# Exploring effects of hypoxia on monocytic differentiation with ATAC-seq and RNA-seq

Report

King Abdullah University of Science and Technology
BESE
Setting bioinformatics pipelines B394A

Team 1
Rohit Satyam, Jiayi Huang, Anel Lopez

March 19, 2024

# Contents

# 1 Introduction

Monocytes (MCs) are immune blood cells, they make up 10% of the total white blood cell population. Their main functions include innate immunity, response to inflammation, and removal of damaged or dead cells. MCs are produced in the bone marrow, and their progenitors are hematopoietic stem cells (HSCs) which later give rise to promonocytes and mature monocytes. Mature MCs can be divided into three subclasses according to their expression patterns of CD14 and CD16: classical monocytes ($CD14^{+}CD16^{-}$), intermediate ($CD14^{+}CD16^{+}$), and non-classical ($CD14^{dim}CD16^{+}$). Classical and intermediate monocytes are also referred to as inflammatory monocytes, whereas non-classical subtypes display crawling behavior. Inflammatory MCs differentiate into M1 inflammatory macrophages whereas the nonclassicals turn into M2 anti-inflammatory macrophages (MACs). The macrophages are the effector cells, meaning they are carrying the function that corresponds to the monocytic predecessor. MCs differentiate into MACs at the same time as they infiltrate into a tissue. Enviromental signals, such as cytokines will determine the functional phenotype of the cell. Therefore, the microenvironment present at the moment of invasion can affect the differentiation process of the monocytes.

During the different life stages of a monocyte, the cell is subjected to very different oxygen concentrations. When in the bone marrow the the oxygen partial pressure is less than 10 mmHg or around 1.3%. During their circulation across the blood vessels, the oxygen availability increases 10 times (6 to 13% or 50 to 100 mmHg). In healthy tissue, the oxygen partial pressure is reduced, coming down to 2.5-6% or 20 to 50 mmHg. When MACs are recruited to damaged tissues, the conditions resemble those of hypoxia. Diseases such as rheumatoid arthritis and lesions, present hypoxic conditions as well. In arthritis, it has been observed a negative relationship between oxygen and synovial oxygen and inflammatory activity. Hypoxia-inducible factors (HIFs) are molecules capable of modifying the transcriptional landscape of a cell, for example, in T-cells HIF-1 promotes cell adaptability to low oxygen conditions.

Given the previous information, we would like to explore how hypoxic conditions can affect the fate of monocytes. Can low oxygen conditions determine the functionality of a monocyte once it differentiates into a macrophage? Can low oxygen initiate the process of differentiation?

# 2 Dataset description

To investigate the role of environmental oxygen levels on monocyte differentiation, we will use the GEO dataset GSE165999.
In this article, the authors treated primary human monocytes with DMOG. DMOG is a substance that mimics hypoxia by activating the HIF pathway. A limitation of this study is that DMOG

fails to recreate an accurate image of constant low-oxygen conditions. The authors harvested the DMOG-treated cells at 0, 2, 10, and 24 hours and performed RNA-seq and ATAC-seq with 7 biological replicates for each assay.

1. Genome-wide chromatic accessibility (Bulk ATAC-seq): GSE165997
2. Transcriptome expression changes (Bulk RNA-seq): GSE165998

# 3 Integrating ATAC-seq and RNA-seq

## 3.1 Preprocessing of the data

**RNA-seq**
Protocol taken from [1]: FASTQ files will be trimmed with Trimmomatic and low quality reads will be filtered out. Mapping to the human genome will be done with a splice aware algorithm such as STAR. Low quality mappers will be filtered out (¡30). Uniquely mapped reads will be kept. PCR duplicates will be removed. Counting will be performed with HTSeq.
**ATAC-seq**
Protocol taken from [1]: FASTQ files will be trimmed with Trimmomatic and low quality reads will be filtered out. Mapping to the human genome will be done with BWA MEM or another non-splice aware algorithm. Low quality mappers will be filtered out (¡30). Uniquely mapped reads will be kept. PCR duplicates will be removed. Peaks will be called with MACS, blacklist are to be applied. deepTools will be used to generate bigwig files.

