



HAL
open science

Using eQTL networks to functionally characterize regulatory mutations

Maud Fagny

► **To cite this version:**

Maud Fagny. Using eQTL networks to functionally characterize regulatory mutations. Doctoral. NCMM PhD Courses in Multi-Omics, NCMM, Oslo, Norway. 2022. hal-04331011

HAL Id: hal-04331011

<https://hal.inrae.fr/hal-04331011>

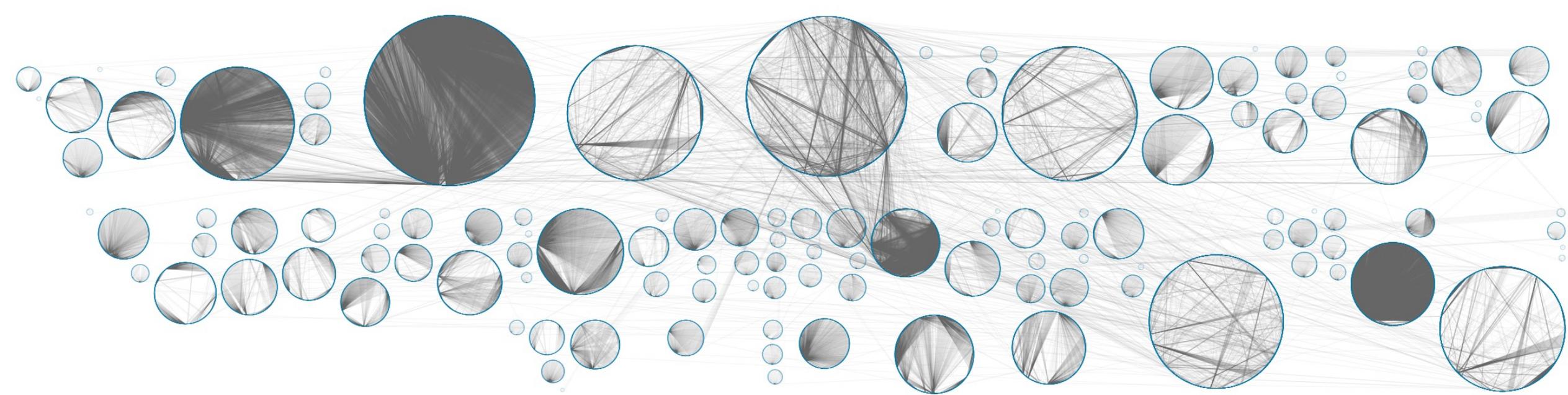
Submitted on 8 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



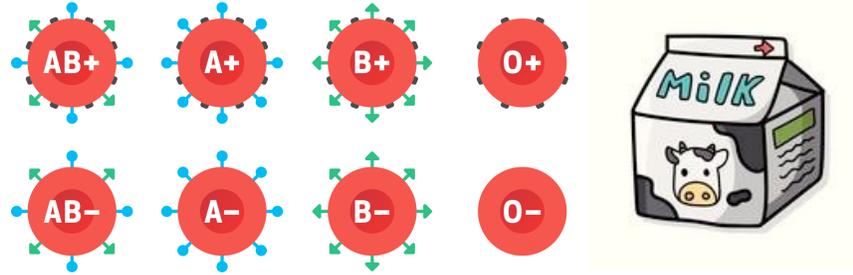
Using eQTL networks to functionally characterize regulatory mutations

Maud Fagny, PhD
INRAE, Gif-sur-Yvette

Mendelian vs. Polygenic traits

Mendelian traits:

Gene \longrightarrow Phenotype

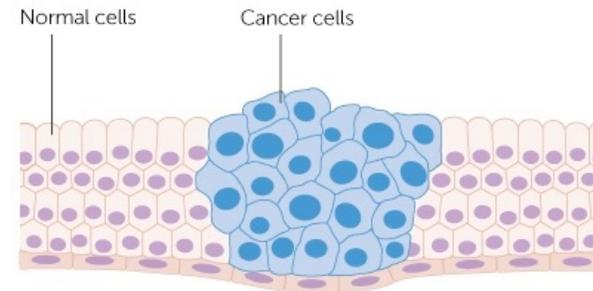


Complex traits:

