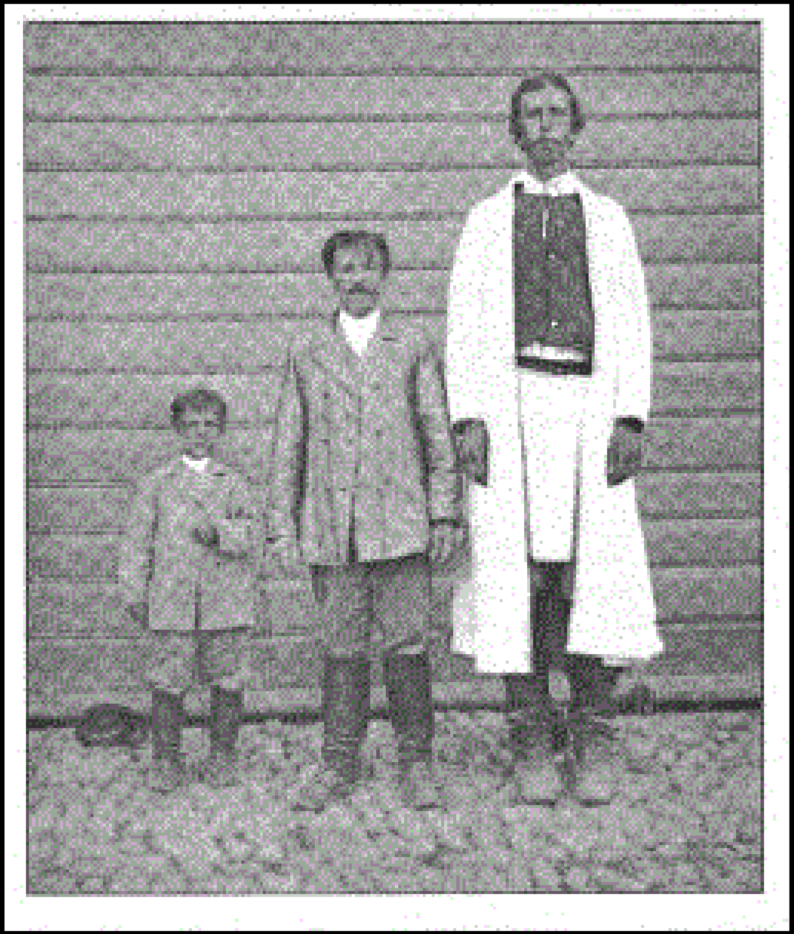
Using Network Flow to Bridge the Gap between Genotype and Phenotype

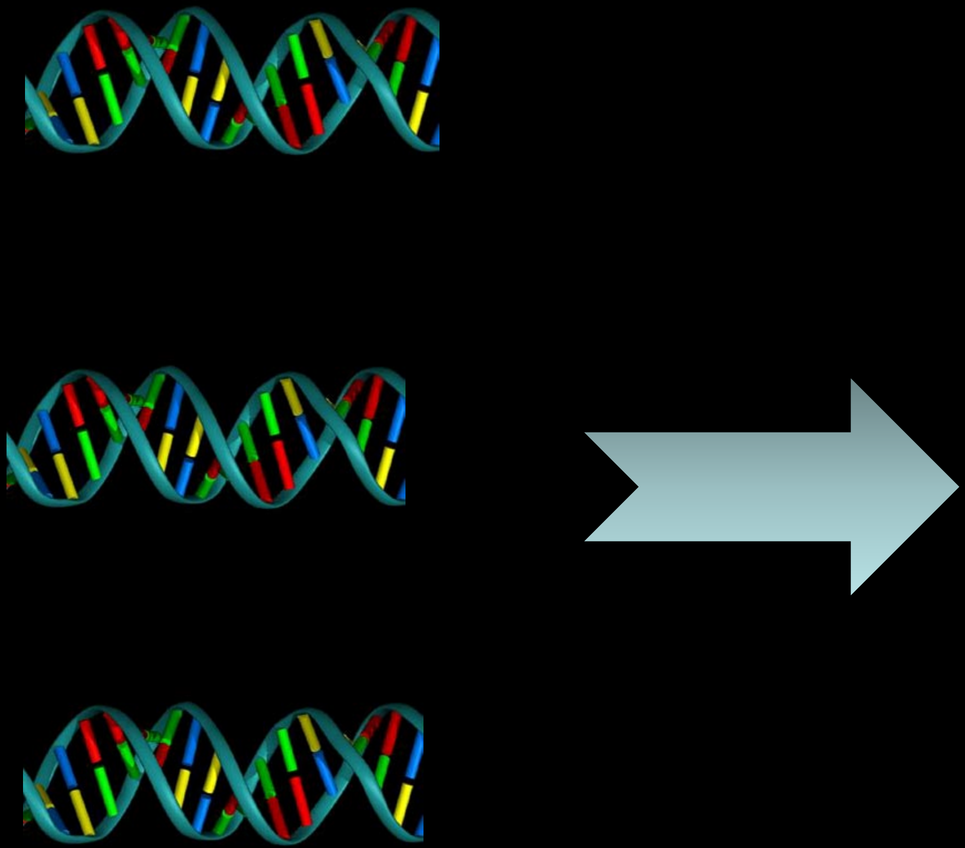
Teresa Przytycka
NIH / NLM / NCBI

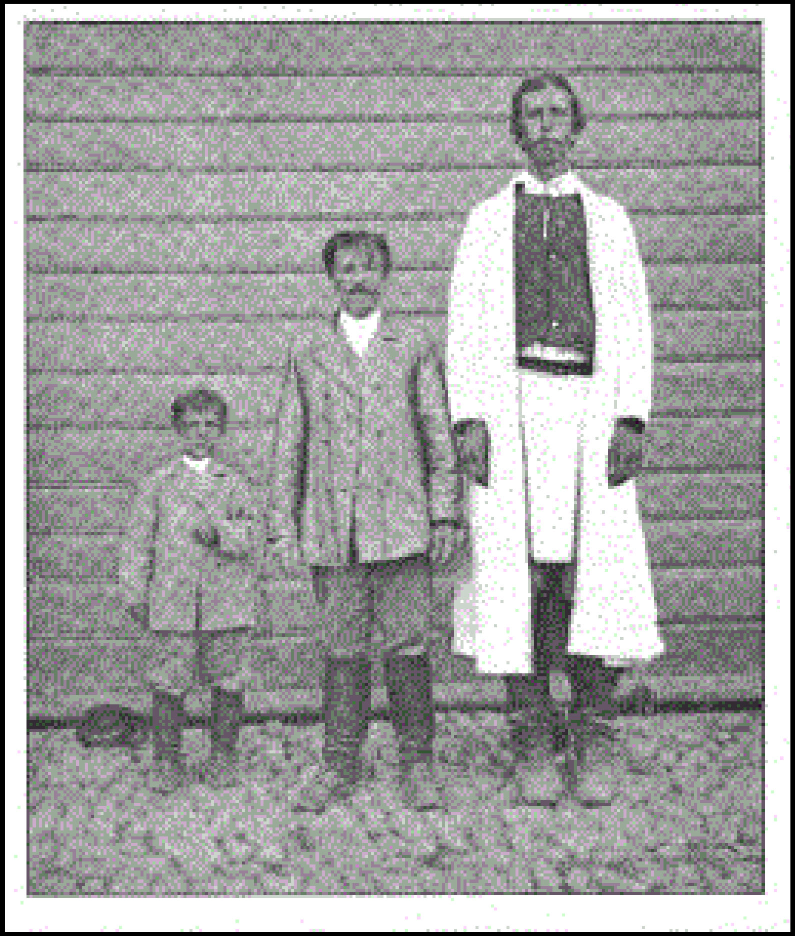


Journal "Wisla" (1902) Picture from a local fare in Lublin, Poland

Genotypes

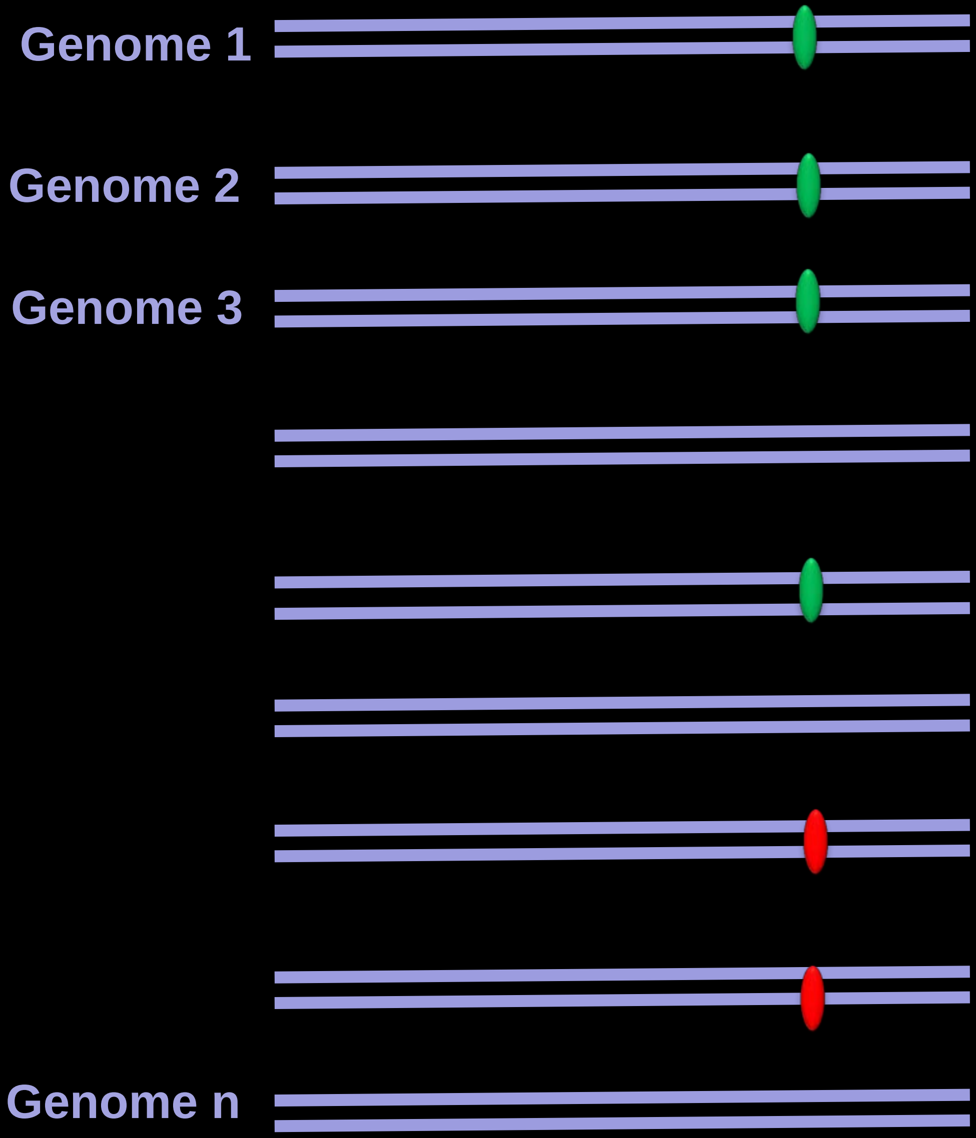
Phenotypes





Journal “Wisla” (1902) Picture from a local fare in Lublin, Poland

Association studies



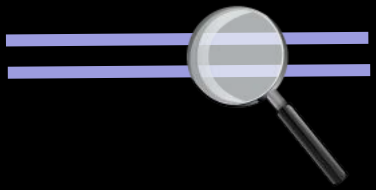
Genome 1

Genome 2

Genome 3

Genome n

Genotype:

