



# Functional genomics of psoriasis

Alicia Lledo Lara  
Hertford College  
University of Oxford

*A thesis submitted in partial  
fulfilment of the requirements for the degree of  
Doctor of Philosophy  
Trinity Term, 2018*

# Abstract

## **Functional genomics of psoriasis**

Alicia Lledo Lara, Hertford College, Trinity Term 2018

A thesis submitted in partial fulfilment of the requirements for the degree of  
Doctor of Philosophy of the University of Oxford

This is my abstract...

# Acknowledgements

Thank you, thank you, thank you.

# Declarations

I declare that unless otherwise stated, all work presented in this thesis is my own. Several aspects of each project relied upon collaboration where part of the work was conducted by others.

# Submitted Abstracts

Title	Year
Authors	

## Associated Publications

Title  
Journal  
Authors

## Other Publications

Title  
Journal  
Authors

# Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgements</b>	<b>ii</b>
<b>Declarations</b>	<b>iii</b>
<b>Submitted Abstracts</b>	<b>iv</b>
<b>Associated Publications</b>	<b>v</b>
<b>Contents</b>	<b>vi</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>viii</b>
<b>Abbreviations</b>	<b>ix</b>
<b>1 Establishment of laboratory methods and analytical tools to assess genome-wide chromatin accessibility in clinical samples</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Results . . . . .	2
1.2.1 Establishment of an ATAC-seq data analysis pipeline based on current knowledge . . . . .	2
1.2.2 Assessment of ATAC-seq transposition times and comparison with FAST-ATAC protocol in relevant cell types	10
1.2.3 Limitations of ATAC-seq and FAST-ATAC to assess chromatin accessibility in KC . . . . .	10
<b>Appendices</b>	<b>16</b>
<b>A Establishment of methods to assess genome-wide chromatin accessibility</b>	<b>17</b>

# List of Figures

1.1	Measurements for quality control assessment in ATAC-seq samples	7
1.2	Peak calling and sequencing depth in ATAC-seq samples . . . . .	8
1.3	Peak calling assessment and IDR filtering in ATAC-seq samples . .	9
1.4	Assessment of the effect of transposition times on the ATAC-seq QC parameters . . . . .	11
1.5	Differences in MT DNA abundance and signal specificity between ATAC-seq and FAST-ATAC protocols . . . . .	12
1.6	QC assessment of ATAC-seq in KC enriched cell suspension derived from a psoriatic lesional skin biopsy . . . . .	12
1.7	QC assessment of FAST-ATAC and Omni-ATAC in cultured NHEK	14
1.8	QC assessment of Omni-ATAC in NHEK and chromatin accessibility signal for the samples generated with the different ATAC-seq protocols . . . . .	15
A.1	FAST-ATAC and Omni-ATAC NHEK tapestation profiles. . . . .	18
A.2	Assessment of TSS enrichment from ATAC-seq and FAST-ATAC in healthy and psoriasis skin biopsies samples. . . . .	19



# List of Tables

1.1	Summary table of ATAC-seq methodology analysis for peak calling, filtering and differential analysis.. . . . .	3
1.2	ATAC-seq percentage of MT reads and fraction of reads in called peaks . . . . .	6
1.3	Description of the most relevant parameter from the ATAC-seq and FAST-ATAC protocols assayed in NHEK and skin biopsies. . .	13

# Abbreviations

Abbreviation	Definition
Ab	Antibody
ATAC-seq	
Atopic dermatitis	AD
ChIPm	
CLE	cutaneous lupus erythematosus
DMARDs	disease-modifying antirheumatic drugs
Fast-ATAC	
IDR	
GWAS	Genome-wide association studies
KC	Keratinocytes
NSAID	nonsteroidal antiinflammatory drug
Omni-ATAC	
PCA	
PI	Protein inhibitor
PsA	
QC	
qPCR	quantitative polymerase chain reaction
RA	Rheumatoid arthritis
SDS	Sodium dodecyl sulfate
SF	Synovial fluid

# Chapter 1

## Establishment of laboratory methods and analytical tools to assess genome-wide chromatin accessibility in clinical samples

### 1.1 Introduction

Previous and current methods to identify the accessible genome in cells and tissues

Implementation of ATAC-seq to define the chromatin landscape

Technical limitations and recent advances in optimisation

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4473780/>

Talk about ATAC being more variable, a native chromatin accessibility assessment without cross-linking. Role of transposase ability in accessing the chromatin, debris and DNA from dead cells adding noise

Paper to justify peak calling: A comparison of peak callers used for DNase-Seq data.

New ATAC but also explanations of the limitations: Characterization of chromatin accessibility with a transposome hypersensitive sites sequencing (THS-seq) assay

## **Challenges of working with clinical samples**

### **1.2 Results**

#### **1.2.1 Establishment of an ATAC-seq data analysis pipeline based on current knowledge**

When the first ATAC-seq publication (**Buenrostro2013**) appeared, there were not well established protocols for the complete processing of the data. Since then, several publications have used ATAC-seq and modifications of this protocol together with a wide range of data analysis strategies to answer different biological questions (Table 1.1). There are several limiting aspects in the process of analysing ATAC-seq data, including QC assessment, peak calling/filtering and differential analysis of chromatin accessibility regions between groups. Using the current knowledge in the field as well as on my own analysis, I agreed on the most appropriate criteria and parameters to implement in our in-house pipeline. For this purpose, I used ATAC data generated with the original protocol (**Buenrostro2013**) in paired CD14<sup>+</sup> monocytes and CD4<sup>+</sup> total T cells from the same three healthy individuals, all of them downsamples to 30 million of reads, in order to facilitate the comparison across all of them.

