

Oct 13, 2020

🌐 Multitissue DNA methylome profiling during onset of salmon maturation

DOI

dx.doi.org/10.17504/protocols.io.bncrmav6

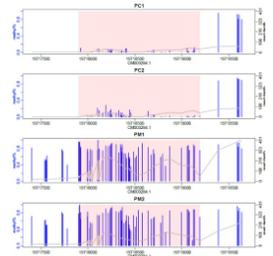
Amin R Mohamed¹, Bradley Evans², Antonio Reverter¹, James Kijas¹

¹CSIRO; ²Tassal

Salmon Multiomics



Amin R Mohamed



OPEN  ACCESS



DOI: dx.doi.org/10.17504/protocols.io.bncrmav6

Protocol Citation: Amin R Mohamed, Bradley Evans, Antonio Reverter, James Kijas 2020. Multitissue DNA methylome profiling during onset of salmon maturation. **protocols.io** <https://dx.doi.org/10.17504/protocols.io.bncrmav6>

Manuscript citation:

Mohamed et al (2020) Integrated transcriptome, DNA methylome and chromatin state accessibility landscapes reveal regulators of Atlantic salmon maturation, bioRxiv 2020.08.28.272286; doi: <https://doi.org/10.1101/2020.08.28.272286>

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: October 12, 2020

Last Modified: October 13, 2020

Protocol Integer ID: 43121

Keywords: epigenetics, DNA methylome, WGBS, CpG, maturation, salmon, aquaculture ,

Abstract

Background:

Atlantic salmon farming promotes growth in conditions which mean animals may complete sexual development at weights below harvest size leading to a reduction in productivity. This prompted us to investigate the biological mechanisms that control the timing of sexual maturation. We performed a time course experiment, whereby animals were manipulated with photoperiod before tissues were collected across the time window when animals commence sexual development. We performed whole genome bisulfite sequencing (WGBS) of three salmon tissues (pituitary, ovary and liver) at both the beginning and end of the experiment (T1 and T4), to take a first look at the patterns of DNA methylation and examine how they change in response to the onset of an important life history trait.

Results:

Comparison across timepoints revealed 6,373 differentially methylated regions (DMRs), of which approximately 50% were located within genes (DMGs). The ovary underwent the most profound remodelling, with a strong bias towards increased methylation levels (hyper-methylation) specially at gene bodies (gene body methylation). The majority of differentially methylated genes (DMGs) in ovary were hypermethylated ($n=1165$; 74%) and significantly enriched for three biological process (GO-BP), one cellular component (GO-CC) and 24 molecular function (GO-MF) terms including those with maturation-related functions such as *asemaphorin/glutamate receptor activity*. We also performed deep transcriptomic profiling (RNA-seq) of the same tissues to explore the relationship between methylation changes and gene expression. Weak correlation was observed considering all available genes, suggesting methylation may not be the key epigenomic regulator of global expression in the context of our experiment. However, the identification of significant transcriptional and methylation changes allowed us to explore the dynamic between these two processes by assessing the overlap of genes declared as both DEG and DMG. The overlap was low and non-significant for liver (38 / 616 or 6% of DMGs were also DEGs) and pituitary (11 / 762 or 1.4% of DMGs were DEGs). Strikingly, 195 or 14% of ovary DMGs (195 / 1357) were also differentially expressed, a number that exceeded random expectation in 83.8% of 1000 permutations tests. This suggests changes in methylation status may directly control gene expression in this subset of genes. If true, we would expect to see correspondence between the directionality of the expression and methylation changes. This appeared to be the case, as 82% of upregulated genes (148 / 179; Binomial P-value = $8.727E-20$) were hyper-methylated at T4 relative to T1, matching the classical expectation of gene body methylation mediated control of gene expression (Neri et al 2017, Arechederra et al 2018). The 148 genes were enriched for 3 GO-CC terms related to chromatin remodelling complexes (SWI/SNF and nBAF) and the associated genes displayed coordinated expression and methylation status.

Conclusion and implications:

Taken together, the results confirmed that while methylation alone does not control genome-wide patterns of gene expression, it plays a key role upregulating a defined set of genes during the maturation process. Co-analysis of transcriptome and DNA methylome in ovary suggests chromatin remodelling genes play a role in the commitment of animals to the sexual maturation pathway. These results also open the way for the identification of functional variants that can be used in advanced breeding approaches to boost productivity in Atlantic salmon farming.

References:

Neri, F. et al. Intragenic DNA methylation prevents spurious transcription initiation. *Nature* 543, 72–77 (2017).

Arechederra, M. et al. Hypermethylation of gene body CpG islands predicts high dosage of functional oncogenes in liver cancer. *Nat. Commun.* 9, 3164 (2018).

Attachments



[Methyl_workflow.pdf](#)

26KB

Guidelines

methylation data could be found here <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE156997>

Perl, shell and R scripts could be found here https://github.com/AminRM/salmon_WGBS

Genomic DNA isolation, WGBS library preparation and sequencing

- 1 Tissue samples were snap frozen in liquid Nitrogen and stored at $-80\text{ }^{\circ}\text{C}$ until genomic DNA (gDNA) was extracted using DNeasy blood and tissue kit (QIAGEN).

Tissues were lysed in $360\text{ }\mu\text{L}$ of lysis solution on a Precellys 24 homogenizer for 30s at 4.0 ms^{-1} .

Samples were incubated with $40\text{ }\mu\text{L}$ of Proteinase K enzyme at $56\text{ }^{\circ}\text{C}$ for 1 h.

