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

List of Figures

List of Tables

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, debri 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 1.2). 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 $CD14^+$ monocytes and $CD4^+$ 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.

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 $CD4^+$ and $CD14^+$) and similar or lower for others (e.g CTL3 $CD4^+$ and $CD14^+$).

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 $CD4^+$ samples and they were much higher(between 17-20) for the $CD4^+$ samples. The lower sample quality of the $CD4^+$compared to $CD14^+$ 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).

Agreement between TSS enrichment values and the background levels in each of the samples.

The fraction of reads mapping to the location of peaks following basic filtering for FDR¡0.01 was calculated for each of the samples using data for 30M reads and a clear positive correlation was found with the TSS enrichment values.

Table for fraction of reads in peaks and % of mitochondrial reads range between 14.9-43.3% and they appear to be higher in CD4 than in CD14, opposite

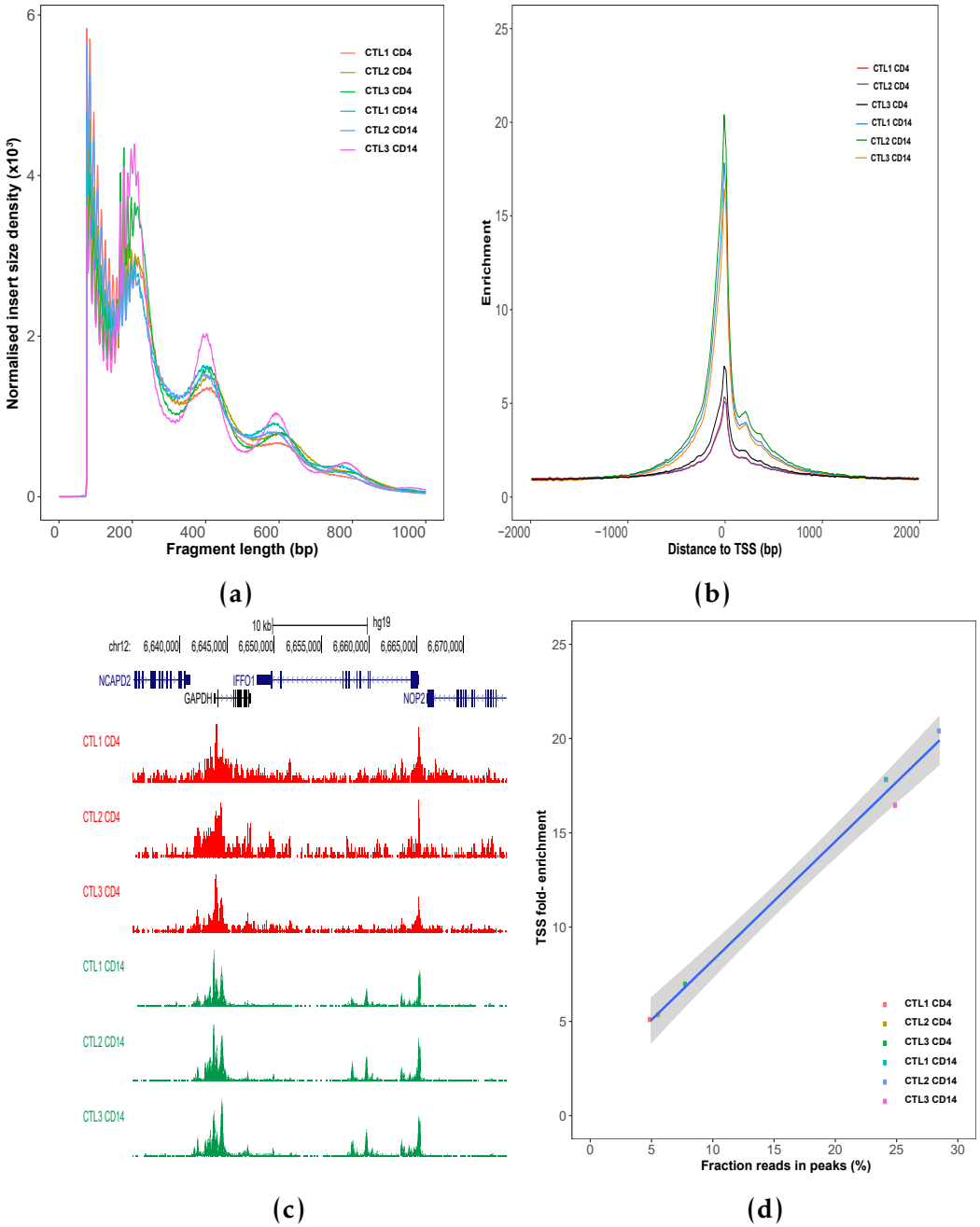
Normalised insert size density (x10³)
Fragment length (bp)
CTL1 CD4
CTL2 CD4
CTL3 CD4
CTL1 CD14
CTL2 CD14
CTL3 CD14
(a)
Enrichment
Distance to TSS (bp)
(b)
10 kb
hg19
chr12:
6,640,000
6,645,000
6,650,000
6,655,000
6,660,000
6,665,000
6,670,000
NCAPD2
IFFO1
GAPDH
NOP2
(c)
TSS fold- enrichment
Fraction reads in peaks (%)
(d)

Figure 1.1:

to the trend observed for the enrichment of signal at TSS and the fraction of reads in called peaks. Therefore, the percentage of MT reads seems to be cell type dependent and not being directly related with quality sample.

Sample	% MT reads	Fraction of reads in peaks
CTL1 CD4	14.9	4.9
CTL2 CD4	30.5	5.6
CTL3 CD4	28.8	7.7
CTL1 CD14	43.3	24.2
CTL2 CD14	36.8	28.5
CTL3 CD14	37.6	24.9

Table 1.1: Details regarding target molecule, fluorochrome, clone, supplier and dilution used for PBMC and SFC staining are provided for each of the antibodies in the panel. In controls only CD3, CD4 and CD4 markers were used.

Table 1.2: Summary table of ATAC-seq methodology analysis for peak calling, filtering and differential analysis..

Publication	Peak calling and filtering	Master list	Differential analysis
(**Corces2016**)	MACS2 -nomodel, peak summit extension +/−250bp, rank summits by pval	Maximally significant non-overlapping peaks.	Quantile normalisation and unsupervised hierarchical clustering.
(**ENCODE**)	MACS2 -nomodel, pairwise IDR analysis, filtering IDR<10%.	Choosing longest pairwise IDR filtered list or only peaks present in the two samples pseudoreplicates.	NA
(**Turner2018**)	MACS2 -nomodel –q 0.01.	Merging all filtered called peaks from the different cell types.	De novo:DiffReps with fragment size 50bp.
Alasoo2018	MACS2 -nomodel -shift -25 -extsize 50 –q 0.01.	Union of peaks from all conditions present in at leats in three samples of the same condition.	Peak based: TMM normalisation and lima voom (FDR<0.01).

(**Qu2017**)	ZINBA PP¿0.99.	Merging of filtered peaks from each individual sample.	Quantile normalisation and peak based in house Pearson correlation method.
(**Rendeiro2016**)	MACS2 -nomodel -extsize 147.	Merge of peaks from all samples in an iterative process including permutations	Peak based: quantile normalisation and Fisher exact text (FDR<0.05).
(**Scharer2016**)	HOMER -style dnase	Merge of all overlapping peaks between all samples using HOMER mergePeaks	Peak based: TMM normalisation and edgeR package (FDR<0.05).