**Table 1.1:** Summary table of ATAC-seq methodology analysis for peak calling, filtering and differential analysis..

Publication	Peak calling and filtering	Master list	Differential analysis
Corces <i>et al.</i> , 2016	MACS2 (-nomodel), peak extension +/-250bp, rank summits by pval	Maximally overlapping peaks.	Quantile normalisation and unsupervised hierarchical clustering.
ENCODE	MACS2 -nomodel, pairwise IDR analysis, filtering IDR<10%	Choosing longest IDR filtered list or only peaks present in the two samples pseudoreplicates.	pairwise NA
Turner <i>et al.</i> , 2018	MACS2 (-nomodel -q 0.01)	Merging all filtered called peaks from the different cell types.	De novo:DiffReps with fragment size 50bp.
Alasoo <i>et al.</i> , 2018	MACS2 (-nomodel -shift -25 -extsize 50 -q 0.0)	Union of peaks present in at least three samples of the same condition.	Peak based: TMM normalisation and lima voom (FDR<0.01).

Qu <i>et al.</i> , 2017	ZINBA PP>0.99.	Merging of filtered peaks from each individual sample.	Quantile normalisation and peak based in house Pearson correlation method.
Rendeiro <i>et al.</i> 2016	MACS2 (-nomodel -extsize 147)	Merge of peaks from all samples in an iterative process including permutations	Peak based: quantile normalisation and Fisher exact test (FDR<0.05).
Scharer <i>et al.</i> 2016	HOMER (-style dnase)	Merge of all overlapping peaks between all samples using HOMER mergePeaks	Peak based: TMM normalisation and edgeR package (FDR<0.05).

### **Sample quality control**

Regarding QC measurements, the variability in performance of the methodology, particularly ATAC-seq and Fast-ATAC, has required to agree on appropriate parameters to determine the quality of the samples before proceeding with downstream differential analysis. After reviewing the different read-outs implemented across different publications, I have identified the most informative ones showing supporting correlation between them.

Firstly, I analysed the fragment size distribution for each of the samples in order to determine if they recapitulated the expected nucleosome periodicity every ~200bp (Figure 1.1a). All the samples showed periodicity up to 600bp, clearly distinguishing chromatin organisation into mono-, di- and tri-nucleosomes. The relative intensity of nucleosome-free DNA fragments (<200pb) compared to nucleosome-bound DNA was greater for some of the samples (e.g CTL1 CD4<sup>+</sup> and CD14<sup>+</sup>) and similar or lower for others (e.g CTL3 CD4<sup>+</sup> and CD14<sup>+</sup>). Nucleosome-free fragments (<147bp) are also clearly distinguished in all of the samples, meeting the ENCODE QC recommendations (ENCODE).

Another QC measurement was based on the enrichment over a random background of ATAC-seq reads across all the TSS for the identified for Ensemble genes (Figure 1.1b). It is well established that nucleosome repositioning and an increase of chromatin accessibility take place at TSS to allow formation of the transcriptional machinery and initiation of transcription. Fold-enrichment signals ranged between 5-7 for the CD4<sup>+</sup> samples and they were much higher (between 17-20) for the CD14<sup>+</sup> samples. The lower sample quality of the CD4<sup>+</sup> compared to CD14<sup>+</sup> shown by the TSS signal were recapitulated by the ATAC-seq genome browser density at the promoter of the constitutively expressed gene *GAPDH* (Figure 1.1c).

As part of the QC assessment I looked at the percentage of mitochondrial reads and the fraction of reads in peaks (FRiP)(Table 1.2).

Sample	% MT reads	Fraction of reads in peaks
CTL1 CD4	14.9	9.8
CTL2 CD4	30.5	11.2
CTL3 CD4	28.8	11.6
CTL1 CD14	43.3	32.2
CTL2 CD14	36.8	57.0
CTL3 CD14	37.6	49.9

**Table 1.2:**

Positive correlation between the TSS fold-change enrichment and FRiP was observed, being both appropriate inter-dependent QC measures to evaluate sample noise (Figure 1.1d). Regarding the cut-off values, Alsoo *et al.*, 2018 and, recently, ENCODE have recommended minimum FRiP between 10-20% and TSS between 6-10. ENCODE has prioritised the use of TSS over FRiP as the measurement to determine the noise in the sample (ENCODE). The mitochondrial content ranged between 14.9-43.3% and, alike FRiP and TSS, it was higher in CD4<sup>+</sup> than in CD4<sup>+</sup> and was cell type dependent and not directly related with any of the other QC measurements.

### **Peak calling and filtering**

As part of the ATAC-seq pipeline implementation, peak calling and the criteria for filtering were another two aspects to determine. Although different peak callers have been used, most of the publications as well as ENCODE have been using MACS2 as the preferred methodology (Table 1.1). MACS2 has been initially developed for ChIP but it has also been used for DHS and ATAC-seq with disabling the model and agreeing in an extension size (`-extsize`) and a shift (`-shift`), which indicate the direction and number of bp for reads to be shifted and the number of bp for them to be extended, respectively. The `-extsize` should correspond to the average fragment size, which in my libraries is ~200bp and the `-shift` is set to -100, as it is recommended to be set to -1/2 of the fragment size for



