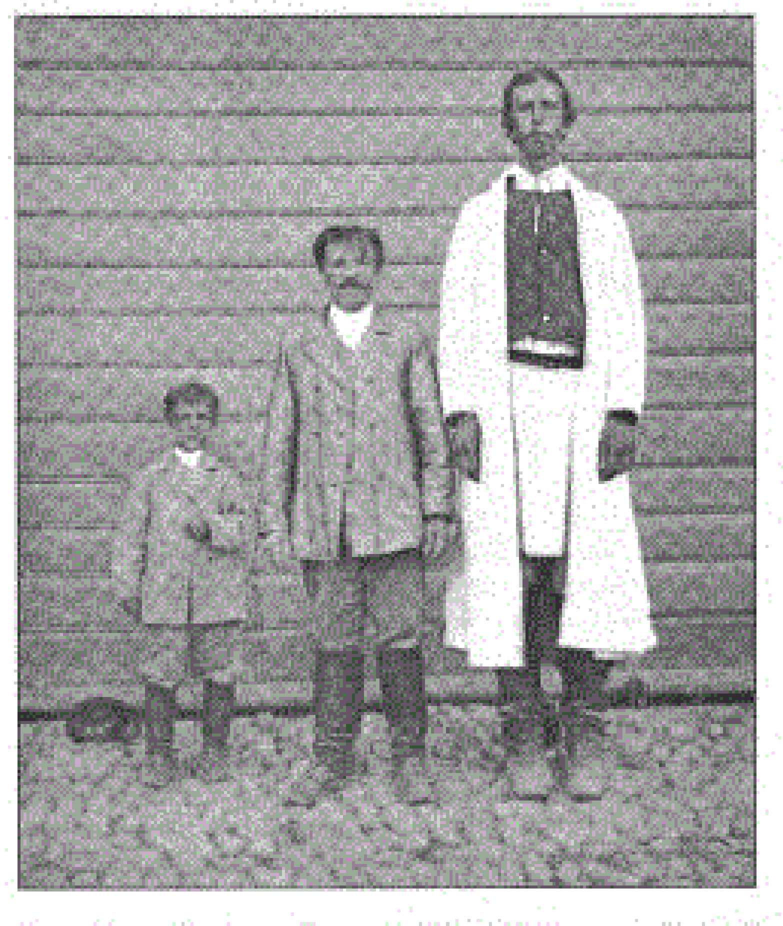
Using Network Flow to Bridge the Gap between Genotype and Phenotype

