

A novel method to identify key factors of gene regulatory network behind salmonids reproductive behavior using directed graphical modeling

Reiichiro Nakamichi¹, Hirohisa Kishino² and Shuichi Kitada¹

¹ Graduate School of Marine Biosciences, Tokyo University of Marine Science and Technology, Japan

² Graduate School of Agricultural and Life Sciences, University of Tokyo, Japan

Abstract

Background

Salmonids play an important role in the social and economic fabric of the world. Huge amount of salmon are yielded by fisheries and captive breeding. However, decreases of genetic divergence and reduction in reproductive fitness are reported for artificially bred fish. Farmed salmon escapees also become an issue of concern for conservation of natural populations. Behind fish behaviors in reproductive process, hormones, among others may play important roles. Furthermore, reports for farmed and transgenic salmon imply that the traits related to reproductive fitness are controlled by expression levels of innumerable genes and their interactions. To investigate the genetic mechanisms underlying this kind of complicated traits, genome-wide study of gene expression and phenotype is required.

Our Approach

Microarray data enables high throughput approach to measure the correlations between transcripts and phenotypes. Biologically meaningful message may be obtained by correlations between the phenotypes and the sets of known pathways. We propose a likelihood-based graphical modeling approach to get the system-biological groundview of molecular mechanism behind the variation of phenotype. The geometry of the graph is selected based on AIC. From the local structure around the phenotype-node, it is possible to estimate the genes and their effects determining the phenotype directly. By tracing a series of the hubs from the phenotype-node towards the center of the graph, it is possible to estimate the hierarchical module structure controlling the phenotype.

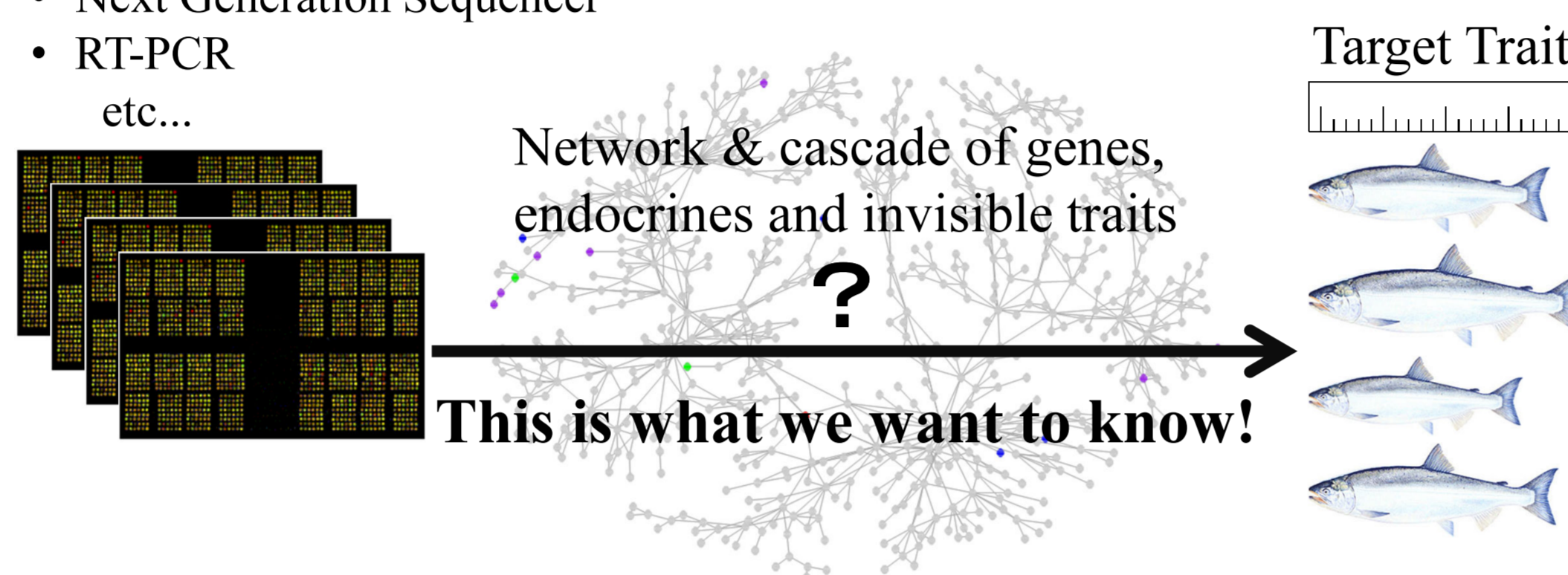
Motivation

Do artificially produced fish show worse performance of survival and reproduction? If true, why?