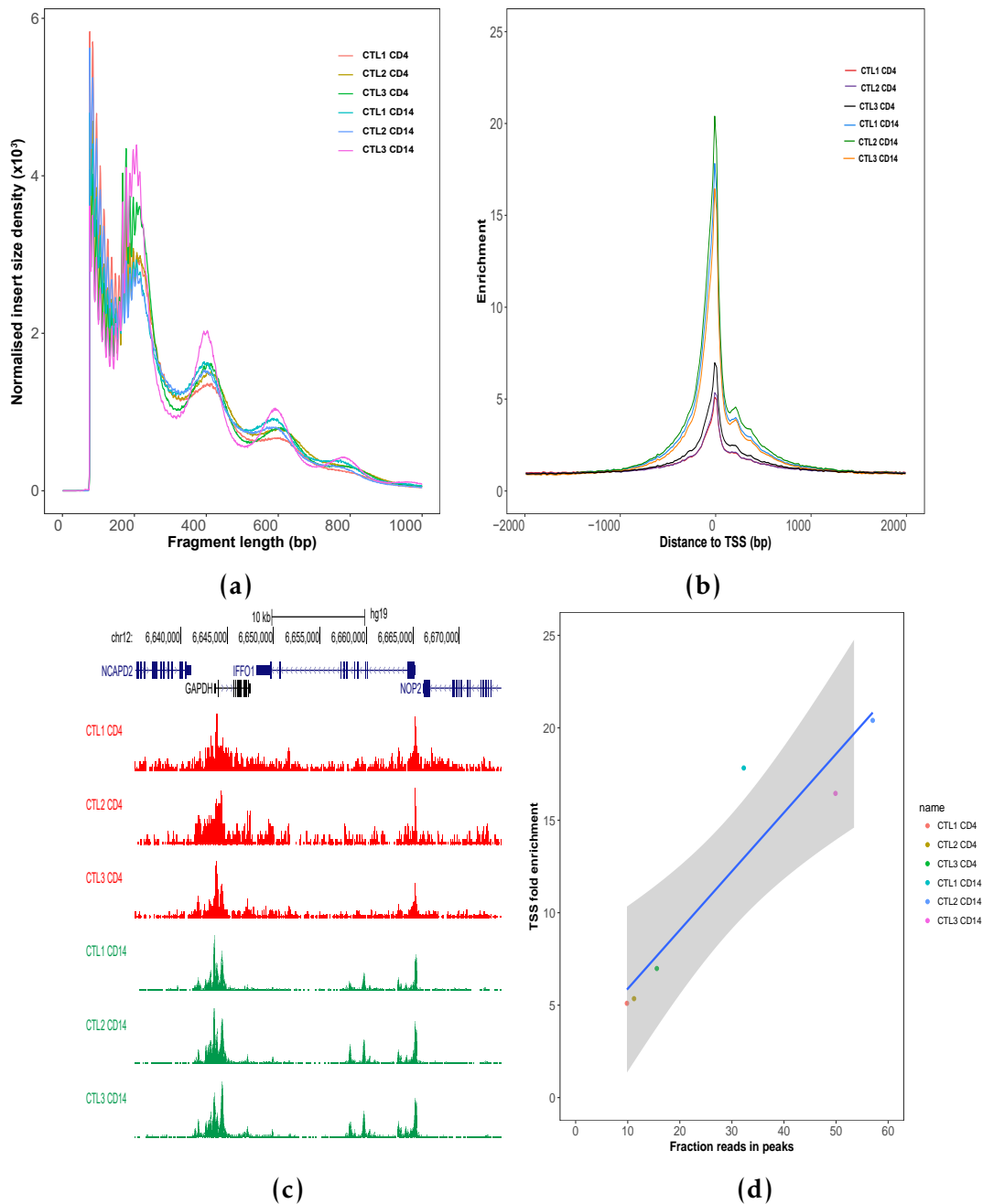


Figure 1.1:

## Establishment of methods to assess genome-wide chromatin accessibility

chromatin accessibility assays. This parameter could be further optimised but it escapes from the aim of this thesis.

I was interesting in understanding the effect of sequencing depth and the sample quality on the peak calling to have a better control of both variables in the downstream analysis. I performed random read sub-sampling every 5M total reads (from 5M to 30M) followed by peak calling with arbitrary filtering for  $FDR < 0.01$  in each of the six aforementioned samples.

Number of reads is dependent of the read depth and sample quality. Lower number of peaks called in CD4 samples compared to CD14, reflecting sample quality effect. However for both set of samples number of called peaks increases with the number of reads and when looking at the increment of number of peaks both reach plateau