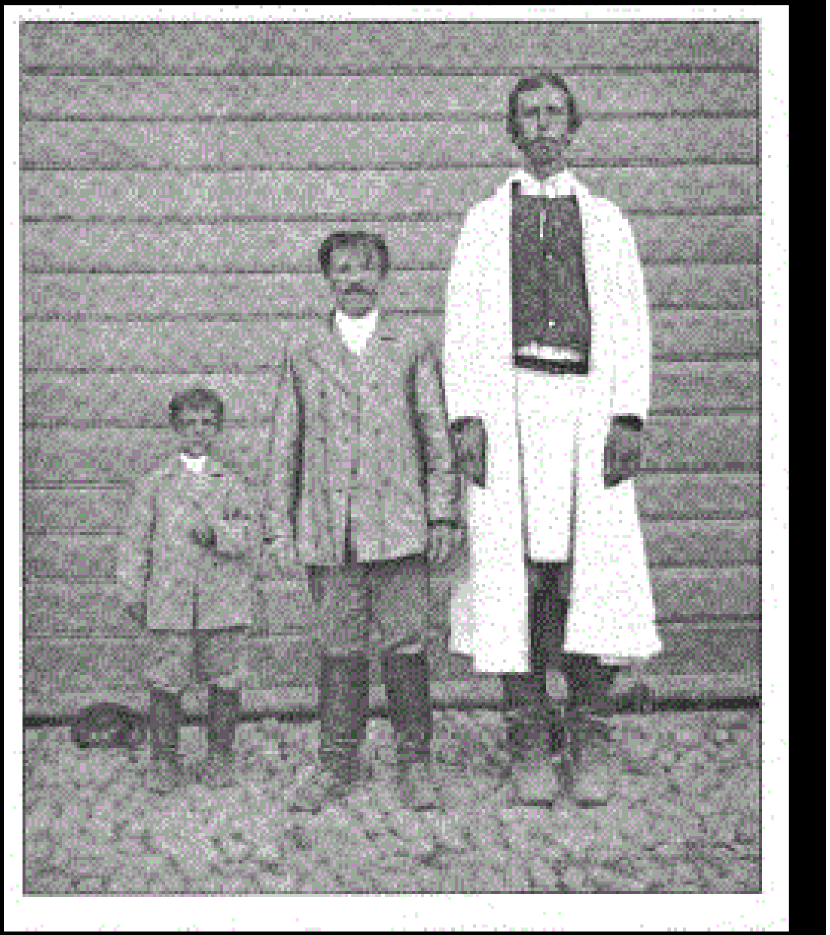
Teresa Przytycka
NIH / NLM / NCBI



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

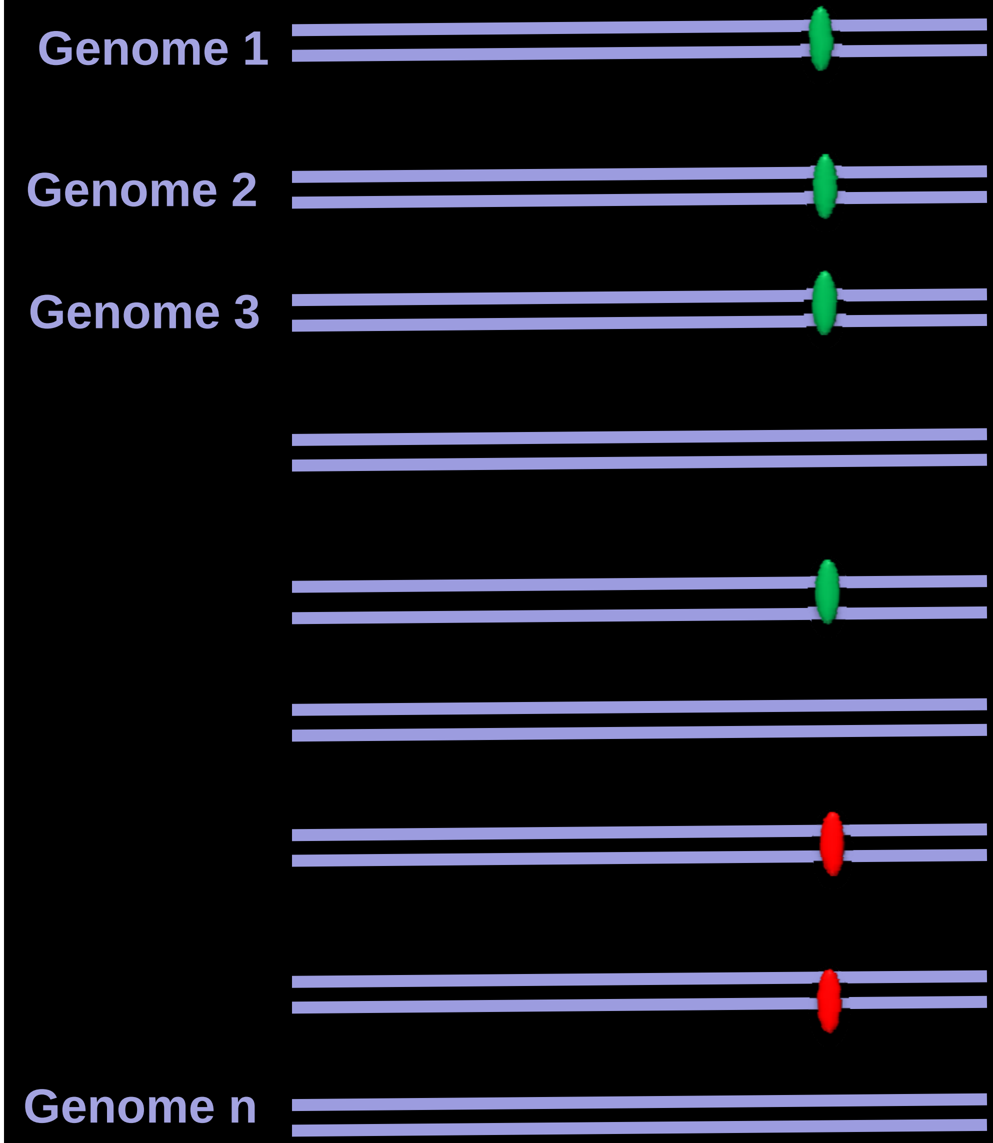
Genotypes

Phenotypes



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

Association studies



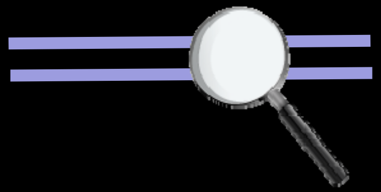
Genome 1

Genome 2

Genome 3

Genome n

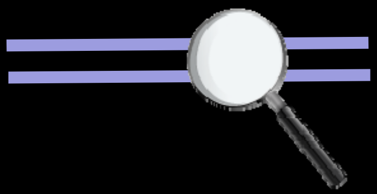
Genotype:

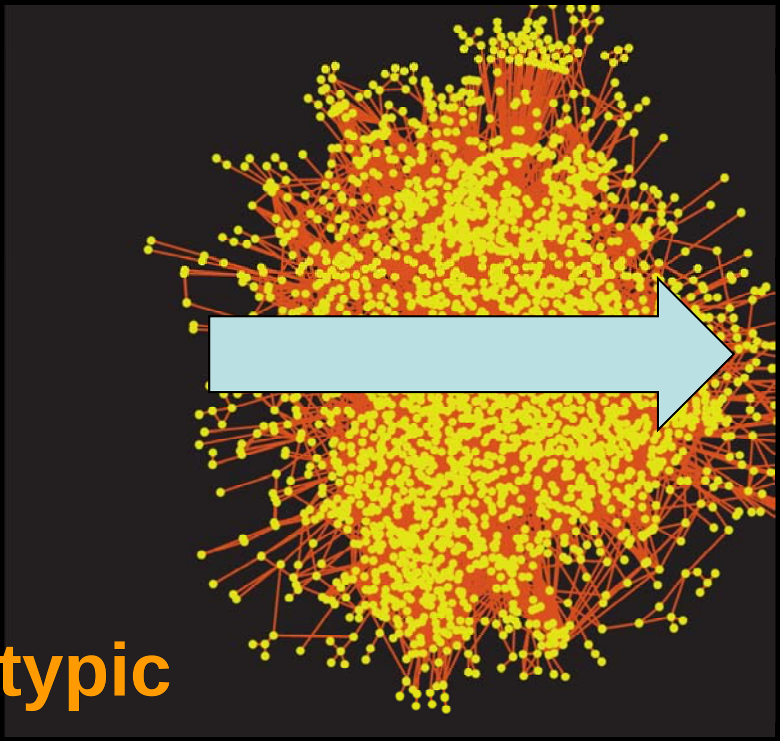


effects of genotypic variation:

- change in amino acid
- change in gene structure
- copy number variations ….

Genotype:





Phenotype (e.g. disease)

effects of genotypic variation:

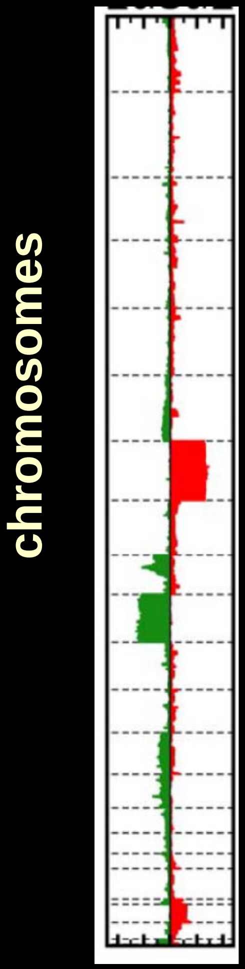
- change in amino acid
- change in gene structure
- copy number variations ….

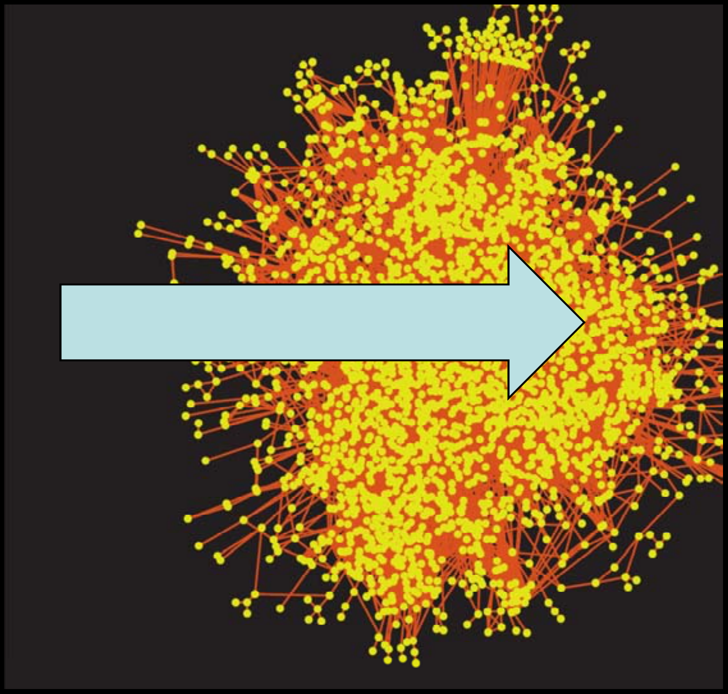
Goals :

