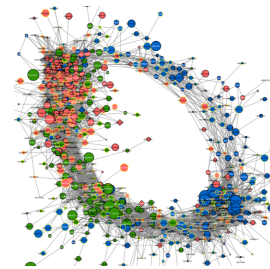


Oct 10, 2020

## Gene Regulatory Network

DOI

[dx.doi.org/10.17504/protocols.io.bm6rk9d6](https://dx.doi.org/10.17504/protocols.io.bm6rk9d6)



Amin R Mohamed<sup>1</sup>, Antonio Reverter<sup>1</sup>, James Kijas<sup>1</sup>

<sup>1</sup>CSIRO

Salmon Multiomics



Amin R Mohamed

### Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.bm6rk9d6>

**Protocol Citation:** Amin R Mohamed, Antonio Reverter, James Kijas 2020. Gene Regulatory Network. **protocols.io**  
<https://dx.doi.org/10.17504/protocols.io.bm6rk9d6>



**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 08, 2020

**Last Modified:** October 10, 2020

**Protocol Integer ID:** 42929

**Keywords:** systems biology, multiomics, regulatory, networks, gene regulatory network gene regulatory network, gene regulatory network gene, regulatory network, gene network, connections between gene node, gene network inference, transcriptional regulation of complex trait, gene node, transcriptional regulation, methylation at gene body, chromatin accessibility, regulatory impact factor, expressed gene, gene, underpinning complex trait, integrating multiomic data, multiomic data, models of biomolecular interaction network, connected gene, biomolecular interaction network, gene body, complex trait, multiomics approach, other complex trait, bioinformatics, genome, biological transitions such as puberty, construction of separate network, separate network,

## Abstract

Gene regulatory networks (GRNs) provide a platform for integrating multiomic data and can be used to characterize the dynamics of perturbations during biological transitions such as puberty and other complex traits. We used a multiomics approach, which has the power to identify the control mechanisms underpinning complex traits (Argelaguet et al 2019; Lloyd-Price et al 2019). We also utilised a Systems Biology approach to co-analyse genes with evidence of differential behaviour using seven categories that included expression (DEGs), changed methylation at gene bodies (DMGs) or promoters (DMPs) and differential chromatin accessibility (DACs) into gene network. To focus the analysis towards investigation of key regulators, we also performed regulatory impact factor (RIF) analysis (Reverter et al 2010). This used co-expression correlation between TFs and their target differentially expressed genes to identify master regulator TFs. For gene network inference, genes were used as nodes and significant connections (edges) between them were identified using the Partial Correlation and Information Theory (PCIT) algorithm (Reverter & Chan 2008) considering all samples. PCIT determinates the significance of the correlation between two nodes after accounting for all the other nodes in the network. Connections between gene nodes were accepted when the partial correlation was greater than two standard deviations from the mean ( $P < 0.05$ ). The output of PCIT was visualized using Cytoscape Version 3.7.2 (Shannon et al 2003). Key regulators are likely to undergo substantial change in their number of connections and identify gene networks driving the transition to maturation. This prompted construction of separate networks for each physiological stage, before identifying those genes that underwent the largest change in connectivity of differentially connected genes (DCGs).

## References:

- Argelaguet, R. et al. Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* <https://doi.org/10.1038/s41586-019-1825-8> (2019)
- Lloyd-Price, J. et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* 569,655–662 (2019)
- Reverter, A. & Chan, E.K.F. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* 24,2491–2497 (2008).
- Reverter, A., Hudson, N. J., Nagaraj, S. H., Perez-Enciso, M. & Dalrymple, B. P. Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. *Bioinformatics* 26,896–904 (2010).
- Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*13, 2498–2504 (2003)



## Guidelines

This analysis requires obtaining results from multiomic experiments.

Genes needs to be filtered to have 2000 genes for network construction using PCIT.

Attribute table is required to assign genes to the different analysis.

## Troubleshooting

## Before start

Prepare gene expression matrix with all samples and normalised (log2FPKM) expression values for all genes from different analyses.

1 Master regulator analysis was performed using regulatory impact factor (RIF) metrics (Reverter et al., 2010) to identify key regulators contributing to the differential expression in the T4 vs T1 comparison in each tissue. Data for potential transcription factors (TFs) in Atlantic salmon were retrieved (Mohamed et al., 2018). RIF exploits the differential co-expression concept where regulators were contrasted against unique lists of genes that were differentially expressed at T4 in each tissue. RIF exploits the differential co-expression concept where regulators were contrasted against unique lists of genes that were differentially expressed at T4 in each tissue. Genes with a mean expression FPKM  $< 0.2$  were excluded. Those scores deviating  $\pm 2.57$  standard deviation from the mean were considered significant at  $P < 0.01$ . We identified a total of 305 significant regulators (113, 68 and 123 in pituitary, ovary and liver, respectively at  $P < 0.01$ ). Most of these regulators ( $n=298$ ; 97.7%) were unique to each tissue leaving only 7 that were shared among tissue pairs. The regulators identified were used as input for construction of gene regulatory networks as described in step # 2.

2 Genes selected for the network analysis originated from different omics analysis (transcriptome, DNA methylome and Chromatin accessibility). genes were used as nodes and significant connections (edges) between them were identified using the Partial Correlation and Information Theory (PCIT) algorithm (Reverter and Chan, 2008), considering all 48 samples.

Genes from different omics analyses (DEGs, DMGs, DMPs, DACs) along with key transcription factors identified by RIF (TFs), as well as information for tissue-specific (TS) genes and gene-harboursing GWAS SNPs (SNPs) were selected based on overlap (at least once) and mean normalised expression (at least 0.2 FPKM) for network construction. The R package UpSetR (<https://cran.r-project.org/web/packages/UpSetR/vignettes/basic.usage.html>) was used to investigate the cross-talk among genes from different sources. For gene network inference, genes were used as nodes and significant connections (edges) between them were identified using the Partial Correlation and Information Theory (PCIT) algorithm, considering all samples. PCIT determinates the significance of the correlation between two nodes after accounting for all the other nodes in the network. Connections between gene nodes were accepted when the partial correlation was greater than two standard deviations from the mean ( $P < 0.05$ ). The output of PCIT was visualized using Cytoscape Version 3.7.2.

3 Differential connectivity



In order to explore differential connectivity during maturation onset, two networks were created; one using 12 samples at T1 (pre-maturation) and a second using 36 samples at T2, T3 and T4 (post-maturation). The number of connections of each gene in each network was computed, making it possible to compare the same gene in the two networks to identify differentially connected genes (DCGs).

From these networks, we explored a series of subnetworks. First subnetworks based on the top trio genes and the top regulators (TFs) based on their differential connectivity between pre-and post-maturation. Pre- and post-maturation networks were constructed from the 12 control samples at T1 and 36 post-maturation (T2-T4) samples.