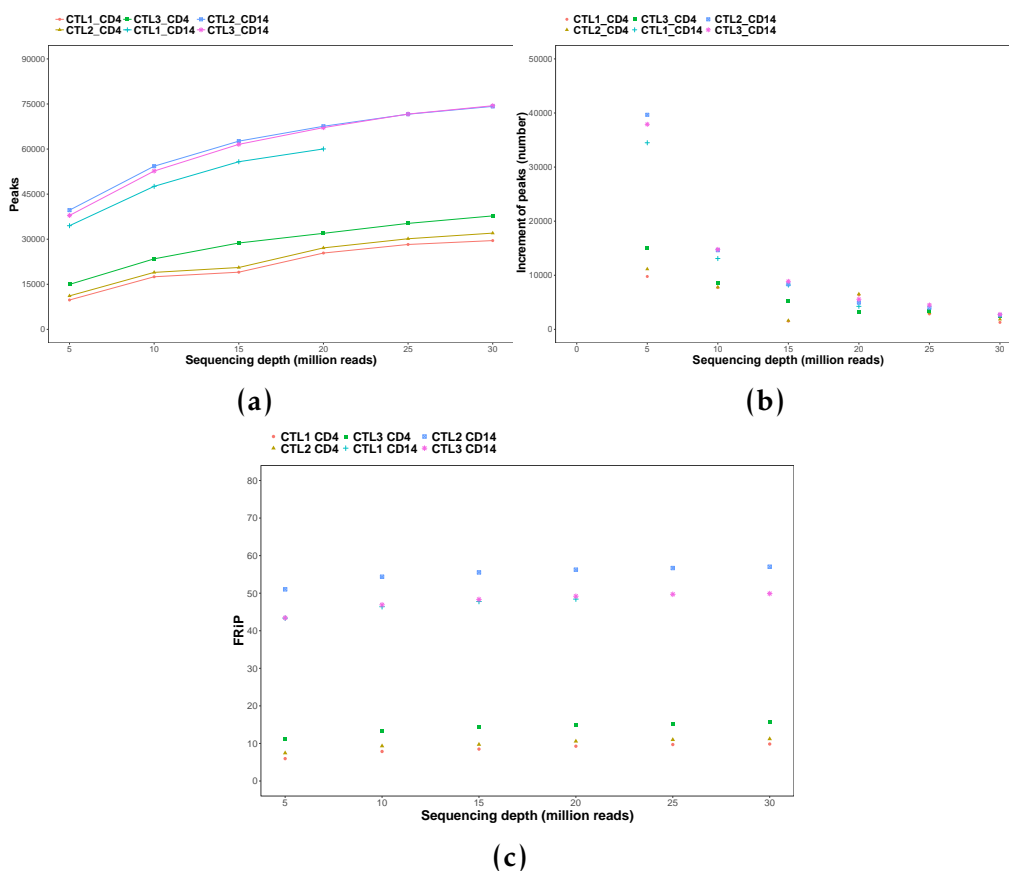


Figure 1.2: Peak calling at different sequencing depth in ATAC-seq samples

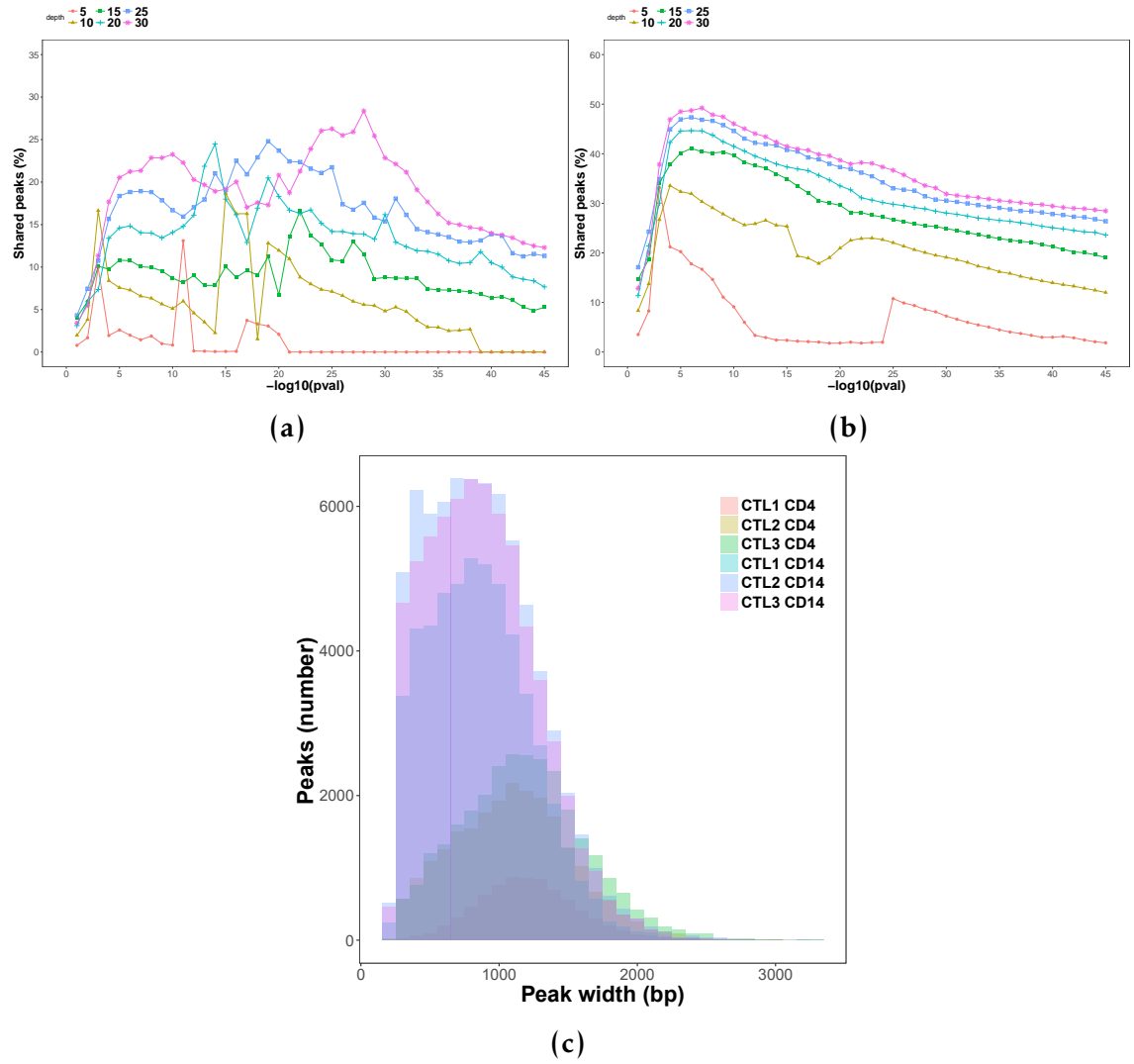


Figure 1.3: Peak calling filtering and assessment of width distribution in ATAC-seq samples

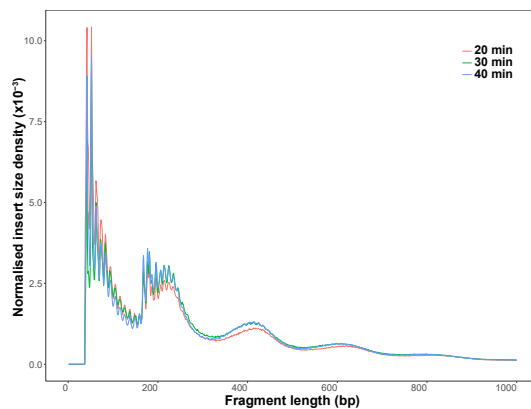
### **1.2.2 Assessment of ATAC-seq transposition times and comparison with FAST-ATAC protocol in relevant cell types**

