



Oct 12, 2020

Multitissue transcriptome profiling during onset of salmon maturation

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

¹CSIRO; ²Tassal

Amin R Mohamed: CSIRO;

Antonio Reverter: CSIRO

James Kijas: CSIRO

1

Works for me

dx.doi.org/10.17504/protocols.io.bnbdmai6

Salmon Multiomics



Amin Mohamed

ABSTRACT

Despite the importance of sexual maturation as a trait of interest it can be difficult to study, as the timing of onset varies widely in response to both genetics and environmental factors and occurs prior to measurable phenotypic change. To overcome this, we chose to investigate sexual maturation in Atlantic salmon where photoperiod manipulation in an experimental system can be used to synchronise animals and access tissues across the time period when animals first commit to the onset of puberty. We describe changes in genome-wide gene expression in pituitary gland, ovary and liver to identify transcriptional landscapes associated with maturation in salmon. We sequenced mRNA from four biological replicates of each tissue before (T1) and after the onset of maturation (T2, T3, T4). A total of 3.2 billion paired-end reads were mapped against the Atlantic salmon reference genome with 72% mapping efficiency to create an average depth of 50 million reads per library. The number of differentially expressed genes (DEGs) increased with elapsed time following the onset of the long light photoperiod for the two BPG axis tissues (pituitary and ovary). Of these, the ovary underwent the most dramatic remodelling over time with 403, 1709 and then 3497 DEGs observed at timepoints T2, T3 and T4 respectively. This increasing trajectory of differential gene expression, coupled with the elevated GSI following the light stimuli, strongly suggests the experimental approach successfully initiated the onset of maturation. Next, we identified clusters of DEGs in each of the analysed tissues, which describe their physiological roles. The identity of the DEGs, in response to the onset of maturation, revealed the key players involved in the earliest triggers into maturation. Among these, upregulation of genes encoding specific pituitary hormones such as gonadotropins along with genes encoding transcription factors (such as GATA2) are significant in controlling onset of maturation. To characterize the transcriptomic remodelling occurring in the ovary and liver, we assessed the DEG sets for GO enrichment. Upregulated genes in the ovary revealed processes related to *cell adhesion*, *immune/inflammatory response* and *development*, while gene families involved in *organic acid metabolic processes* and *mitochondrial transport* were enriched among liver upregulated genes. This multitissue transcriptome paper provides novel insights into the transcriptional signature associated with onset of salmon maturation and is the basis of subsequent epigenetic studies aimed at understanding epigenetic mechanisms underlying maturation.

THIS PROTOCOL ACCOMPANIES THE FOLLOWING PUBLICATION

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>

ATTACHMENTS

[Exp_workflow.pdf](#)

DOI

dx.doi.org/10.17504/protocols.io.bnbdmai6

PROTOCOL CITATION

Amin R Mohamed, Bradley Evans, Antonio Reverter, James Kijas 2020. Multitissue transcriptome profiling during onset of salmon maturation. **protocols.io**



MANUSCRIPT CITATION please remember to cite the following publication along with this protocol

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>

KEYWORDS

salmon, maturation, puberty, transcriptome, pituitary, gonadotropins, RNA-seq, aquaculture, genomics

LICENSE

————— This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

CREATED

Oct 12, 2020

LAST MODIFIED

Oct 12, 2020

PROTOCOL INTEGER ID

43077

ATTACHMENTS

[Exp_workflow.pdf](#)

GUIDELINES

shell and R scripts could be found here https://github.com/AminRM/salmon_mat_transcriptomes
Raw/normalised data could be found here <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE157003>

ABSTRACT

Despite the importance of sexual maturation as a trait of interest it can be difficult to study, as the timing of onset varies widely in response to both genetics and environmental factors and occurs prior to measurable phenotypic change. To overcome this, we chose to investigate sexual maturation in Atlantic salmon where photoperiod manipulation in an experimental system can be used to synchronise animals and access tissues across the time period when animals first commit to the onset of puberty. We describe changes in genome-wide gene expression in pituitary gland, ovary and liver to identify transcriptional landscapes associated with maturation in salmon. We sequenced mRNA from four biological replicates of each tissue before (T1) and after the onset of maturation (T2, T3, T4). A total of 3.2 billion paired-end reads were mapped against the Atlantic salmon reference genome with 72% mapping efficiency to create an average depth of 50 million reads per library. The number of differentially expressed genes (DEGs) increased with elapsed time following the onset of the long light photoperiod for the two BPG axis tissues (pituitary and ovary). Of these, the ovary underwent the most dramatic remodelling over time with 403, 1709 and then 3497 DEGs observed at timepoints T2, T3 and T4 respectively. This increasing trajectory of differential gene expression, coupled with the elevated GSI following the light stimuli, strongly suggests the experimental approach successfully initiated the onset of maturation. Next, we identified clusters of DEGs in each of the analysed tissues, which describe their physiological roles. The identity of the DEGs, in response to the onset of maturation, revealed the key players involved in the earliest triggers into maturation. Among these, upregulation of genes encoding specific pituitary hormones such as gonadotropins along with genes encoding transcription factors (such as GATA2) are significant in controlling onset of maturation. To characterize the transcriptomic remodelling occurring in the ovary and liver, we assessed the DEG sets for GO enrichment. Upregulated genes in the ovary revealed processes related to *cell adhesion*, *immune/inflammatory response* and *development*, while gene families involved in *organic acid metabolic processes* and *mitochondrial transport* were enriched among liver upregulated genes. This multitissue transcriptome paper provides novel insights into the transcriptional signature associated with onset of salmon maturation and is the basis of subsequent epigenetic studies aimed at

Induction of maturation through photoperiod manipulation

- 1 Animals were managed using photoperiod manipulation to synchronise the timing of commitment into maturation. A population of female brood stock were used that were ~ 36 months post fertilization in April 2017.