Target trait is not clear. Problem of survival/reproduction → hormones, maybe... But, which hormone? Another traits?

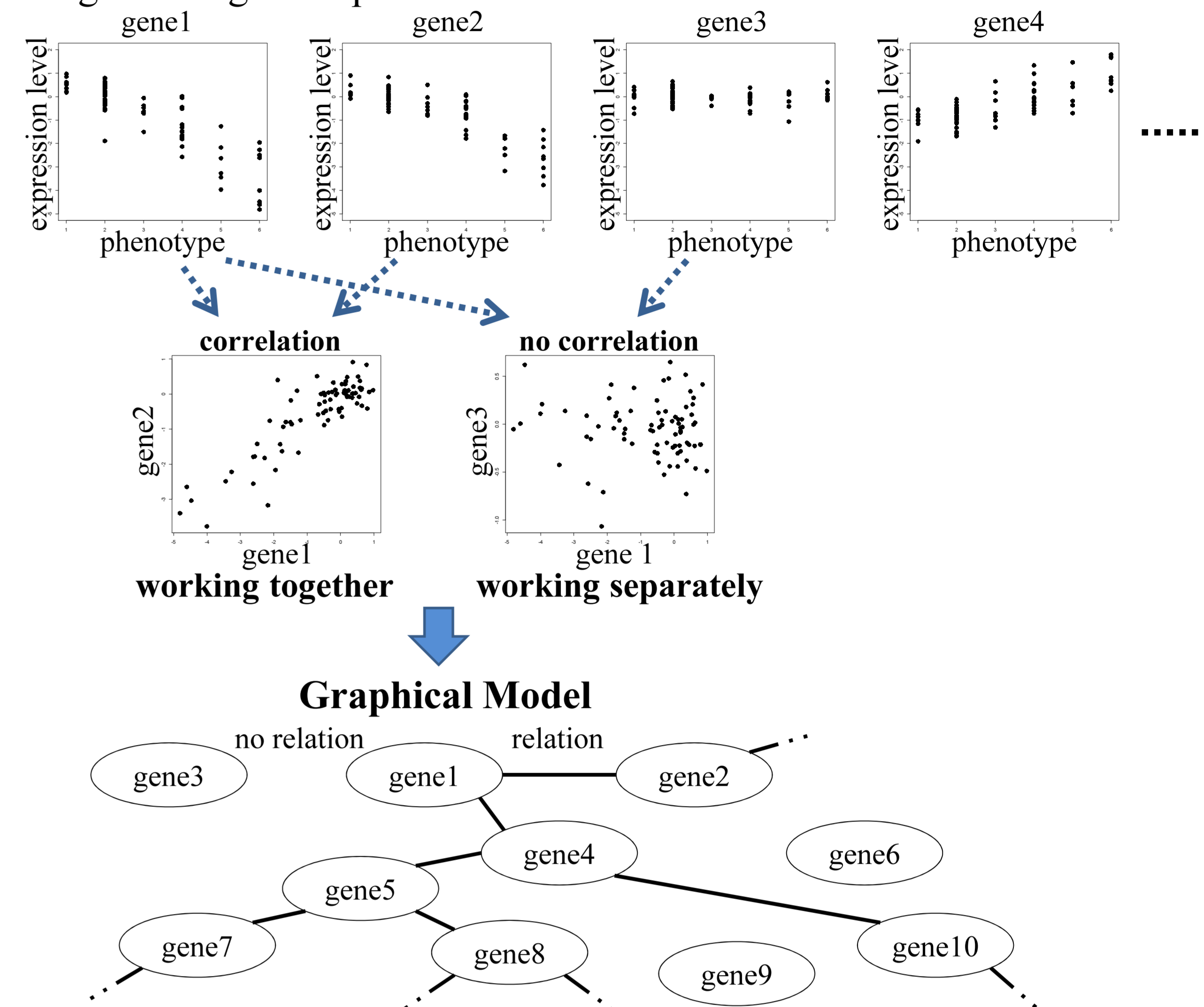
Simultaneous analysis of many traits and genes without limiting target

- Microarray
- Next Generation Sequencer
- RT-PCR
- etc...