### **1.2.3 Limitations of ATAC-seq and FAST-ATAC to assess chromatin accessibility in KC**

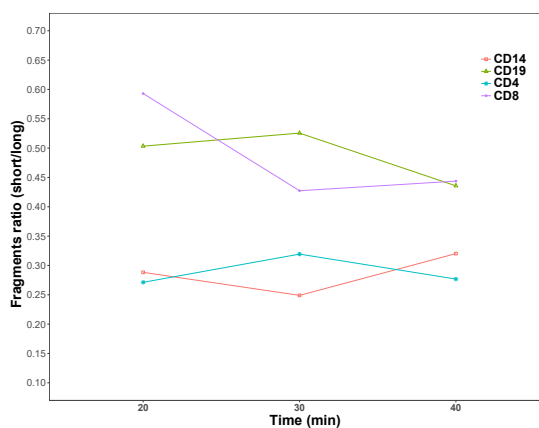
Due to the fact that KC is one of the most relevant cell types in psoriasis pathophysiology, ATAC-seq as described in Buenroostro *et al.*, 2013 (named as ATAC-seq 1 here) was performed in 50,000 cells of a suspensions isolated from a psoriasis lesional skin biopsy. Two different transposition times (30 and 40 min) were tested. Since biopsy handling and lesional epidermal KC are particularly challenging this was considered the best system to test the performance of the standard protocol in the clinical setting of interest for the study. Two transposition times (30 and 40 min) were tested.

Although cell suspension obtained from biopsies using trypsinisation of the epidermal sheet are 90% enriched in KC, they also contain significant amounts of dead cells and free-DNA releases by apoptotic cells. In order to overcome this problem and the impact that it may have over ATAC-seq background signal, viable KC were selected by adherence assay. Biopsy cell suspensions were cultured for 3h in a 96-well plate and washed afterwards to ensure that only the viable and less differentiated KC would remain for down stream analysis. In parallel cultured NHEK were also used to assess the performance of the different ATAC-seq protocols.

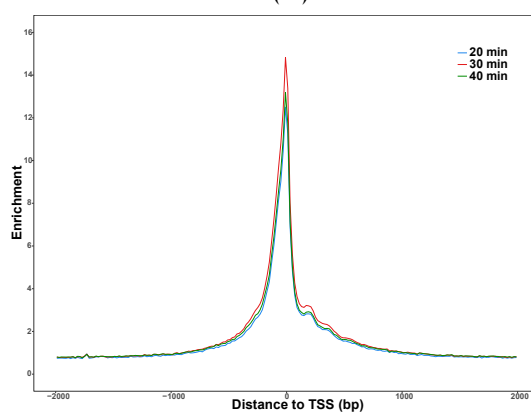
Table for the conditions: done Tapestation profiles of the the chosen condition. done Send the others to supplementary. QC measurements: for ATAC1, ATAC2 and NHEK, mention frag size distribution done DHS enrichment for p and q done but not convincing. The complex network of keratin filaments in stratified epithelia is tightly regulated during squamous cell differentiation. Keratin 14 (K14) is expressed in mitotically active basal layer cells, along with



(a)

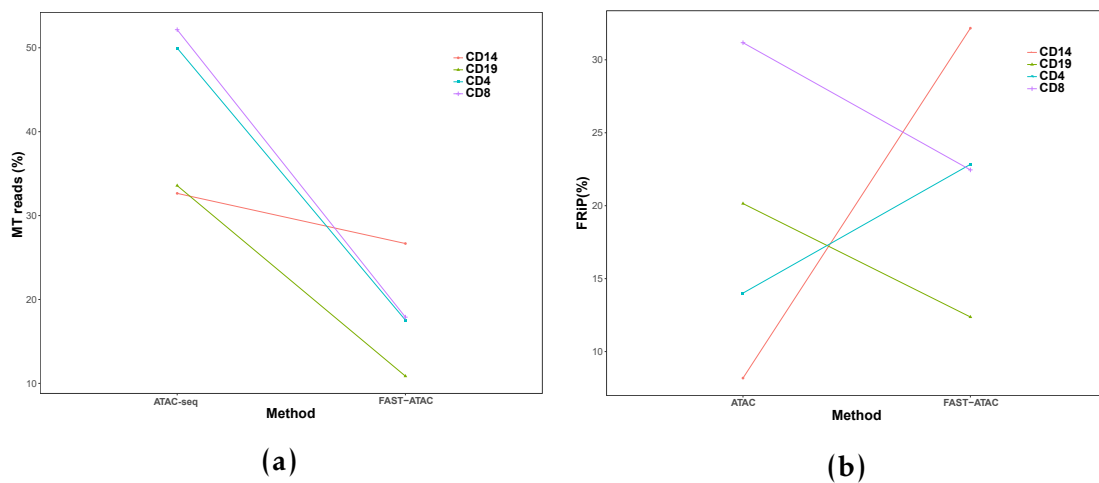


(b)

