (Yan *et al.*, 2020). One of the advantages of ATAC-seq is its time-efficient protocol. It requires as few as 500 to 50,000 cells, unlike other techniques like DNase-seq and FAIRE-seq, which need millions of cells. Furthermore, ATAC-seq does not require rigid size selection during library preparation, allowing for the analysis of nucleosome positioning alongside chromatin accessibility (Yan *et al.*, 2020). For instance, a previous ATAC-seq study revealed subtle differences in gene expression regulation specific to each synovial pathotype, highlighting its potential in identifying therapeutic targets for RA (Hughes, 2023). The identification of unique epigenetic patterns in patients with RA can offer insights into disease mechanisms and aid in pinpointing therapeutic targets.

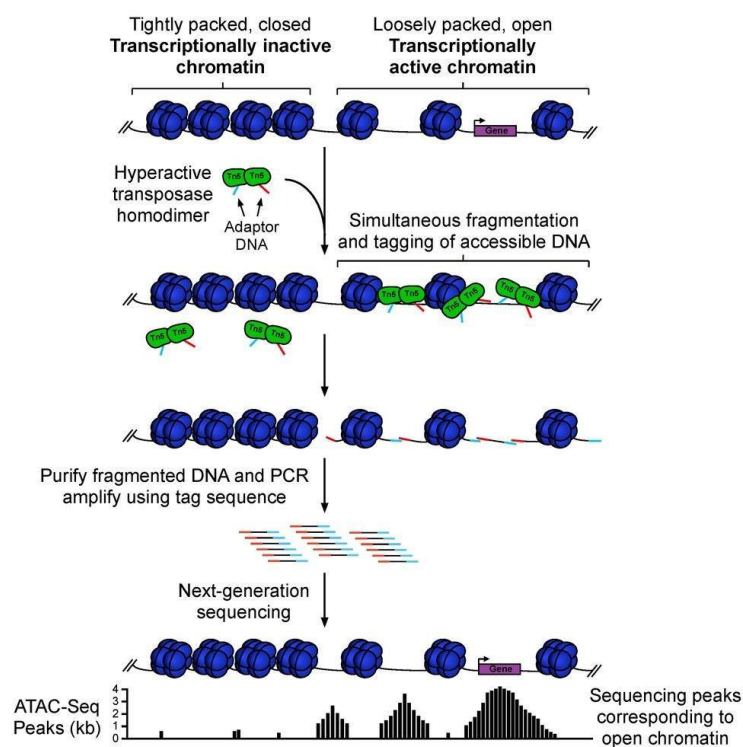


Figure 2. Overview of the Mechanism of Assay for Transposase-Accessible Chromatin Sequencing (ATAC-seq).

Figure 2 is a schematic of the ATAC-seq workflow for identifying chromatin accessibility. The Tn5 transposase (green tags) inserts sequencing adapters (light blue and red) into accessible chromatin regions, marking nucleosome-depleted areas. The blue shapes represent histone octamers. This process enables the mapping of open chromatin regions, which are crucial for understanding gene regulation.

Figure adapted from Davis (no date).

Role of Nucleosome Positioning in RA: Biomarker for Early Diagnosis

Despite extensive research on cell-free DNA (cfDNA), which consists of extracellular fragments of primarily double-stranded nuclear and mitochondrial DNA circulating in body fluids such as blood plasma (Kustanovich *et al.*, 2019), significant gaps remain in understanding its role in disease. cfDNA's in vivo nucleosome footprint reveals its tissue of origin (Qi *et al.*, 2023); yet, while research has primarily focused on using cfDNA concentration to assess its correlation with disease activity in RA, inconsistent results have suggested that cfDNA alone may not be a reliable biomarker. For instance, Dunaeva *et al.* (2015) and Jørgensen *et al.* (2011) reported lower cfDNA levels in the sera of patients with established RA compared to those in healthy samples. In contrast, other studies (Bartoloni *et al.*, 2011; Zhong *et al.*, 2007; Leon *et al.*, 1981) observed increased cfDNA concentrations in patients with RA. These differences may arise from variations in techniques and DNA targets, but they suggest that cfDNA is not a reliable diagnostic biomarker or tool for tracking disease progression (Dunaeva *et al.*, 2015). Nucleosome positioning, which has been linked to various diseases such as cancer, metabolic disorders, autoimmune diseases, and neurodegenerative conditions, is emerging as a promising biomarker (Penny *et al.*, 2024; Fahmueller *et al.*, 2012; Yehya, Thomas and Margulies, 2016). While it has been extensively studied in colorectal, lung, breast, and ovarian cancers (Fahmueller *et al.*, 2012; Jacob *et al.*, 2024; Vanderstichele *et al.*, 2022; Wei *et al.*, 2015), its role in autoimmune diseases like RA remains underexplored.

Evidence suggests that altered chromatin accessibility and nucleosome positioning contribute to RA. In a study by Jadhav *et al.* (2022), chromatin accessibility in CD4 T cells revealed changes in nucleosome positioning at RA susceptibility loci, suggesting its potential relevance in RA pathogenesis. Since nucleosome positioning influences gene expression and immune regulation, analysing its patterns in RA could serve as a biomarker for early diagnosis and treatment response, improving precision medicine approaches. RA, a complex autoimmune disease with an unclear aetiology, impacts quality of life and can lead to disability. Early diagnosis is essential, but research on nucleosome positioning in RA is limited compared to that on autoimmune diseases like systemic lupus erythematosus and lupus flares (Ghiggeri *et al.*, 2019), highlighting a significant research gap. This study aims to investigate chromatin accessibility by analysing nucleosome positioning in RA using ATAC-seq data from the NCBI GEO Database and bioinformatics tools such as Bowtie, MACS2, BEDTools, OriginPro, the UCSC Genome Browser, and WebGestalt. It focuses on the potential of altered nucleosome positioning in patients with RA as a biomarker for early diagnosis, which could also help track treatment progress and lead to personalised medicine and targeted therapies (Cribbs, Feldmann and Oppermann, 2015).

Methods

Figure 3 shows an outline of the methods used in this study.

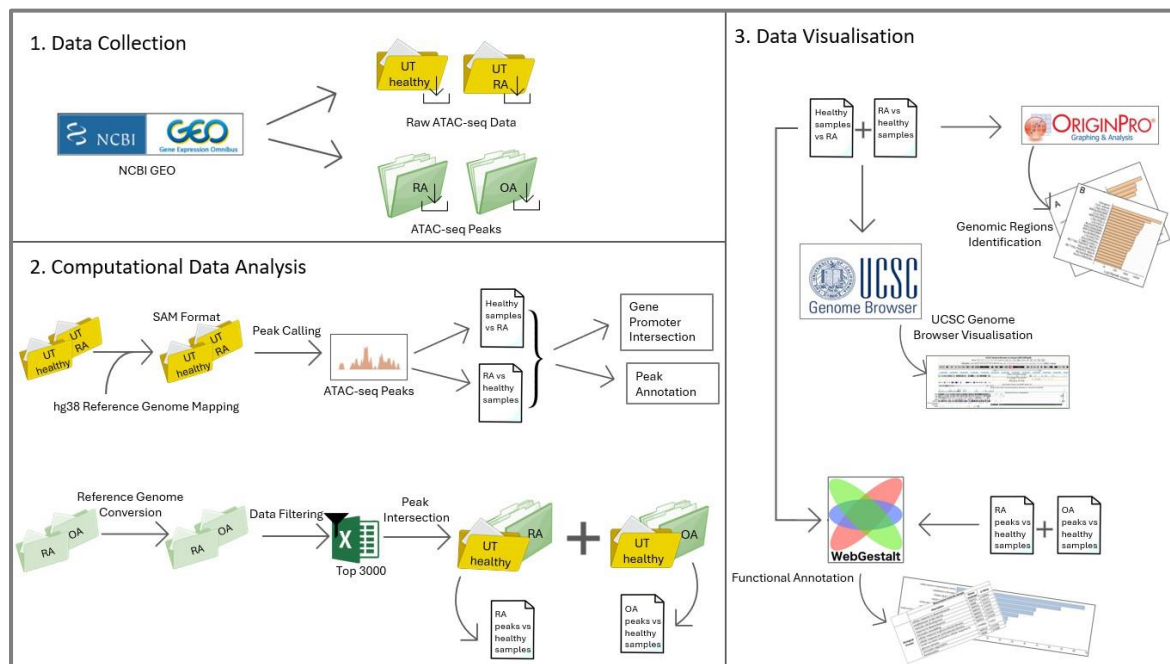


Figure 3. Overview of the Methods for Studying RA and Nucleosome Positioning.

This flow chart shows the steps involved in investigating RA and the role of nucleosome positioning in early diagnosis. It outlines the key stages of data collection, processing using computational analysis, and data visualisation using different bioinformatics tools to analyse chromatin accessibility and nucleosome positioning in healthy samples vs RA.

ATAC-seq Datasets

Two datasets were downloaded from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) (Edgar, Domrachev and Lash, 2002), a publicly accessible repository for functional genomics data, which includes high-throughput gene expression and epigenomic datasets database as detailed below.