- 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

Propagation of the effects of Copy number aberrations in Glioma

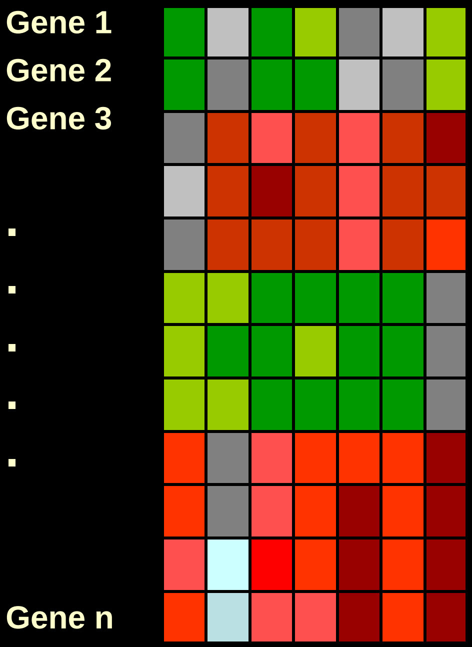
CNV

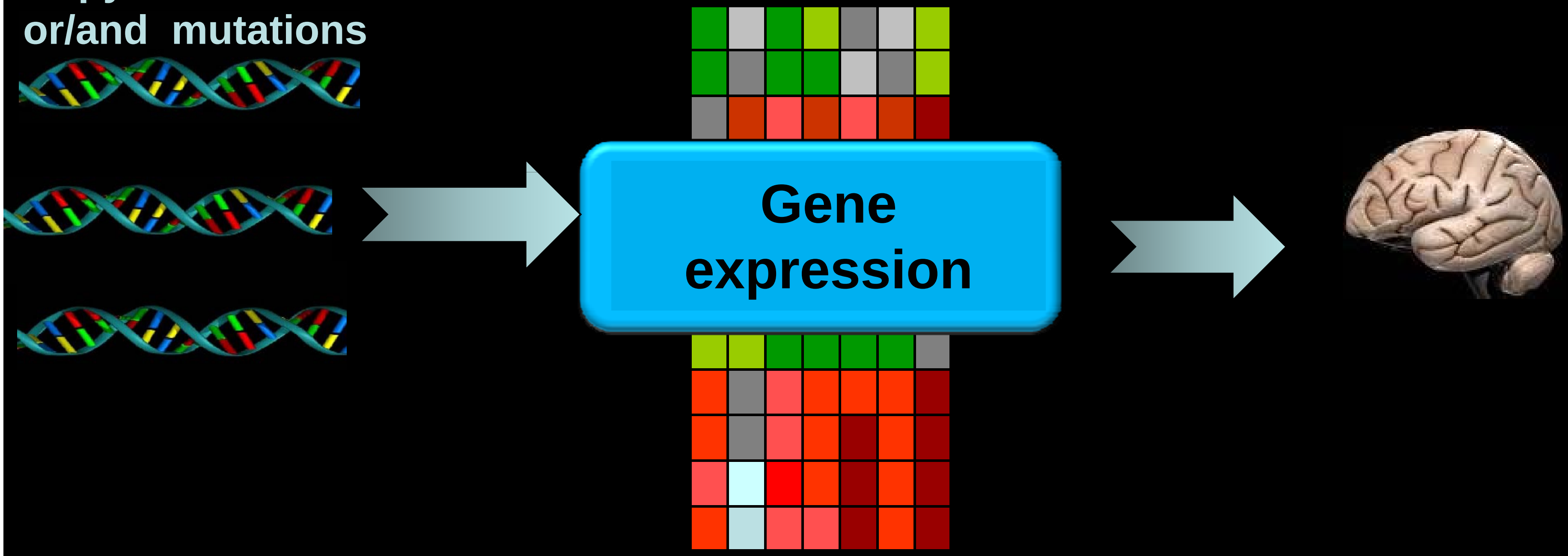




Integrated
Protein-protein, protein-DNA
phosphorylation network

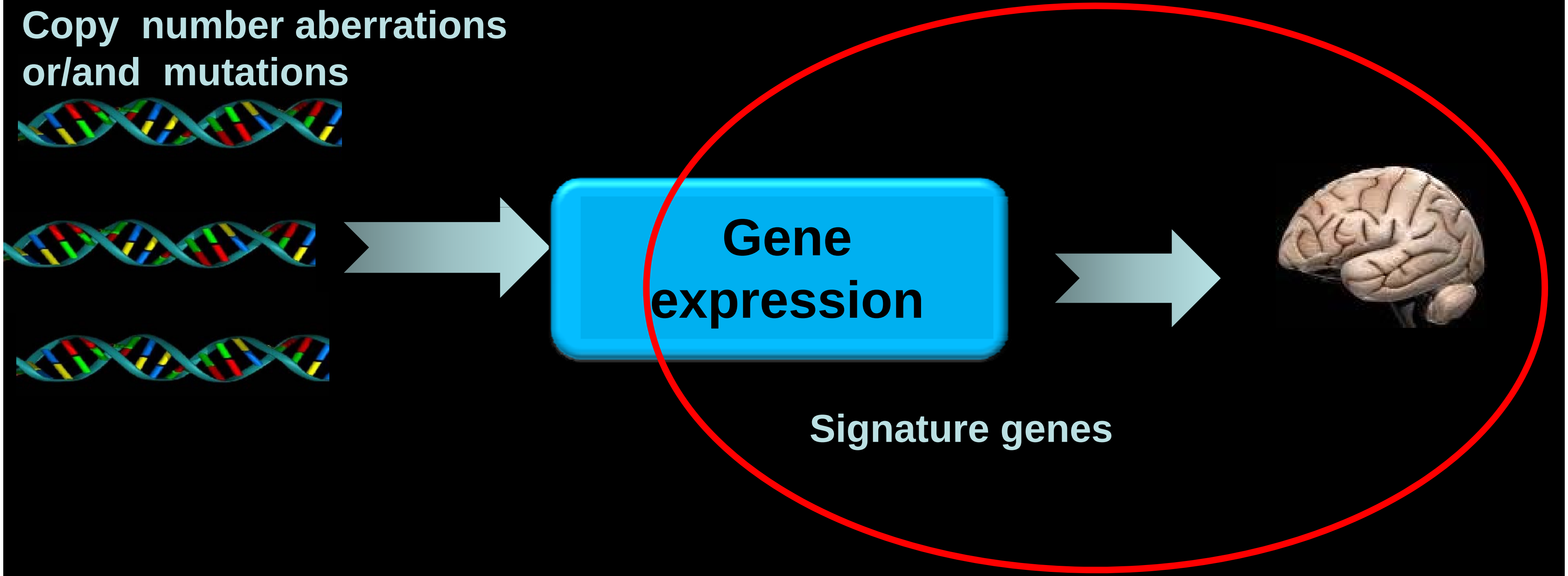
Cancer Cases
Gene expression data





Copy number aberrations
or/and mutations

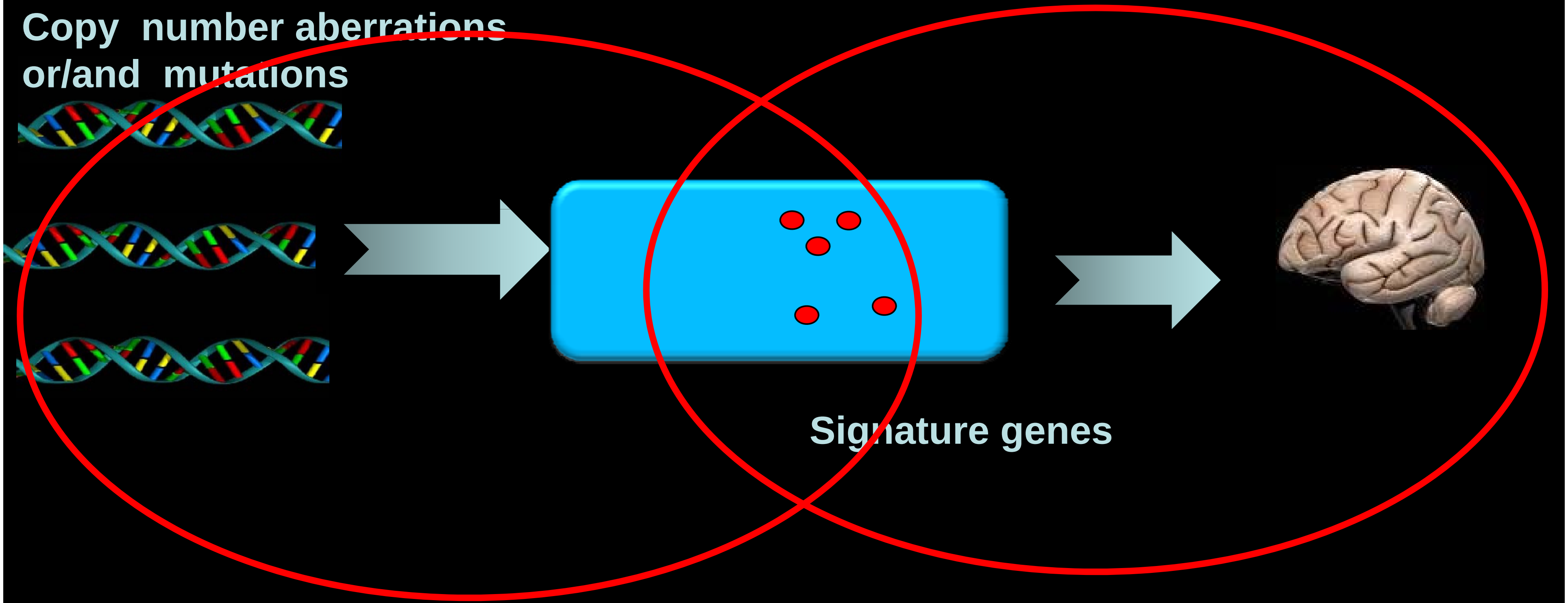
Gene expression



Copy number aberrations