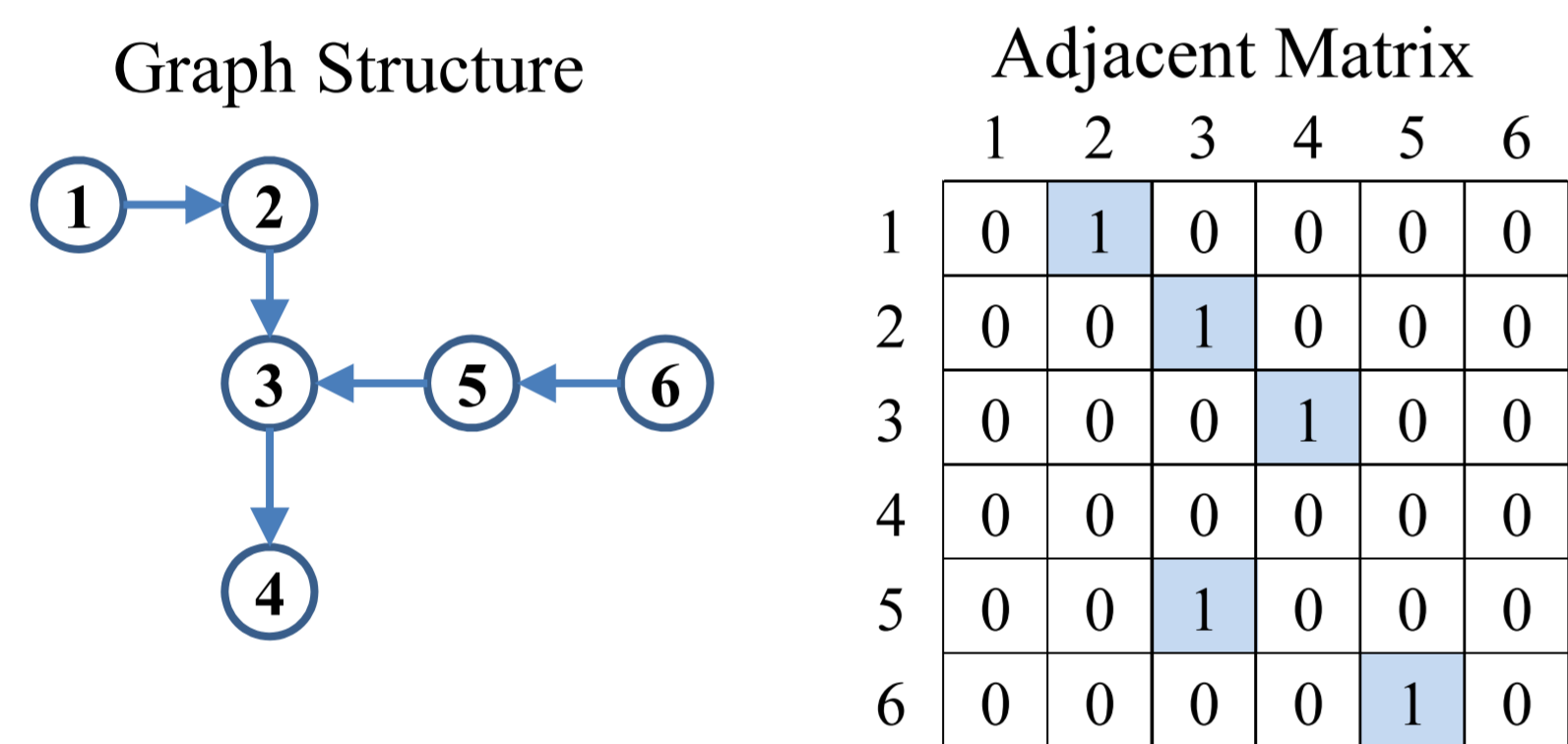
Outline: from list to network

Huge list of gene expression is available.