The first dataset (GEO accession: GSE148403) used in this study contains ATAC-seq data from Fibroblast-Like Synoviocytes (FLS) isolated from three RA-FLS patients and three healthy donors (Krishna *et al.*, 2021). Only untreated RA and untreated healthy samples were used for this analysis. Experimental data of this dataset was generated according to the protocol described by Buenrostro *et al.* (2013). ATAC-seq was performed by Epinomix (San Francisco, CA), and the resulting FASTQ

files were processed in-house using a custom analysis pipeline. Raw data, provided in the Sequence Read Archive (SRA) format, were used for this analysis.

The second dataset (GEO accession: GSE112658) used in this study included 22 open chromatin datasets, which were profiled by ATAC-seq to identify regions of accessible chromatin (Ai *et al.*, 2018). This dataset is based on synovial tissues obtained from patients with RA and patients with OA during total joint replacement or synovectomy, with diagnoses confirmed based on American College of Rheumatology criteria (Ai *et al.*, 2018). The mean ages of the RA and patients with OA were 55 ± 9 and 66 ± 12 years, respectively, with joint locations including knee, hip, and hand for RA, and knee and hip for OA. ATAC-seq sample preparation followed a previously established protocol (Buenrostro *et al.*, 2013). Raw reads were mapped by the authors to the hg19 genome using Bowtie, followed by the normalisation of read depth. Differentially modified epigenetic regions were identified using the DiffBind package with a q-value threshold of <0.05 . Peak calling was performed using MACS2 (version 2.1.0.20150420.1) on the processed data (Ai *et al.*, 2018).

Computational Data Analysis

The datasets obtained from NCBI GEO were analysed using Ceres, the University of Essex's High-Performance Computing Cluster (HPCC). Remote access to the cluster was established through UNIX commands via PuTTY v0.83 (Tatham, 2011), a secure shell and Telnet client, ensuring secure connections to the remote servers for data analysis. PuTTY was used to run files, make directories, and download datasets from GEO. WINSXP v6.3.5 (<https://winscp.net/>) was used to transfer and organise files.

The raw data from the dataset published by Krishna *et al.* (2021), in SRA format, was mapped using Bowtie2 v2.5.1 (Langmead and Salzberg, 2012) to the human genome assembly hg38, converting them to SAM format. The SAM files were then used to call peaks using MACS2 v2.2.9.1 (Zhang *et al.*, 2008). Two types of peaks were called: (1) untreated RA vs untreated healthy (referred to as RA vs healthy samples) and (2) untreated healthy vs untreated RA (referred to as healthy samples vs RA).

The resulting NarrowPeak files were intersected with the gene promoters on hg38, where the promoters were defined as the $[-1000, 1000]$ regions around transcription start sites (based on the RefSeq annotation provided by my supervisor) using the command BEDTools v2.29.2 (Quinlan and Hall, 2010). Peak annotation was then performed using HOMER 5.1 (Heinz *et al.*, 2010) with the annotatePeaks.pl script based on the hg38 genome. This generated a text file containing information about nearby genes, genomic features, and other relevant annotations.

For the second dataset published by Ai *et al.* (2018), ATAC-seq peaks were directly available on GEO and downloaded using the wget command. Since the dataset was mapped to hg19, the peaks

were first converted to hg38 using the liftOver tool (Hinrichs *et al.*, 2006) of the UCSC Genome Browser before any further analysis (Kent *et al.*, 2002). Since the peaks were in a non-standard format (merged multiple peaks and reporting genomic windows with peak scores), filtering was applied to select the top 3000 genomic regions with the highest peak scores. The top 3000 peaks from each RA and OA sample were intersected with the untreated healthy dataset, showing peaks present in (i) RA and not in healthy (RA peaks vs healthy samples) and (ii) OA and not in healthy (OA peaks vs healthy samples). These resulting peaks were further intersected with the gene promoters on hg38, defined as the [-1000, 1000] regions around transcription start sites, to connect them to the gene annotations.

Data Visualisation

The files generated from the peak annotation were visualised using OriginPro (<https://www.originlab.com/>), a software for data analysis, graphing, and visualisation. For each dataset (RA vs healthy samples and healthy sample vs RA), a bar graph was constructed to visualise the regions where the peaks were most enriched. A logarithmic scale was applied to the x-axis to better visualise small changes. The y-axis displayed the genomic regions, with the most enriched regions plotted at the top. The genomic locations of these peaks were further visualised using the UCSC Genome Browser, focusing on the centromeric region of chromosome 1. Custom tracks for the healthy samples vs RA and RA vs healthy samples datasets were added to complement the bar graph visualisations. The files generated from the intersection with promoter regions provided a list of overlapping genes, which were functionally annotated using the WEB-based Gene Set Analysis Toolkit (WebGestalt) (Elizarraras *et al.*, 2024). Gene Ontology labels were assigned, and enriched terms were identified for these genes. For the RA vs healthy samples dataset, no terms were enriched at an FDR <0.05. Therefore, the top 10 terms across three categories—biological processes, cellular components, and molecular functions—were selected. For the healthy samples vs RA dataset, an FDR <0.05 was applied across the same categories. To facilitate visualisation, tables with gene names were created, and bar graphs from WebGestalt were imported as they were. For comparative analysis, functional annotation was also conducted for RA peaks vs healthy samples and OA peaks vs healthy samples. For RA peaks vs healthy samples, the top 10 terms across all categories were used. For OA peaks vs healthy samples, the thresholds were set as follows: FDR <0.05 for biological processes, <0.1 for cellular components, and <0.9 for molecular functions. For all functional annotations, the genome hg38 was selected as the reference set.

Results

Genomic Distribution of ATAC-seq Peaks in RA and Healthy Samples

To understand how chromatin accessibility differs between RA and healthy samples, a comparative analysis was conducted. To investigate these differences, ATAC-seq data from both groups, as reported by Ai *et al.* (2018) and Krishna *et al.* (2021), were analysed. The goal was to identify regions with variations in chromatin accessibility between conditions. To do this, two subsets of ATAC-seq peaks were defined: (1) those which are not present in healthy but appear in RA and (2) those which are present in healthy but disappear in RA.

Firstly, the types of genomic regions enriched in each of these two subsets of ATAC-seq peaks were examined. Figure 4 shows the distribution of these two subsets of ATAC-seq peaks across the genome, with RA compared to healthy samples in panel A and healthy samples compared to RA in panel B. The x-axis represents the number of peaks on a log scale, and the y-axis indicates the types of genomic regions where these peaks are found.

In Figure 4A, which details the annotation of peaks unique to RA (not present in healthy), the most enriched peaks were found in repetitive DNA regions, specifically the ALR/Alpha satellite and SST1 satellite regions, followed by intergenic regions. The high peak frequency in intergenic regions suggests these non-coding DNA stretches may influence chromatin remodelling. Simple repeat elements, like (AATGGAATGG)_n and (AATGG)_n, also showed high peak frequency. These repeats, often found in promoters, untranslated regions, or coding sequences, are prone to mutations that can alter repeat numbers and affect gene function (Fondon *et al.*, 2008), suggesting a potential role in disease progression. Enrichment was also observed in Alu elements, like AluSc8, which can influence gene expression when located near a gene (Deininger, 2011). Alu elements are known to promote stable DNA–histone interactions, which can affect nucleosome positioning, chromatin accessibility, and gene regulation (Englander and Howard, 1995). Peaks were further enriched in CpG islands and genic regions, including exons and introns.

Overall, the enrichment of peaks in satellite regions and simple repeat elements suggests that these regions may contribute to chromatin organisation and genome stability in RA compared to healthy samples. In contrast, ATAC-seq peaks present in healthy samples but not present in RA were mostly enriched in intergenic regions and CpG islands (Figure 4B). The enrichment of intergenic regions suggests that in the absence of disease, chromatin is more accessible in non-coding regions, potentially facilitating regulatory activity under normal physiological conditions. CpG islands, which play a role in DNA methylation and epigenetic regulation, were the second most enriched. Several LINE and SINE elements, including L2, L1, Alu, MIR, and CR1, were also enriched in ATAC-seq peaks that are present in healthy samples but lost in RA. These elements regulate gene expression by modifying chromatin structure.

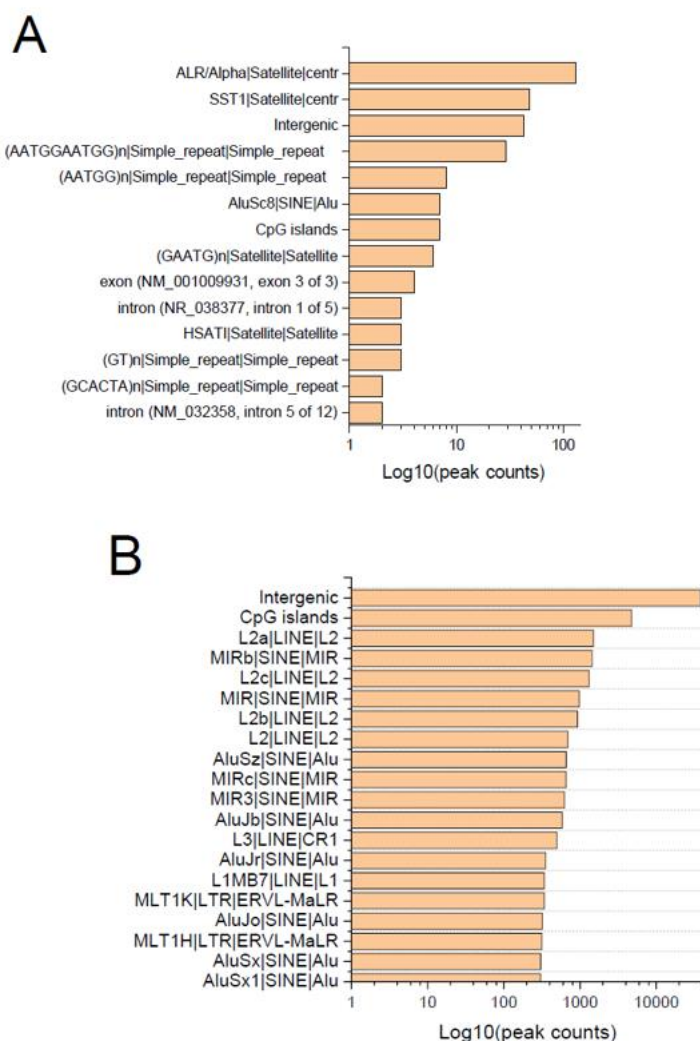


Figure 4. Distribution of ATAC-seq Peaks in RA vs Healthy Samples.