or/and mutations

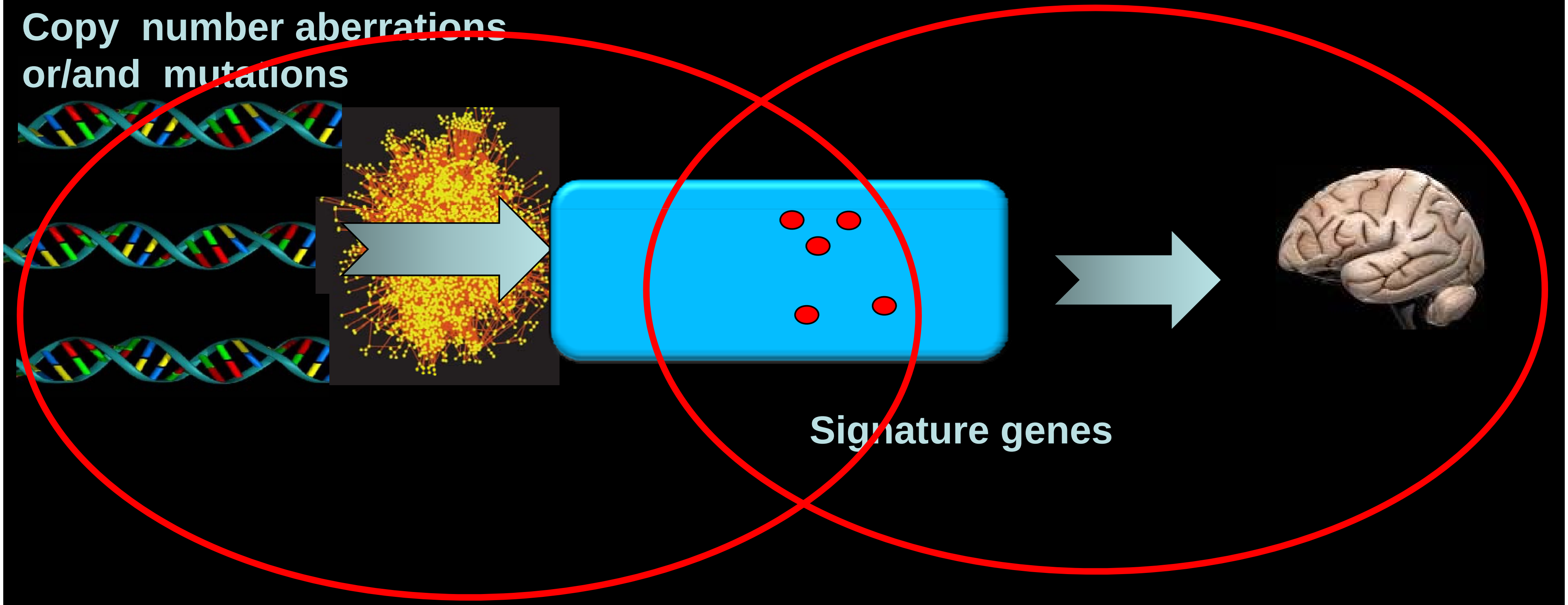
Gene expression

Signature genes



Copy number aberrations
or/and mutations

Signature genes



Copy number aberrations
or/and mutations

Signature genes

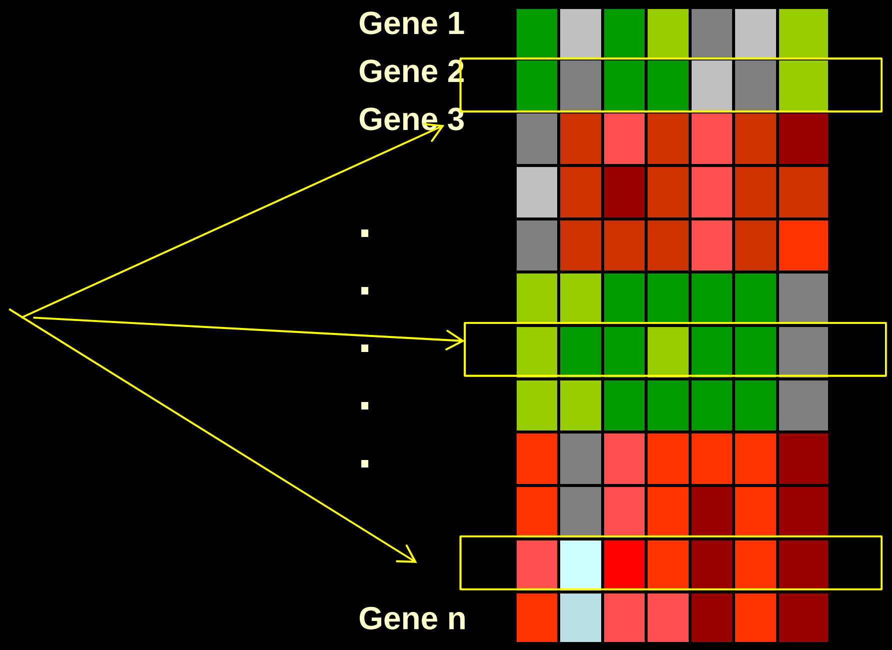
Method outline

1. Selecting marker genes to be used as “phenotype”
2. Genotype-phenotype association
3. Uncovering information flow between genotype and phenotype
4. Inferring dys- regulated, genes, pathways, and causal mutations

