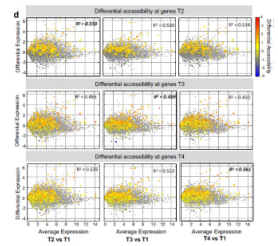Oct 13, 2020

🌐 Chromatin accessibility landscapes during onset of salmon maturation

DOI

dx.doi.org/10.17504/protocols.io.bndxma7n

Amin R Mohamed[1], Marina Naval-Sanchez[2], Moira Menzies[1], Bradley Evans[3], Harry King[1], Antonio Reverter[1], James Kijas[1]

[1]CSIRO; [2]University of Queensland; [3]Tassal

Salmon Multiomics

Amin R Mohamed

Protocol status: Working
We use this protocol and it's working

Created: October 13, 2020

Last Modified: October 13, 2020

Protocol Integer ID: 43159

Abstract

Background:
Despite sexual development being ubiquitous to vertebrates, the epigenetic mechanisms controlling this fundamental transition remain largely undocumented in many organisms. Our previous own work showed that despite DNA methylations played a key role upregulating a defined set of genes during the maturation process, methylation alone does not control genome-wide patterns of gene expression. This prompted us to to characterise the epigenomic features during onset of maturation at the chromatin level.

Results:
We performed ATAC-seq (assay for transposase-accessible chromatin sequencing) to produce genome-wide maps of chromatin accessibility changes. ATAC-seq was performed for multiple tissues and peak enrichment around transcription start sites used as the key quality control metric (TSS). Following data pruning, we took 12 liver libraries (3 replicates across all 4 timepoints) and a total of 699 million uniquely mapped paired-end reads forward into joint analysis with RNA-seq and WGBS data. To characterise changes in chromatin state following long light initiation, we defined differentially accessible regions (DARs) where mapping counts differed significantly between T1 and other time points. This revealed a strong early remodelling in the chromatin state landscape, as most DARs were observed at T2 (n=1501) before decreasing in stepwise fashion at T3 (n=477) and T4 (n=148). The direction of change was approximately balanced between DARs with increased and decreased accessibility, broadly matching the balance between up and down regulated global gene expression changes observed for liver. We next asked if the early changes in chromatin state persisted throughout the time course using hierarchical clustering. The majority of DARs (n=1036; 57%) exhibit reduced accessibility at T2 compared with T1 and subsequently remained unchanged at later time points. Similarly, regions that gained accessibility at T2 (n=696; 38%) also remained unchanged at later timepoints. This left less than 10% of DARs (n=99) that displayed an oscillating pattern following the onset of the maturation. Together, this revealed the ATAC-seq signatures were predominantly stable chromatin state changes, as opposed to pulsatile epigenomic changes that snapped back after a small number of days or weeks. We observed high correlation between chromatin state changes and altered gene expression for the single tissue with ATAC-seq data (liver). The correlation was highest for differentially accessible regions immediately adjacent to coding genes, implicating *cis*-regulatory elements (CREs). To examine the biological consequence of chromatin state changes, we focused on CREs given their established role on transcriptional regulation via transcription factors (TFs) binding. We focussed on the subset of CREs that underwent a change in accessibility during the time course to evaluate i) the expression behaviour of their closet gene; ii) the biological function of those genes, and iii) any enrichment for transcription factor binding sites (motifs). We found a small subset of CREs (n = 65) underwent increased accessibility early in the time course and the majority (n = 46; 79%$'X^2 p < 8.028^{-08}$) upregulated their nearest gene in a tightly coordinated manner. It also appears CREs more tightly controlled the downregulation of genes compared to DARs located in gene bodies, downstream regions or within intergenic regions. The gene set associated with coordinated up regulation exhibited significant GO enrichment related to lipid metabolism and energy metabolism (AAcyl-CoA biosynthesis). Acyl-CoA are coenzymes involved in energy synthesis, consistent with the expectation of liver function through an energetically costly transition such as maturation.

Conclusions:

The results clearly demonstrated chromatin state changes played a dominant role in directing global changes in gene expression and strongly suggest that chromatin state changes at CREs directly control gene expression in liver and upregulate energy metabolism genes via changes in TF activity.

Guidelines

scripts could be found here https://github.com/AminRM/Salmon-Chromatin-Accessibility
data could be found here https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE156998

Nuclei extraction, ATAC-seq library preparation and sequencing

1   ATAC-seq libraries were prepared from frozen tissues using the Omni-ATAC method with the following modifications.

Frozen tissue (20 mg) was ground in liquid nitrogen using a mortar and pestle.