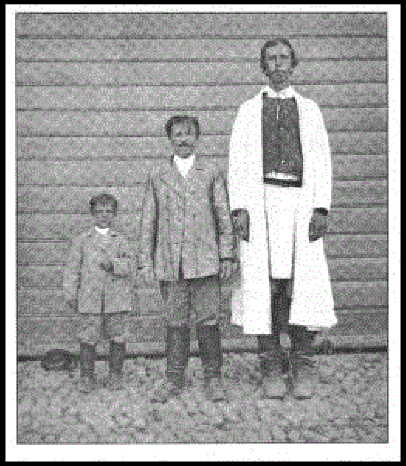


Phenotype

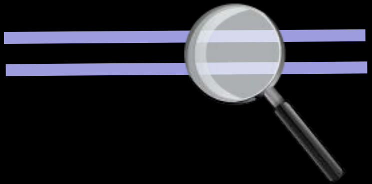
- height
- disease
- drug response

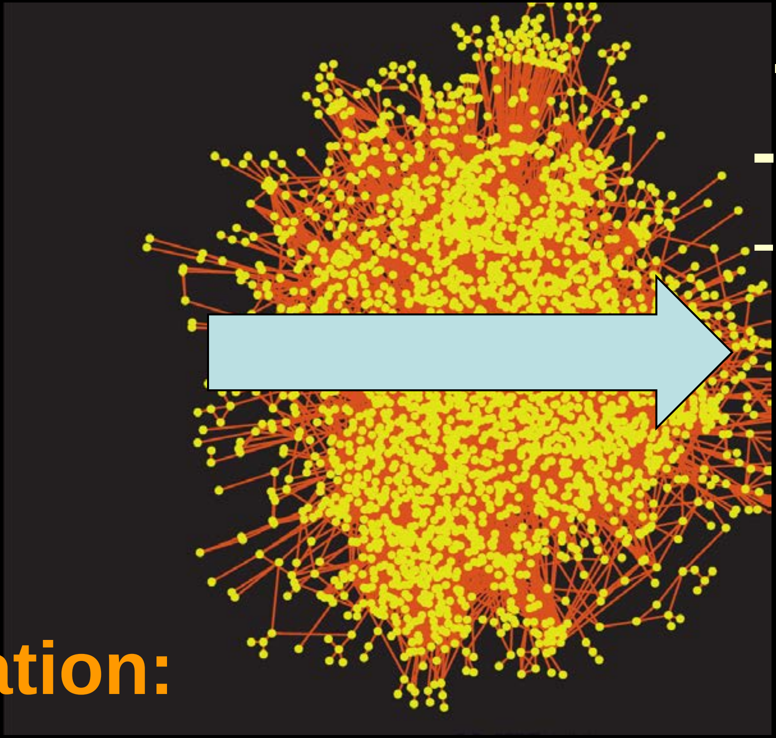
genotypic variation:

- Single nucleotide polymorphism (SNP)
- change in gene structure
- copy number variations ….



Genotype:



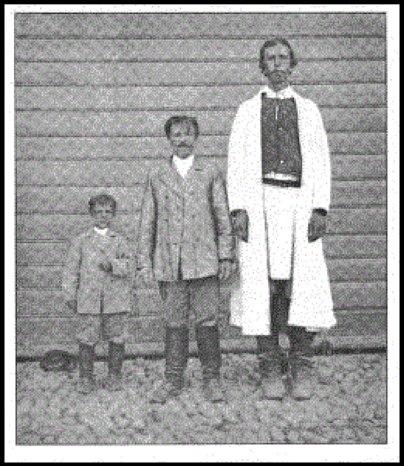


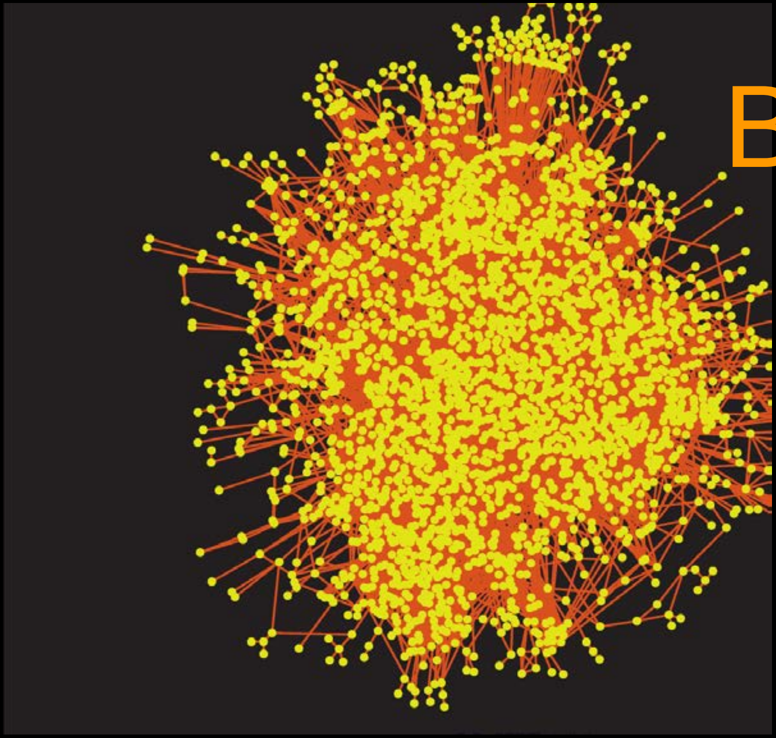
Phenotype

- **height**
- **disease**
- **drug response**

genotypic variation:

- Single nucleotide polymorphism (SNP)
- change in gene structure
- copy number variations ….





Biological networks

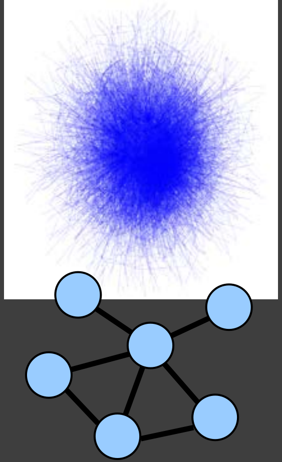
Nodes:

genes/ gene products

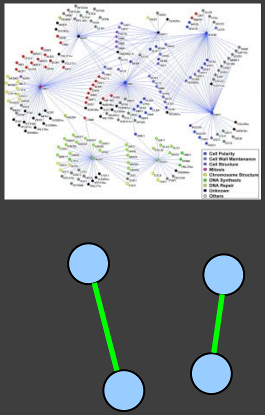
Edges:

Experimentally (or computationally) detected interactions

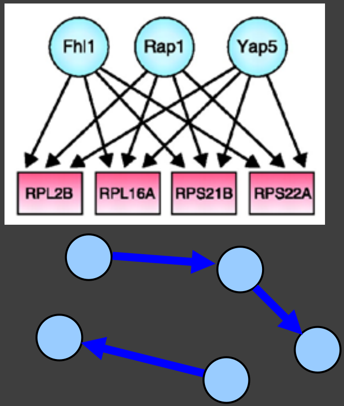
Interaction Networks



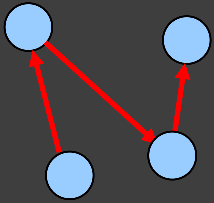
Physical interaction network



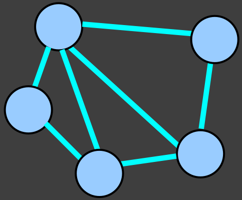
Genetic interaction network



Regulatory network



Kinome network

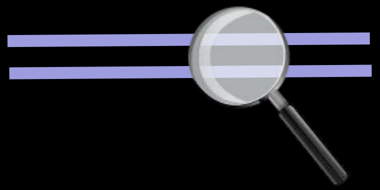


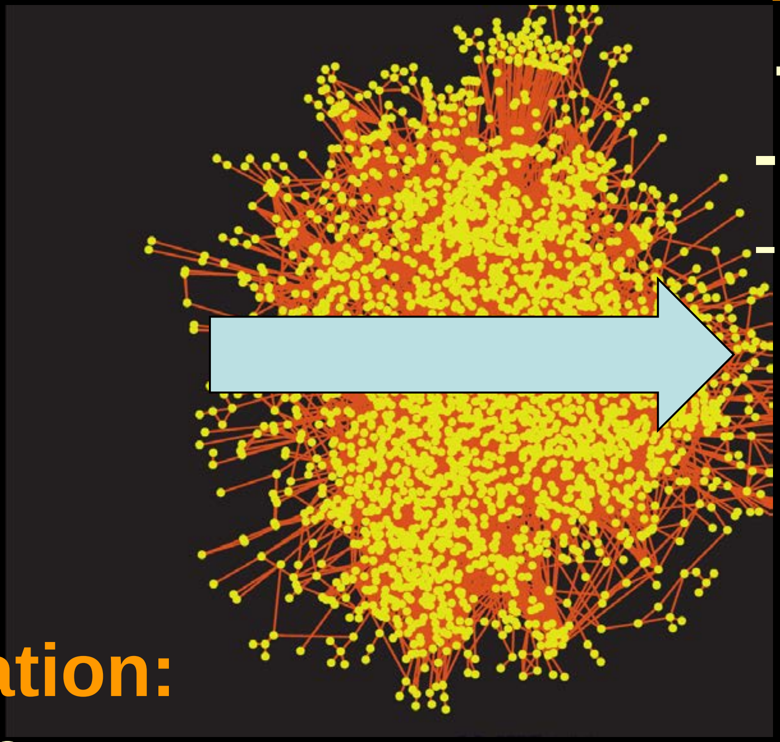
Coexpression network

System Biology

- Integrate biological data as an attempt to understand how biological systems function as whole.
- Study the relationships and interactions between various parts of a biological system
- Develop a model of the whole system, predict the behavior of a system upon perturbation.