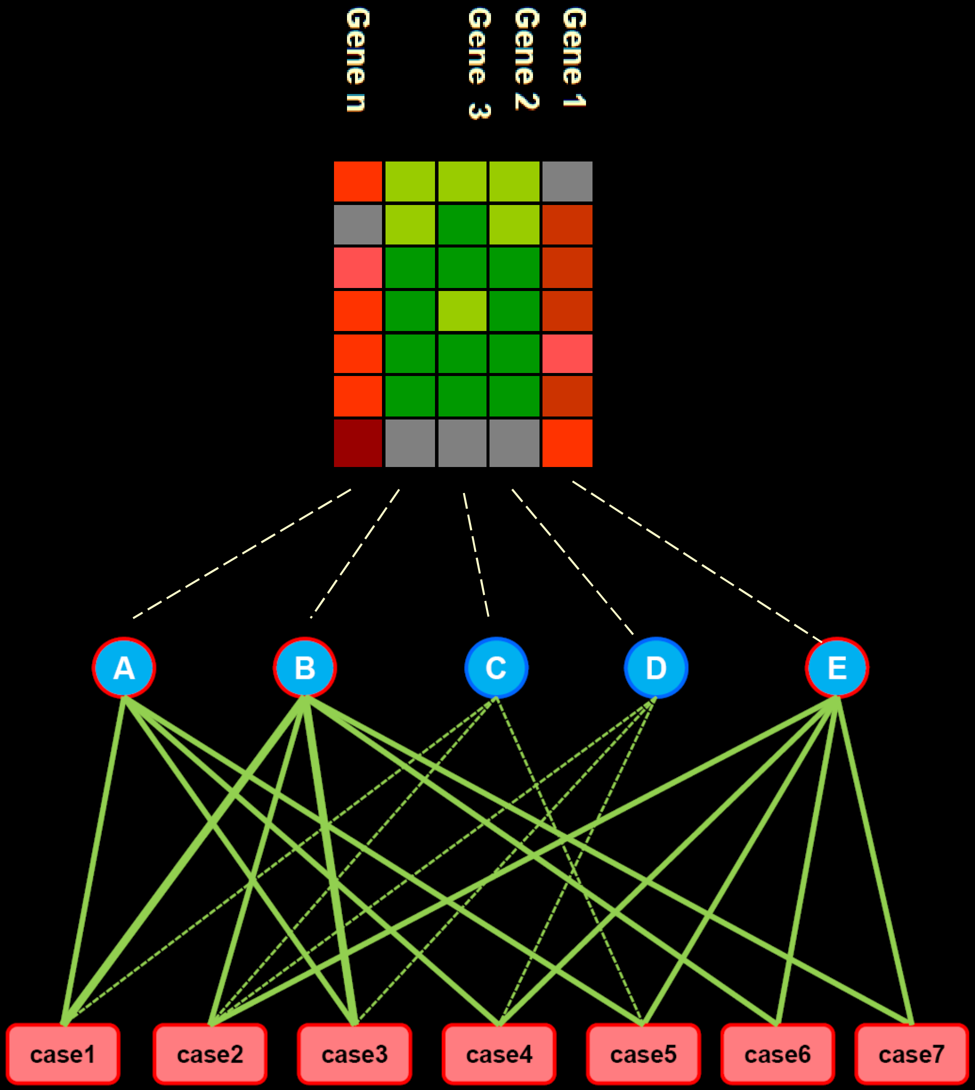
Selecting “phenotype” genes

Cancer Cases
Gene expression data

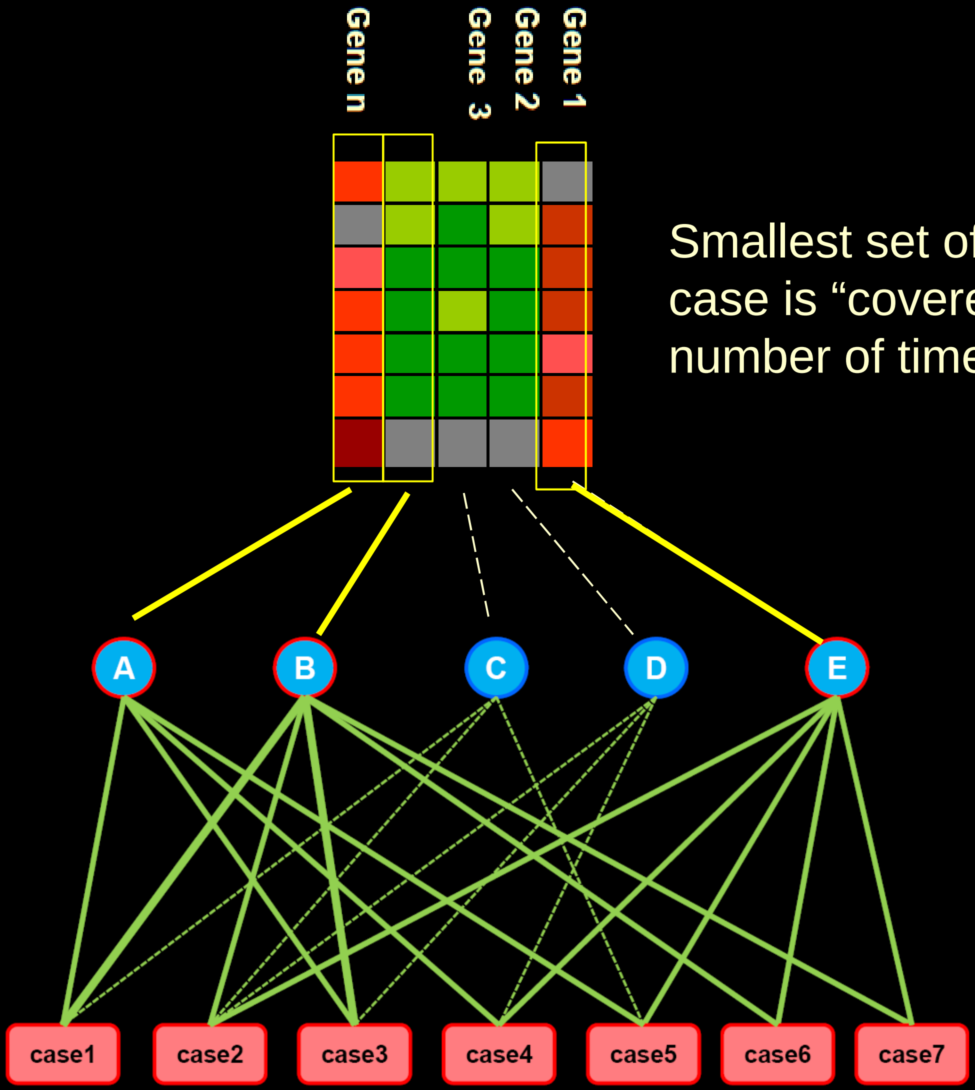
target genes



Selecting “phenotype” genes

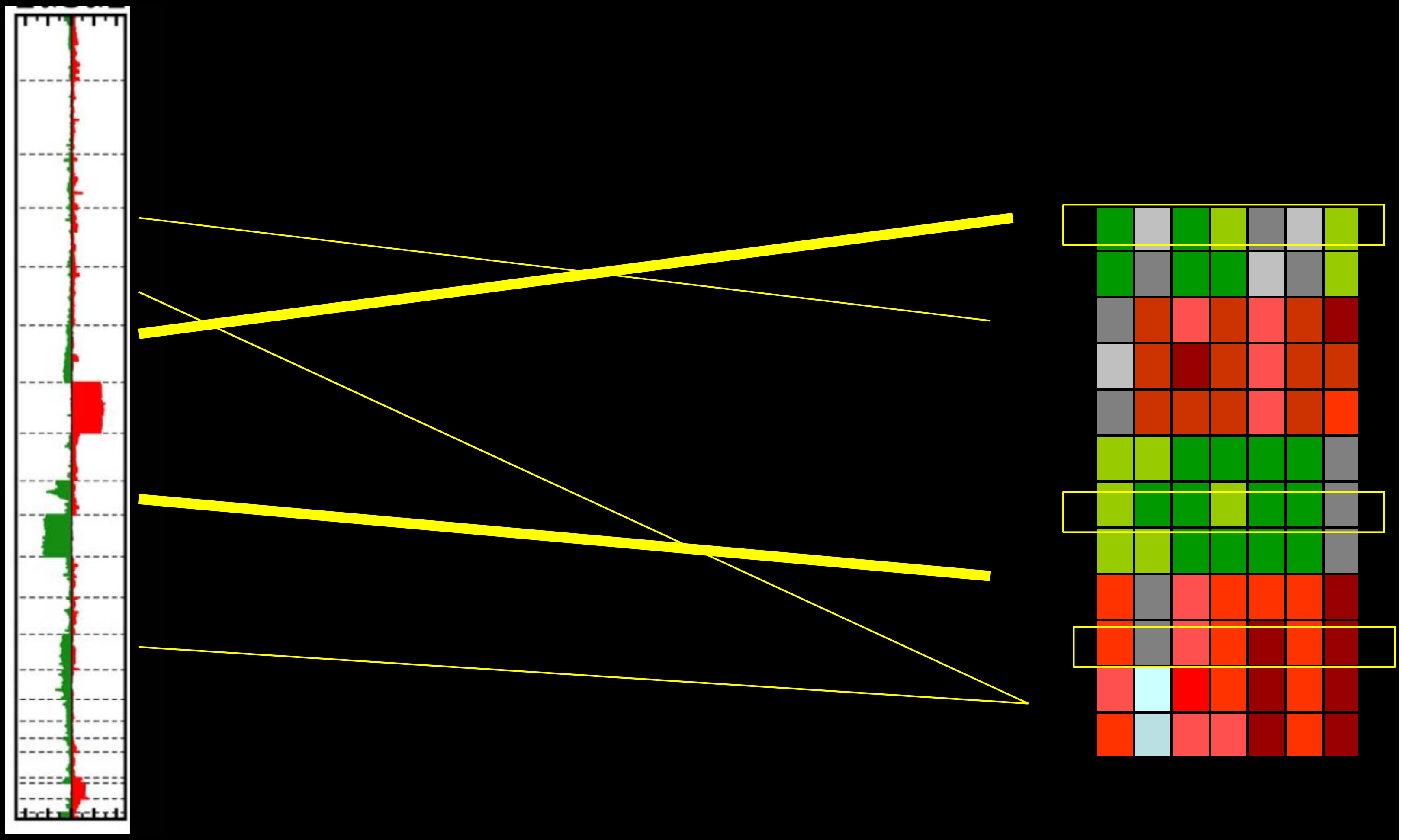