The pulverized tissue was transferred to a pre-chilled 2 ml dounce homogenizer containing 1mL cold 1x homogenisation buffer and homogenised with the pestle to form a uniform suspension (10-20 strokes).

The homogenate was filtered with a 40uM nylon cell strainer (BD Falcon) before layering onto the iodixanol solution as described previously.

The ratio of nuclei to enzyme concentration was optimised for each sample by performing transposition reactions containing 50000, 100000 and 200000 nuclei with 2.5ul of tagment enzyme in 50ul of transposition mix.

The transposed DNA was amplified with custom primers as described elsewhere. before libraries were purified using Agencourt AMPure XP beads (Beckman Coulter) and quality controlled using a Bioanalyser High Sensitivity DNA Analysis kit (Agilent).

Twelve liver ATAC-seq libraries arising from 3 biological replicates x 4 time points (T1-T4) were sequenced at the IMB sequencing facility (University of Queensland) on an Illumina NextSeq 150 cycle (2 × 75 bp). Sequencing produced a total of 1.2 billion individual paired-end reads.

ATAC-seq data QC, genome mapping and peak calling

2   Raw reads were mapped to the Atlantic salmon reference genome ICSASG_v2 using BOWTIE2 version 2.3.5.1 with the --very-sensitive parameter.

Duplicate reads were removed using the MarkDuplicates function in Picard (http://broadinstitute.github.io/picard/).

Multi-mapped reads and mitochondrial reads were filtered out and only uniquely mapped reads (MAPQ > 10) were extracted from alignment files using SAMTOOLS for downstream analyses.

For peak calling, the model-based analysis of ChIP-seq (MACS2) (https://github.com/macs3-project/MACS) was used to identify read enrichment regions

"peaks" using default parameters.

Only peaks detected in at least two replicates per condition were used for downstream analyses, and peaks across timepoints were merged to generate a unique peak list per tissue.

The number of raw reads mapped to each peak was quantified using the Python package HTSeq version 0.11.1

Raw accessibility counts for all samples were obtained and a gene expression matrix was prepared in R

Differential accessibility and clustering analyses

3   Samples from the long photoperiod time points (T2, T3 and T4) were compared to control samples (T1) for each tissue. Raw counts were analysed using the R package edgeR and P-values were corrected for multiple testing using the Benjamini and Hochberg algorithm. Peaks with FDR < 0.05 and log2FC > ± 1 were considered significantly differentially accessible regions (DARs).

PCA of significant DARs used normalised accessibility data (log2CPM) prepared using the function *--prin_comp* within trinity. Hierarchical clustering analysis was conducted using *analyze_diff_expr.pl* where mean-centred normalized accessibility (log2CPM+1) were compared across time points.

Gene clusters with similar accessibility patterns were obtained using the Perl script *define_clusters_by_cutting_tree.pl* t o cut the hierarchically clustered gene tree into clusters with similar accessibility patterns as described above.

Hierarchical clustering identified both accessible and inaccessible DAR clusters. DARs per cluster were annotated in a genomic context (genic, promoter, 5 kb downstream or intergenic)

ATAC-seq and RNA-seq correspondence analysis

4   Only DARs co-located with genes and promoters were used for co-analysis with gene expression data. The relationship between accessibility of DARs and gene expression was visualised by overlying information of significant DARs to genome-wide normalised expression estimates in liver samples and plotted as a MA-biplot.

A linear regression analyses were performed to assess correlations between accessibility and expression abundance and the effect of changes in accessibility and changes in gene expression across time.

Chromatin accessibility and gene expression data were visualised using Gnuplot version 5.0.7 (http://www.gnuplot.info) by overlying accessibility data of significant DARs at genes and promoters to genome-wide normalised expression estimates at each timepoint.

Motif enrichment analyses

5   The function *findMotifsGenome.pl* within Homer software version 4.11 (**http://homer.ucsd.edu/homer/**) was used with default parameters to find sequence motifs significantly enriched among accessible DARs against background of inaccessible DARs located within promoter regions.

TF motifs that are highly enriched (Pvalue $< 1 \times 10^{-10}$) were selected.

Multiomic heatmap analysis per time and genomic regions

6   All heatmaps were produced using the R package pheatmap.

GO enrichment analyses have been conducted on the set of nearest genes to accessible promoters using the R package clusterProfiler as described above.

The integrated genome viewer (IGV) was used to visualise the relationship between accessibility and gene expression in a 15kb region that contains the *hmgrc* gene and its promoter region.