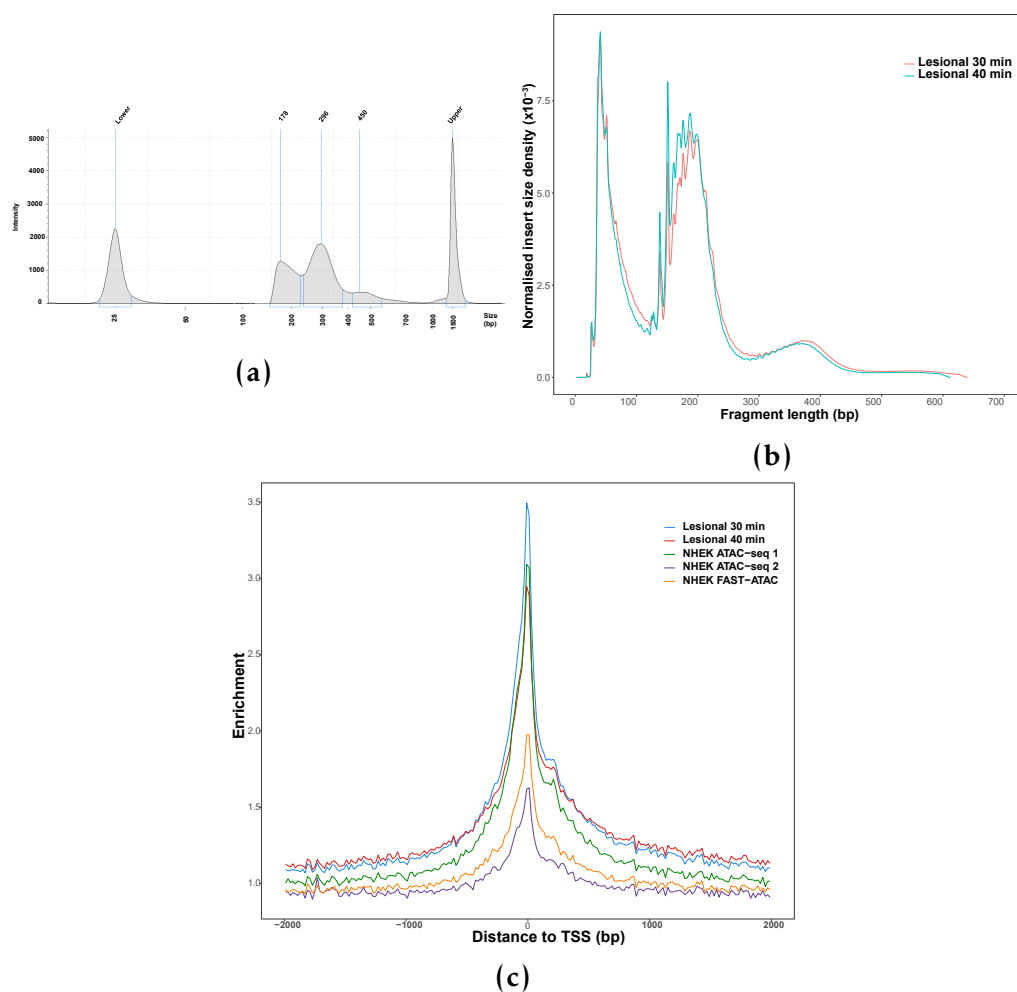


(c)

Figure 1.4: Assessment of the effect of transposition times on the ATAC-seq QC parameters



**Figure 1.5: Differences in MT DNA abundance and signal specificity between ATAC-seq and FAST-ATAC protocols**



**Figure 1.6: QC assessment of ATAC-seq in KC enriched cell suspension derived from a psoriatic lesional skin biopsy. Two transposition times (30 and 40 min) were tested using the standard ATAC-seq protocol (Buenrostro *et al.*, 2013 in 50,000 cells from the same suspension.**

Establishment of methods to assess genome-wide chromatin accessibility

its partner keratin 5 (K5), and their expression is down-regulated as cells differentiate.

	Protocol	Lysis and transposition	
	Buenrostro et al., 2013	Two steps	
only to stretch the columns if you want	Bao et al., 2015	Two steps	
	Corces et al., 2016	One step	C C C3 C4

**Table 1.3: Description of the most relevant parameter from the ATAC-seq and FAST-ATAC protocols assayed in NHEK and skin biopsies.**Transposition for all the different protocols was 30 min.

Omni-ATAC Tapestation profiles of the the chosen condition include it with the supplementary that includes all other tapestation profiles.done QC measurements: frag size distribution and TSS done Track including all skin samples