Selecting "phenotype" genes



Smallest set of genes so that each case is "covered" at least specified number of times

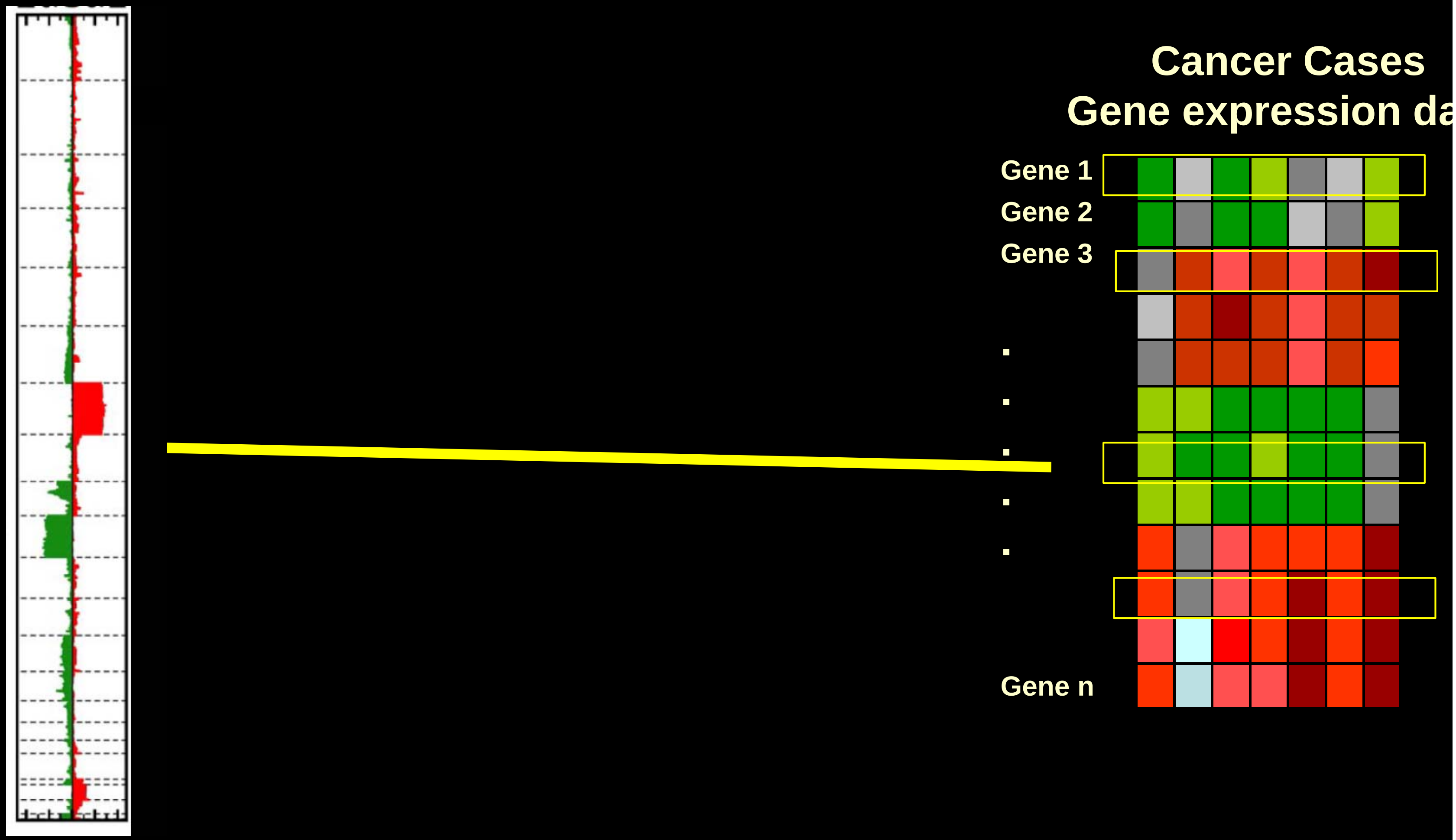
Associations between copy number variations and gene expression of selected target genes