Genotype:



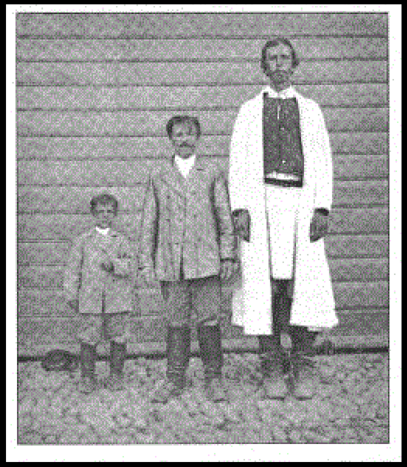


Phenotype

- height
- disease
- drug response

genotypic variation:

- Single nucleotide polymorphism (SNP)
- change in gene structure
- copy number variations



Copy number variations (CNV) (gene dosage)

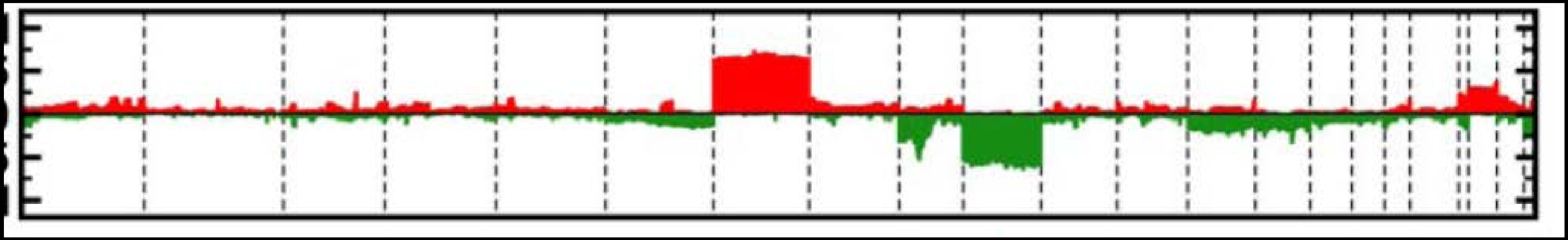
- implicated in large number of human diseases (cancer, Crohn's disease, autism)

- 28,025 structural variants identified in 1000 genome study (2,000 changes affecting full genes or exons)