Think of what to include about the biopsies in supplementary done

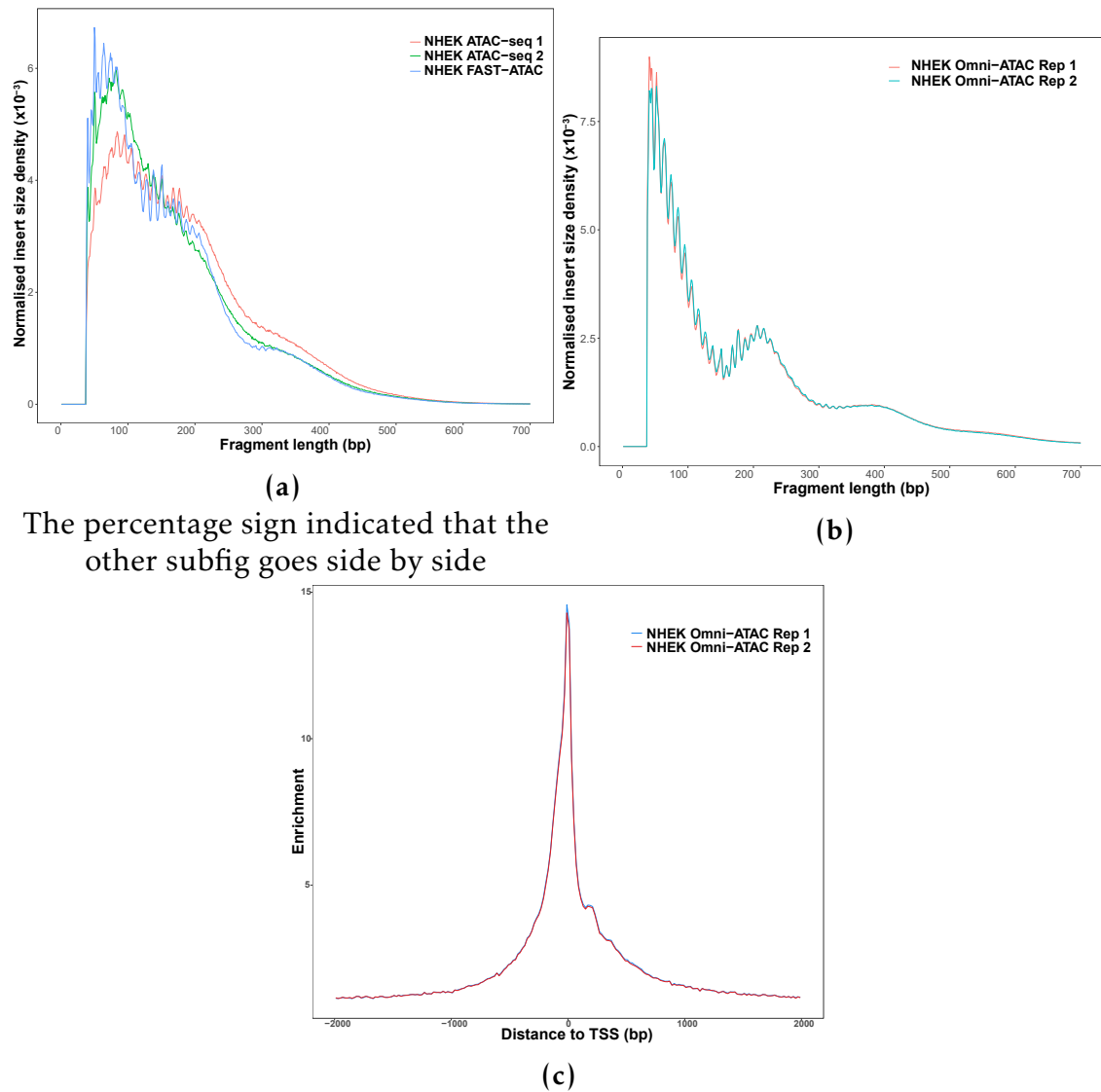


Figure 1.7: QC assessment of FAST-ATAC and Omni-ATAC in cultured NHEK.



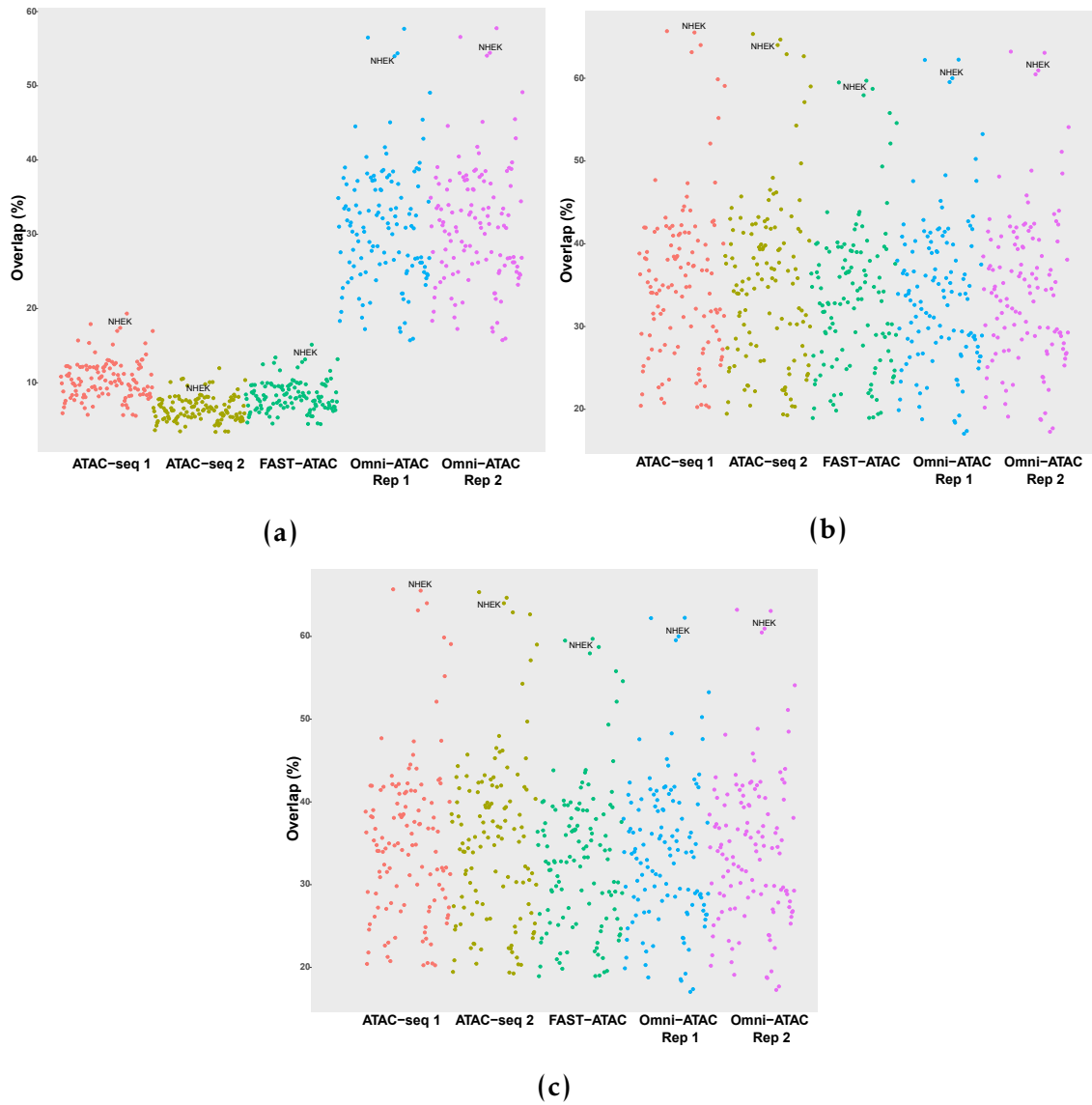


Figure 1.8: QC assessment of Omni-ATAC in NHEK and chromatin accessibility signal for the samples generated with the different ATAC-seq protocols.