Cancer Cases
CNV data

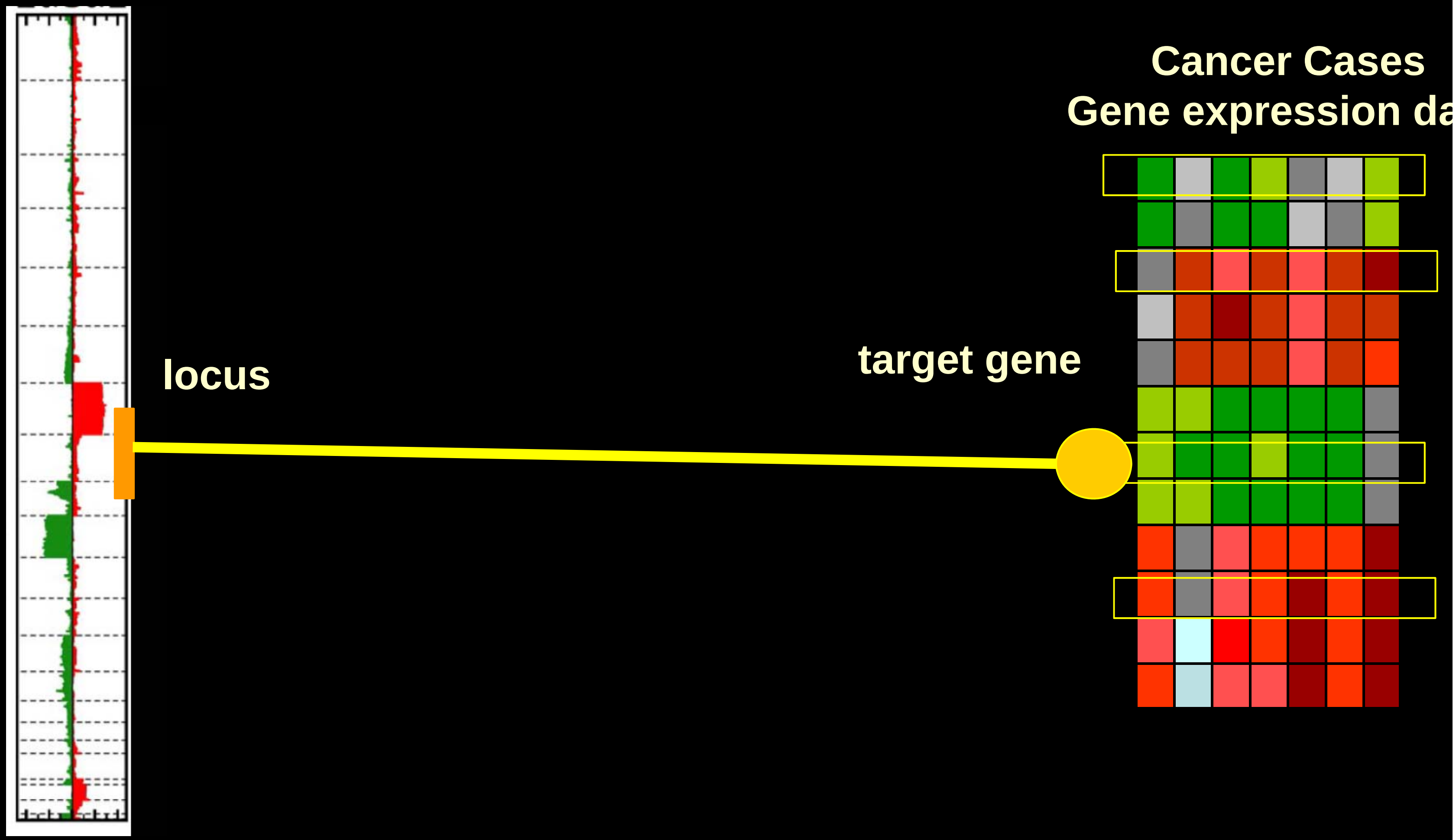
Cancer Cases
Gene expression data

Significant correlation between CNV and expression



Cancer Cases
Gene expression da

Significant correlation between CNV and expression



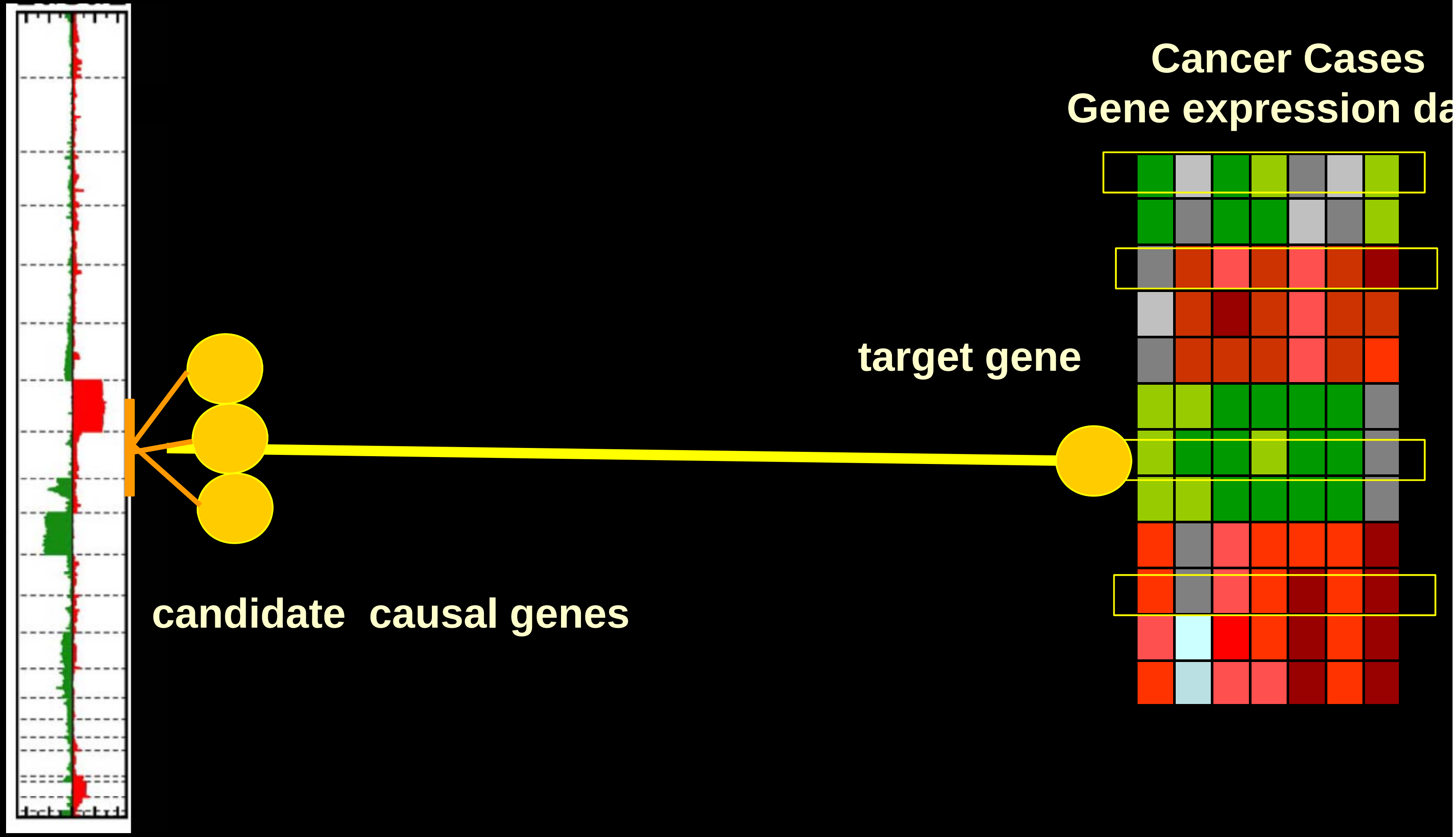
Cancer Cases

Gene expression da