Gene 1
Gene 2
Gene 2

\longrightarrow Phenotype

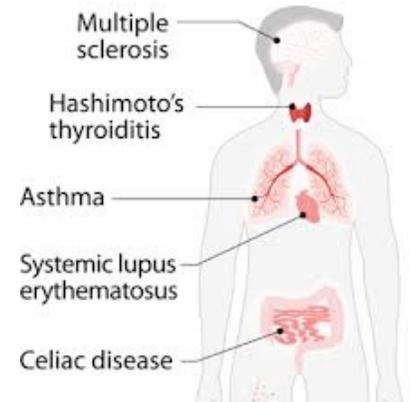
Environment \uparrow



Cancer Research UK



© Comica 2018



Genetic architecture of some often-studied traits

Monogenic

Mendelian

1 trait = 1 gene

Risk to develop a cancer

Adult size

Skin/eye/hairs color

Oligogenic

1 trait = 2-10 genes

Risk to develop
schizophrenia

Lactose tolerance

Polygenic

1 trait = many genes

Risk to develop type II
diabetes

Sickle-cell disease

Genetic architecture of some often-studied traits

Monogenic

Oligogenic

Polygenic

Risk to develop a cancer

Risk to develop
schizophrenia

Risk to develop type II
diabetes

Adult size

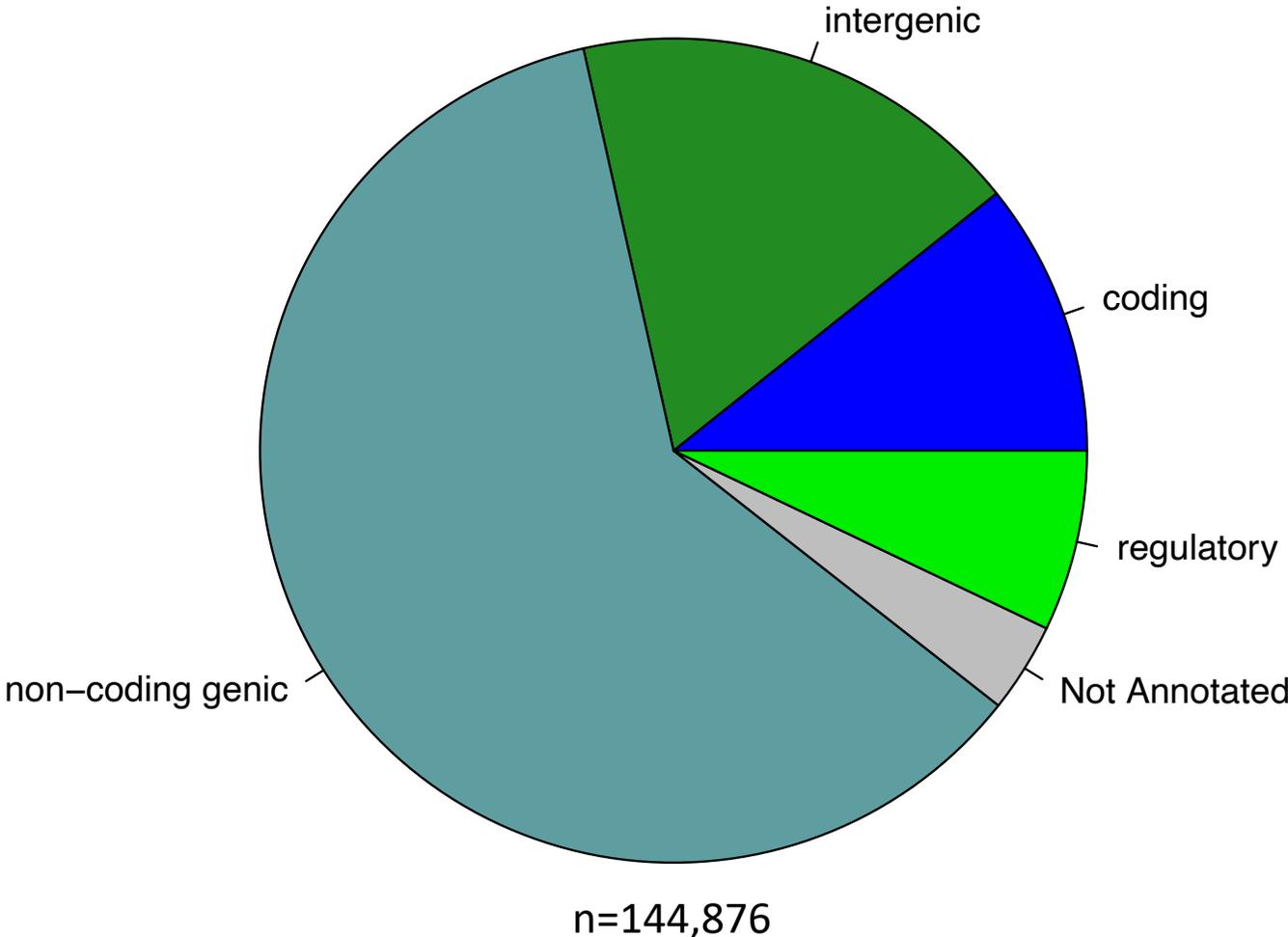
Lactose tolerance

Skin/eye/hairs color

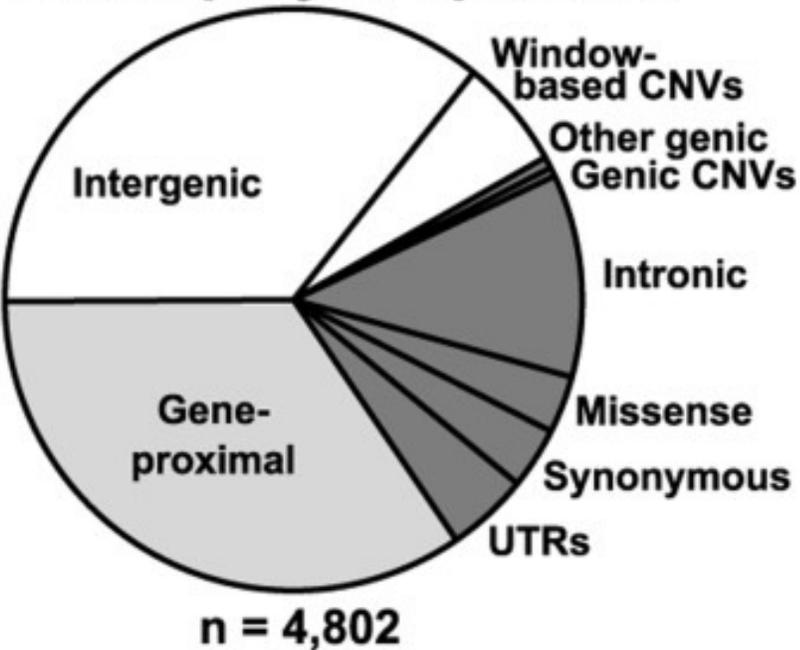
Sickle-cell disease

The importance of gene expression regulation in polygenic phenotypes

GWAS in humans

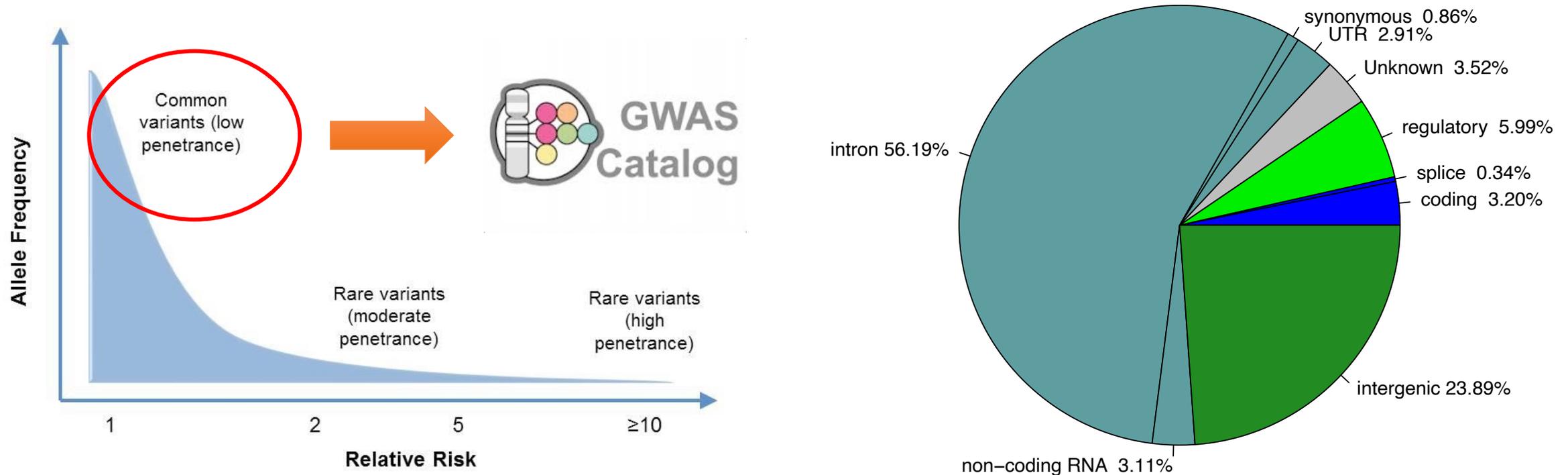


GWAS for developmental traits in maize



A particularly complex trait: the risk to develop cancer(s)

Genetic Architecture of Cancer Risk



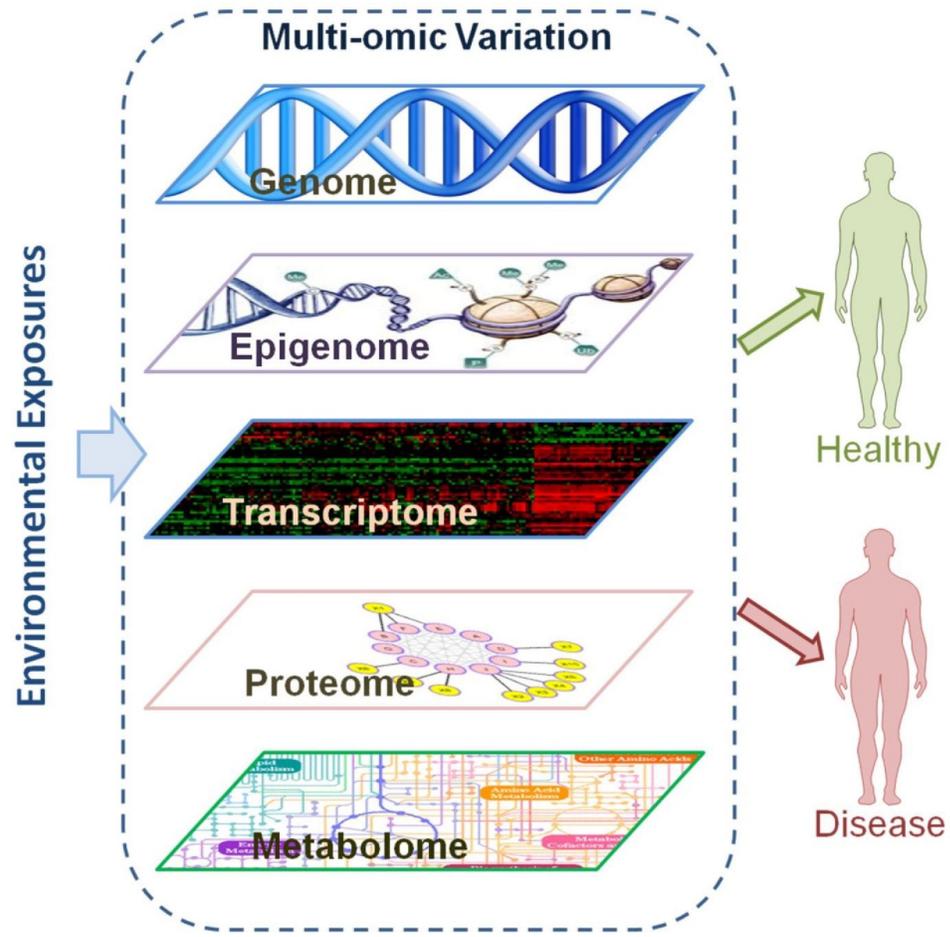
Adapted from *Cancer Genetics Overview (PQD®)*

- 4,587 SNPs associated to 265 cancer-related traits
- 87% of cancer-risk SNPs with an odds ratio under 3

How can we functionally annotate regulatory SNPs associated with traits & diseases?



Exploring the genotype-phenotype gap



Cancer-risk SNPs:

Intermediate frequencies

&

Small effect size on phenotype

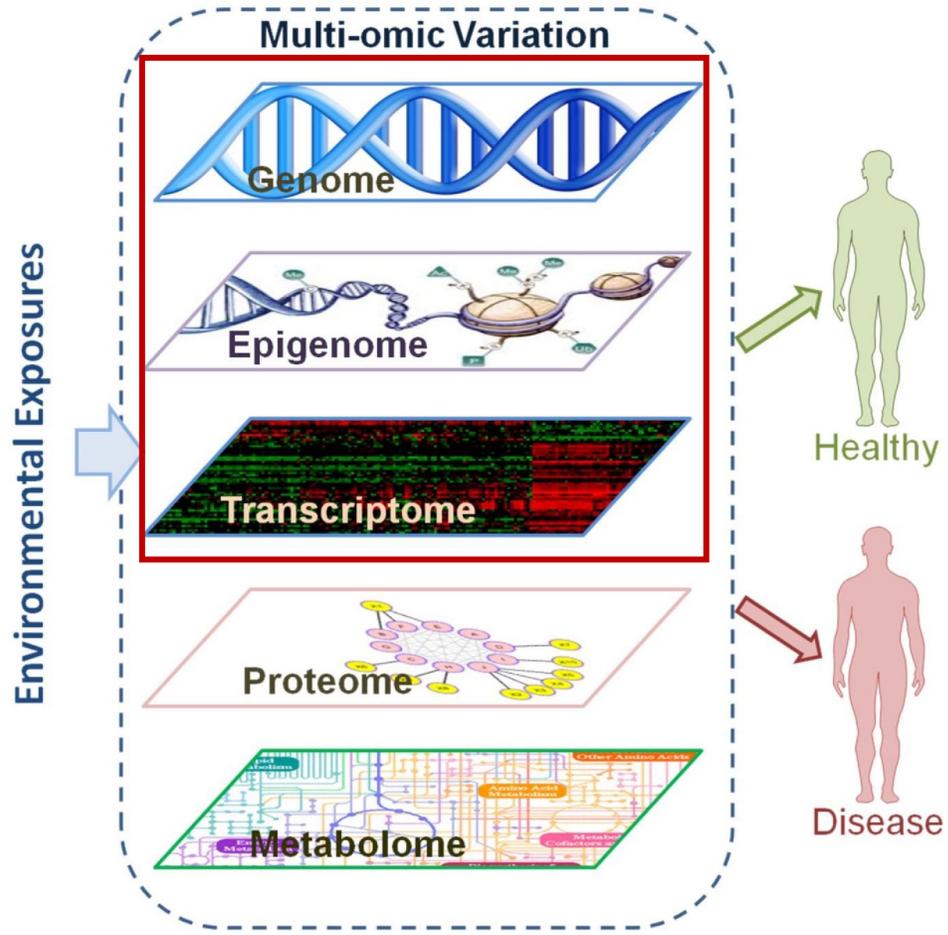
&

Tissue-specific effect

&

Located in regulatory regions

Exploring the genotype-phenotype gap



Cancer-risk SNPs:

Intermediate frequencies

&

Small effect size on phenotype

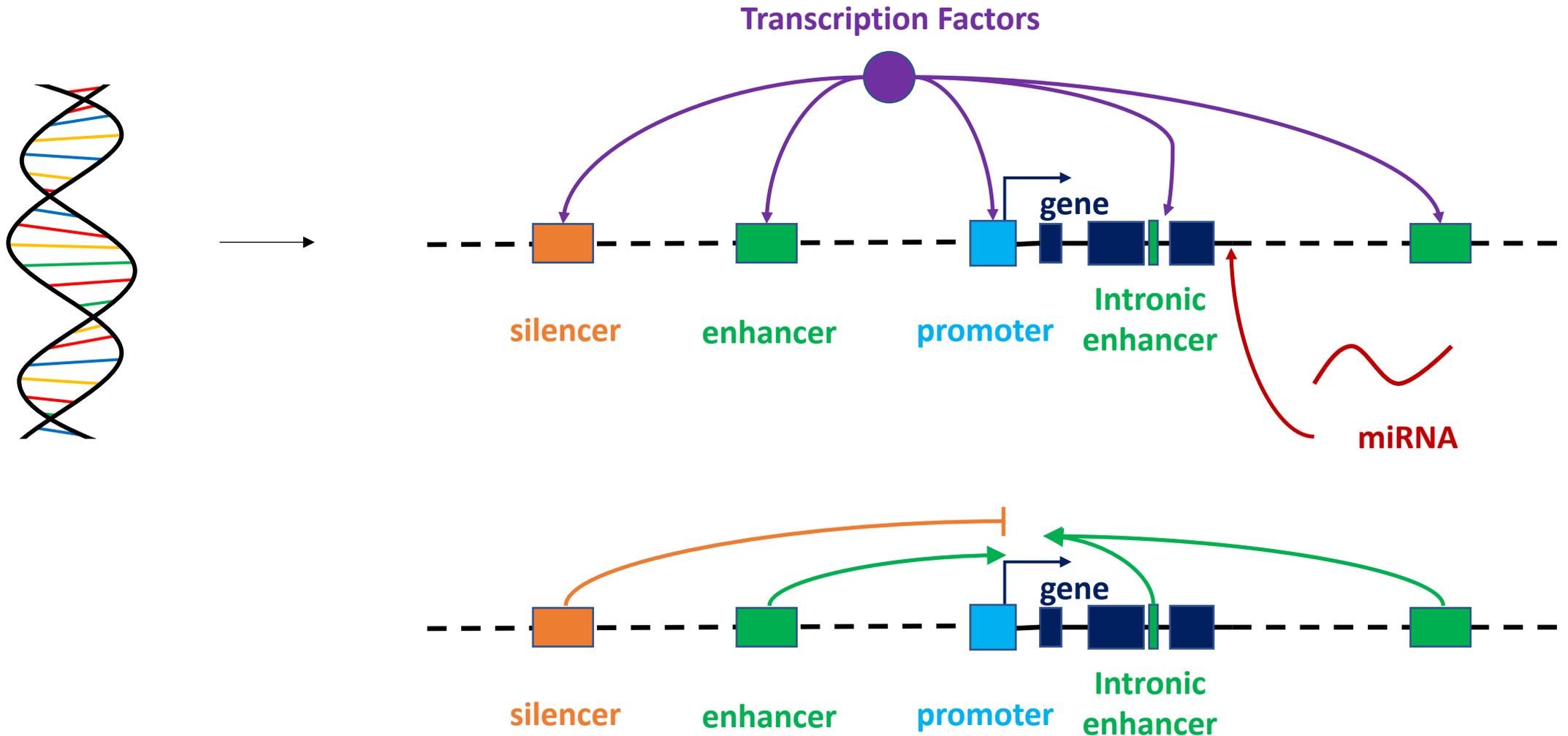
&

Tissue-specific effect

&

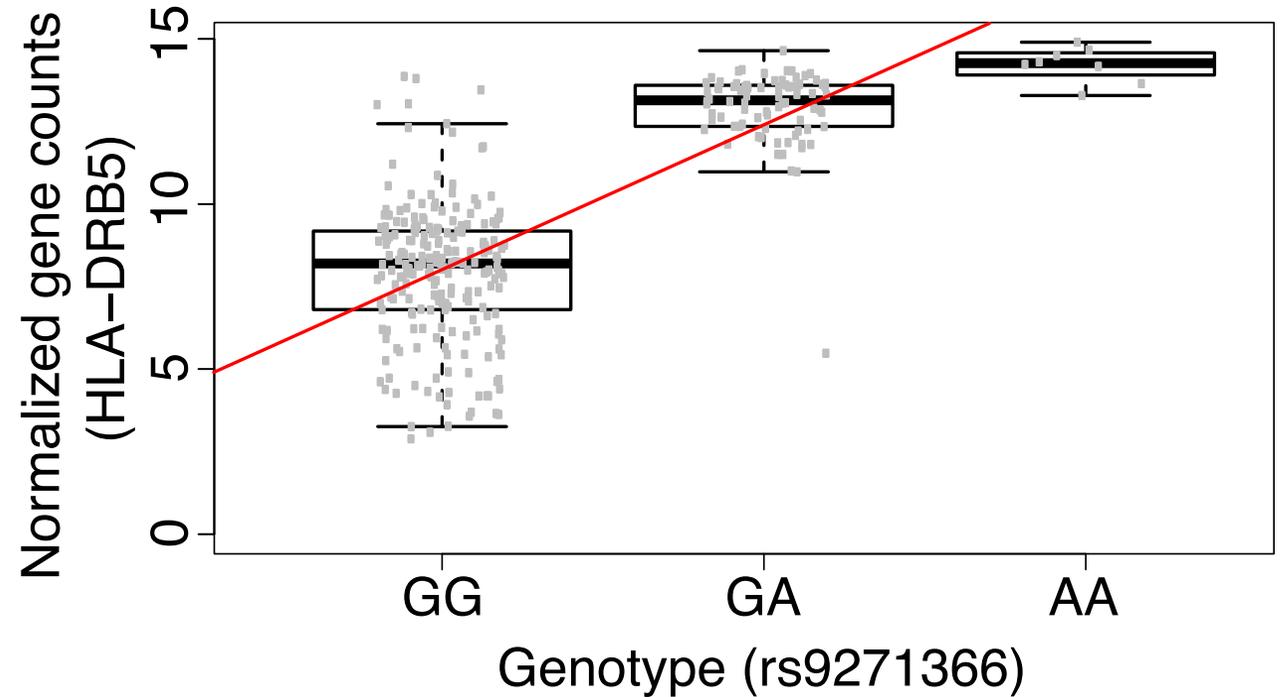
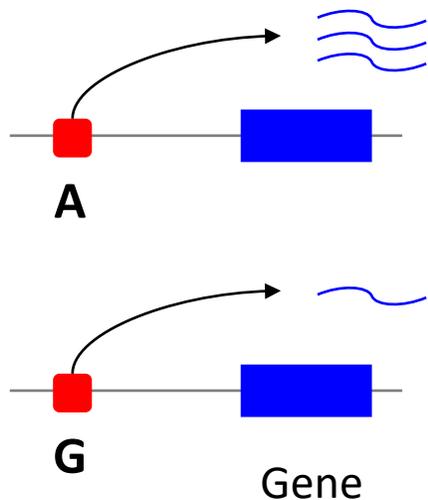
Located in regulatory regions

Intergenic and non-coding genic regions regulate gene expression



Identifying the gene targeted by regulatory SNPs: eQTLs

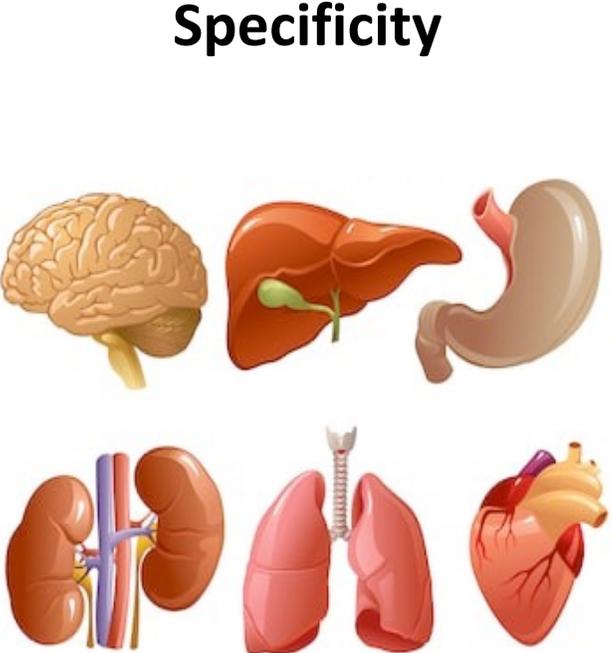
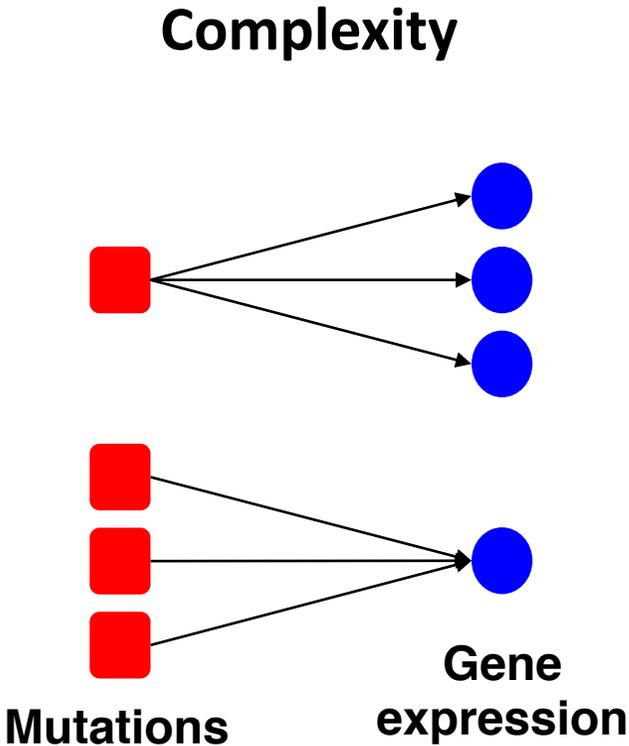
Expression quantitative trait locus (eQTL)



- Expression quantitative traits loci are potential regulatory SNPs
- SNPs associated to traits or diseases are enriched for eQTLs

A complex relationship between mutations and gene expression

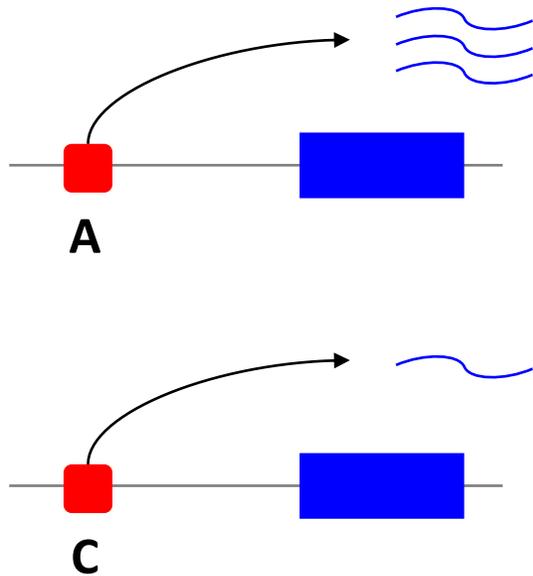
A complex relationship between mutations and gene expression



How to handle the complexity?

How to handle the complexity?