In order to measure and control for variation between individuals, 4 fish (biological replicates) at each of the four time points (T1-T4) were used. The maturation status of animals (leading up to the long day photoperiod initiation) was monitored by ultrasound.

Control samples at T1 time point were collected on mid-June 2017 before induction of maturation occurred late-June 2017.

Following the application of the long photoperiod, tissues were sampled at different three time points in 2 weeks intervals (T2-T4).

At each sampling event, the gonadosomatic index GSI was calculated from the ovary mass as a proportion of the total body mass as follows: $GSI = [\text{ovary weight} / \text{total body weight}] \times 100$.

RNA isolation

- 2 Tissue samples were preserved in RNA-Later at -80°C and total RNA was isolated using RNeasy mini kit (QIAGEN).

Tissues were lysed twice in 450 μL of lysis solution on a Precellys 24 homogenizer for 30s at 4.0 ms^{-1} .

RNA was bound to a column and washed twice before elution with 40 μL at room temperature.

RNA quantity and quality were assessed using a NanoDrop ND-1000 spectrometer, Qubit 2.0 fluorometer and Agilent 2100 bioanalyzer. Messenger RNA (mRNA) was isolated from 1 μg of total RNA.

64 RNA-Seq libraries (4 time points x 4 tissues x 4 biological replicates) were prepared using the TruSeq RNA Sample Preparation Kit (Illumina).

Libraries were sequenced using a whole illumina flow cell on the Nova-Seq 6000 sequencing platform at the Australian Genome Research Facility (AGRF) in Melbourne, Australia.

RNA-seq analysis

- 3 Illumina reads were checked for quality using FastQC software. Sequencing produced a total of 4.4 billion individual 150 bp paired-end reads and ~ 70 million PE reads per library

High quality reads ($Q>30$) were mapped to the Atlantic salmon genome ICSASG_v2 using TopHat2 version 2.1.1 with default parameters

A total of 3.2 billion paired-end reads were mapped against the Atlantic salmon reference genome with 72% mapping efficiency to create an average depth of 50 million reads per library

Alignment files in BAM format were sorted by read name and converted into SAM format using SAMtools version 1.4

The Python package HTSeq version 0.7.2 was applied to count unique reads mapped to exons using default parameters except for "*reverse*" with the strandedness

Gene expression matrix that contains raw counts (obtained from HTSeq) for all samples in R

Exploratory data analyses were conducted through many functions within trinity to perform PCA and hierarchical clustering to check on the relative amount of variation between replicates within a time point per tissue.

Pairwise Spearman correlations revealed low variation among biological replicates and separation between samples from different timepoints for the ovary, liver and pituitary. Analysis of brain samples revealed much higher variability

among biological replicates within timepoint, suggesting a low quality dataset that was excluded from subsequent analysis.

We conclude the sampling of brain tissue was non uniform between fish, capturing different brain regions and generating the highly variable patterns in gene expression observed. Consequently, brain samples were excluded and 48 libraries from pituitary, ovary and liver were used for subsequent differential expression and clustering analyses.

differential expression analyses

- 4 Raw counts were analysed using the edgeR package in the R statistical computing environment to infer differential gene expression among tissues.

The four tissues at the long photoperiod time points (T2, T3 and T4) were compared to the control samples at T1. *P*-values for differential gene expression were corrected for multiple testing using the Benjamini and Hochberg algorithm.

For further analyses of differential expression, only genes with a false discovery rate (FDR) of < 0.05 and have at least absolute $\log_2(\text{fold change}) > 1$ were considered significant.

Volcano plots were used to visualise DEGs to explore the dynamics of expression changes across tissue and time

PCA was conducted on the lists of significant DEGs using normalised expression data ($\log_2\text{FPKM}$) using the function "*-prin_comp*" within trinity

Hierarchical clustering analysis was conducted using trinity's utility *analyze_diff_expr.pl* on significant DEGs in each tissue where mean-centred normalized expression (\log_2 -transformed FPKM+1) were compared across time points

Gene clusters with similar expression patterns were obtained using the Perl script "*define_clusters_by_cutting_tree.pl*" within trinity to cut the hierarchically clustered gene tree into clusters with similar expression using the *-Ptree* option.

Functional Profile

- 5 To infer the functions of the gene clusters, gene ontology (GO) enrichment was performed to identify the enriched biological themes using the R package clusterProfiler version 3.9 using default settings.

The ENTREZ gene identifiers of up- and downregulated clusters per tissue were used as query gene list against the background genes in each tissue.

For the purpose of the enrichment analysis, GO categories with a corrected *P*-value of < 0.05 were considered significant.

Categories of candidate genes implicated in maturation were visualised as heatmaps using their normalised expression values with the R package pheatmap version 1.0.12

<https://cran.r-project.org/web/packages/pheatmap/pheatmap.pdf>.