# Appendices

A	Establishment of methods to assess genome-wide chromatin accessibility	17
---	--	----

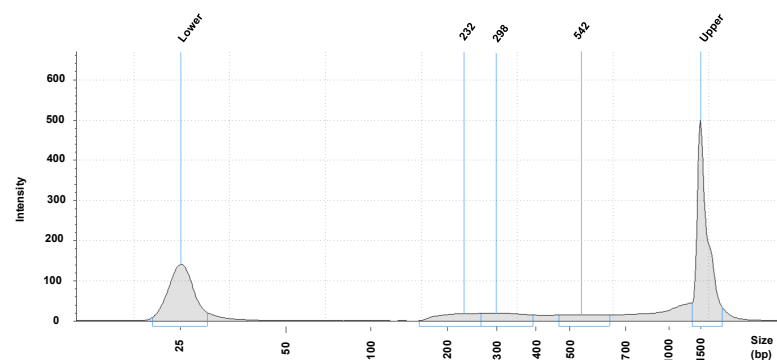
## List of Figures

A.1	FAST-ATAC and Omni-ATAC NHEK tapestation profiles. . . . .	18
A.2	Assessment of TSS enrichment from ATAC-seq and FAST-ATAC in healthy and psoriasis skin biopsies samples. . . . .	19

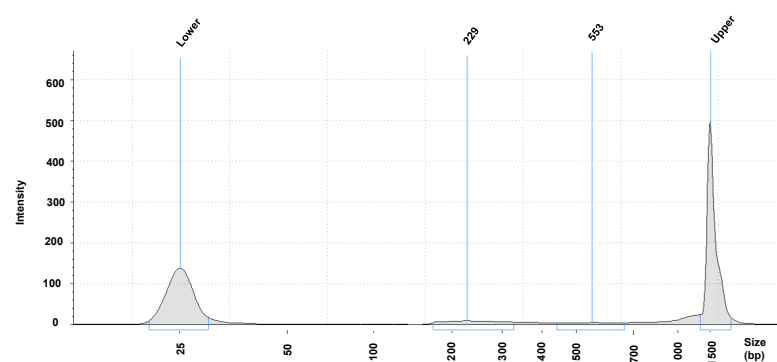
## List of Tables

# **Appendix A**

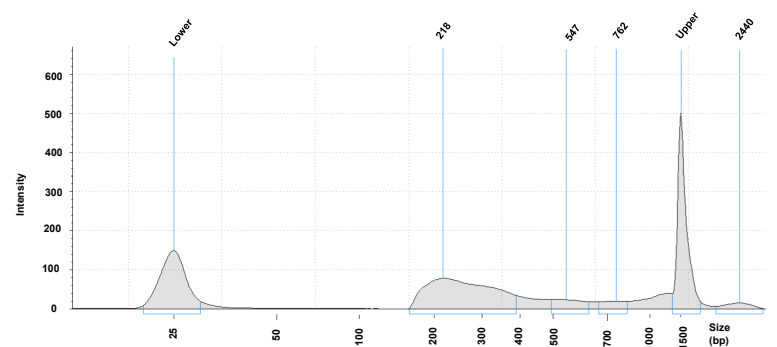
## **Establishment of methods to assess genome-wide chromatin accessibility**



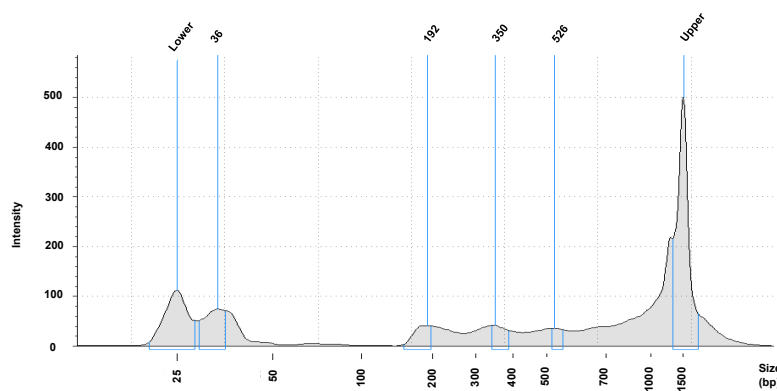
(a)



(b)



(c)



(d)

Figure A.1: FAST-ATAC and Omni-ATAC NHEK tapestation profiles.

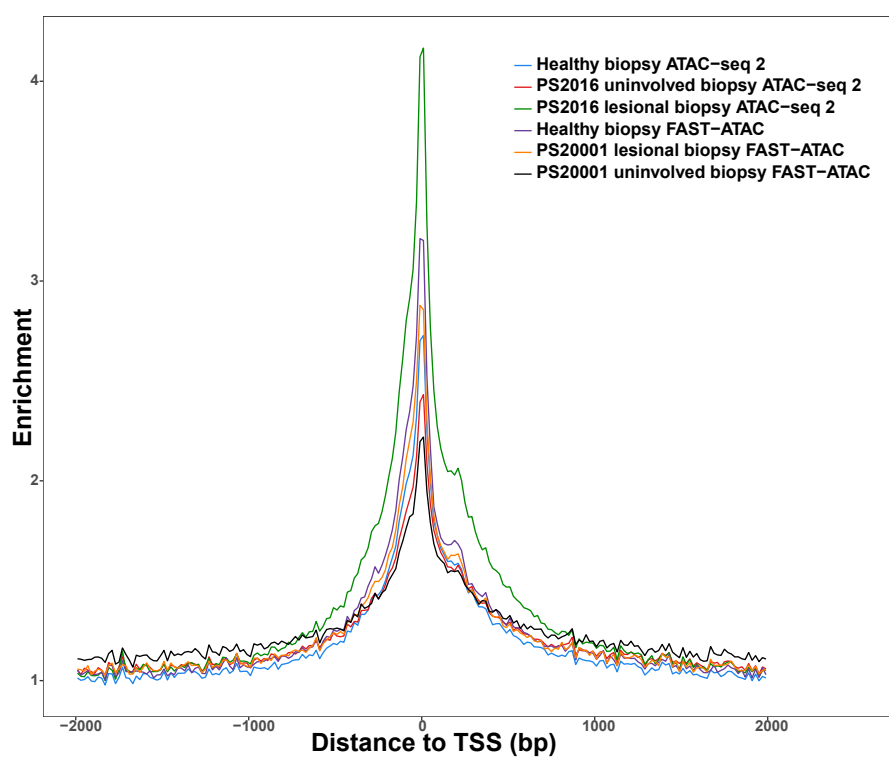


Figure A.2: Assessment of TSS enrichment from ATAC-seq and FAST-ATAC in healthy and psoriasis skin biopsies samples.