Statistical Model: maximal connected graph by maximum entropy method

A graph consists of the set of nodes (V) and edges (E).



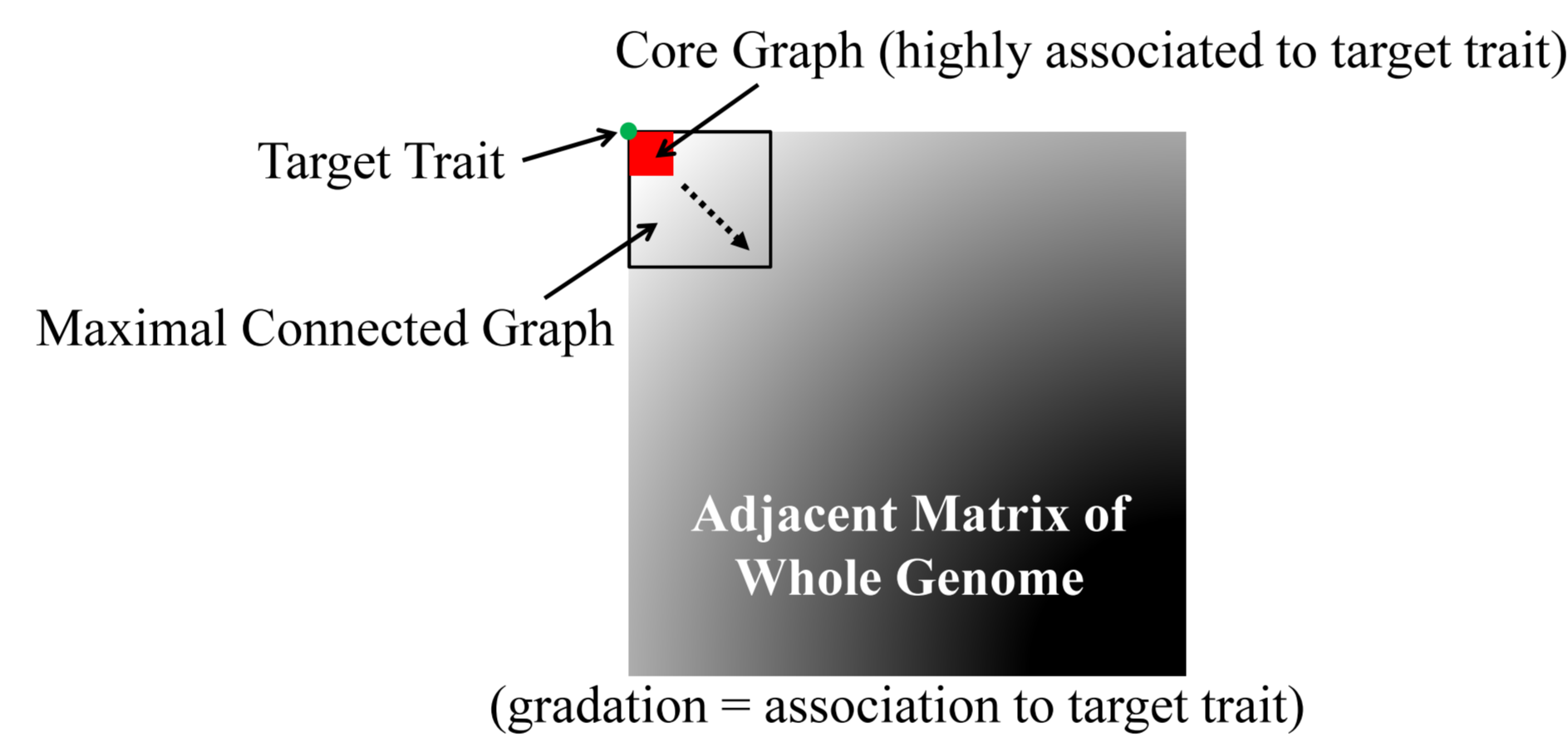
Likelihood of expression profile $\{X_v; v \in V\}$ given (V, E)