- Frequent type of somatic mutations in cancer

Somatic copy number aberrations in brain cancer

CNV

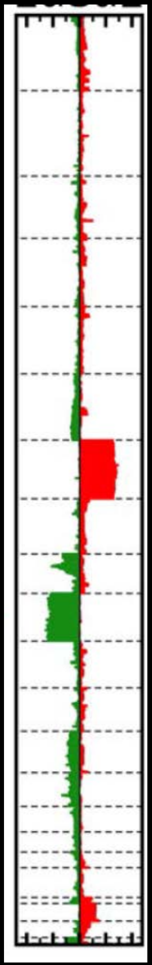


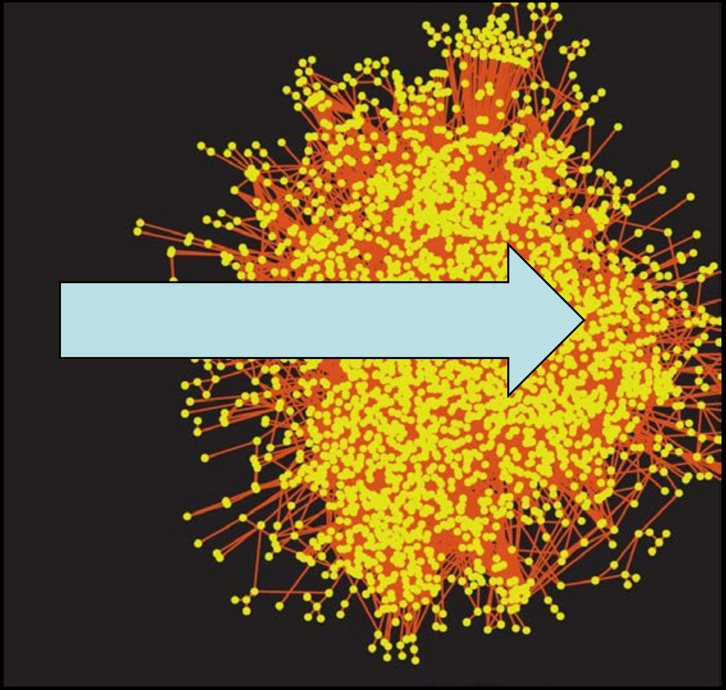
chromosomes

Propagation of the effects of Copy number aberrations in Glioma

CNV

chromosomes

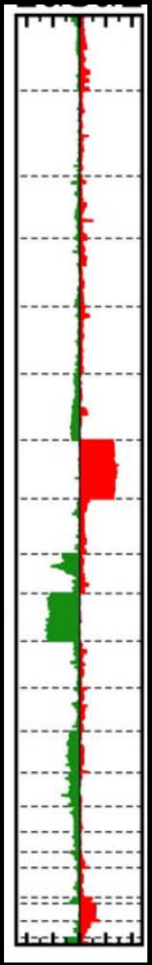


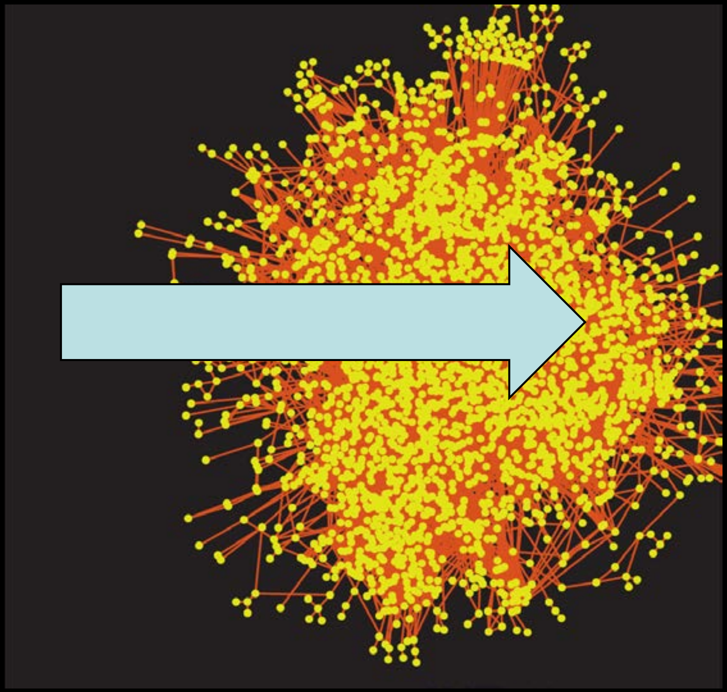


Propagation of the effects of Copy number aberrations in Glioma

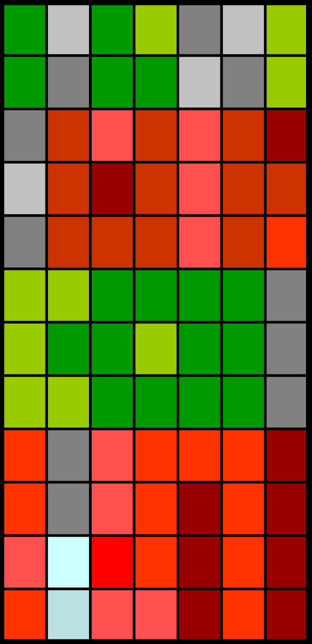
CNV

chromosomes



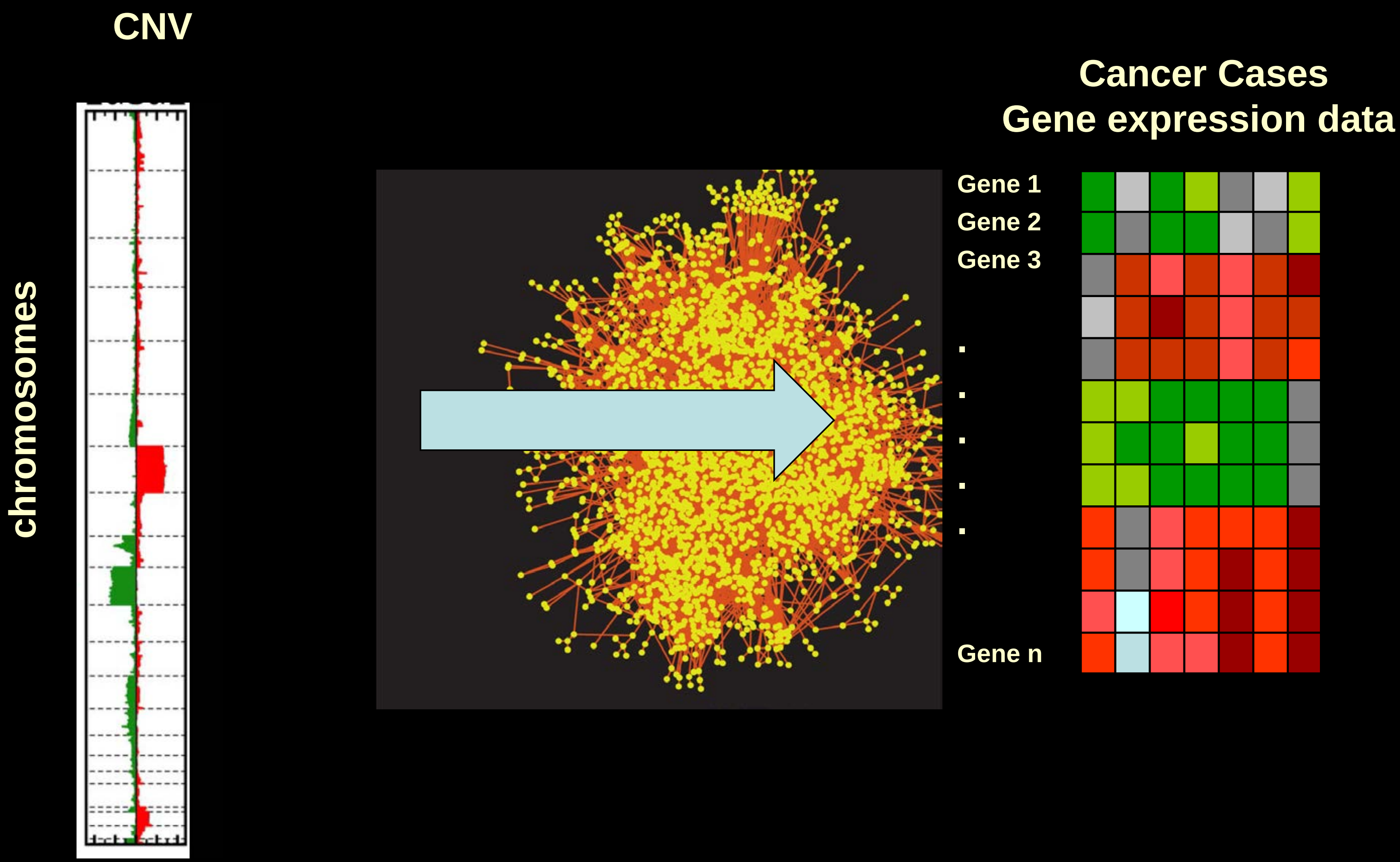


Gene 1
Gene 2
Gene 3
.
.
.
.
.
Gene n



Cancer Cases
Gene expression data

Propagation of the effects of Copy number aberrations in Glioma



CNV

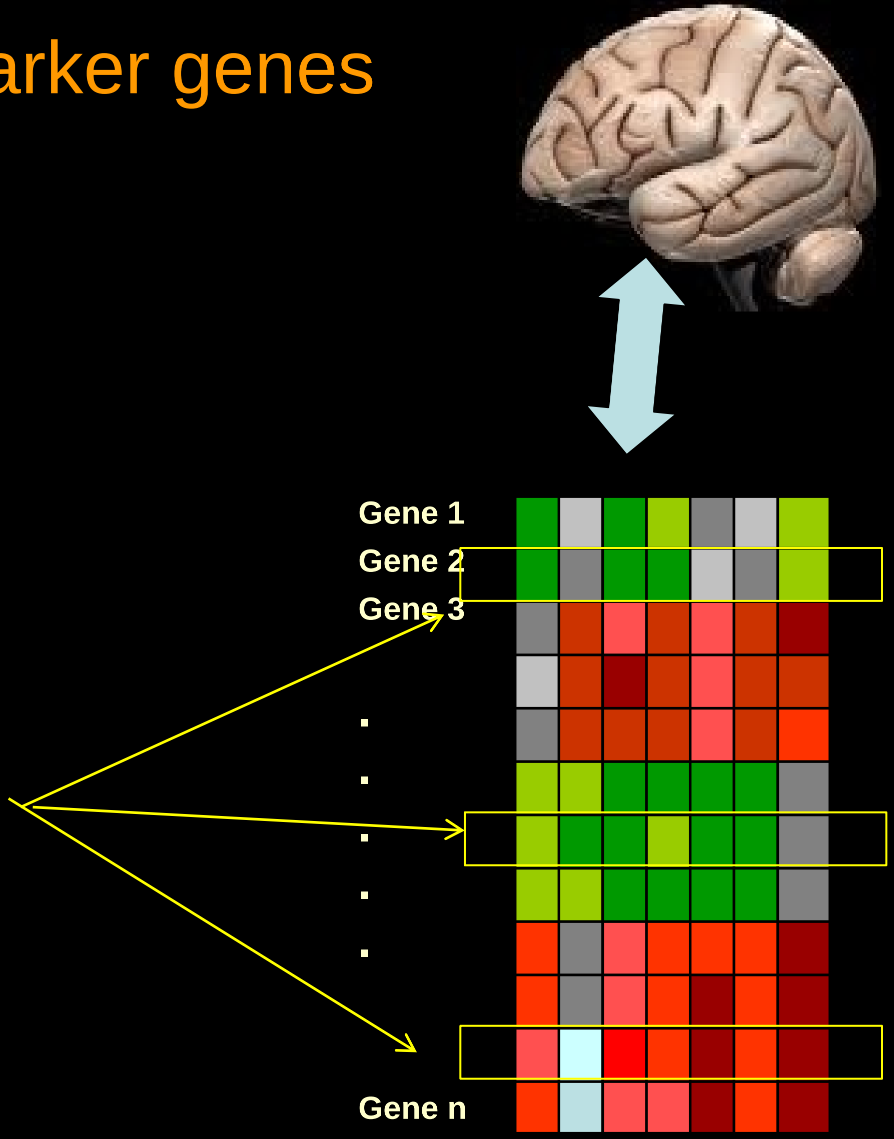
chromosomes