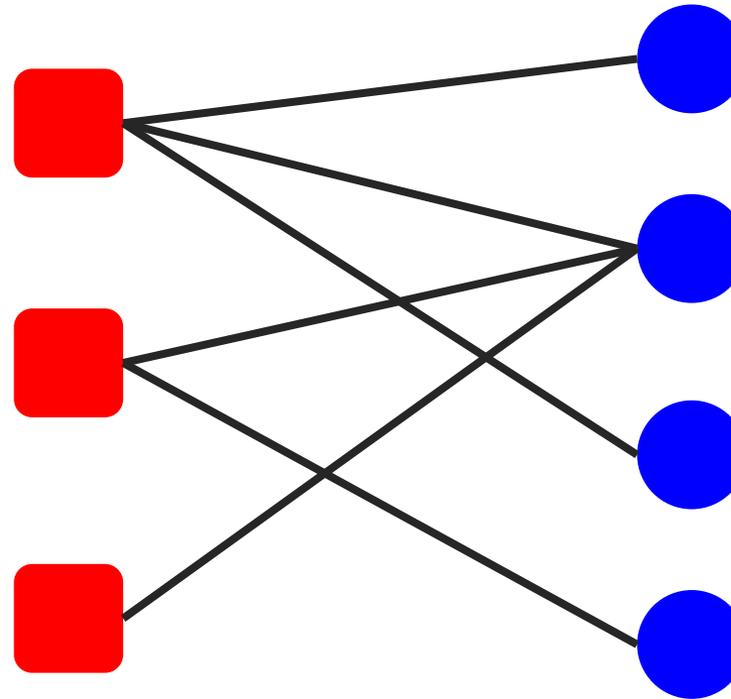
Expression quantitative trait locus



Cis- and trans-eQTL bipartite networks

SNPs

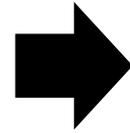
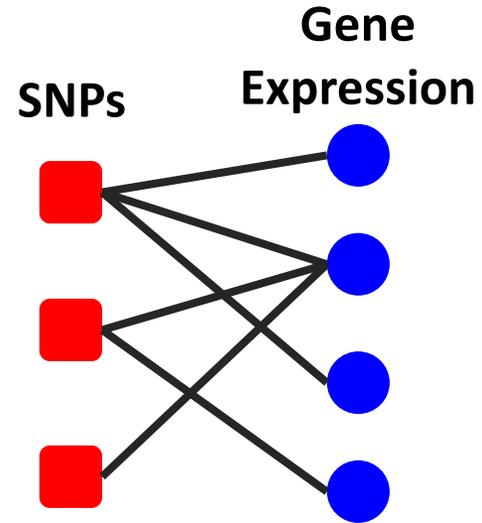
Gene Expression



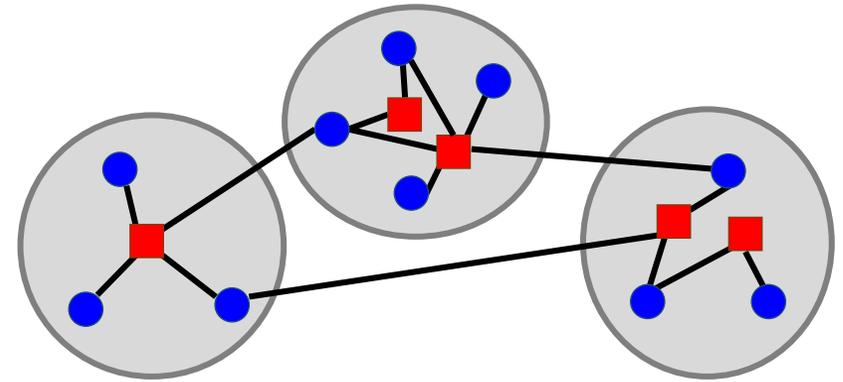
- Using a systems biology network approach to groups SNPs influencing the expression of the same genes

Grouping SNPs using a network property : modularity

Cis- and trans-eQTL bipartite network

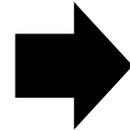
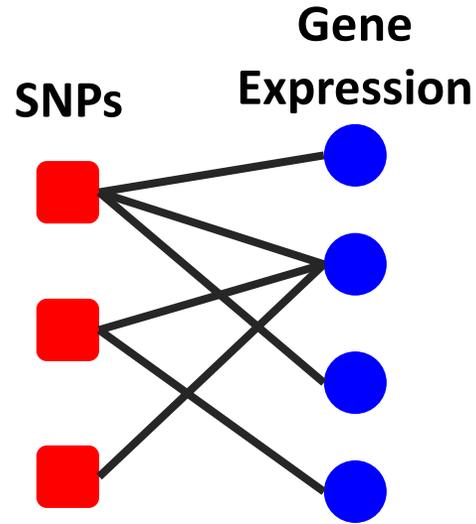


Community detection

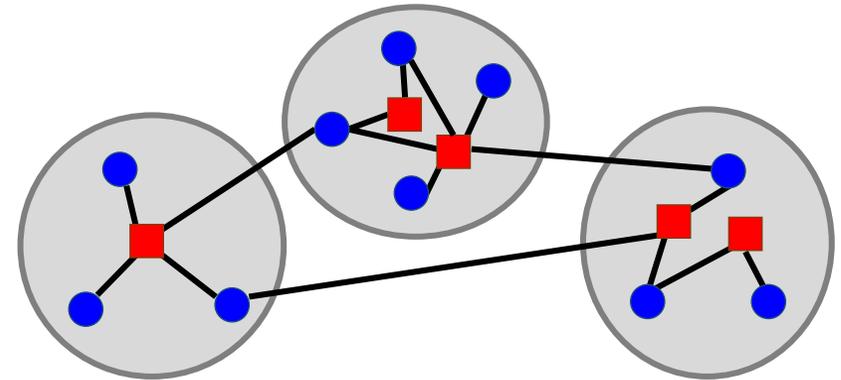


Grouping SNPs using a network property : modularity

Cis- and trans-eQTL bipartite network



Community detection



Bipartite Modularity Maximization

$$Q = \sum_i (e_{ii} - a_i^2)$$

Fraction of network links in community i

Fraction of links expected by chance



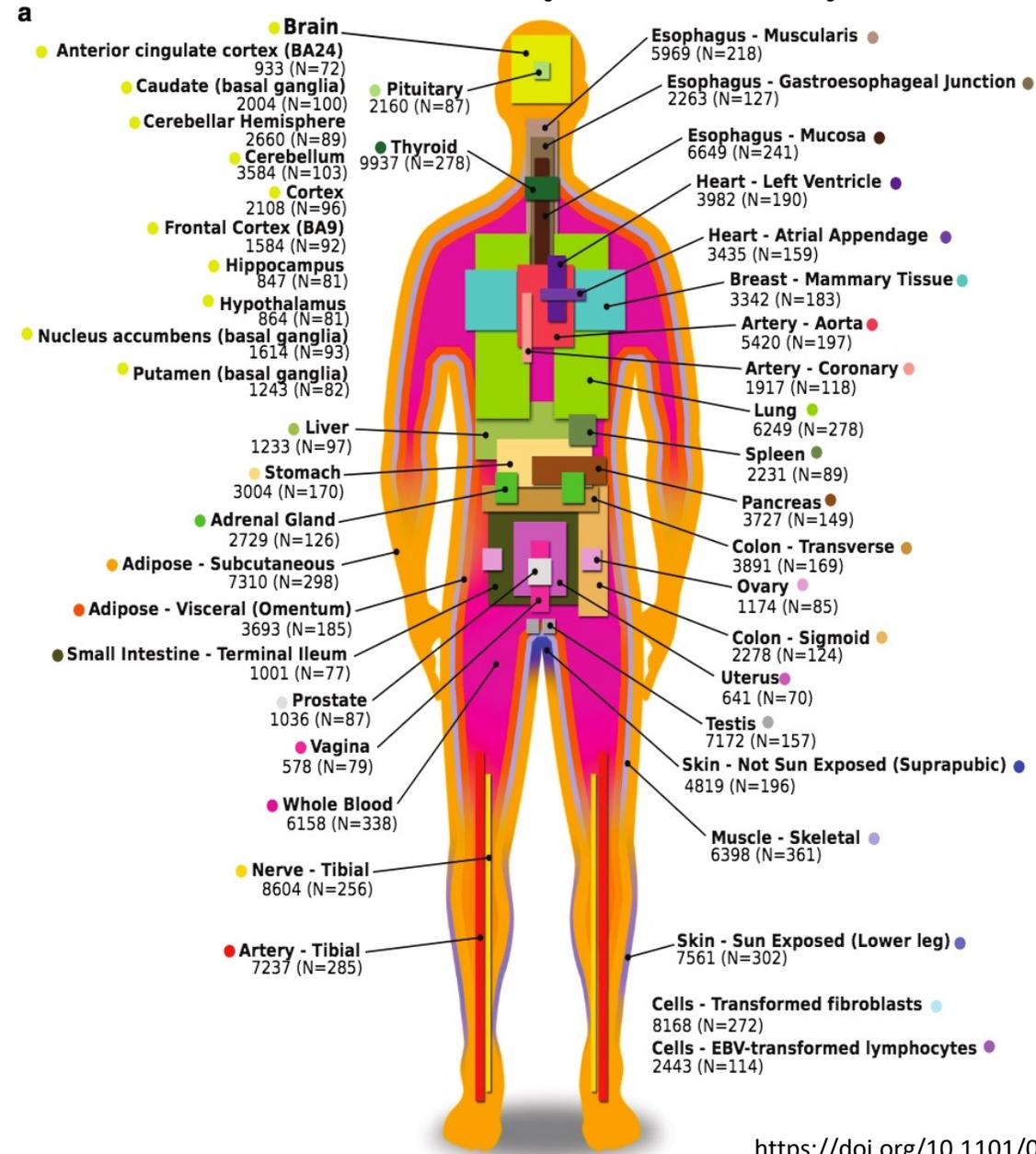
How to handle the tissue-specificity?

How to handle the tissue-specificity?