A) ATAC-seq peaks that are present in RA but not in healthy samples. B) ATAC-seq peaks that are present in healthy samples but not in RA. The x-axis represents the log₁₀ (peak counts), while the y-axis lists the type of annotated genomic regions. The regions are ordered by peak frequency, with the most enriched regions at the top. In RA samples, peaks are most enriched in satellite regions, whereas peaks in healthy samples are predominantly in intergenic regions and CpG islands.

To follow up on the finding that most of the peaks unique to RA are enriched in repetitive DNA regions rather than gene bodies, peaks were visualised using the UCSC Genome Browser. Figure 5 shows a screenshot of the UCSC Genome Browser displaying the centromere region on chromosome 1.

In the blue box, the healthy vs RA ATAC-seq peaks are shown. Most of these peaks align with reference sequence genes from NCBI (RefSeq genes), represented by dark blue boxes and connecting lines that indicate exons and introns. In contrast, the yellow box shows the RA vs healthy ATAC-

seq peaks, where most of the peaks are located within repeating elements, highlighted in the green box. These elements are predominantly satellite regions, with one peak aligning with a known genetic variant. This pattern reinforces the idea that chromatin remodelling in RA may be driven by structural changes in repeat elements rather than alterations in gene-coding regions.

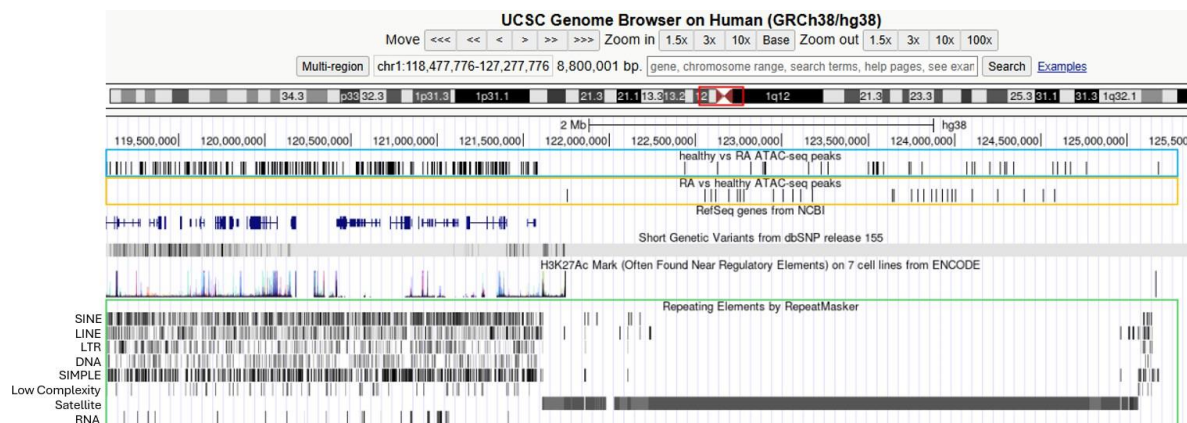


Figure 5. UCSC Genome Browser Visualisation of ATAC-seq Peaks in RA and Healthy Samples.

Screenshot of the UCSC Genome Browser showing the centromeric region on chromosome 1. This figure illustrates the genomic locations of peaks in each sample, highlighting their enrichment in repetitive regions. The blue box highlights ATAC-seq peaks from healthy samples compared to those in RA, while the yellow box shows ATAC-seq peaks from RA compared to those in healthy samples. The green box indicates DNA sequence repeats elements, where most RA-specific peaks are located.

Functional Annotation Analysis of RA, OA, and Healthy Samples

ATAC-seq peaks were mapped to the hg38 reference genome to analyse genes marked by accessible chromatin in RA but absent in healthy samples. Subsets of peaks appearing or disappearing between RA, OA, and healthy conditions were defined, then intersected with annotated gene promoters to identify RA-specific peaks and explore their potential role in disease onset. The results of this analysis are presented in Table 1, which lists nine genes found only in patients with RA, along with their chromosomal locations. Most of these genes are located on the Y chromosome, while the others are distributed across autosomal chromosomes. This suggests that both sex chromosomes and autosomal genes may contribute to the regulation of chromatin accessibility in RA.

Table 1. Genes Whose Promoters are Marked by ATAC-seq Peaks Appearing in RA vs Healthy Samples and their Chromosomal Locations.

Chromosome No.	Gene
----------------	------

chr Y	OFD1P1Y
chr Y	UTY
chr Y	DDX3Y
chr 22	FAM118A
chr 13	LOC124903130
chr 7	PTPRN2
chr 4	USP17L22
chr 1	SMYD3
chr 1	HRNR

To better understand how these genes influence RA progression, functional annotation using WebGestalt was performed to explore the biological processes, molecular functions, and cellular components associated with genes influencing RA progression. Table 2 summarises the most enriched processes, components, and functions for the RA vs healthy samples and healthy vs RA datasets.

Table 2. Summary of Functional Annotation in RA vs Healthy Samples and Healthy Samples vs RA Datasets.

	RA vs Healthy Samples		Healthy Samples vs RA	
	Description	p-value	Description	p-value
Biological Process	cellular response to dexamethasone	0.00961	cellular response to stress	1.8179e-55
Cellular Component	MLL3/4 complex	0.00472	catalytic complex	1.30E-79
Molecular Function	histone H3K27me2/H3K27me3 demethylase activity	0.00201	RNA binding	7.69E-67

The table summarises the most enriched biological processes, cellular components, and molecular functions identified from the functional annotation of genes present in RA but absent in healthy samples and vice versa. The p-values show the statistical significance of the enrichment.

Figure 6 illustrates the biological processes associated with ATAC-seq genes present in promoter regions in RA but absent in healthy samples. The graph reveals that the most enriched biological process is the cellular response to dexamethasone stimulus, with an enrichment ratio of 103.69. Genes in this category are also enriched in pathways related to glucocorticoid (51.85) and

corticosteroid stimulus (43.75), myotube cell development (65.11), and skin barrier establishment (84.84). These findings suggest that these pathways may play a role in immune regulation, inflammation, and muscle maintenance, potentially contributing to disease progression in RA.

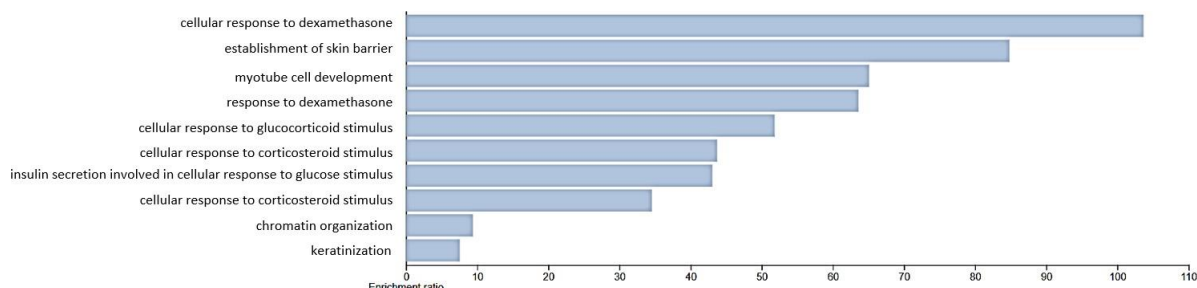


Figure 6. Biological Processes Enriched in Genes Present in RA and Not in Healthy Samples.

The bar chart shows the biological processes associated with genes present in RA but not in healthy samples. Cellular response to dexamethasone has the highest enrichment ratio, followed by establishment of skin barrier. Chromatin remodelling and chromatin organisation has the lowest enrichment ratio.

The genes involved in each of these processes are listed in Table 3, along with their corresponding p-values. As shown in the table, SMYD3 is consistently associated with processes related to steroid response and chromatin regulation, suggesting its potential role in transcriptional regulation in RA.

Table 3. Genes Associated with Enriched Biological Processes in RA Identified from Functional Annotation.