Cancer Cases
Gene expression data

Gene 1
Gene 2
Gene 3
.
.
.
.
.
Gene n

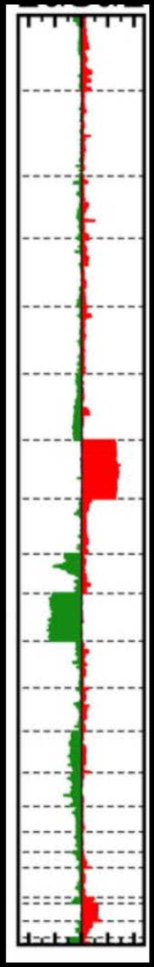
Selecting marker genes

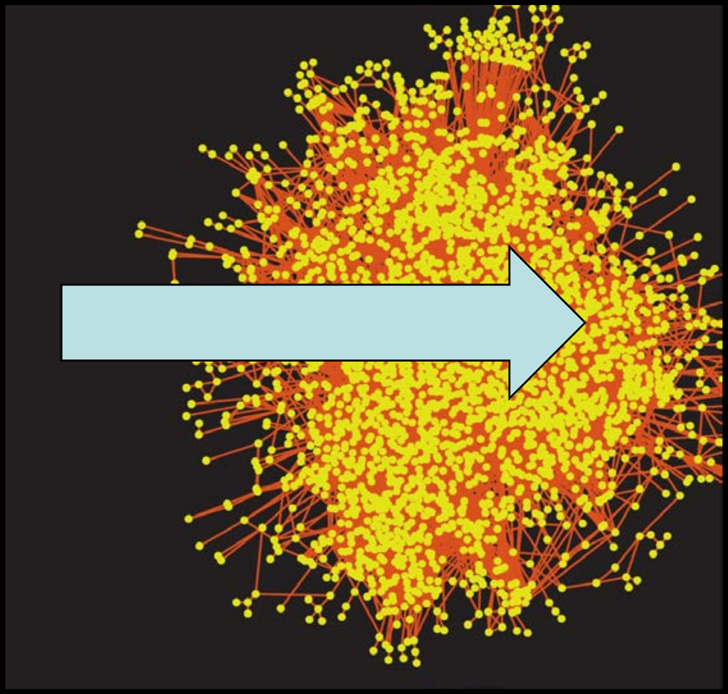
target genes

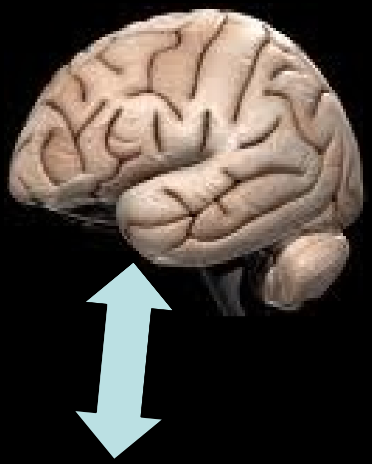


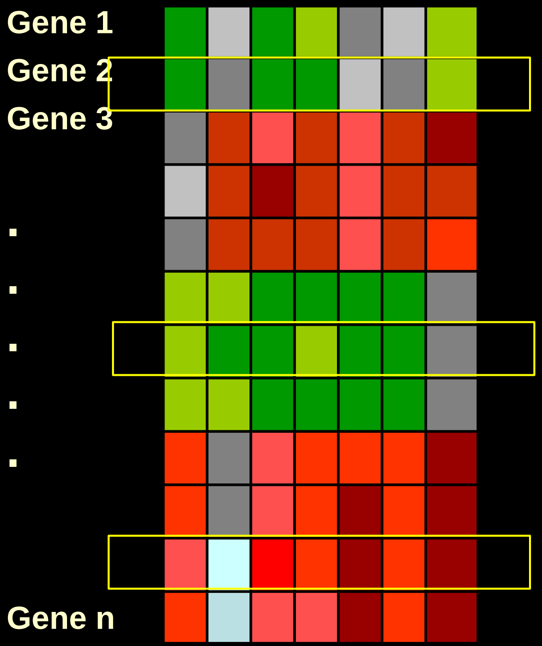
Cancer Cases
Gene expression data

Selecting marker genes



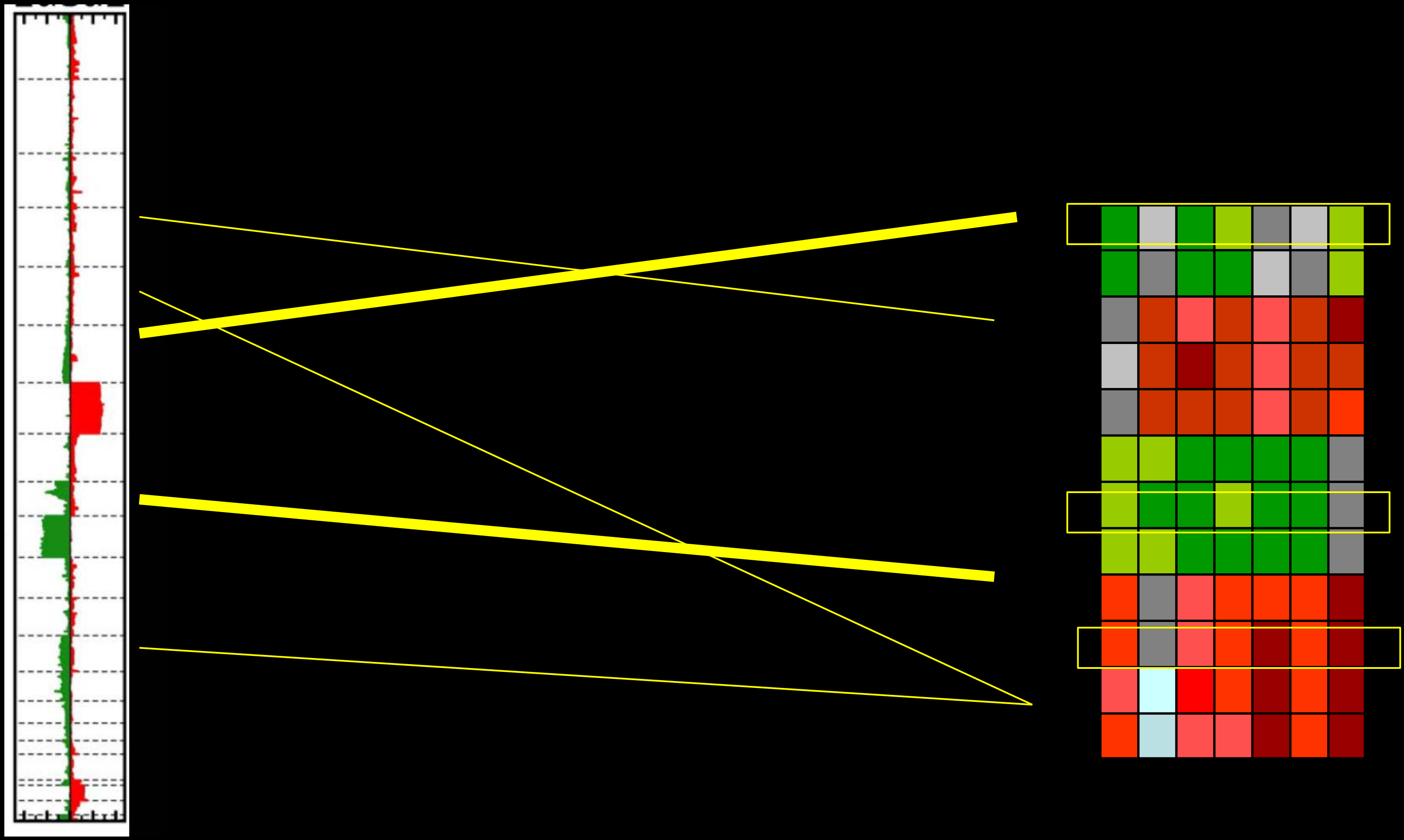






Cancer Cases
Gene expression data

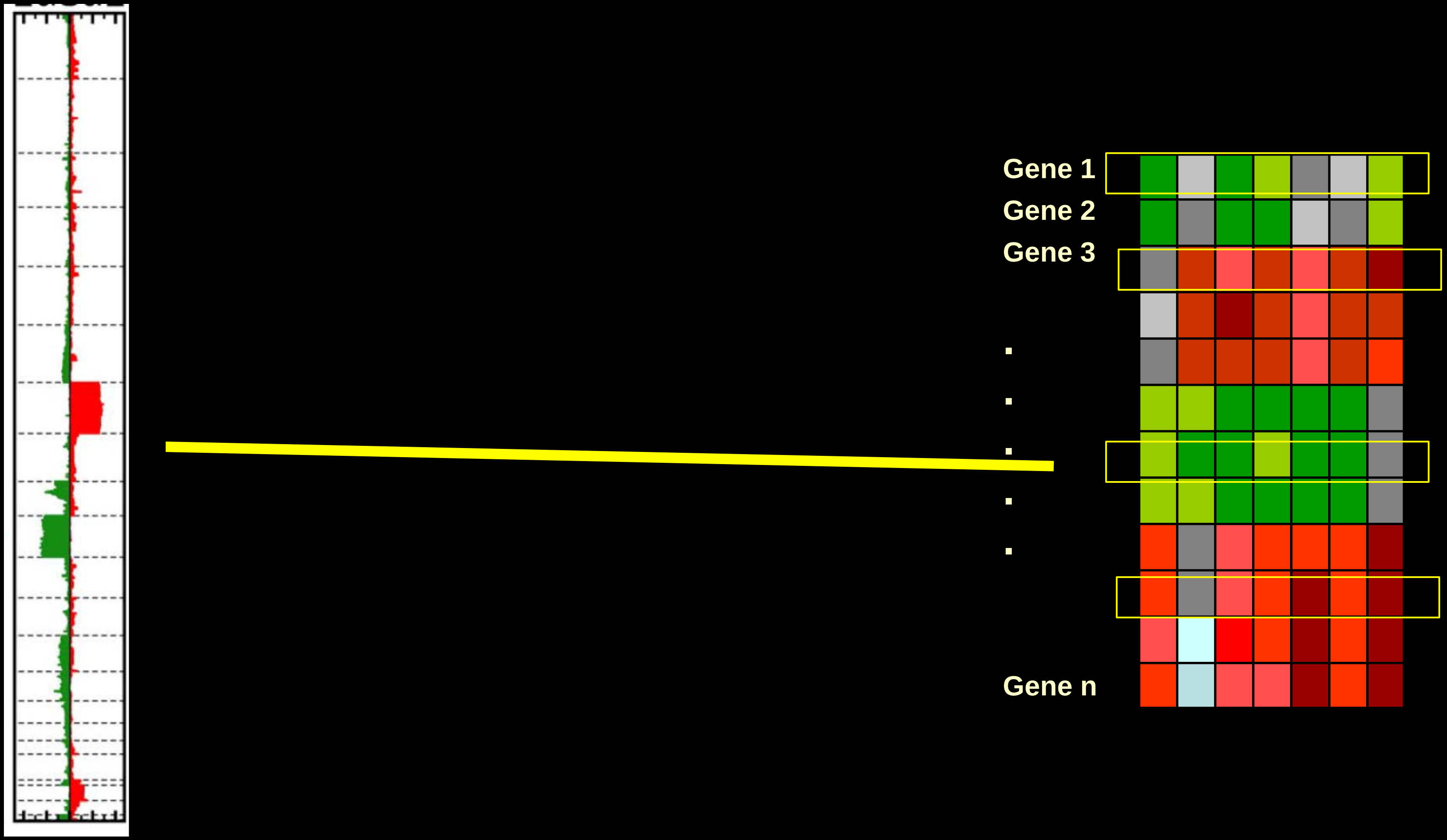
Associations between copy number variations and gene expression of selected target genes



Cancer Cases
CNV data

Cancer Cases
Gene expression data

Significant correlation between CNV and expression



Gene 1

Gene 2

Gene 3

.

.

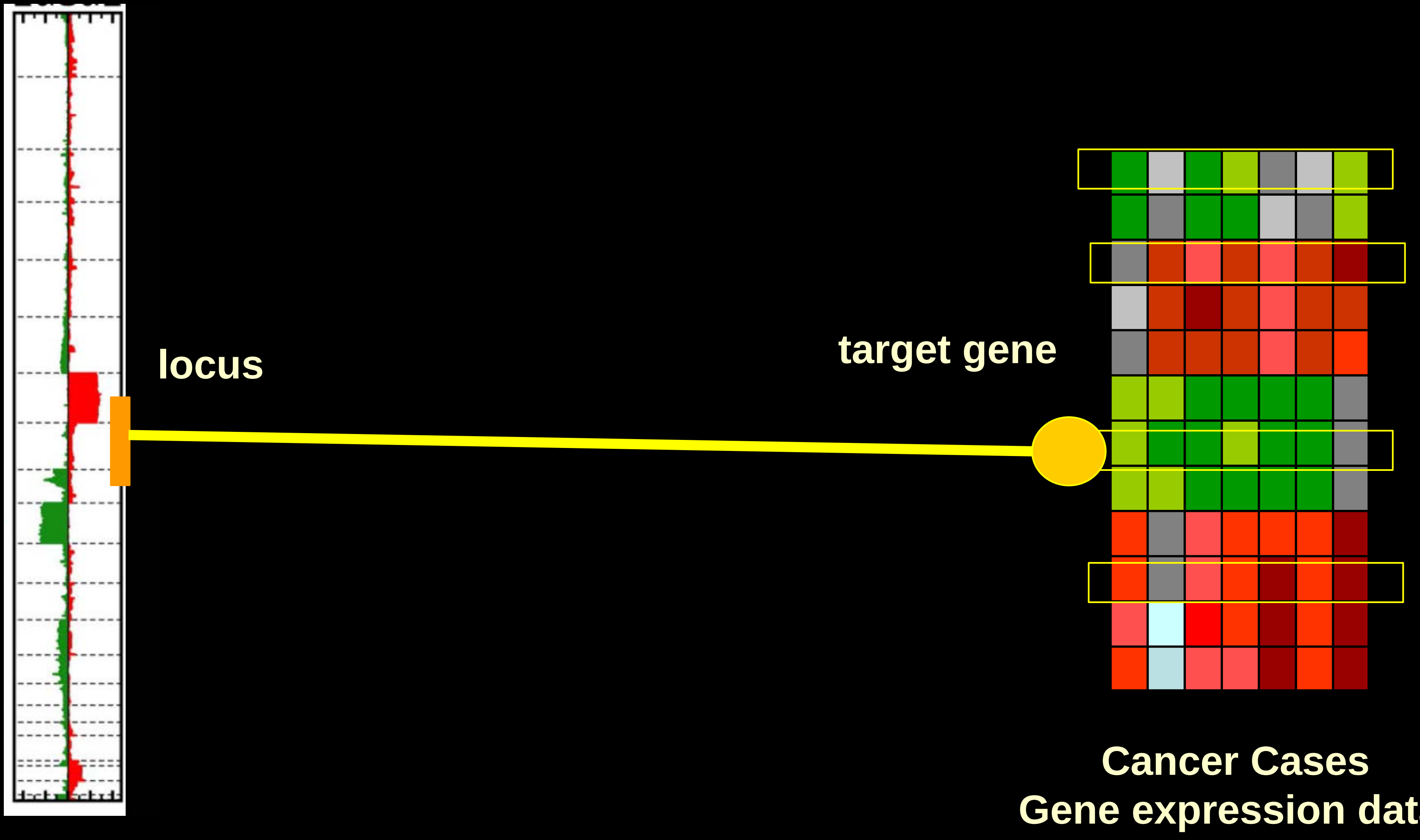
.

.

.