$$L = \prod_{v \in V} P(X_v | \{X_{v'}; v' \in \text{parent}(v)\})$$

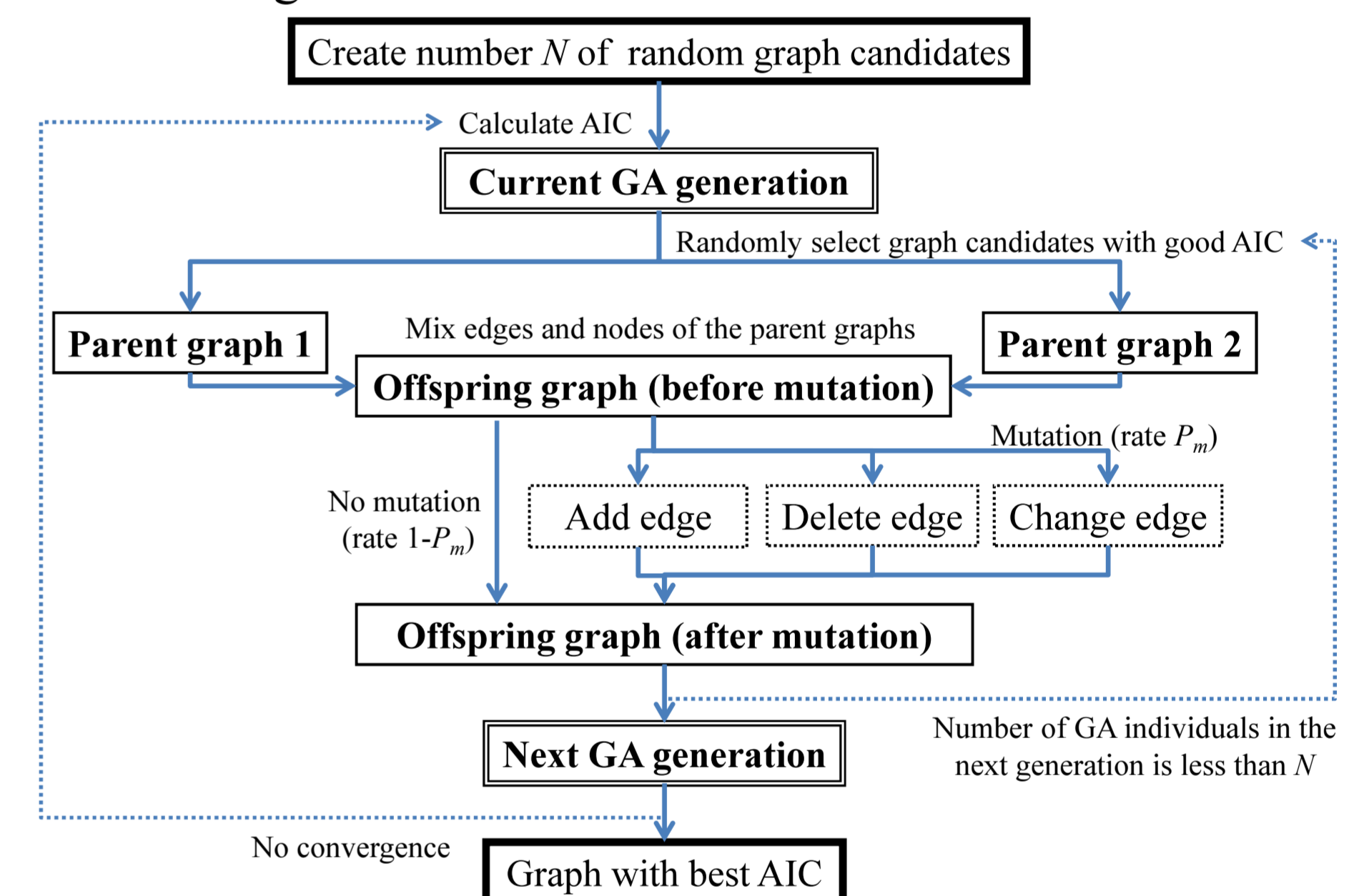
Microarray data is too huge compared with the sample size.