	Rheumatoid Arthritis Sample		
	Description	Genes	p-value
		Involved	
Biological Process	cellular response to dexamethasone	SMYD3	0.00961
	establishment of skin barrier	HRNR	0.0117
	myotube cell development	SMYD3	0.0153
	response to dexamethasone	SMYD3	0.0156
	chromatin remodelling	SMYD3	0.0171
	cellular response to glucocorticoid stimulus	SMYD3	0.0191
	cellular response to corticosteroid stimulus	SMYD3	0.0226
	insulin secretion involved in cellular response to glucose stimulus	PTPRN2	0.023

chromatin organisation	SMYD3	0.0258
keratinisation	HRNR	0.0286

The table lists the specific genes involved in the enriched biological processes, along with their p-values identified through a functional annotation analysis of ATAC-seq peaks present in RA and not in healthy samples.

Figure 7 displays the molecular functions of these genes and their enrichment ratios. The most enriched molecular function is histone H3K27me2/H3K27me3 demethylase activity, with an enrichment ratio of 496.03. Methylation of H3K27 is typically associated with gene repression; therefore, demethylation could either activate or repress genes involved in immune function, potentially influencing RA progression (Pan *et al.*, 2018). Other enriched functions include histone modifying activity (23.96) and RNA polymerase II intronic transcription regulatory region sequence-specific DNA binding (310.02). Histone modifications can alter chromatin structure, increasing chromatin accessibility to transcription factors, which could potentially elevate the expression of genes involved in inflammation, contributing to disease progression (Nemtsova *et al.*, 2019). Most of the enriched molecular functions are linked to histone modifications, which may alter chromatin structure and influence the activation and repression of genes impacting RA progression.

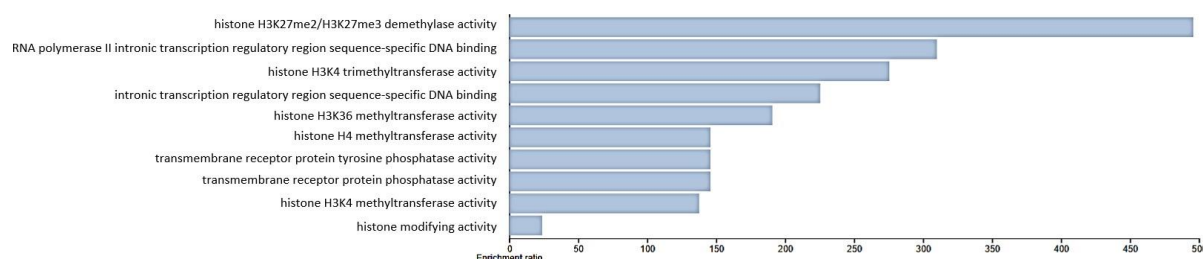


Figure 7. Molecular Functions of Genes Marked by ATAC-seq Peaks Present in RA and Not in Healthy Samples.

The bar chart shows the molecular functions of genes present in RA but not in healthy samples. Histone H3K27me2/H3K27me3 demethylase activity, RNA polymerase II intronic transcription regulatory region sequence-specific DNA binding, and histone H3K4 trimethyltransferase activity are the most enriched functions, whereas histone modifying activity has the lowest enrichment ratio.

Table 4 lists the genes involved in each process, along with their p-values. Most histone modifications are linked to *UTY* and *SMYD3*, with *PTPRN2* associated with transmembrane receptor protein phosphatase activities. *SMYD3* shows higher molecular function activity, and both *UTY* and *SMYD3* are linked to histone-modifying activities, suggesting altered chromatin structure and accessibility that may contribute to RA progression.

Table 4. Genes Associated with Enriched Molecular Functions in RA Identified from Functional Annotation.

	Rheumatoid Arthritis Sample		
	Description	Genes Involved	p-value
Molecular Function	histone H3K27me2/H3K27me3 demethylase activity	UTY	0.00201
	histone modifying activity	SMYD3 and UTY	0.00286
	RNA polymerase II intronic transcription regulatory region sequence-specific DNA binding	SMYD3	0.00322
	histone H3K4 trimethyltransferase activity	SMYD3	0.00362
	intronic transcription regulatory region sequence-specific DNA binding	SMYD3	0.00443
	histone H3K36 methyltransferase activity	SMYD3	0.00523
	histone H4 methyltransferase activity	SMYD3	0.00684
	transmembrane receptor protein tyrosine phosphatase activity	PTPRN2	0.00684
	transmembrane receptor protein phosphatase activity	PTPRN2	0.00684
	histone H3K4 methyltransferase activity	SMYD3	0.00724

The table lists the specific genes involved in the enriched molecular functions and their p-values identified through a functional annotation analysis of ATAC-seq peaks present in RA and not in healthy samples.

Figure 8 displays the cellular components of these genes and their enrichment ratios. The MLL3/4 complex shows the highest enrichment ratio (211.43), followed by P granules, pole plasm, and germ plasm (97.58), with secretory granules having the lowest ratio (5.69). The MLL3/4 complexes are known to be involved in histone modifications. They are recruited to target enhancers by interacting with sequence-specific transcription factors, including ligand-dependent nuclear receptors and pioneer factors, which initiate the activation of de novo enhancers (Wang *et al.*, 2021). These complexes may activate enhancers that regulate genes involved in immune signalling and inflammation, potentially leading to increased transcription of pro-inflammatory cytokines.

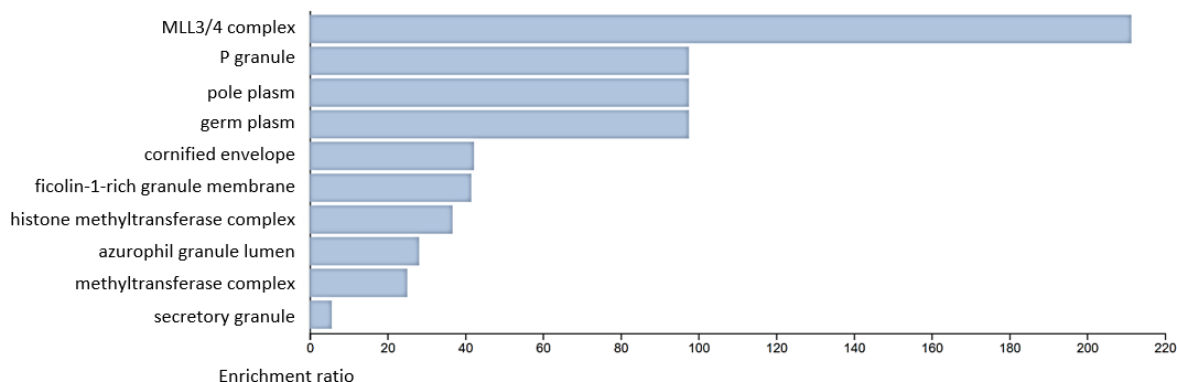


Figure 8. Cellular Components of Genes Marked by ATAC-seq Peaks Present in RA and Not in Healthy Samples.

The bar chart shows the cellular components associated with genes present in RA but not in healthy samples. MLL3/4 complex, P granule, germ plasma, and pole plasma are the most enriched cellular components, whereas secretory granule has the lowest enrichment ratio.

The genes involved in each of these processes are listed in Table 5, along with their p-values indicating the statistical significance of their enrichment. Gene *UTY* is linked to the MLL3/4 complex, histone methyltransferase membrane, and methyltransferase complex, confirming its role in histone modification and enhancer activation. *DDX3Y* is associated with P granules, pole plasm, and germ plasm, while *HRNR* is linked to the cornified envelope, azurophil granule lumen, and secretory granule. *PTPRN2* is associated with cellular components such as the ficolin-1-rich granule membrane and secretory granule.

Table 5. Genes Associated with Enriched Cellular Components in RA Identified from Functional Annotation.

	Rheumatoid Arthritis Sample		
	Description	Genes	p-value
		Involved	
Cellular Components	MLL3/4 complex	UTY	0.00472
	P granule	DDX3Y	0.0102
	pole plasm	DDX3Y	0.0102
	germ plasm	DDX3Y	0.0102
	cornified envelope	HRNR	0.0234
	ficolin-1-rich granule	PTPRN2	0.0238
	membrane		
	histone methyltransferase	UTY	0.0269
	complex		
	azurophil granule lumen	HRNR	0.0349
	methyltransferase complex	UTY	0.0391
	secretory granule	HRNR and PTPRN2	0.0446

The table lists the specific genes involved in the enriched cellular functions, along with their p-values, identified through a functional annotation analysis of ATAC-seq peaks present in RA and not in healthy samples.

For the other dataset used for this analysis, genes from ATAC-seq peak mapping were functionally annotated to compare their functions with those observed in RA. Table 6 presents a comparison between the RA peaks vs healthy samples dataset and the OA peaks vs healthy samples dataset. The table outlines the enriched biological processes, molecular functions, and cellular components for both datasets. The RA peaks vs healthy samples dataset was used to support the RA vs healthy samples results, while the OA peaks vs healthy samples dataset was used to compare how the functions of OA differ from those observed in RA.

The biological processes in both RA datasets show some overlap, such as responses to stimuli. The genes in both datasets are involved in responses to various hormones and stimuli. While both datasets also include developmental processes, they differ in biological contexts: the RA peaks vs healthy samples dataset is linked to brainstem development, while the RA vs healthy samples

dataset is associated with myotube cell development and skin barrier establishment. Despite these differences, both datasets are involved in processes related to chromatin organisation or modification. For example, chromatin remodelling and chromatin organisation are found in the RA vs healthy samples dataset, while the RA peaks vs healthy samples dataset has an indirect connection to chromatin organisation.