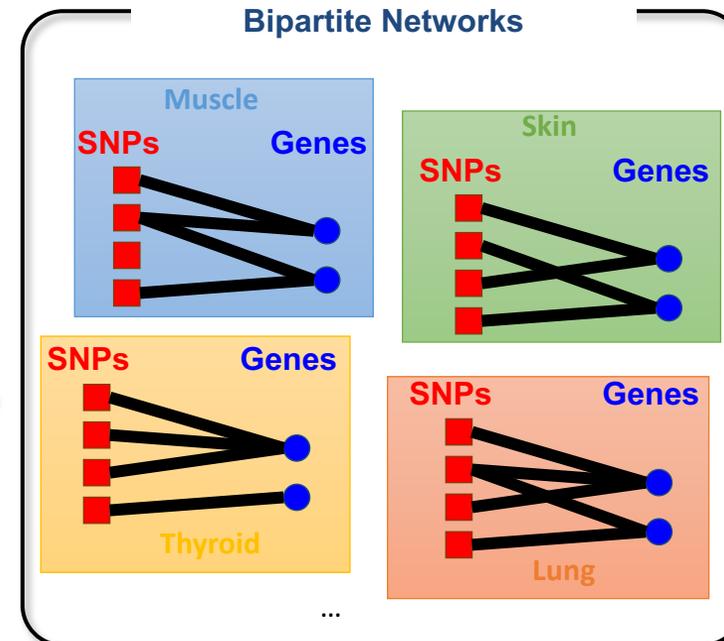
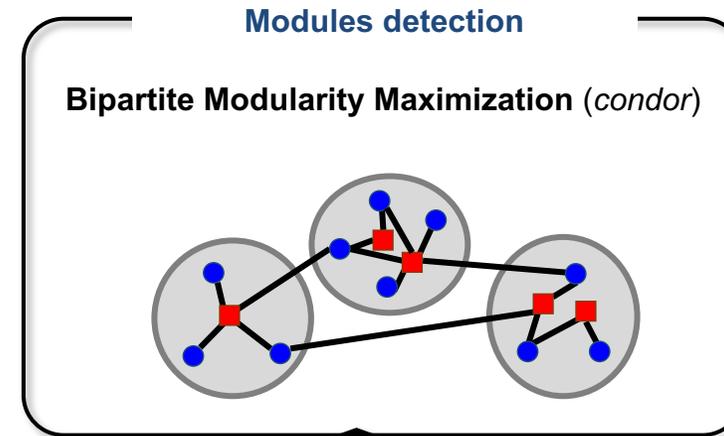
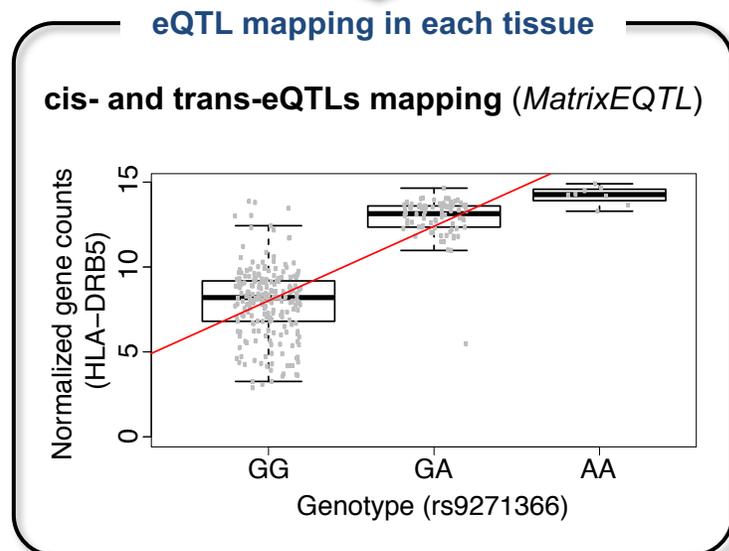
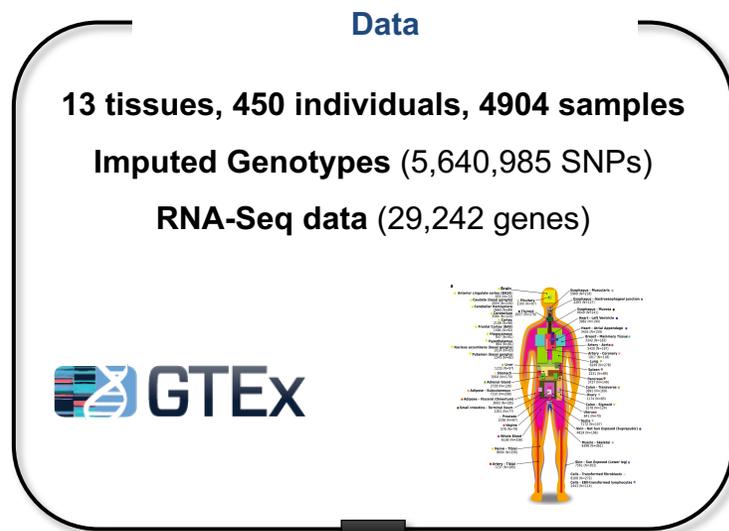
Get tissue-specific expression data !



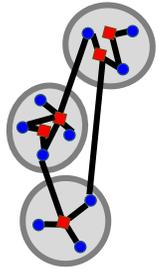
- 449 individuals.
- Genotyping data:
 - 84.3% European Am.
 - 13.7% African Am.
 - 1% Asian Am.
- RNA sequencing data:
 - 13 tissues.



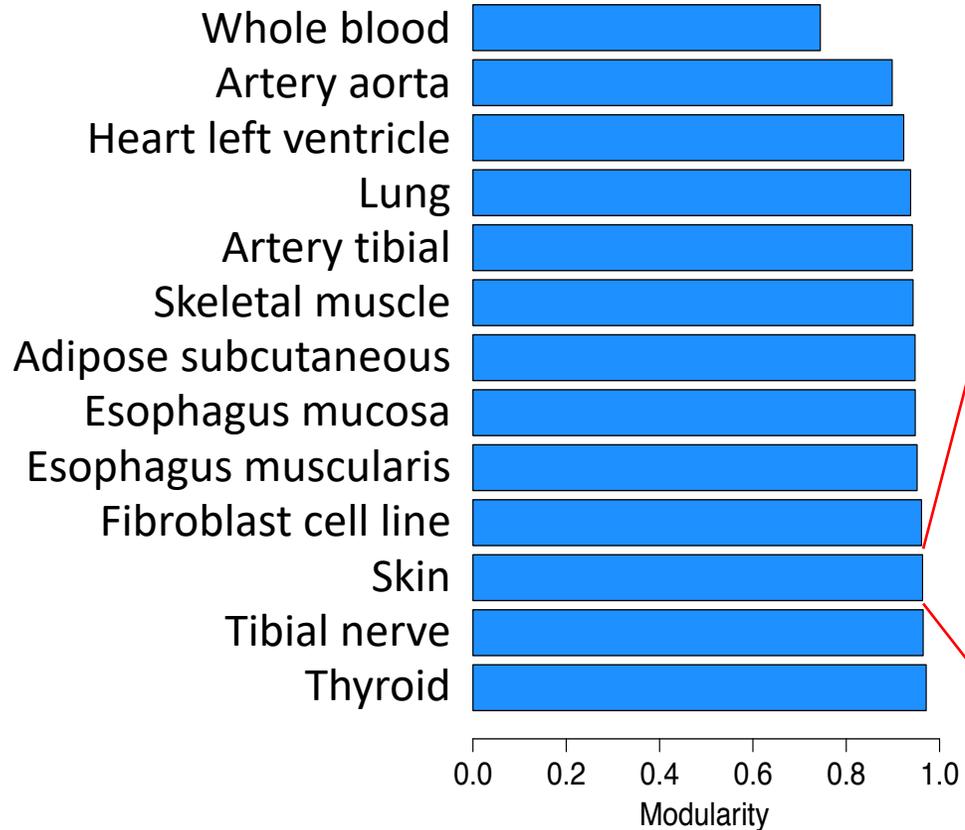
Approach summary: building tissue-specific eQTL networks