→ maximal connected subgraph including target trait

Starting with the core graph, we expand the graph as far as the additional edges have significant mutual information.



Genetic algorithm with the fitness measure of AIC



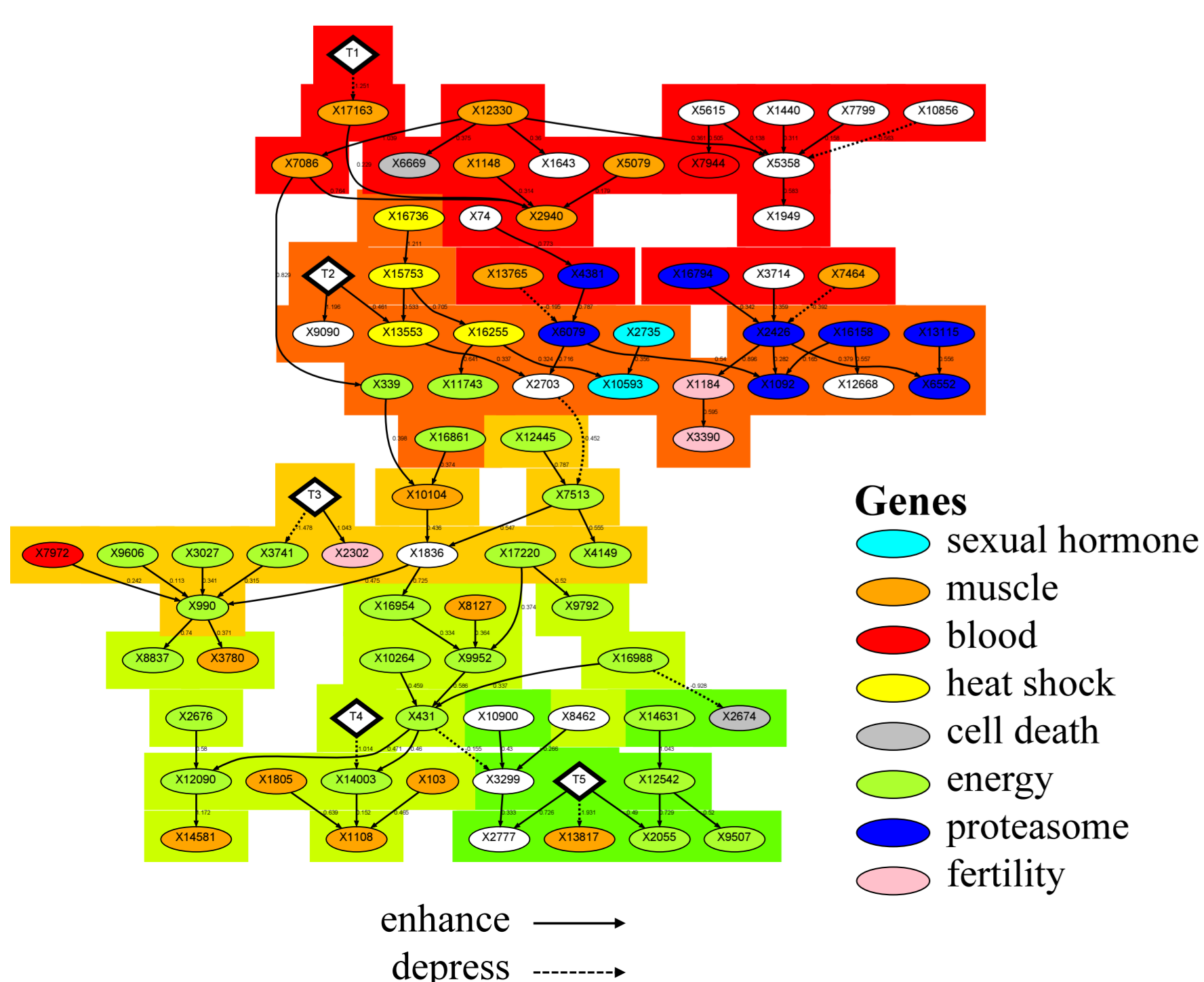
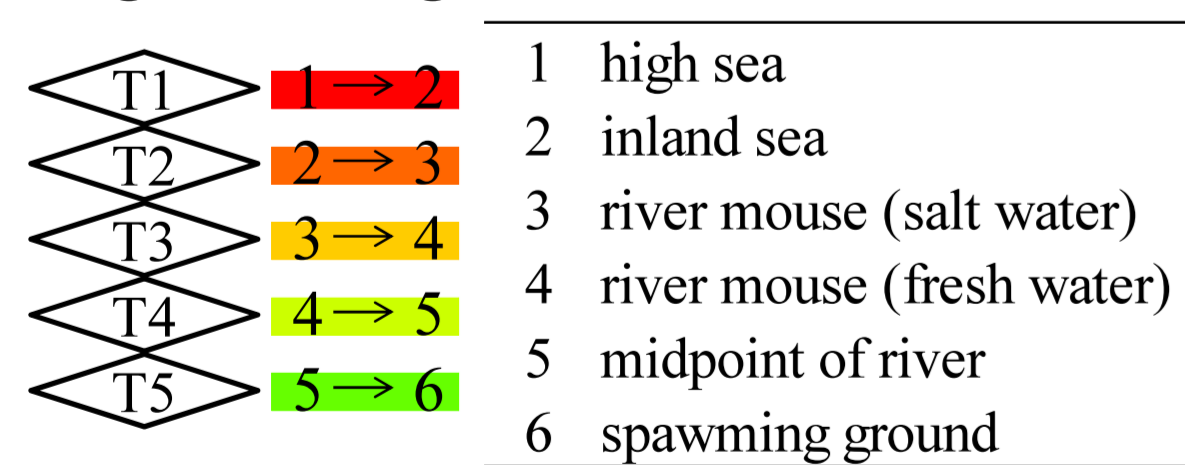
Ex1: spawning migration of sockeye salmon