Similarly, overlaps exist in the enriched cellular components and molecular functions. Both datasets are enriched for terms related to vesicle trafficking, exocytosis, and protein secretion. Additionally, histone modification functions in the RA vs healthy samples dataset overlap with Formyl- and Related Transferase Activity, with both datasets involved in enzymatic transfer activities that could potentially impact the methylation and chemical modification of molecules, influencing gene regulation and chromatin remodelling. These overlaps validate the peaks identified in our analysis, demonstrating consistency with a dataset where peaks had been previously identified. The similarities between the results and the established dataset further support the reliability of the peak calling process.

Table 6 further displays the enriched biological processes, cellular components, and molecular functions for the OA peaks vs healthy samples, highlighting the differences in the processes involved in OA compared to those in RA. For instance, OA is associated with negative regulation of apoptosis, suggesting protective mechanisms in joint tissue, whereas RA exhibits increased apoptosis due to autoimmune inflammation (Xu *et al.*, 2015). Another key difference is that OA is more related to structural and extracellular matrix processes, such as the smooth endoplasmic reticulum and cytoplasmic vesicle lumen. In contrast, RA involves more pronounced immune cell activation, cytokine secretion, and synovial inflammation, indicating a stronger focus on immune responses and synovial cell granules in RA. The major distinction between the OA and RA datasets is that OA primarily involves the structural degradation of cartilage, without the autoimmune-driven destruction that characterises RA.

Table 6. Comparative Analysis of the Enriched Biological Processes, Cellular Components, and Molecular Functions Between RA and OA Peaks vs Healthy Samples.

	RA Peaks vs Healthy Samples		OA Peaks vs Healthy Samples	
	Description	p-value	Description	p-value
Biological Process	tetrahydrofolate biosynthetic process	0.002	negative regulation of apoptotic process	2.20E-06
	response to corticotropin-releasing hormone	0.00208		
	cellular response to corticotropin-releasing hormone stimulus	0.00208	negative regulation of programmed cell death	2.74E-06
	de novo IMP biosynthetic process	0.0025		
	XMP metabolic process	0.00291	negative regulation of extrinsic apoptotic signalling pathway via death domain receptors	4.40E-06
	XMP biosynthetic process	0.00291		
	de novo XMP biosynthetic process	0.00291		
	folate-containing compound biosynthetic process	0.00291	regulation of extrinsic apoptotic signalling pathway via death domain receptors	0.0000181
	de novo AMP biosynthetic process	0.00333		
	brainstem development	0.00333		
	Cul2-RING ubiquitin ligase complex	0.0112	platelet alpha granule	0.000119
	receptor complex	0.0133	secretory granule	0.000242
	microvillus membrane	0.0148	cytosolic ribosome	0.000251
	brush border membrane	0.0267		0.000324
Cellular Component	clathrin-coated pit	0.0342	secretory granule	

			lumen	
	microvillus	0.0429	cytoplasmic vesicle	0.000335
	brush border	0.0468	lumen	
	axon terminus	0.0498	vesicle lumen	0.000339
	neuron projection	0.0567	smooth endoplasmic	0.000433
	terminus		reticulum	
	cluster of actin-based	0.0711	secretory vesicle	0.000635
	cell projections			
Molecular Function	cyclohydrolase activity	0.00231	structural molecule	0.00103
	LBD domain binding	0.00322	activity	
	hydroxymethyl-, formyl- and related transferase activity	0.00368	peptidase regulator activity	0.00172
	zinc ion binding	0.00504		
	oestrogen response	0.00506	enzyme regulator	0.00181
	element binding		activity	
	cadherin binding	0.00913		
	transition metal ion binding	0.0111	protein serine/threonine kinase activator activity	0.00199
	TBP-class protein binding	0.0115		
	nuclear steroid receptor activity	0.0115	glycosaminoglycan binding	0.00209
	metallocarboxypeptidase activity	0.0137		

This table presents the significant biological processes, cellular components, and molecular functions identified in the ATAC-seq peak calling comparing RA peaks vs healthy samples and OA peaks vs healthy samples. The p-value shows the statistical significance of the comparisons.

Discussion

This study investigated whether nucleosome positioning can serve as a biomarker for the early diagnosis of RA, which is crucial as it can alter the disease course, prevent joint erosion, and slow the progression of erosive disease. It can also improve overall disease outcomes, potentially leading to remission (Heidari, 2011). Timely diagnosis helps prevent irreversible joint damage, which typically occurs within the first two years of onset when the rate of erosive joint disease is highest (Visser, 2005). However, many cases of early arthritis are self-limiting, making it difficult to distinguish RA from other types of arthritis. Early differentiation is important to avoid the unnecessary use of disease-modifying antirheumatic drugs, thereby reducing the risk of harm from inappropriate treatment (Visser, 2005; Heidari, 2011).

Nucleosome positioning plays a critical role in regulating gene expression by influencing chromatin accessibility. Altered nucleosome positioning can disrupt gene expression, potentially driving disease onset and progression. Changes in nucleosome positioning, such as nucleosome sliding, can lead to the activation of disease-causing genes, contributing to the development of disease (Jiang and Pugh, 2009). This study therefore investigated nucleosome positioning as a potential biomarker for the early diagnosis of RA. Two ATAC-seq datasets were used: the first focused on FLS from three patients with RA and three healthy donors (Krishna *et al.*, 2021), and the second contained ATAC-seq data from 22 patients with RA and OA (Ai *et al.*, 2018).

ATAC-seq data were used in this analysis, as ATAC-seq identifies open chromatin regions associated with active transcription and regulatory activity. This technique efficiently maps both chromatin accessibility and nucleosome positioning, making it a powerful tool for studying epigenetic regulation (Yan *et al.*, 2020). Additionally, ATAC-seq requires a low number of cells and can provide insights into chromatin structure and gene regulation. These advantages make it ideal for investigating chromatin changes that may influence gene expression.

The tetrahydrofolate biosynthetic process is indirectly linked to chromatin structure through its role in one-carbon metabolism and histone methylation. It contributes to the synthesis of S-adenosylmethionine, a key methyl donor for histone methylation, influencing gene expression and chromatin structure (Serefidou, Venkatasubramani and Imhof, 2019).

Nucleosome Positioning and Chromatin Accessibility in RA

The analysis revealed a significant difference in the number of peaks between RA vs healthy samples and healthy samples vs RA. Specifically, the RA vs healthy samples dataset had a lower number of peaks (n=335) compared to the healthy samples vs RA dataset (n=157184), indicating potential differences in chromatin accessibility. A significant reduction in peaks was observed in intergenic regions and CpG islands, while satellite regions showed an increased enrichment of ATAC-seq peaks. These findings suggest that nucleosome positioning is altered in RA, with

reduced accessibility in gene-regulatory regions (CpG islands and intergenic regions) and an increased number of peaks in satellite repeat regions, which may reflect chromatin remodelling associated with disease progression. Additionally, a functional analysis of genes overlapping with the peaks revealed that most were associated with histone modifications, highlighting epigenetic regulation in RA. A consistent involvement of the SMYD3 gene was observed, suggesting a potential role in RA-associated chromatin alterations. SMYD3 is known for its histone methyltransferase activity (Yang *et al.*, 2021). Its persistent presence in the dataset may indicate a link between histone modifications and nucleosome positioning in RA.

The functional annotation revealed a significant enrichment of histone H3K27me2/H3K27me3 demethylation activity within genes enriched in RA but not in healthy samples. This suggests that DNA demethylation may occur in RA, potentially altering gene expression. One possible explanation for this hypomethylation is the absence of *DNMT1* in RA samples, as indicated by the lack of *DNMT1* peaks in the RA vs healthy samples dataset, whereas *DNMT1* peaks were present in the healthy vs RA dataset. This aligns with previous studies that suggest the absence of *DNMT1* is associated with DNA hypomethylation in RA (Turek-Plewa and Jagodzinski, 2005; Szyf, 2001; Wilson, Power and Molloy, 2007). Additionally, another study confirmed that RA is characterised by DNA hypomethylation and decreased *DNMT1* levels (Payet *et al.*, 2021). Interestingly, even though *DNMT3A* and *DNMT3B* peaks were observed in the healthy samples, they were absent in the RA samples. These enzymes are responsible for de novo methylation during primary development (Gujar, Weisenberger and Liang, 2019). The absence of *DNMT1*, *DNMT3A*, and *DNMT3B* peaks is shown in Figure 9. This further supports the idea that the absence of DNA methylation genes may contribute to the epigenetic alterations observed in RA, providing more evidence for the potential role of DNA methylation changes in disease progression.