Gene n

Significant correlation between CNV and expression

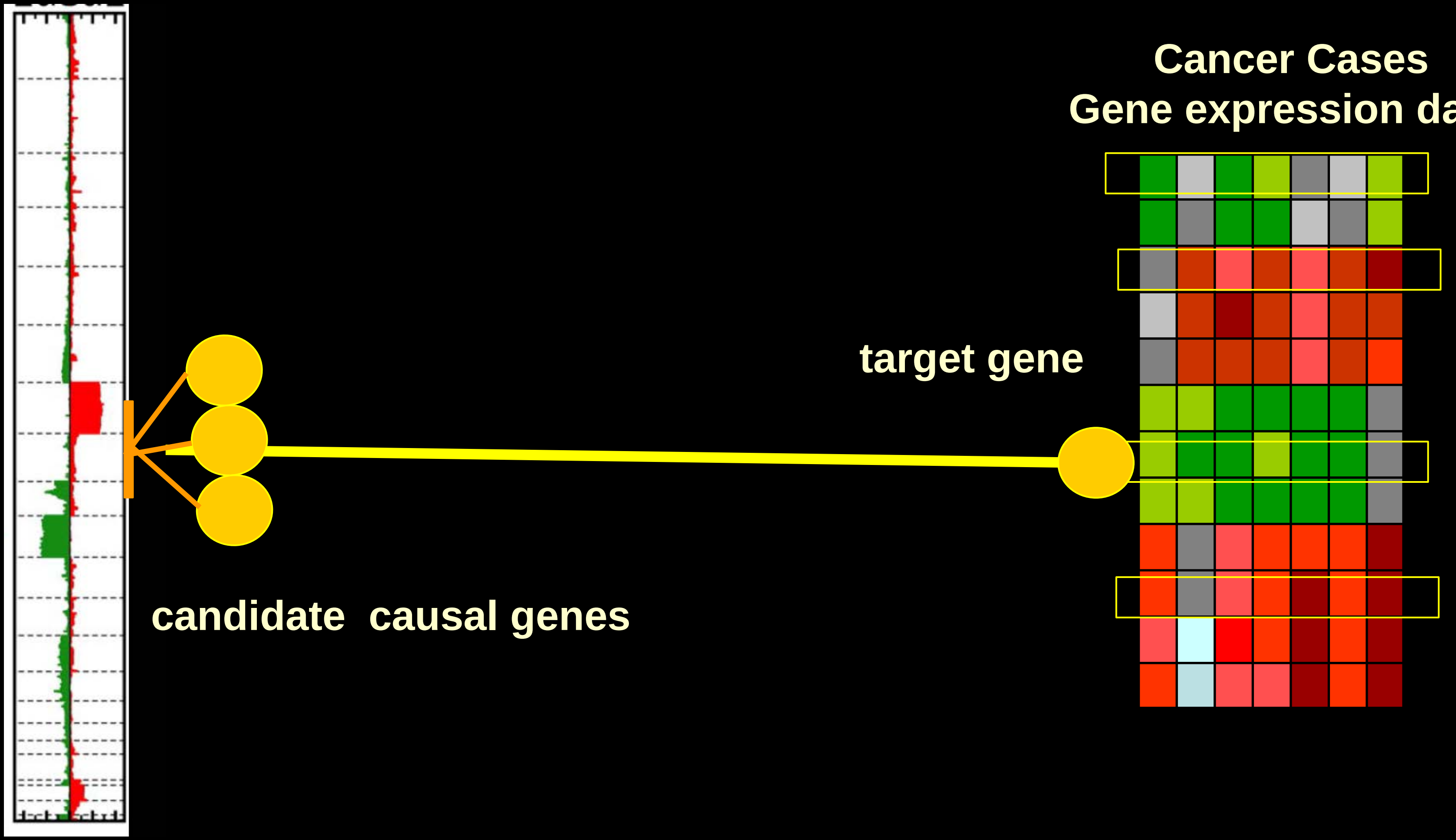


locus

target gene

Cancer Cases
Gene expression dat

Significant correlation between CNV and expression

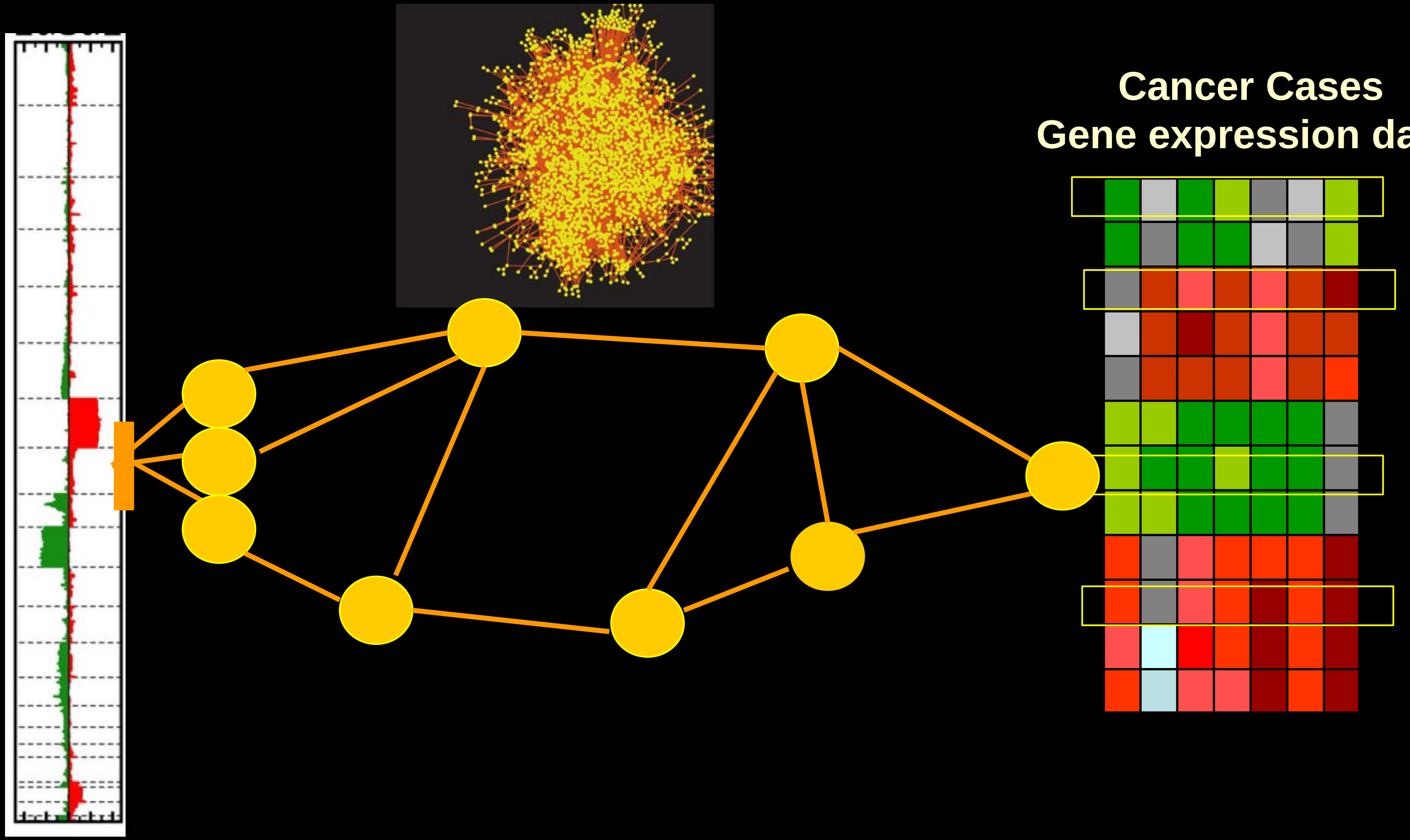


Cancer Cases
Gene expression da

target gene

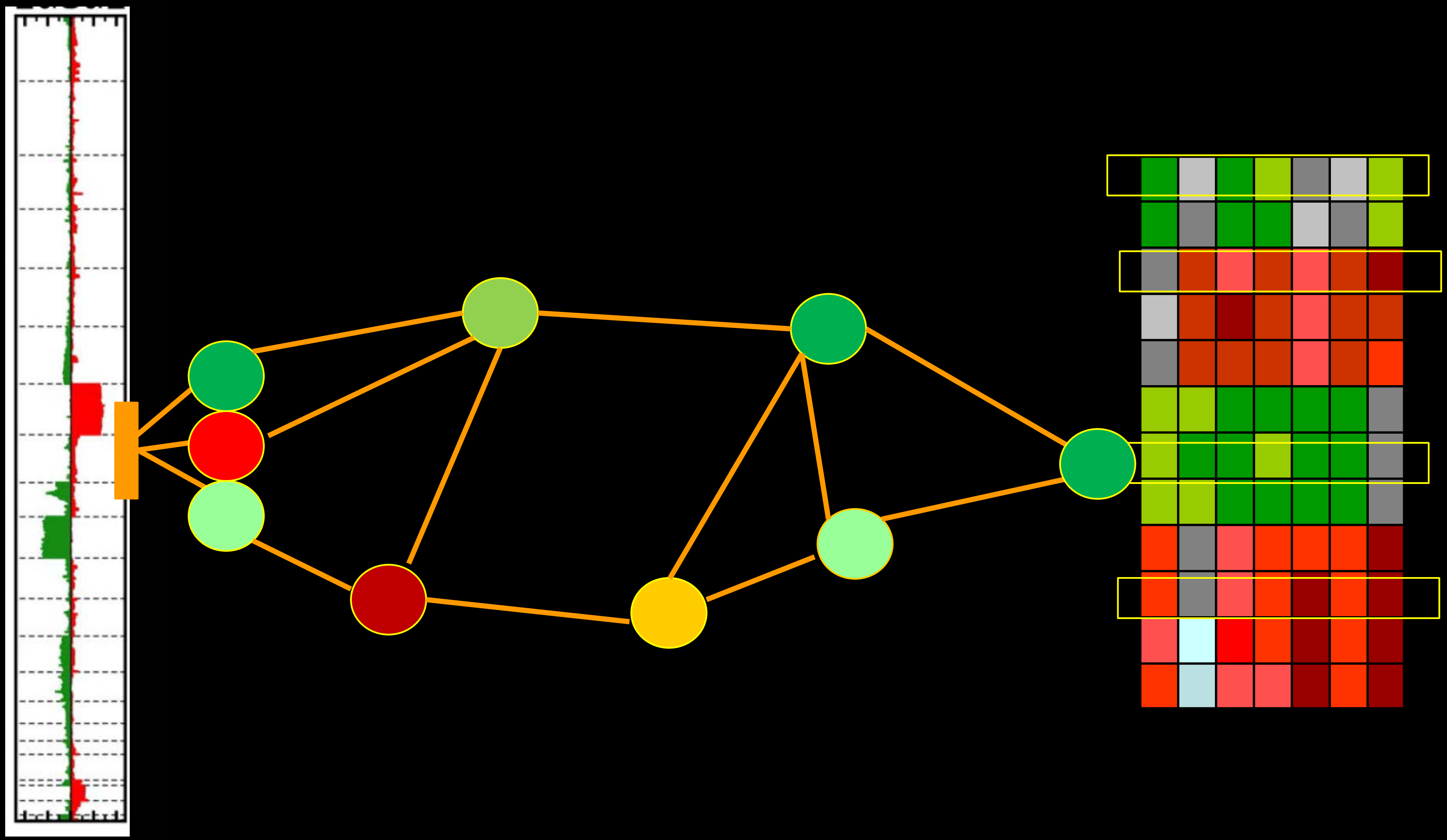
candidate causal genes

Uncovering pathways of information flow between CNV and target gene

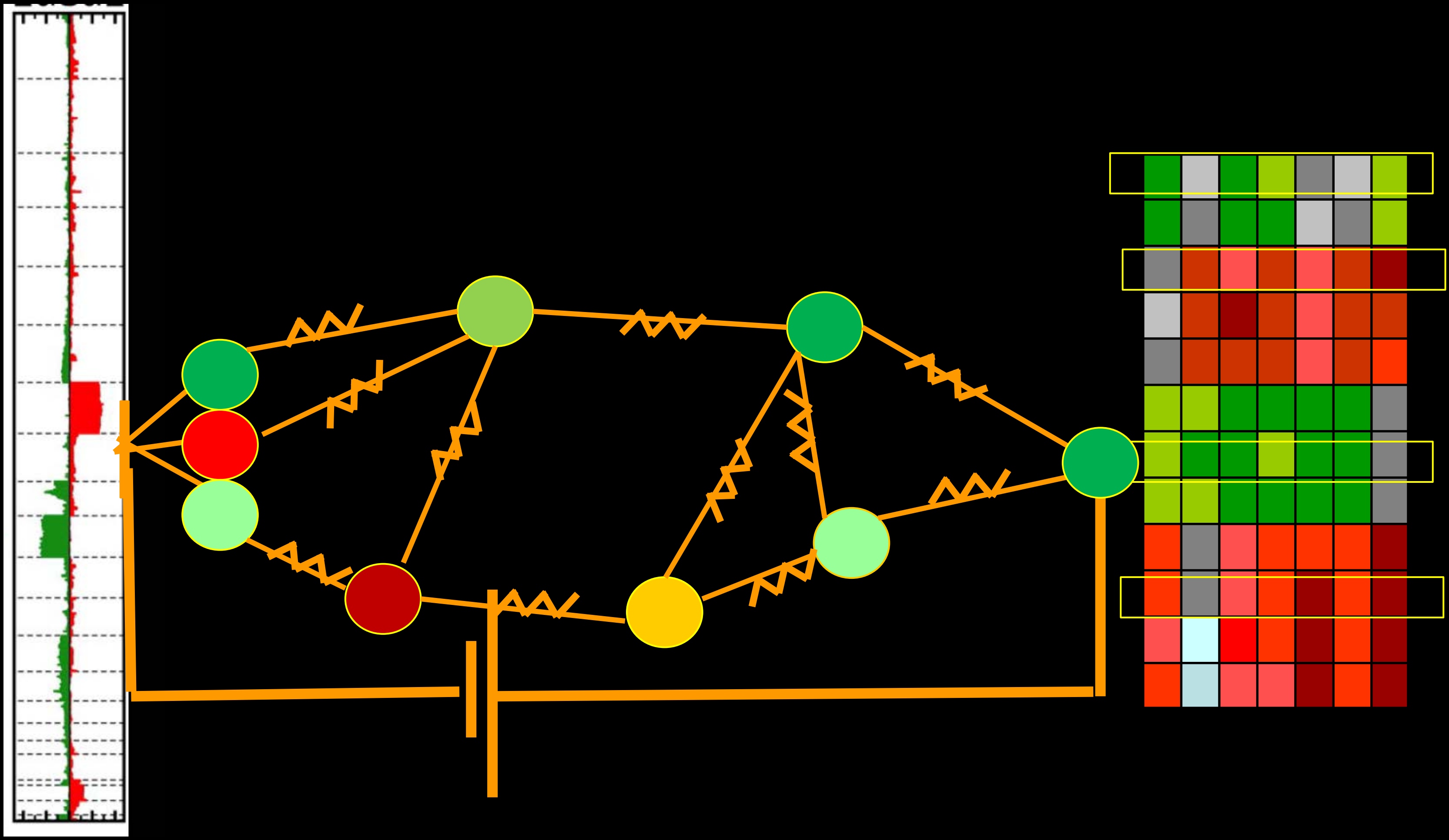


Cancer Cases
Gene expression da

Using expression to guide path discovery

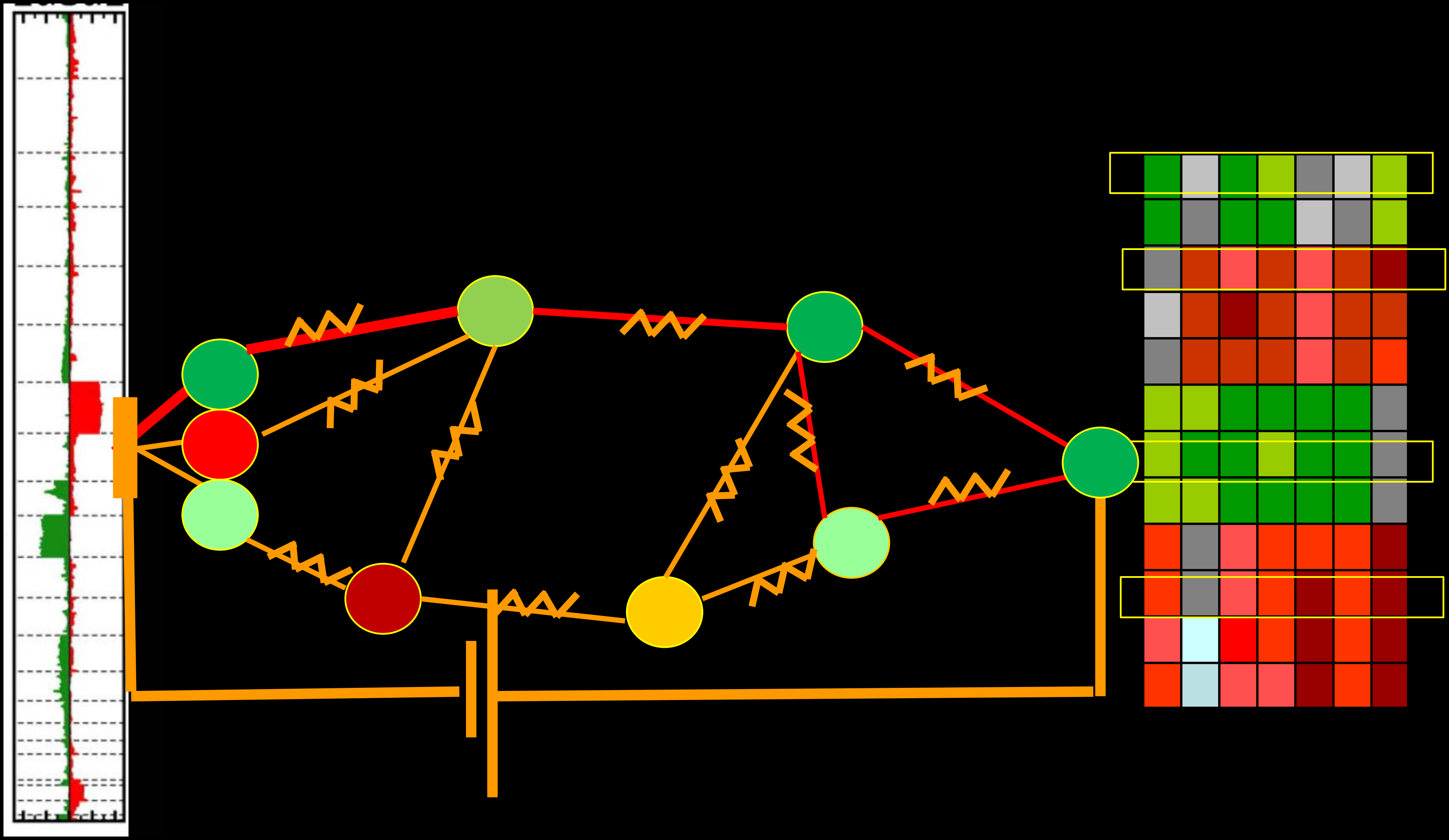