Sockeye salmon (*Oncorhynchus nerka*)
(GEO public data, Miller et al. 2009)

80 samples Fraser River, Canada, August ~ October
Migrating sockeye salmon
taken at 6 locations from sea to spawning ground

16,006 genes Gene expressions measured of muscle tissue

Migration Stage and Environmental Change



Ex2: early development of Atlantic salmon

Atlantic salmon (*Salmo salar*)
(GEO public data, Jantzen et al. 2011)

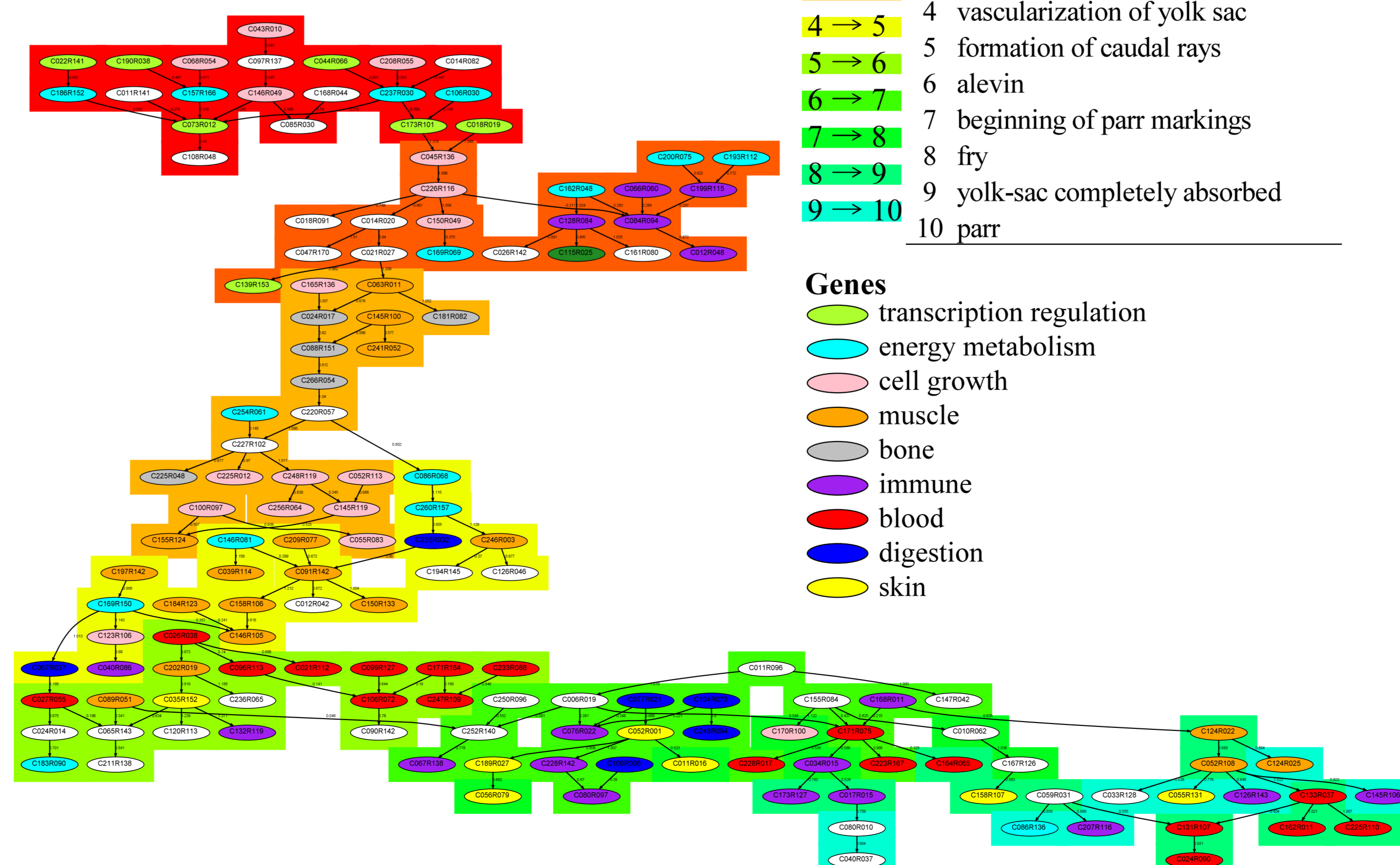
51 samples Marine Harvest United Hatchery, Canada, November ~ January
Egg/parr of Atlantic salmon
taken at 10 stages from 2 to 89 days after fertilization

43,689 genes Gene expression of whole body

Development Stage

- 1 → 2 blastulation
- 2 → 3 gastrulation
- 3 → 4 somitogenesis
- 4 → 5 vascularization of yolk sac
- 5 → 6 formation of caudal rays
- 6 alevin
- 7 beginning of parr markings
- 8 fry
- 9 yolk-sac completely absorbed
- 10 parr

- Genes**
- transcription regulation (red)
 - energy metabolism (orange)
 - cell growth (yellow)
 - muscle (green)
 - bone (blue)
 - immune (purple)
 - blood (brown)
 - digestion (pink)
 - skin (grey)



Result

Spawning migration of sockeye salmon

We estimated expression network of 79 genes related to migration behavior. At the high sea, activity of muscle synthesis and energy metabolism were high. Entering to river mouse, osmotic shock induced heat shock proteins which in turn activate sexual hormone receptors and gonads. Finally, from the river mouse to the spawning ground, activity of muscle synthesis and energy metabolism both decrease.

Early development of Atlantic salmon

We estimated expression network of 147 genes working in early development. Soon after fertilization, transcription regulation, energy metabolism and cell growth began to work. Since these genes are suggested to have relation with oncogenes, concordant activity of tumor repressor is interesting. Prior to hatching, genes were highly activated that generate bones, muscles and blood. Immune system started working after hatching in response to exposure to out-of-egg environment.

Conclusion

Our method shows high potential to help understanding the gene network mechanism behind complicated traits. It gives not only the association among genes, but directions and effects of gene expression cascade.