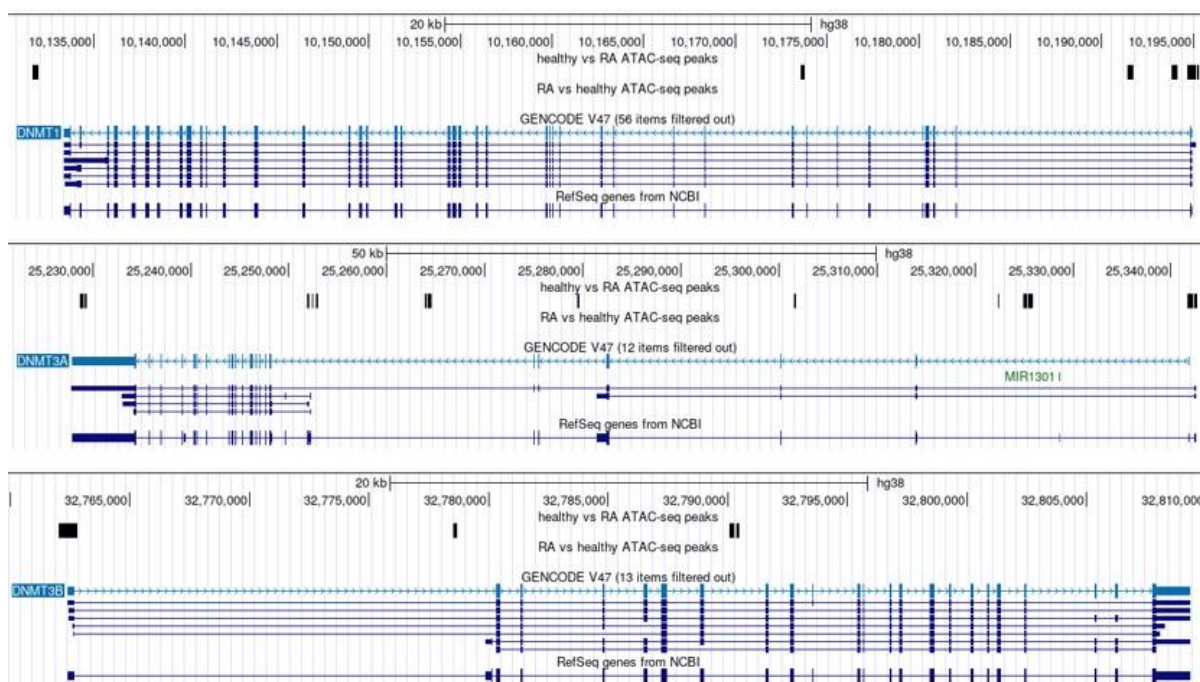


Figure 9. Snapshot of the UCSC Genome Browser Showing the Absence of *DNMT1*, *DNMT3A*, and *DNMT3B* Peaks in RA Compared to Healthy Samples.

This figure displays ATAC-seq peaks from the UCSC Genome Browser, comparing RA and healthy samples. It highlights the differences in peaks present for the *DNMT1*, *DNMT3A*, and *DNMT3B* genes, showing the absence of peaks in RA compared to those in healthy samples.

Furthermore, the results show an enrichment of methyltransferase activity in RA samples, suggesting that the methylation of specific genes may occur in RA but not in healthy individuals. A study by Nakano, Boyle and Firestein (2013) demonstrated that DNA methylation plays a key role in repressing regulatory genes by methylating cytosines in CpG islands; this supports our findings, as reduced peaks were observed at CpG sites in the RA vs healthy samples dataset. Nakano, Boyle and Firestein (2013) also showed that the expression of *DNMT* genes is suppressed by proinflammatory cytokines such as IL-1 β , TNF, and the TLR4 ligand LPS. Since *DNMTs* are responsible for methylating CpG sites, the suppression of *DNMT* function may lead to reduced methylation at CpG sites, which is consistent with the hypomethylation observed in our RA samples. While CpG islands exhibit fewer enriched peaks compared to other genomic regions such as satellite regions, intergenic regions, and simple repeats, they remain statistically significant. This reduction in methylation could be attributed to the decreased levels of *DNMTs* in RA, contributing to the altered DNA methylation landscape observed in our study.

Nucleosome Positioning and Hypomethylation in Repetitive Regions

The enrichment of ALR/Alpha satellite and SST1 satellite regions in RA compared to healthy samples suggests that these regions may be more accessible in RA due to hypomethylation. In healthy samples, these regions are typically silenced by methylation to maintain genome integrity

(Papin et al., 2017). However, in RA, hypomethylation could lead to their activation, resulting in chromatin instability. Retrotransposons, which play a critical role in maintaining chromatin structure, are also influenced by methylation status, as their hypomethylation can disrupt chromatin organisation (Liu and Leung, 2025).

The findings demonstrate clear alterations in chromatin structure in RA compared to those in healthy samples, suggesting that in the presence of disease, CpG sites and genic regions become less accessible, leading to dysregulated gene expression. The enrichment of ATAC-seq peaks in repetitive regions indicates a more open chromatin state in these areas, likely due to nucleosome repositioning and hypomethylation, facilitating their activation. Conversely, the loss of peaks in intergenic regions and CpG islands suggests a shift toward a more closed chromatin conformation, restricting access to regulatory elements and contributing to altered gene regulation. One interesting finding in the results is the presence of three Y-linked genes, specifically *UTY*, which is associated with histone demethylase activity. This is notable since RA is more common in women, partly due to the role of oestrogen in immune regulation (Gerosa et al., 2008). Gaining a deeper understanding of the dataset, including factors such as sample composition and sequencing biases, could help clarify the significance of this observation.

Future Directions for Establishing Nucleosome Positioning as a Biomarker for RA

Nucleosome positioning plays a crucial role in shaping chromatin accessibility, making it a key factor in understanding chromatin changes. This highlights its potential as a biomarker for the early diagnosis of RA. Inflammatory cytokines in RA can alter chromatin structure and accessibility, potentially driving disease progression (Kondo, Kuroda and Kobayashi, 2021). Investigating the interplay between nucleosome positioning and inflammatory pathways may uncover new therapeutic targets. Further exploration through gene expression analysis may provide valuable insights into the molecular mechanisms driving RA onset, potentially aiding in the development of targeted therapies. To build on this study, incorporating more diverse datasets would be beneficial. The dataset used herein focuses on individuals with already-developed RA, which provides valuable insights but does not capture nucleosome positioning changes at different disease stages. Analysing datasets from various RA stages and performing comparative analyses may reveal how nucleosome positioning evolves over time. Furthermore, ensuring a larger sample size, high-quality data, and standardised ATAC-seq methodologies would strengthen the reliability and reproducibility of the findings. Nevertheless, this study confirms that nucleosome positioning changes in fully developed RA, opening the door for future research into its potential as a biomarker for early disease onset.

References

- Ai, R., Laragione, T., Hammaker, D., Boyle, D.L., Wildberg, A., Maeshima, K., Palescandolo, E., Krishna, V., Pocalyko, D., Whitaker, J.W., Bai, Y., Nagpal, S., Bachman, K.E., Ainsworth, R.I., Wang, M., Ding, B., Gulko, P.S., Wang, W. and Firestein, G.S. (2018) 'Comprehensive epigenetic landscape of rheumatoid arthritis fibroblast-like synoviocytes', *Nature Communications*, 9(1), article 1921. Available at: <https://doi.org/10.1038/s41467-018-04310-9>
- Araki, Y. and Mimura, T. (2016) 'The mechanisms underlying chronic inflammation in rheumatoid arthritis from the perspective of the epigenetic landscape', *Journal of Immunology Research*, 2016(1), article 6290682. Available at: <https://doi.org/10.1155/2016/6290682>
- Bai, L. and Morozov, A.V. (2010) 'Gene regulation by nucleosome positioning', *Trends in Genetics*, 26(11), pp. 476-483. Available at: <https://doi.org/10.1016/j.tig.2010.08.003>
- Bartoloni, E., Ludovini, V., Alunno, A., Pistola, L., Bistoni, O., Crinò, L. and Gerli, R. (2011) 'Increased levels of circulating DNA in patients with systemic autoimmune diseases: a possible marker of disease activity in Sjögren's syndrome', *Lupus*, 20(9), pp. 928-935. Available at: <https://doi.org/10.1177/0961203311399606>
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. and Greenleaf, W.J. (2013) 'Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position', *Nature Methods*, 10(12), pp. 1213-1218. Available at: <https://doi.org/10.1038/nmeth.2688>
- Bullock, J., Rizvi, S.A.A., Saleh, A.M., Ahmed, S.S., Do, D.P., Ansari, R.A. and Ahmed, J. (2018) 'Rheumatoid arthritis: a brief overview of the treatment', *Medical Principles and Practice*, 27(6), pp. 501-507. Available at: <https://doi.org/10.1159/000493390>
- Cribbs, A., Feldmann, M. and Oppermann, U. (2015) 'Towards an understanding of the role of DNA methylation in rheumatoid arthritis: therapeutic and diagnostic implications', *Therapeutic Advances in Musculoskeletal Disease*, 7(5), pp. 206-219. Available at: <https://doi.org/10.1177/1759720X15598307>
- Davis, S. (no date) *ATAC-Seq with Bioconductor*. Available at: <https://seandavi.github.io/AtacSeqWorkshop/articles/Workflow.html#overview> (Accessed: 10/12/2025).
- Deininger, P. (2011) 'Alu elements: know the SINEs', *Genome Biology*, 12, article 236. Available at: <https://doi.org/10.1186/gb-2011-12-12-236>
- Dunaeva, M., Buddingh', B.C., Toes, R.E.M., Luime, J.J., Lubberts, E. and Pruijn, G.J.M. (2015) 'Decreased serum cell-free DNA levels in rheumatoid arthritis', *Autoimmunity Highlights*, 6, pp. 23-30. Available at: <https://doi.org/10.1007/s13317-015-0066-6>