Translating probabilities it resistances



Resistance **- set to favor most likely path -based on gene expression values**
(reversely proportional to the average correlation of the expression of the adjacent genes with expression of the target gene)

Finding subnetworks with significant current flow



Resistance - **set to favor most likely path -based on gene expression values**

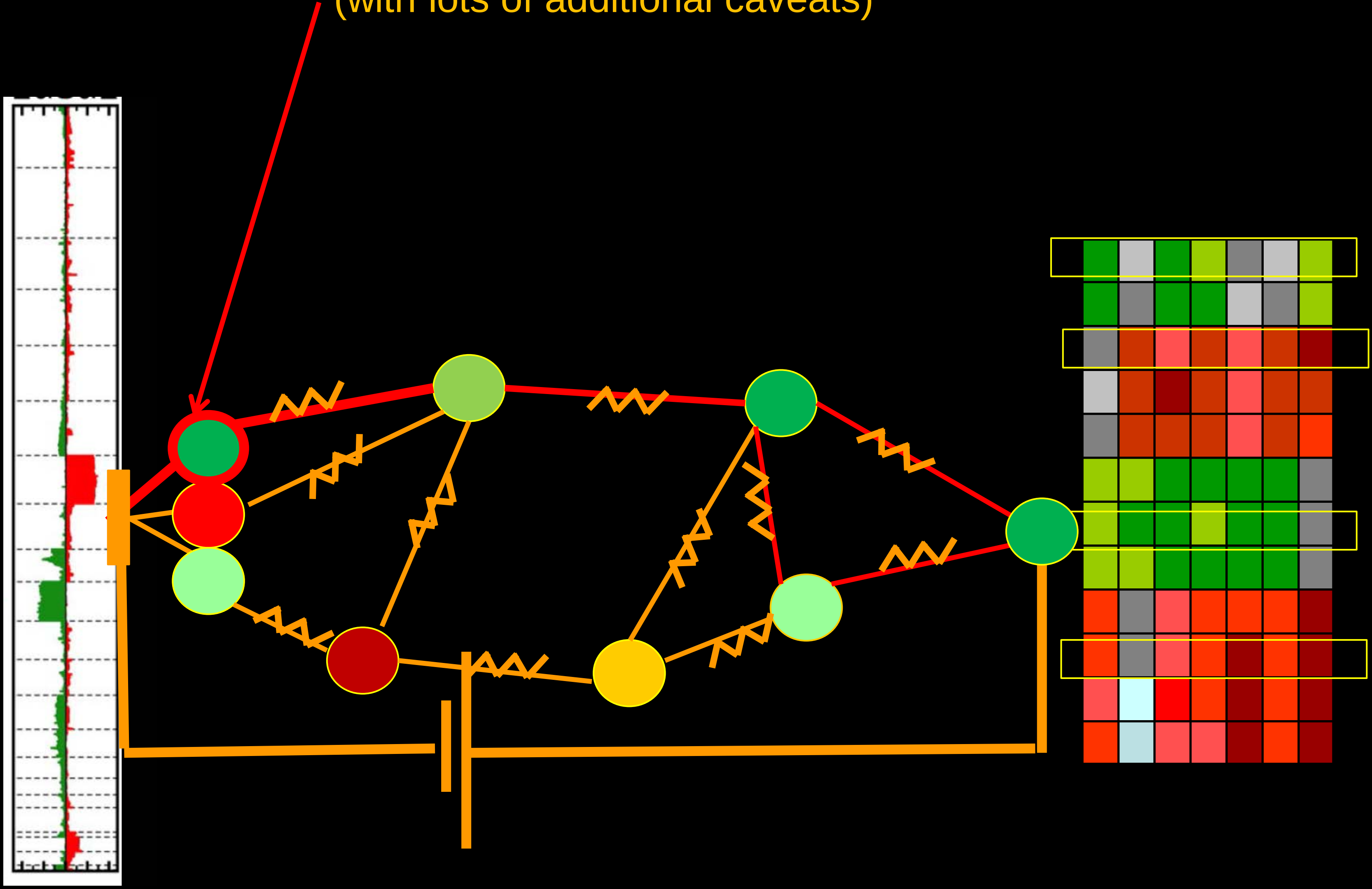
(reversely proportional to the average correlation of the expression of the adjacent genes with expression of the target gene)