Samples were treated with RNase ($8\text{ }\mu\text{L}$ of RNase A incubated for 2 min at room temperature).

DNA was bound to the provided columns, washed twice and eluted in $100\text{ }\mu\text{L}$ at room temperature.

gDNA purity were assessed by gel electrophoresis and NanoDrop ND-1000 spectrometer. DNA concentration and integrity were assessed using Agilent 2100 bioanalyzer.

gDNA was fragmented (200-400bp) by sonication using Covaris S220, followed by end repair/adenylation and adapter ligation.

Bisulfite modification was performed to the DNA fragments using the EZ DNA Methylation-Gold™ Kit (Zymo Research, Inc.).

Twelve libraries prepared from 3 tissues (pituitary, ovary and liver), 2 time points (T1 and T4) and 2 biological replicates.

Libraries were sequenced on HiSeq 2500 sequencing platform at Novogene, Hong Kong. Sequencing produced a total of 2.2 billion individual 150 bp paired-end reads and 185 M PE reads per library. Bisulfite conversion rates (percentage of C changed to T after bisulfite treatment) were consistently $>99.8\%$.

WGBS data QC, genome mapping and methylation calling

- 2 Raw data quality control was performed using Trim Galore v0.5 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to filter bases (Q scores < 30) and remove both universal and indexed adapter sequences.

Processed high-quality data were mapped to into a bisulfite-converted version of the Atlantic salmon reference genome ICSASG_v2 using BSseeker2 v2.1.8 with default parameters for aligning paired-end libraries using Bowtie2 (Langmead & Salzberg 2012).

PCR duplicates were detected and removed using Picard MarkDuplicates (<http://broadinstitute.github.io/picard/>).

Filtered (duplicates-free) reads (110 M PE reads) were retained for downstream methylation analysis with an average genome coverage of 11x in pituitary, ovary and liver.

Methylation calling was conducted using the Python script *call-methylation.py* within BSseeker2.

CGmap files were used for subsequent exploratory and differential methylation analyses.

DNA methylome exploratory analyses

- 3 The *mstat* command within CGmap tools was used to generate global and CG context (CG, CHG, CHH) DNA methylation levels.

Among the different dinucleotide contexts, CpG methylation contributed the vast majority (~ 99.5% on average) compared to CHH or CHG methylation which were excluded from further analysis.

As CG methylation contributed to the bulk of methylated Cs, average methylation levels of genome-wide CpG positions were calculated in 50 kb bins across the genome using *mbin* command within CGmap tools and plotted as Violin plots using the R package *vioplot* (<https://cran.r-project.org/web/packages/vioplot/index.html>).

To begin assessment of the quality of our libraries, common CpGs with minimum 10x coverage among the 12 samples were used in PCA using *prcomp* implemented in R.

Correlation matrices (based on Pearson coefficient) were prepared using the R package *corrplot* (<https://cran.r-project.org/web/packages/corrplot/index.html>).

Hierarchical clustering analysis was conducted with *hclust* implemented in R using compute linkage and Euclidean distances.

Differential CpG methylation analysis

- 4 The R package DSS was used to identify differential methylation regions using common CpGs.

In each tissue, two replicates at T4 were compared to the control samples at T1 based on CpG methylation levels. At each CpG site, the methylation (M) level was calculated as a proportion of the total counts (coverage) as follows: $M \text{ levels} = [\text{methylated counts} / \text{total counts}] \times 100$. DSS was selected as it takes into account the biological variation among replicates (characterized by a dispersion parameter) and the sequencing depth.

Differentially methylated loci (DMLs) were identified by estimating mean methylation for all CpG sites followed by estimating dispersion at each site and conducting a Wald test ($P < 0.001$). Smoothing (combining information from nearby CpG sites to improve the estimation of methylation levels) was utilised to obtain mean methylation estimates in WGBS data where the CpG sites are dense.

Based on the DML results, regions with statistically significant CpG sites were identified as a differentially methylated regions (DMRs) with minimum length/distance of 50 bp and minimum CpG coverage of 3.

Mean methylation between groups of greater than 10 % ($\text{delta} = 0.1$) and $P < 0.001$ was considered significant.

A circos plot was produced to visualize multi-tissue genome-wide DMRs using Circos (<http://circos.ca/software/>).

Individual DMRs were also visualized using the *showOneDMR* function within the DSS package to plot both the methylation percentages (including a smoothed curve) as well as the coverage depths at each CpG site.

DMR genomic annotation

- 5 Differentially methylated regions were compared against the protein coding gene set annotated on reference ICSASG_v2 using a custom Perl script

This classified DMRs as overlapping a gene body (genic), 5kb upstream of a transcription start site TSS (putative promoter), 5 kb downstream of TSS (5kb downstream), or otherwise intergenic.

The distance between each DMR and nearest gene is provided.

DMGs and DEGs correspondence analysis



- 6 The overlap between significant genes from differential expression and methylation was checked using the *intersect* function within bedtools

The relationship between gene body methylation and gene expression was visualised by overlying information of significant DMGs to genome-wide normalised expression estimates in ovary samples and plotted as a MA-biplot using Gnuplot version 5.0.7 (<http://www.gnuplot.info>).

A linear regression analyses were performed to assess correlations between methylation and expression abundance.

DMGs functional profile

- 7 GO enrichment analyses were conducted on both the sets of hypermethylated genes (n= 1,156) and genes found to be hypermethylated and upregulated in ovary (n= 148) using the R package clusterProfiler.

Genes driving GO enrichment were plotted as a heatmap using the R package pheatmap