Edgar, R., Domrachev, M. and Lash, A.E. (2002) 'Gene Expression Omnibus: NCBI gene expression and hybridization array data repository', *Nucleic Acids Research*, 30(1), pp. 207-210. Available at: <https://doi.org/10.1093/nar/30.1.207>

Elizarraras, J.M., Liao, Y., Shi, Z., Zhu, Q., Pico, A.R. and Zhang, B. (2024) 'WebGestalt 2024: faster gene set analysis and new support for metabolomics and multi-omics', *Nucleic Acids Research*, 52(W1), pp. W415-W421. Available at: <https://doi.org/10.1093/nar/gkae456>

Englander, E.W. and Howard, B.H. (1995) 'Nucleosome Positioning by Human Alu Elements in Chromatin (*)', *Journal of Biological Chemistry*, 270(17), pp. 10091-10096. Available at: <https://doi.org/10.1074/jbc.270.17.10091>

Fahmueller, Y.N., Nagel, D., Hoffmann, R.T., Tatsch, K., Jakobs, T., Stieber, P. and Holdenrieder, S. (2012) 'Predictive and prognostic value of circulating nucleosomes and serum biomarkers in patients with metastasized colorectal cancer undergoing Selective Internal Radiation Therapy', *BMC Cancer*, 12, article 5. Available at: <https://doi.org/10.1186/1471-2407-12-5>

Fondon, J.W., Hammock, E.A.D., Hannan, A.J. and King, D.G. (2008) 'Simple sequence repeats: genetic modulators of brain function and behavior', *Trends in Neurosciences*, 31(7), pp. 328-334. Available at: <https://doi.org/10.1016/j.tins.2008.03.006>

Fyodorov, D.V., Zhou, B.R., Skoultschi, A.I. and Bai, Y. (2018) 'Emerging roles of linker histones in regulating chromatin structure and function', *Nature Reviews. Molecular Cell Biology*, 19(3), pp. 192-206. Available at: <https://doi.org/10.1038/nrm.2017.94>

Gerosa, M., De Angelis, V., Riboldi, P. and Meroni, P.L. (2008) 'Rheumatoid arthritis: a female challenge', *Women's Health*, 4(2), pp. 195-201. Available at: <https://doi.org/10.2217/17455057.4.2.195>

Ghiggeri, G.M., D'Alessandro, M., Bartolomeo, D., Degl'Innocenti, M.L., Magnasco, A., Lugani, F., Prunotto, M. and Bruschi, M. (2019) 'An update on antibodies to nucleosome components as biomarkers of systemic lupus erythematosus and of lupus flares', *International Journal of Molecular Sciences*, 20(22), article 5799. Available at: <https://doi.org/10.3390/ijms20225799>

Grandi, F.C., Modi, H., Kampman, L. and Corces, M.R. (2022) 'Chromatin accessibility profiling by ATAC-seq', *Nature Protocols*, 17(6), pp. 1518-1552. Available at: <https://doi.org/10.1038/s41596-022-00692-9>

Gujar, H., Weisenberger, D.J. and Liang, G. (2019) 'The roles of human DNA methyltransferases and their isoforms in shaping the epigenome', *Genes*, 10(2), article 172. Available at: <https://doi.org/10.3390/genes10020172>

- Handy, D.E., Castro, R. and Loscalzo, J. (2011) 'Epigenetic modifications: basic mechanisms and role in cardiovascular disease', *Circulation*, 123(19), pp. 2145-2156. Available at: <https://doi.org/10.1161/CIRCULATIONAHA.110.956839>
- Heidari, B. (2011) 'Rheumatoid Arthritis: Early diagnosis and treatment outcomes', *Caspian Journal of Internal Medicine*, 2(1), pp. 161-170.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) 'Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities', *Molecular Cell*, 38(4), pp. 576-589. Available at: <https://doi.org/10.1016/j.molcel.2010.05.004>
- Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F. and Hillman-Jackson, J. (2006) 'The UCSC genome browser database: update 2006', *Nucleic Acids Research*, 34(suppl_1), pp. D590-D598. Available at: <https://doi.org/10.1093/nar/gkj144>
- Hughes, S. (2023) *Surveying the epigenetic landscape of murine synovitis to explore the heterogeneity of rheumatoid arthritis*. PhD Thesis. Cardiff University and Monash University.
- Jacob, D.R., Guiblet, W.M., Mamayusupova, H., Shtumpf, M., Ciuta, I., Ruje, L., Gretton, S., Bikova, M., Correa, C., Dellow, E. and Agrawal, S.P. (2024) 'Nucleosome reorganisation in breast cancer tissues', *Clinical Epigenetics*, 16, article 50. Available at: <https://doi.org/10.1186/s13148-024-01656-4>
- Jadhav, R.R., Hu, B., Ye, Z., Sheth, K., Li, X., Greenleaf, W.J., Weyand, C.M. and Goronzy, J.J. (2022) 'Reduced chromatin accessibility to CD4 T cell super-enhancers encompassing susceptibility loci of rheumatoid arthritis', *EBioMedicine*, 76, article 103825. Available at: <https://doi.org/10.1016/j.ebiom.2022.103825>
- Jiang, C. and Pugh, B.F. (2009) 'Nucleosome positioning and gene regulation: advances through genomics', *Nature Reviews Genetics*, 10(3), pp. 161-172. Available at: <https://doi.org/10.1038/nrg2522>
- Jørgensen, M.H., Rekvig, O.P., Jacobsen, R.S., Jacobsen, S. and Fenton, K.A. (2011) 'Circulating levels of chromatin fragments are inversely correlated with anti-dsDNA antibody levels in human and murine systemic lupus erythematosus', *Immunology Letters*, 138(2), pp. 179-186. Available at: <https://doi.org/10.1016/j.imlet.2011.04.006>
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) 'The human genome browser at UCSC', *Genome Research*, 12(6), pp. 996-1006. Available at: <https://doi.org/10.1101/gr.229102>

- Kondo, N., Kuroda, T. and Kobayashi, D. (2021) 'Cytokine networks in the pathogenesis of rheumatoid arthritis', *International Journal of Molecular Sciences*, 22(20), article 10922. Available at: <https://doi.org/10.3390/ijms222010922>
- Kornberg, R.D. and Lorch, Y. (1999) 'Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome', *Cell*, 98(3), pp. 285-294. Available at: [https://doi.org/10.1016/S0092-8674\(00\)81958-3](https://doi.org/10.1016/S0092-8674(00)81958-3)
- Krishna, V., Yin, X., Song, Q., Walsh, A., Pocalyko, D., Bachman, K., Anderson, I., Madakamutil, L. and Nagpal, S. (2021) 'Integration of the transcriptome and genome-wide landscape of BRD2 and BRD4 binding motifs identifies key superenhancer genes and reveals the mechanism of bet inhibitor action in rheumatoid arthritis synovial fibroblasts', *The Journal of Immunology*, 206(2), pp. 422-431. Available at: <https://doi.org/10.4049/jimmunol.2000286>
- Kustanovich, A., Schwartz, R., Peretz, T. and Grinshpun, A. (2019) 'Life and death of circulating cell-free DNA', *Cancer Biology & Therapy*, 20(8), pp. 1057-1067. Available at: <https://doi.org/10.1080/15384047.2019.1598759>
- Langmead, B. and Salzberg, S.L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nature Methods*, 9(4), pp. 357-359. Available at: <https://doi.org/10.1038/nmeth.1923>
- Lee, J.E., Kim, I.J., Cho, M.S. and Lee, J. (2017) 'A case of rheumatoid vasculitis involving hepatic artery in early rheumatoid arthritis', *Journal of Korean Medical Science*, 32(7), pp. 1207-1210. Available at: <https://doi.org/10.3346/jkms.2017.32.7.1207>
- Leon, S.A., Revach, M., Ehrlich, G.E., Adler, R., Petersen, V. and Shapiro, B. (1981) 'Dna in synovial fluid and the circulation of patients with arthritis', *Arthritis & Rheumatology*, 24(9), pp. 1142-1150. Available at: <https://doi.org/10.1002/art.1780240905>
- Li, J., Ding, Y. and Zheng, L. (2014) 'Histone-mediated transgenerational epigenetics', in T. Tollefsbol (ed), *Transgenerational Epigenetics*. Academic Press, pp. 87-103.
- Li, S., Peng, Y. and Panchenko, A.R. (2022) 'DNA methylation: Precise modulation of chromatin structure and dynamics', *Current Opinion in Structural Biology*, 75, article 102430. Available at: <https://doi.org/10.1016/j.sbi.2022.102430>
- Liu, K.Y. and Leung, D. (2025) 'Epigenetic dysregulation of retrotransposons in cancer', *Molecular Cancer Research*, 23(5), pp. 369-378. Available at: <https://doi.org/10.1158/1541-7786.MCR-24-0744>
- Liu, R., Wu, J., Guo, H., Yao, W., Li, S., Lu, Y., Jia, Y., Liang, X., Tang, J. and Zhang, H. (2023) 'Post-translational modifications of histones: Mechanisms, biological functions, and therapeutic targets', *MedComm*, 4(3), e292. Available at: <https://doi.org/10.1002/mco2.292>