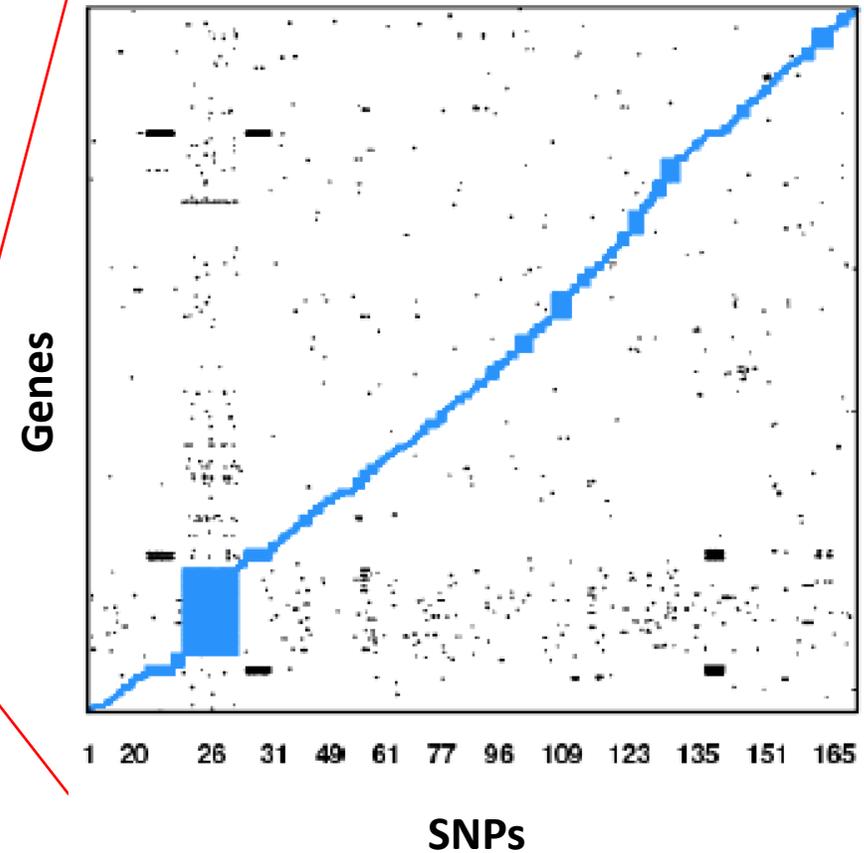
eQTL networks have high modularity



Network Modularity

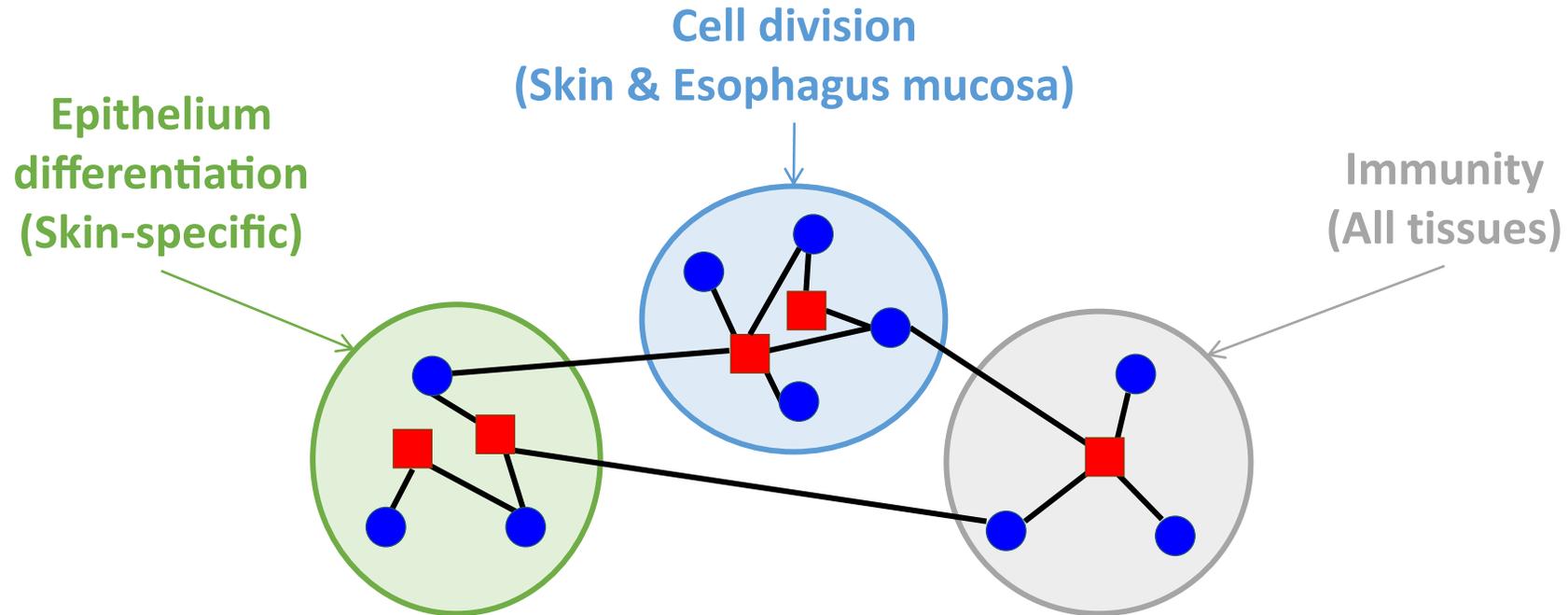


Skin network



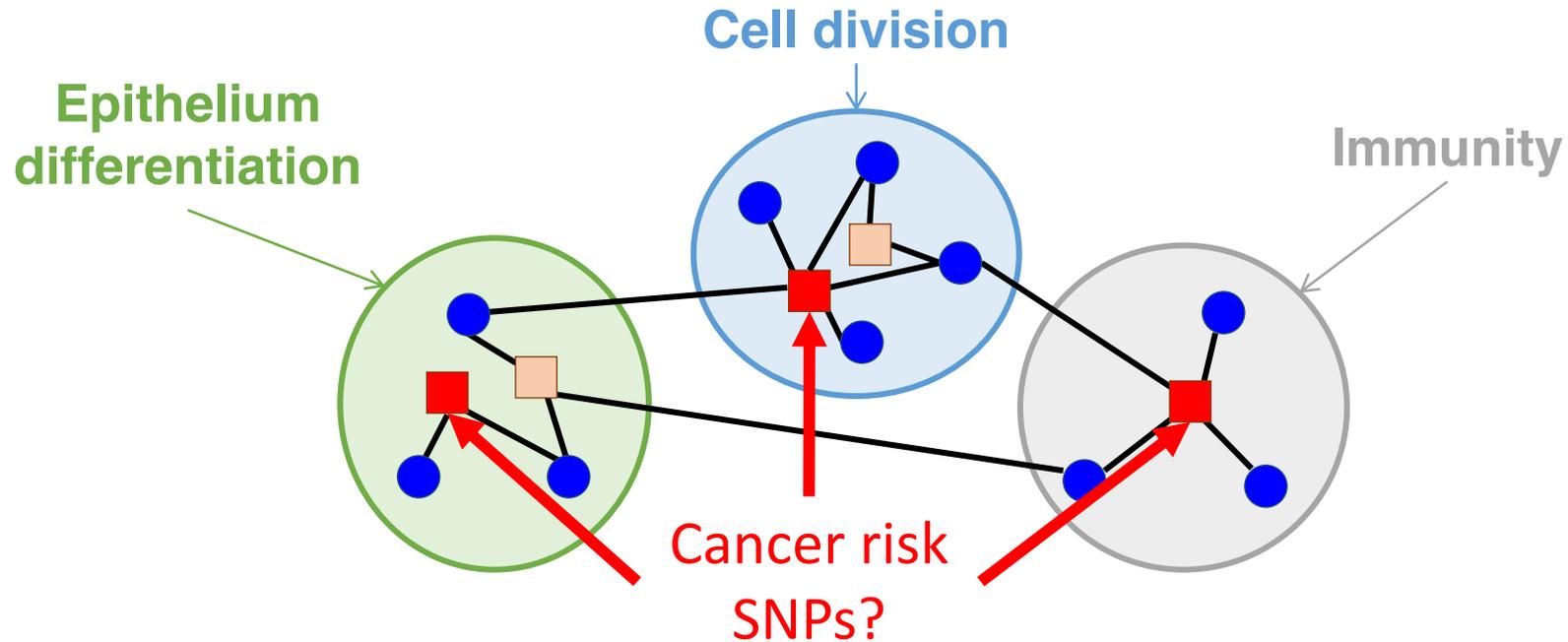
Communities correspond to biological functions of tissues

Example of communities identified in skin eQTL network



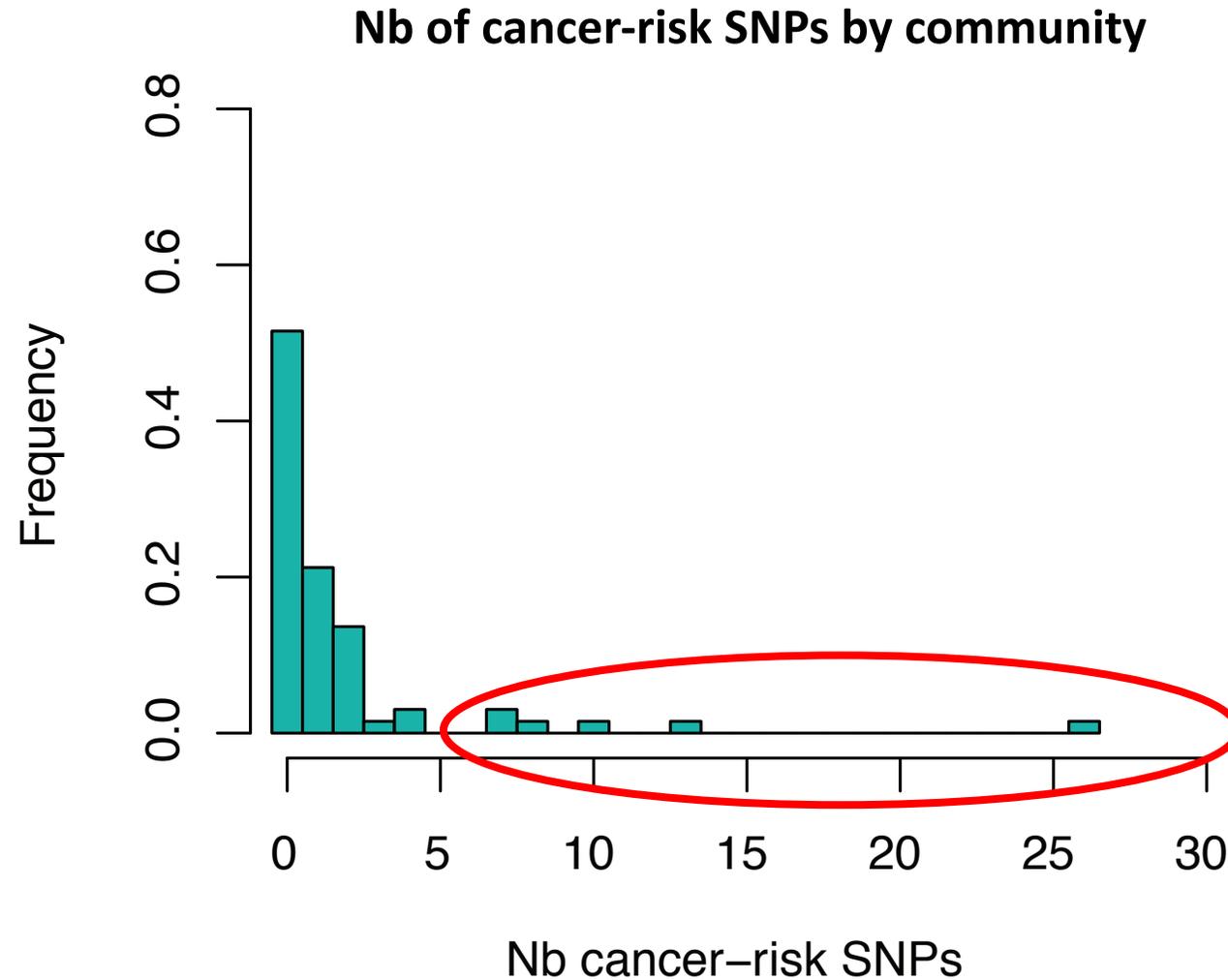
- **Groups of SNPs regulate groups of genes involved in similar functions.**
- **Communities can be tissue-specific or shared across tissues**

Mapping cancer-risk SNPs to eQTL networks



- **Where do the cancer-risk SNPs map in these networks?**
- **In which community are they located?**
- **What are their properties?**

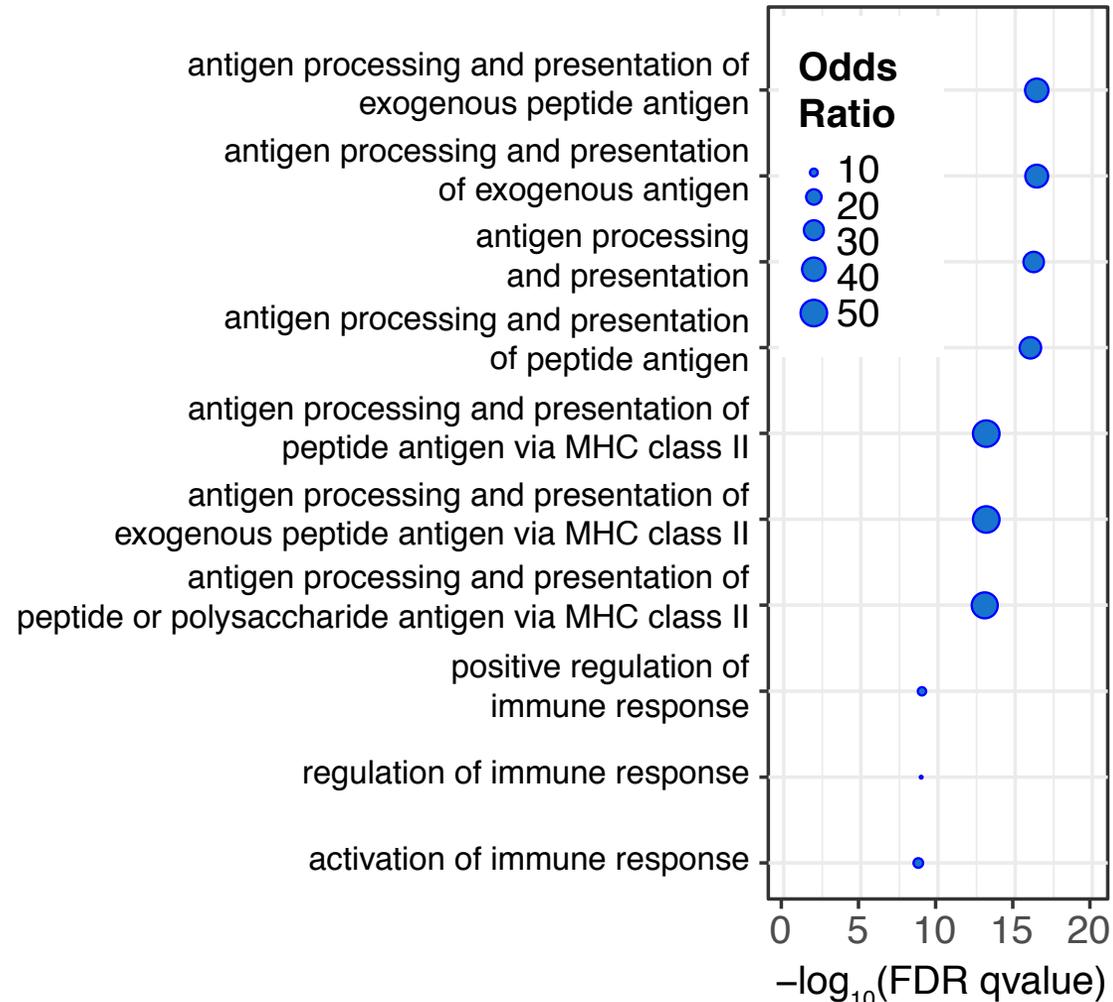
Mapping cancer-risk SNPs to eQTL networks