## 3.2 Differential Gene Expression Analysis

DGEA will be performed with DESeq2 likelihood ratio test (LTR). The model should be $\sim treatment + time + treatment : time$ for full model and $\sim treatment + time$ for the reduced model.

```
#Make DESeq2 object

dds <- DESeqDataSetFromMatrix(countData = raw_gene_counts, colData = metadata, design = ~
    treatment + time + treatment:time)

dds_lrt_time <- DESeq(dds, test="LRT", reduced = ~ treatment + time)
```

## 3.3 Differential accessibility analysis

Protocol taken from [2]:

Peaks corresponding to each time point and condition will be obtained with bedtools intersect. Then merge each dataset to obtain a global peak set. DiffBind will be used to generate the count matrix using the function dba.count(bUseSummarizedOverlaps=TRUE) and normalized with dba.normalized(normalize=DBA_NORM_TMM). Differential accessibility analysis will be performed with dba.analyze(method=DBA_DESEQ2).

```
#Peakset per condition and time
bedtools intesect -a B_0_untreated.narrowPeaks \
    -b C_0_untreated.narrowPeaks [...] H_0_untreated.narrowPeaks -f 0.30 >
    consensus_0_untreated.narrowPeak

bedtools intesect -a B_0_treated.narrowPeaks \
    -b C_0_treated.narrowPeaks [...] H_0_treated.narrowPeaks -f 0.30 >
    consensus_0_treated.narrowPeak

[...]

bedtools intesect -a B_24_untreated.narrowPeaks \
    -b C_24_untreated.narrowPeaks [...] H_24_untreated.narrowPeaks -f 0.30 >
    consensus_24_untreated.narrowPeak

bedtools intesect -a B_24_treated.narrowPeaks \
    -b C_24_treated.narrowPeaks [...] H_24_treated.narrowPeaks -f 0.30 >
    consensus_24_treated.narrowPeak

#Global peak set
cat consensus_0_untreated.narrowPeak consensus_0_treated.narrowPeak [...]
    consensus_24_treated.narrowPeak > unsorted_global.bed
bedtools sort -i unsorted_global.bed > sorted_global.bed
bedtools merge -i sorted_global.bed > global_consensus.bed
```

```
#DiffBind
samples <- read.csv('meta/samplesheet.csv') #This file contains the path to the bam files
    and the MACS peaks
dbObj <- dba(sampleSheet=samples)

#make counts object
dbObj <- dba.count(dbObj, bUseSummarizeOverlaps=TRUE)

#normalize counts
dbObj <- dba.count(dbObj, normalize=DBA_NORM_TMM)
```

```
#make the differential accessibility analysis
dbObj <- dba.analyze(dbObj, method=DBA_ALL_METHODS)

#set contrast
dbObj <- dba.contrast(dbObj, categories=DBA_FACTOR, minMembers = 7)

#retrieve the results for a specific contrast
res_deseq <- dba.report(dbObj, method=DBA_DESEQ2, contrast = 1, th=1)
```

### 3.4 Integrative analysis of RNA-seq and ATAC-seq data

This protocol was taken from [1] and [2]:

Peaks will be assigned to the nearest TSS, we can use ChIPseeker library for this. Pearson correlation can evaluate the relationship between the promoter and the gene they regulate; promoters that have greater accessibility should display greater expression, and genes for which regulatory regions are closed, their expression should also be down-regulated.

```
#annotate peaks with ChIPseeker
peaks_DBS_annotated <- annotatePeak(res_deseq, tssRegion=c(-3000,3000),
    TxDb=TxDb.Hsapiens.UCSC.hg38.refGene, annoDb="org.Hs.eg.db")
```

Quantitative matrices of both genes and peaks are to be provided. Lowly expressed genes are to be filtered out (TMP ¿= 0.1 across all samples). Pearson correlation will be calculated for all open chromatin regions and their possible target gene. As background, $\tilde{1}0$ 000 randomly selected peaks will be used and Pearson correlation coefficient will be calculated. Z-test will be used to test the association of each pair. FDR correction will be performed.

Footprinting analysis can help build a gene regulatory network by establishing TF-to-gene relationships. Differentially expressed genes can be analyzed with DAVID, and for differentially accessible regions they can be analyzed with GREAT.

# References


[1] Tao Zhu et al. "cisDynet: An integrated platform for modeling gene-regulatory dynamics and networks". *iMeta* 2 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/imt2.152, e152, 2023. ISSN: 2770-596X. DOI: 10.1002/imt2.152. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/imt2.152 (visited on 03/19/2024).

[2] Min Ding et al. "Integration of ATAC-Seq and RNA-Seq reveals FOSL2 drives human liver progenitor-like cell aging by regulating inflammatory factors". *BMC Genomics* 24, p. 260, 2023. ISSN: 1471-2164. DOI: 10.1186/s12864-023-09349-7. URL: https://doi.org/10.1186/s12864-023-09349-7 (visited on 03/18/2024).