- Luger, K., Dechassa, M.L. and Tremethick, D.J. (2012) 'New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?', *Nature Reviews Molecular Cell Biology*, 13(7), pp. 436-447. Available at: <https://doi.org/10.1038/nrm3382>
- Moore, L.D., Le, T. and Fan, G. (2013) 'DNA methylation and its basic function', *Neuropsychopharmacology*, 38(1), pp. 23-38. Available at: <https://doi.org/10.1038/npp.2012.112>
- Nakano, K., Boyle, D.L. and Firestein, G.S. (2013) 'Regulation of DNA methylation in rheumatoid arthritis synoviocytes', *The Journal of Immunology*, 190(3), pp. 1297-1303. Available at: <https://doi.org/10.4049/jimmunol.1202572>
- Nemtsova, M.V., Zaletaev, D.V., Bure, I.V., Mikhaylenko, D.S., Kuznetsova, E.B., Alekseeva, E.A., Beloukhova, M.I., Deviatkin, A.A., Lukashev, A.N. and Zamyatnin Jr., A.A. (2019) 'Epigenetic changes in the pathogenesis of rheumatoid arthritis', *Frontiers in Genetics*, 10, article 570. Available at: <https://doi.org/10.3389/fgene.2019.00570>
- Pan, M.R., Hsu, M.C., Chen, L.T. and Hung, W.C. (2018) 'Orchestration of H3K27 methylation: mechanisms and therapeutic implication', *Cellular and Molecular Life Sciences*, 75, pp. 209- 223. Available at: <https://doi.org/10.1007/s00018-017-2596-8>
- Papin, C., Ibrahim, A., Le Gras, S., Velt, A., Stoll, I., Jost, B., Menoni, H., Bronner, C., Dimitrov, S. and Hamiche, A. (2017) 'Combinatorial DNA methylation codes at repetitive elements', *Genome Research*, 27(6), pp. 934-946. Available at: <https://doi.org/10.1101/gr.213983.116>
- Payet, M., Dargai, F., Gasque, P. and Guillot, X. (2021) 'Epigenetic regulation (including micro-RNAs, DNA methylation and histone modifications) of rheumatoid arthritis: a systematic review', *International Journal of Molecular Sciences*, 22(22), article 12170. Available at: <https://doi.org/10.3390/ijms222212170>
- Penny, L., Main, S.C., De Michino, S.D. and Bratman, S.V. (2024) 'Chromatin- and nucleosome-associated features in liquid biopsy: implications for cancer biomarker discovery', *Biochemistry and Cell Biology*, 102(4), pp. 291-298. Available at: <https://doi.org/10.1139/bcb-2024-0004>
- Qi, T., Pan, M., Shi, H., Wang, L., Bai, Y. and Ge, Q. (2023) 'Cell-Free DNA fragmentomics: the novel promising biomarker', *International Journal of Molecular Sciences*, 24(2), article 1503. Available at: <https://doi.org/10.3390/ijms24021503>
- Quinlan, A.R. and Hall, I.M. (2010) 'BEDTools: a flexible suite of utilities for comparing genomic features', *Bioinformatics*, 26(6), pp. 841-842. Available at: <https://doi.org/10.1093/bioinformatics/btq033>
- Serefidou, M., Venkatasubramani, A.V. and Imhof, A. (2019) 'The impact of one carbon metabolism on histone methylation', *Frontiers in Genetics*, 10, article 764. Available at: <https://doi.org/10.3389/fgene.2019.00764>

- Struhl, K. and Segal, E. (2013) 'Determinants of nucleosome positioning', *Nature Structural & Molecular Biology*, 20(3), pp. 267-273. Available at: <https://doi.org/10.1038/nsmb.2506>
- Szyf, M. (2001) 'The role of DNA methyltransferase 1 in growth control', *Frontiers in Bioscience-Landmark (FBL)*, 6(3), pp. 599- 609. Available at: <https://doi.org/10.2741/szyf>
- Tatham, S. (2011) *PuTTY: A free Telnet/SSH client*. Available at: <http://www.chiark.greenend.org.uk/~sgtatham/putty/> (Accessed: 4 December 2025).
- Teif, V.B. and Clarkson, C.T. (2019) 'Nucleosome positioning', in S. Ranganathan, M. Gribskov and C. Schönbach (eds), *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier, pp. 308–317.
- Turek-Plewa, J. and Jagodzinski, P.P. (2005) 'The role of mammalian DNA methyltransferases in the regulation of gene expression', *Cellular and Molecular Biology Letters*, 10(4), p. 631-647.
- Vanderstichele, A., Busschaert, P., Landolfo, C., Olbrecht, S., Coosemans, A., Froyman, W., Loverix, L., Concin, N., Braicu, E.I., Wimberger, P. and Van Nieuwenhuysen, E. (2022) 'Nucleosome footprinting in plasma cell-free DNA for the pre-surgical diagnosis of ovarian cancer', *NPJ Genomic Medicine*, 7(1), article 30. Available at: <https://doi.org/10.1038/s41525-022-00300-5>
- Venetsanopoulou, A.I., Alamanos, Y., Voulgari, P.V. and Drosos, A.A. (2023) 'Epidemiology and Risk Factors for Rheumatoid Arthritis Development', *Mediterranean Journal of Rheumatology*, 34(4), pp. 404-413. Available at: <https://doi.org/10.31138/mjr.301223.eaf>
- Visser, H. (2005) 'Early diagnosis of rheumatoid arthritis', *Best Practice & Research Clinical Rheumatology*, 19(1), pp. 55-72. Available at: <https://doi.org/10.1016/j.berh.2004.08.005>
- Wang, L.H., Aberin, M.A.E., Wu, S. and Wang, S.P. (2021) 'The MLL3/4 H3K4 methyltransferase complex in establishing an active enhancer landscape', *Biochemical Society Transactions*, 49(3), pp. 1041-1054. Available at: <https://doi.org/10.1042/BST20191164>
- Wei, F., Yang, F., Jiang, X., Yu, W. and Ren, X. (2015) 'High-mobility group nucleosome-binding protein 1 is a novel clinical biomarker in non-small cell lung cancer', *Tumor Biology*, 36, pp. 9405-9410. Available at: <https://doi.org/10.1007/s13277-015-3693-7>
- Wilson, A.S., Power, B.E. and Molloy, P.L. (2007) 'DNA hypomethylation and human diseases', *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1775(1), pp. 138-162. Available at: <https://doi.org/10.1016/j.bbcan.2006.08.007>
- Xu, Y., Huang, Y., Cai, D., Liu, J. and Cao, X. (2015) 'Analysis of differences in the molecular mechanism of rheumatoid arthritis and osteoarthritis based on integration of gene expression profiles', *Immunology Letters*, 168(2), pp. 246-253. Available at: <https://doi.org/10.1016/j.imlet.2015.09.011>

Yan, F., Powell, D.R., Curtis, D.J. and Wong, N.C. (2020) 'From reads to insight: a hitchhiker's guide to ATAC-seq data analysis', *Genome Biology*, 21, article 22. Available at: <https://doi.org/10.1186/s13059-020-1929-3>

Yang, D., Su, Z., Wei, G., Long, F., Zhu, Y.C., Ni, T., Liu, X. and Zhu, Y.Z. (2021) 'H3K4 methyltransferase Smyd3 mediates vascular smooth muscle cell proliferation, migration, and neointima formation', *Arteriosclerosis, Thrombosis, and Vascular Biology*, 41(6), pp. 1901-1914. Available at: <https://doi.org/10.1161/ATVBAHA.121.314689>

Yehya, N., Thomas, N.J. and Margulies, S.S. (2016) 'Circulating nucleosomes are associated with mortality in pediatric acute respiratory distress syndrome', *American Journal of Physiology- Lung Cellular and Molecular Physiology*, 310(11), pp. L1177-L1184. Available at: <https://doi.org/10.1152/ajplung.00067.2016>

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. and Liu, X.S. (2008) 'Model-based analysis of ChIP-Seq (MACS)', *Genome Biology*, 9, article R137. Available at: <https://doi.org/10.1186/gb-2008-9-9-r137>

Zhao, J. and Li, Z. (2018) 'The challenges of early diagnosis and therapeutic prediction in rheumatoid arthritis', *International Journal of Rheumatic Diseases*, 21(12). Available at: <https://doi.org/10.1111/1756-185X.13459>

Zhong, X., von Mühlengen, I., Li, Y., Kang, A., Gupta, A.K., Tyndall, A., Holzgreve, W., Hahn, S. and Hasler, P. (2007) 'Increased concentrations of antibody-bound circulatory cell-free DNA in rheumatoid arthritis', *Clinical Chemistry*, 53(9), pp. 1609-1614. Available at: <https://doi.org/10.1373/clinchem.2006.084509>

Zong, D., Huang, B., Li, Y., Lu, Y., Xiang, N., Guo, C., Liu, Q., Sha, Q., Du, P., Yu, Q. and Zhang, W. (2021) 'Chromatin accessibility landscapes of immune cells in rheumatoid arthritis nominate monocytes in disease pathogenesis', *BMC Biology*, 19, article 79. Available at: <https://doi.org/10.1186/s12915-021-01011-6>

Copyright Statement

©Vaidika Goyani. This article is licensed under a [Creative Commons Attribution 4.0 International Licence \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).