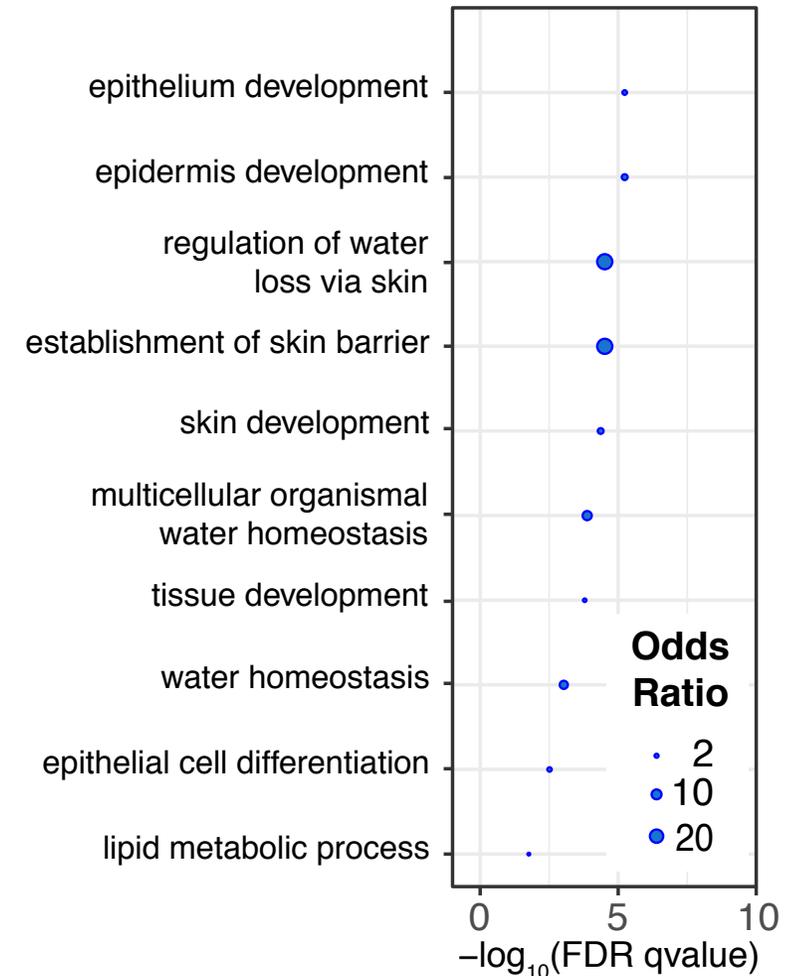
➤ **2-6 communities enriched in cancer risk SNPs by tissue**

Biological function of cancer-risk SNPs-enriched communities

Enriched in multiple cancer risk SNPs (all tissues)

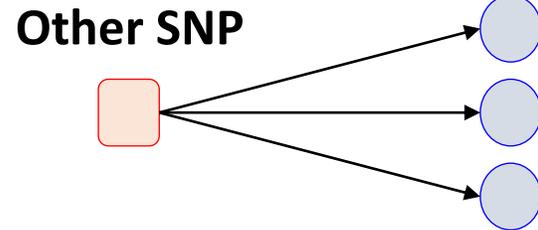
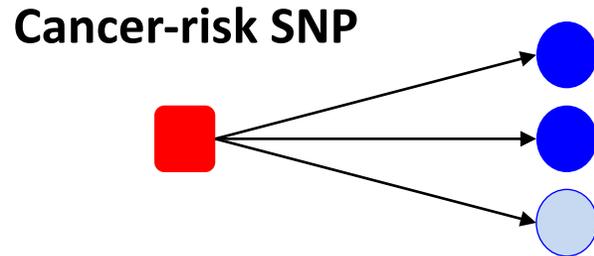


Enriched in breast cancer risk SNPs (skin)



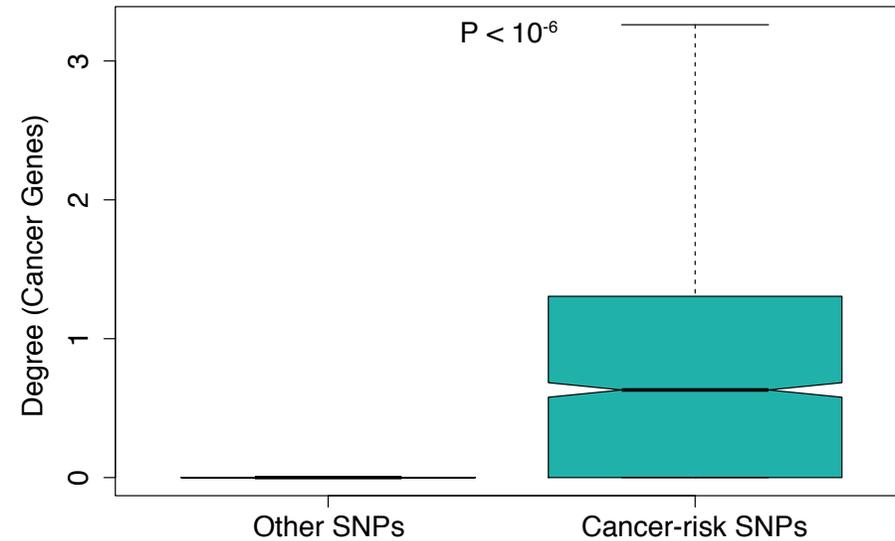
Cancer-risk SNPs preferentially target oncogenes

Genes targeted by cancer risk SNPs?