Goals :

- A method for system level analysis of propagation of such perturbation in the network
- Prediction of “causal” mutations
- Identification master regulators (network hubs) involved in disease
- Identification pathways dys-regulated in disease

Putative causal variation

(with lots of additional caveats)



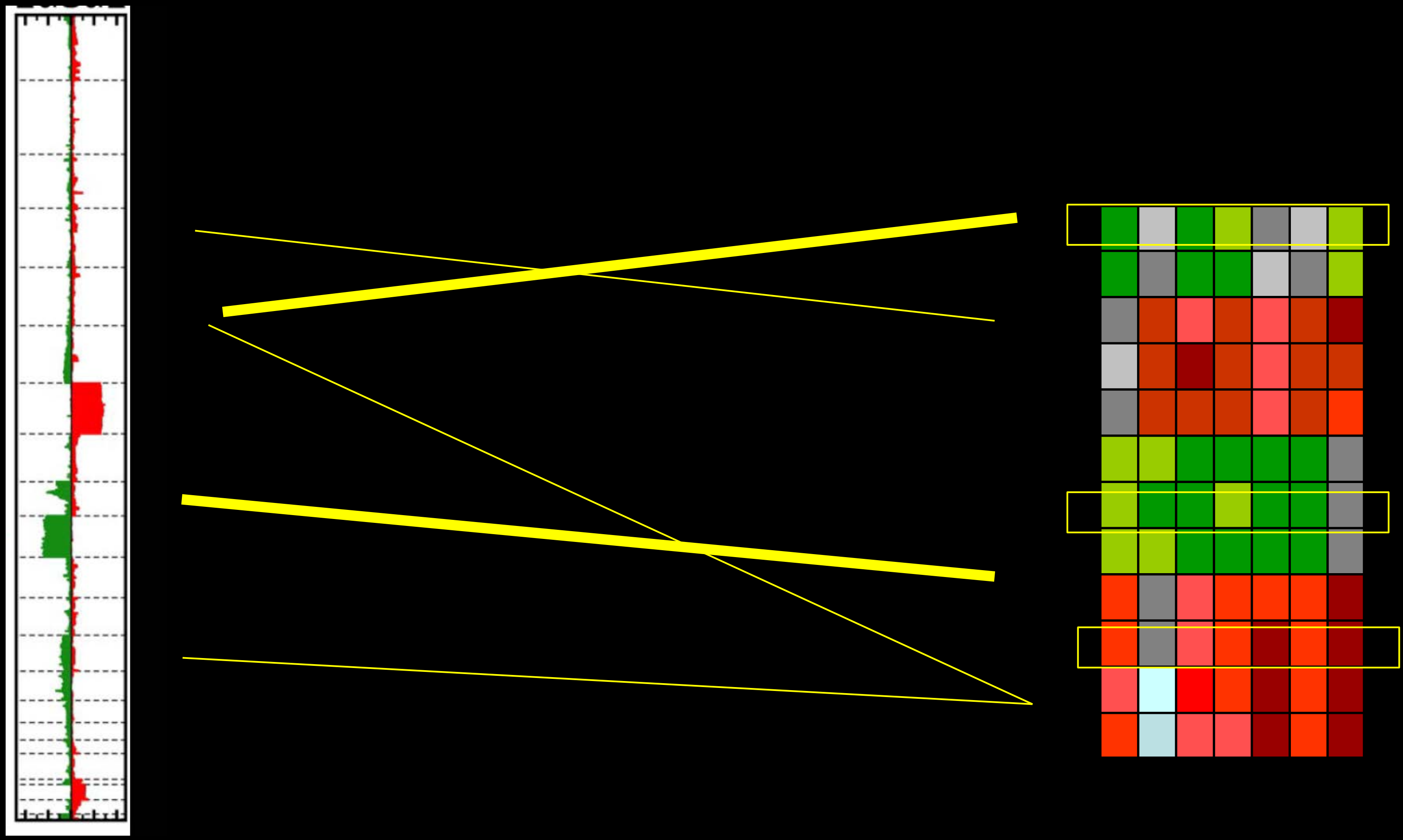
Resistance **- set to favor most likely path -based on gene expression values**

(reversely proportional to the average correlation of the expression of the adjacent genes with expression of the target gene)

Goals :

- ✓ A method for system level analysis of propagation of such perturbation in the network
- ✓ Prediction of “causal” mutations
- Prediction “master regulators” (network hubs) involved in disease
- Prediction pathways dys-regulated in disease

Solve current flow for all pairs and find nodes belonging to many paths



Cancer Cases
CNV data

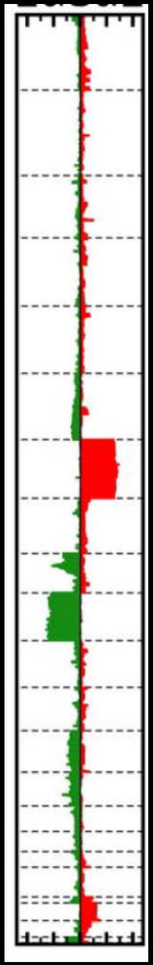
Cancer Cases
Gene expression data

Goals :

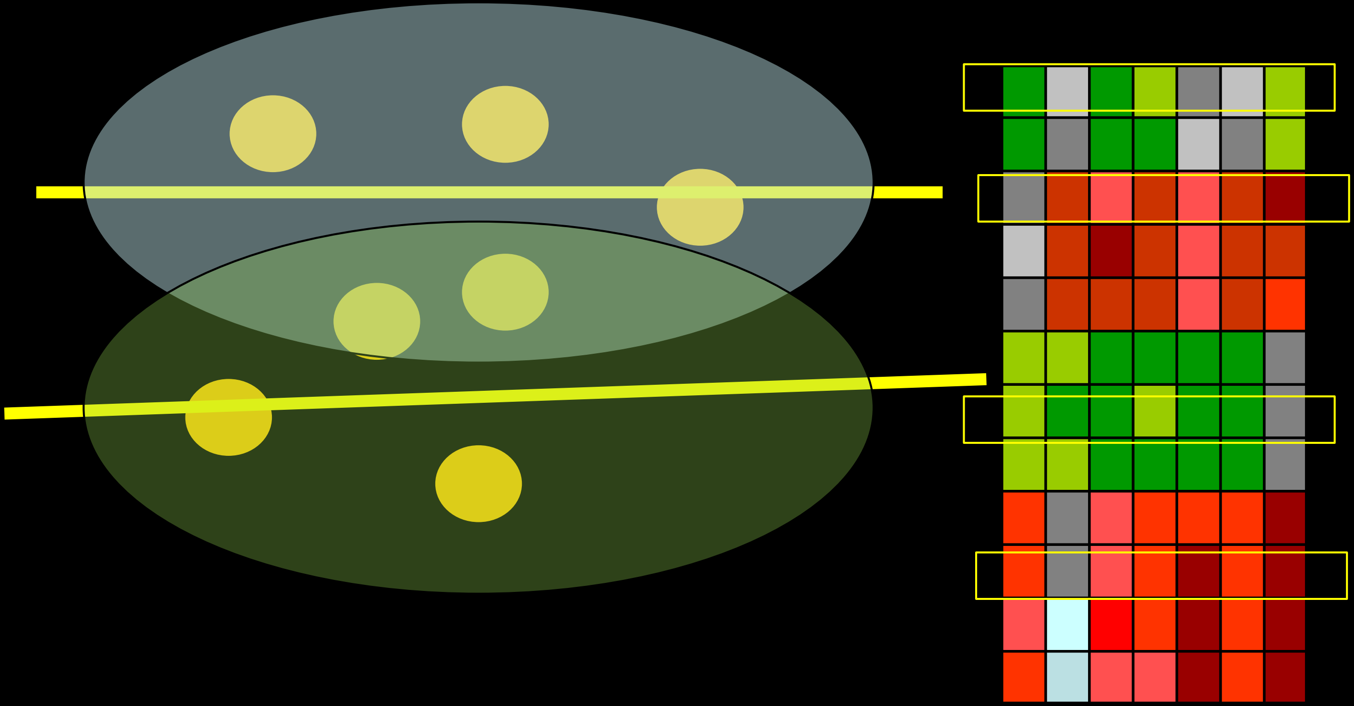
- ✓ A method for system level analysis of propagation of such perturbation in the network
- ✓ Prediction of “causal” mutations
- ✓ Prediction of “master regulators” (network hubs) involved in disease
- Prediction of pathways dys-regulated in disease

Are there common functional pathways?

Cancer Cases
CNV data



Cancer Cases
Gene expression dat



Goals :

- ✓ A method for system level analysis of propagation of such perturbation in the network
- ✓ Prediction of “causal” mutations
- ✓ Prediction of “master regulators” (network hubs) involved in disease
- ✓ Prediction of pathways dys-regulated in disease

Systems biology of diseases:

- methods for system level analysis perturbations related to dieses in the network
- Prediction of master regulators (network hubs) involved in disease
- Prediction drug targets / interventions
- Prediction of pathways dys-regulated in diseases
- Disease classification / prognosis

Challenging Design questions

- Network size 20,000 nodes – efficient approaches are needed
- Dealing with network noise
- Statistical issues
- Heterogeneity of the “information” type

Acknowledgments

Group members:

Yoo-Ah Kim (Cancer)

DongYeon Cho (Fly)

Xiangjun Du

Jan Hoinka

Yang Huang

Raheleh Salari

Damian Wojtowicz

Collaborators in glioma project:

Stefan Wuchty (NCBI)

Jozef Przytycki (GWU)

Journal "Wisla" (1902) Picture from a local fare in Lublin, Poland