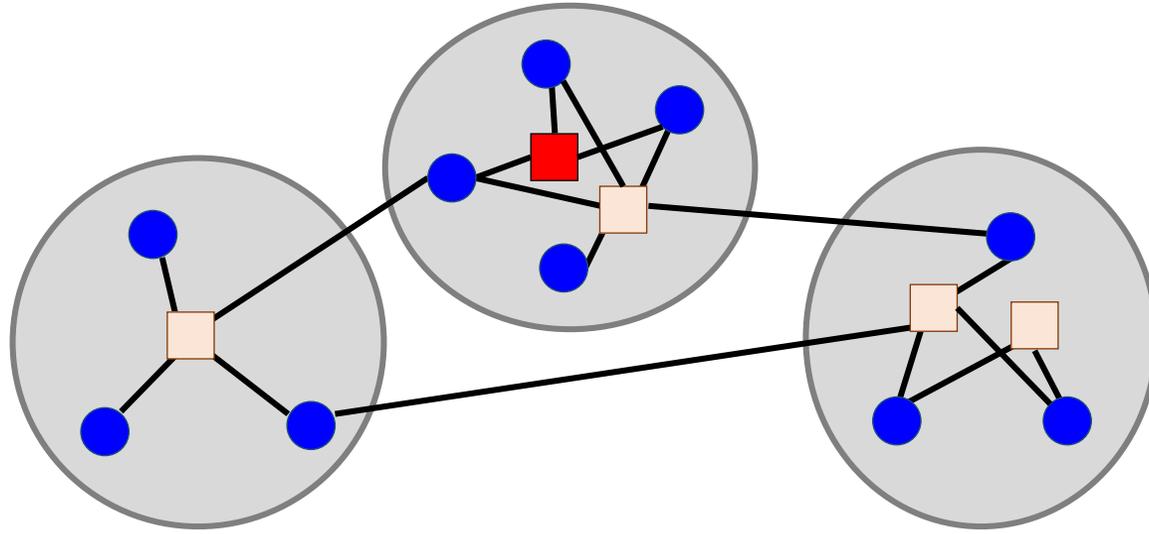


○ Others

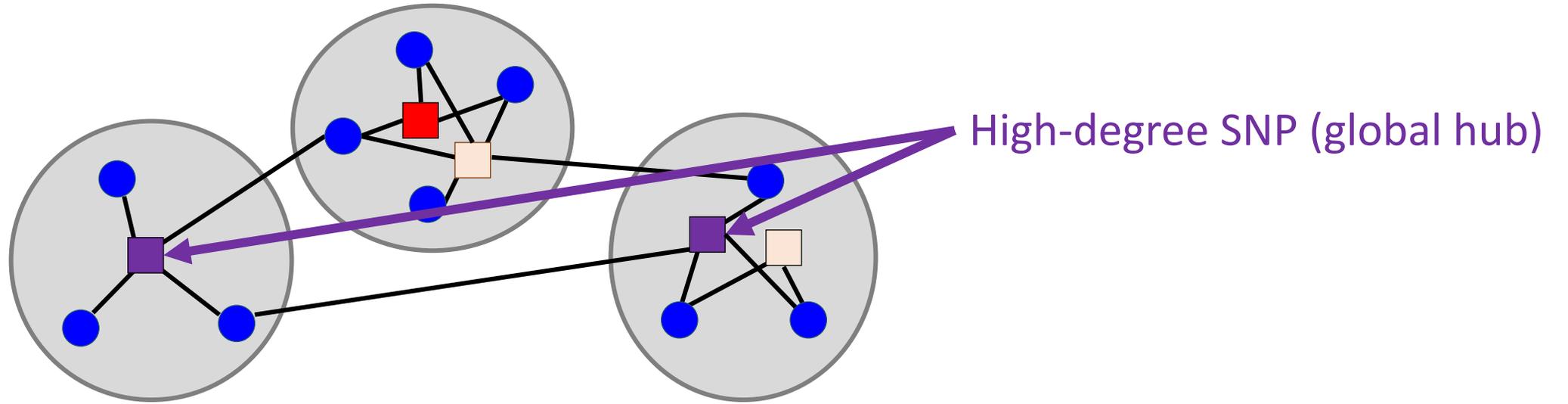
● Oncogenes
Tumor suppressors



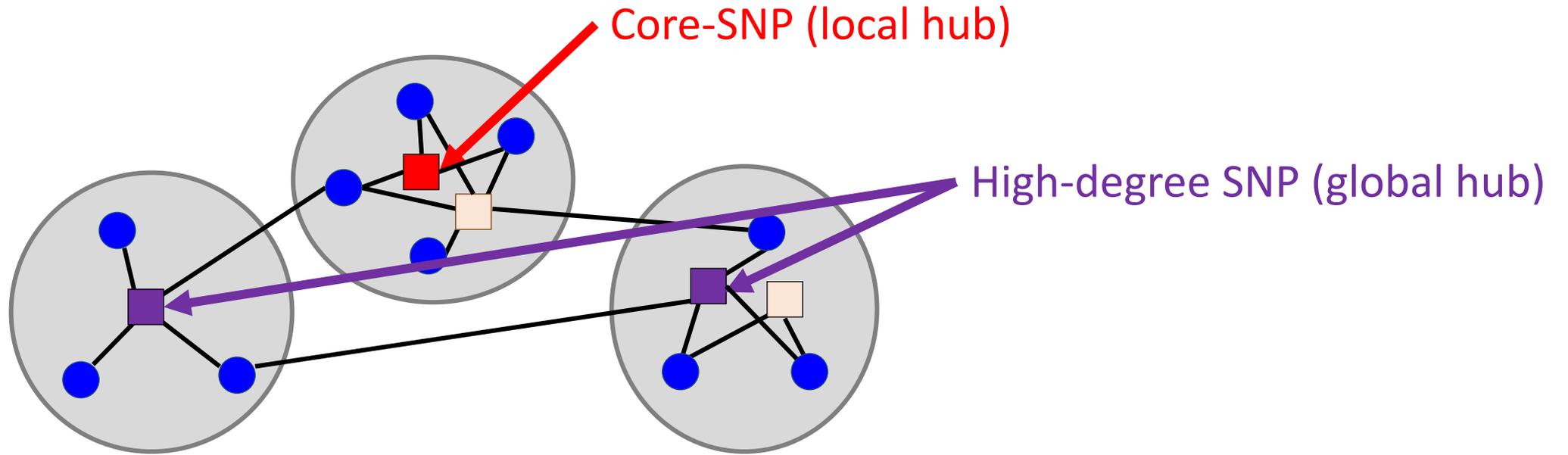
What are the properties of cancer-risk SNPs in the eQTL networks?



2 hypotheses: high-degree?



Or high core-score?

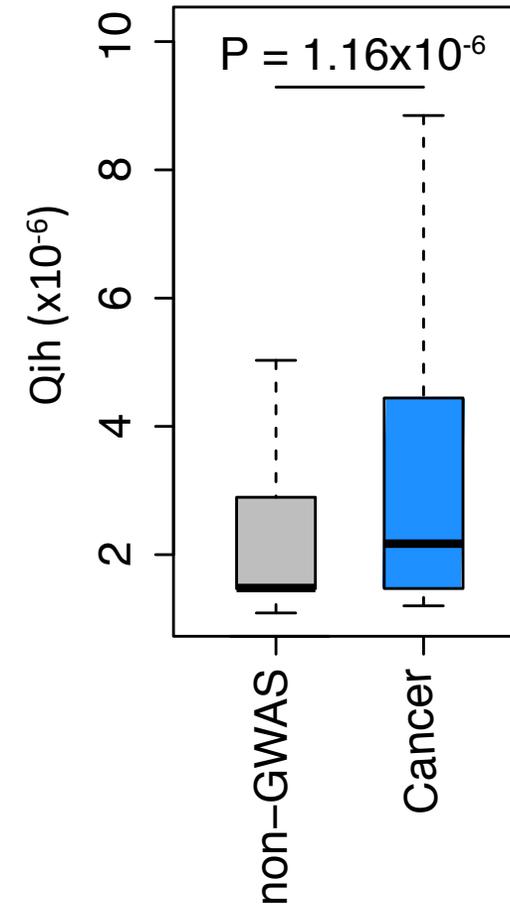
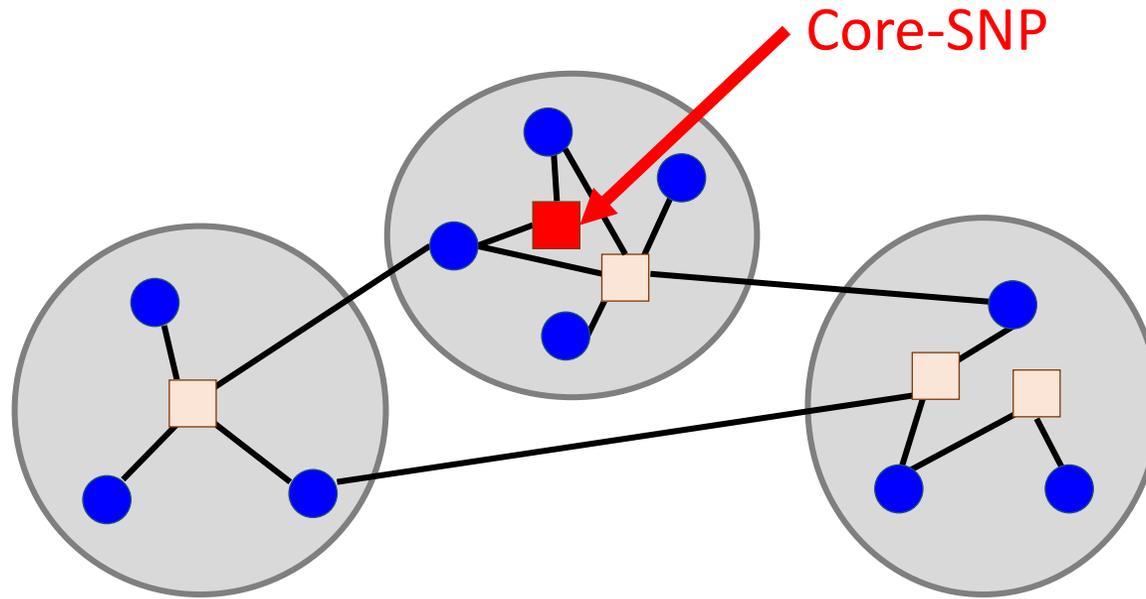


Core-score for SNP i in community h

$$Q_{ih} = \frac{1}{m} \sum_j \left(\tilde{A}_{ij} - \frac{k_i d_j}{m} \right) \delta(C_i, h) \underbrace{\delta(C_j, h)}_{\text{Is gene } j \text{ in community } h?}$$

Observed – expected edge
between SNP i and gene j

Cancer-risk SNPs are local hubs

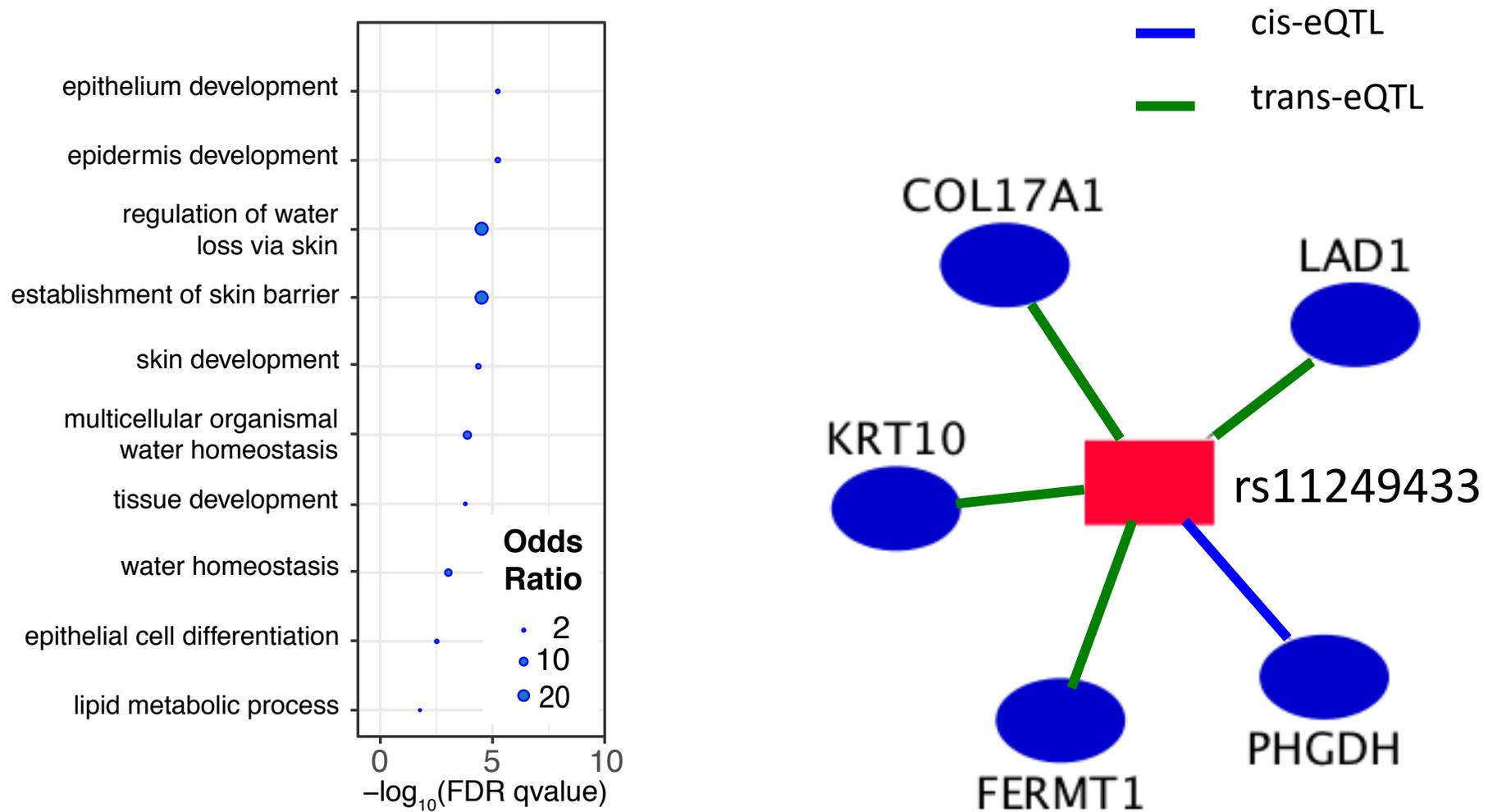


Core-score for SNP i in community h

$$Q_{ih} = \frac{1}{m} \sum_j \left(\underbrace{\tilde{A}_{ij}}_{\text{Observed - expected edge between SNP } i \text{ and gene } j} - \frac{k_i d_j}{m} \right) \underbrace{\delta(C_i, h) \delta(C_j, h)}_{\text{Is gene } j \text{ in community } h?}$$

Observed – expected edge
between SNP i and gene j

Example of a core, cancer-risk SNP



➤ **rs11249433 is associated to breast cancer**

➤ **Target genes deregulated in epithelium cancers (extra-cellular matrix secretion).**

Take-home message

- Most disease-associated SNPs are located outside of coding regions, and likely impact regulatory functions.
- Expression quantitative trait loci (eQTL) bipartite networks can help functionally annotating them
- Cancer risk SNPs:
 - Impact the expression of oncogenes and tumor suppressor genes
 - Collectively deregulates groups of genes involved in recognition of damaged cells and immune response.
 - Are more likely to be “core-SNPs” impacting the expression of many genes involved in the same biological process.