locus

target gene

Significant correlation between CNV and expression



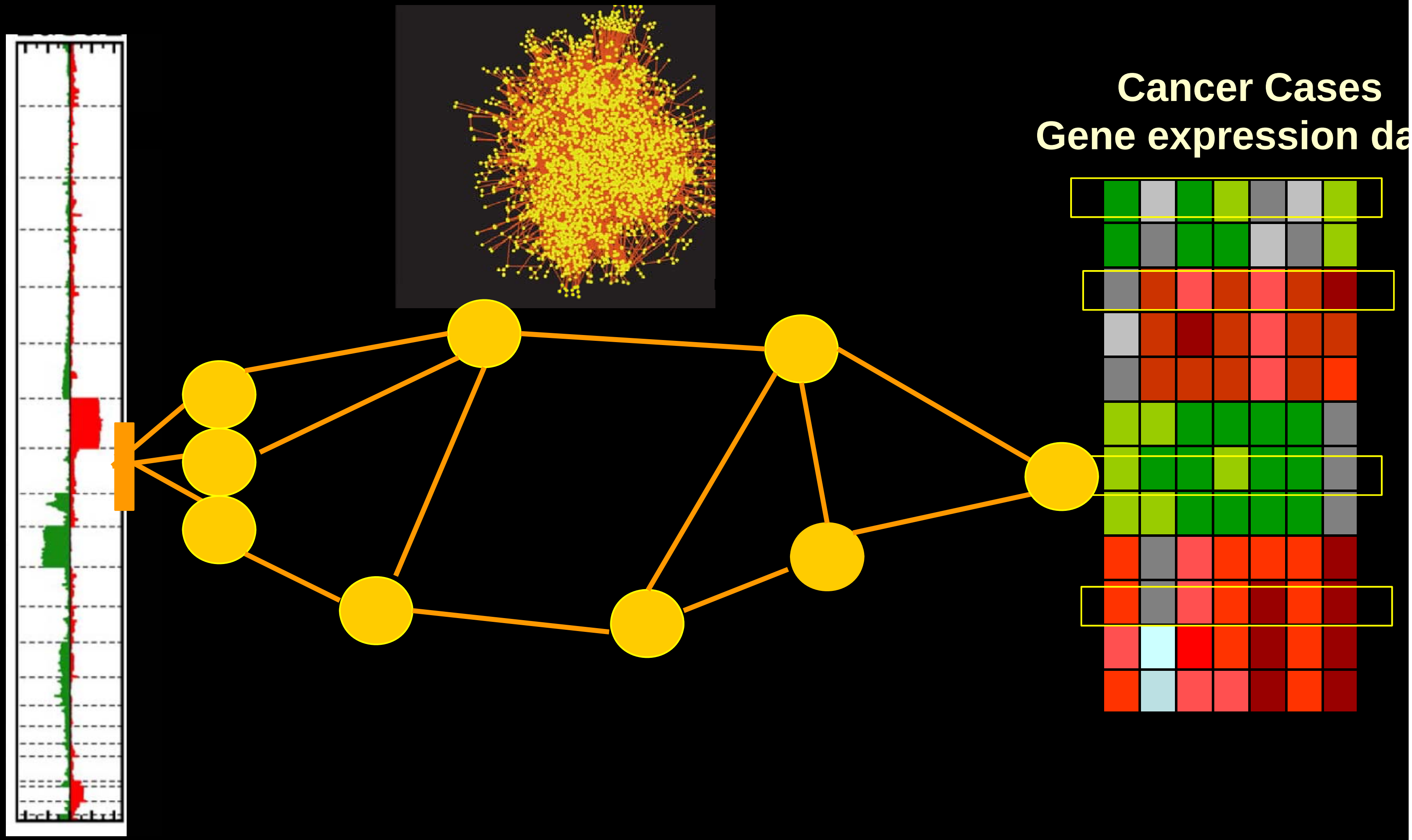
Cancer Cases

Gene expression da

target gene

candidate causal genes

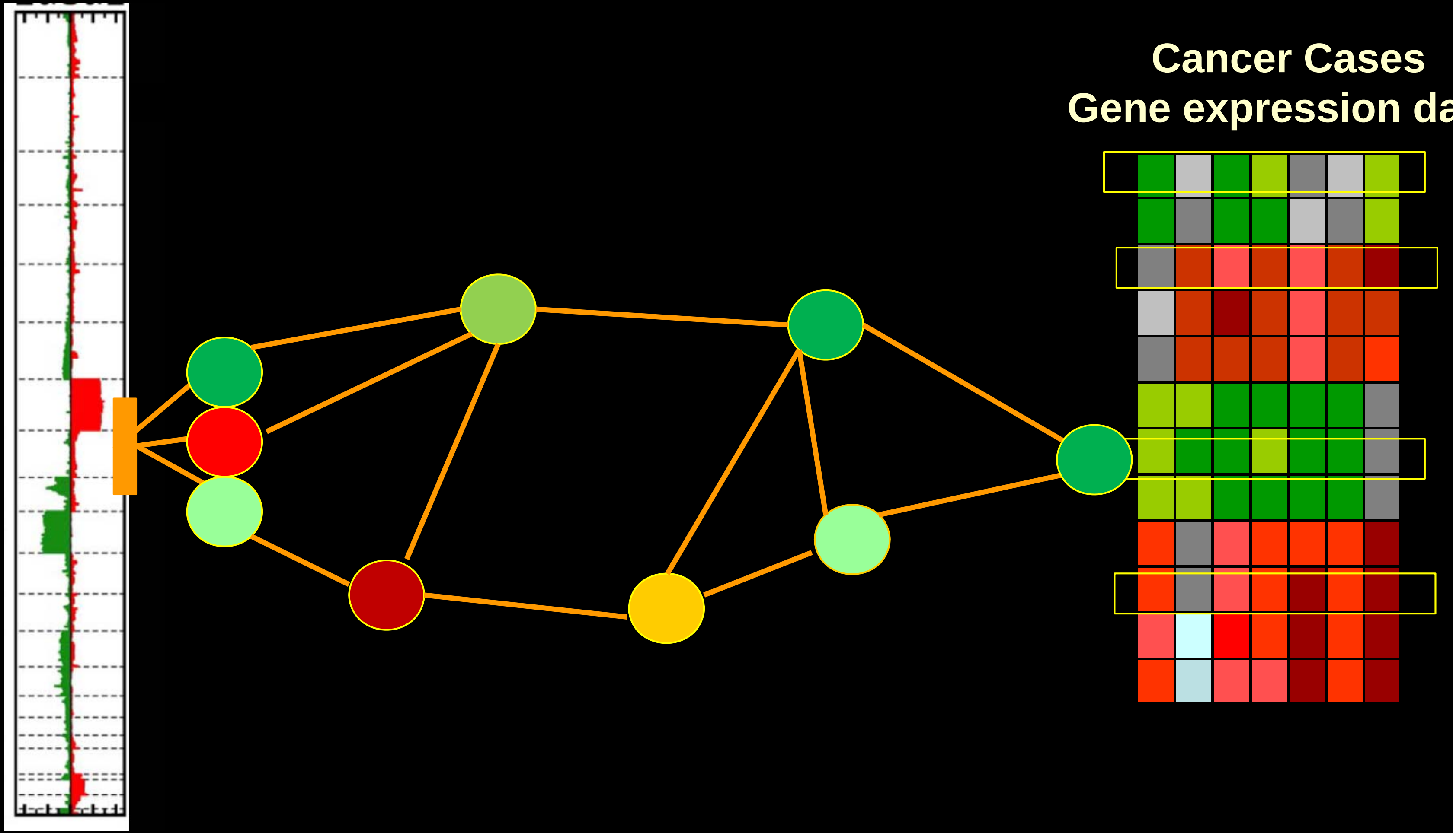
Uncovering pathways of information flow between CNV and target gene



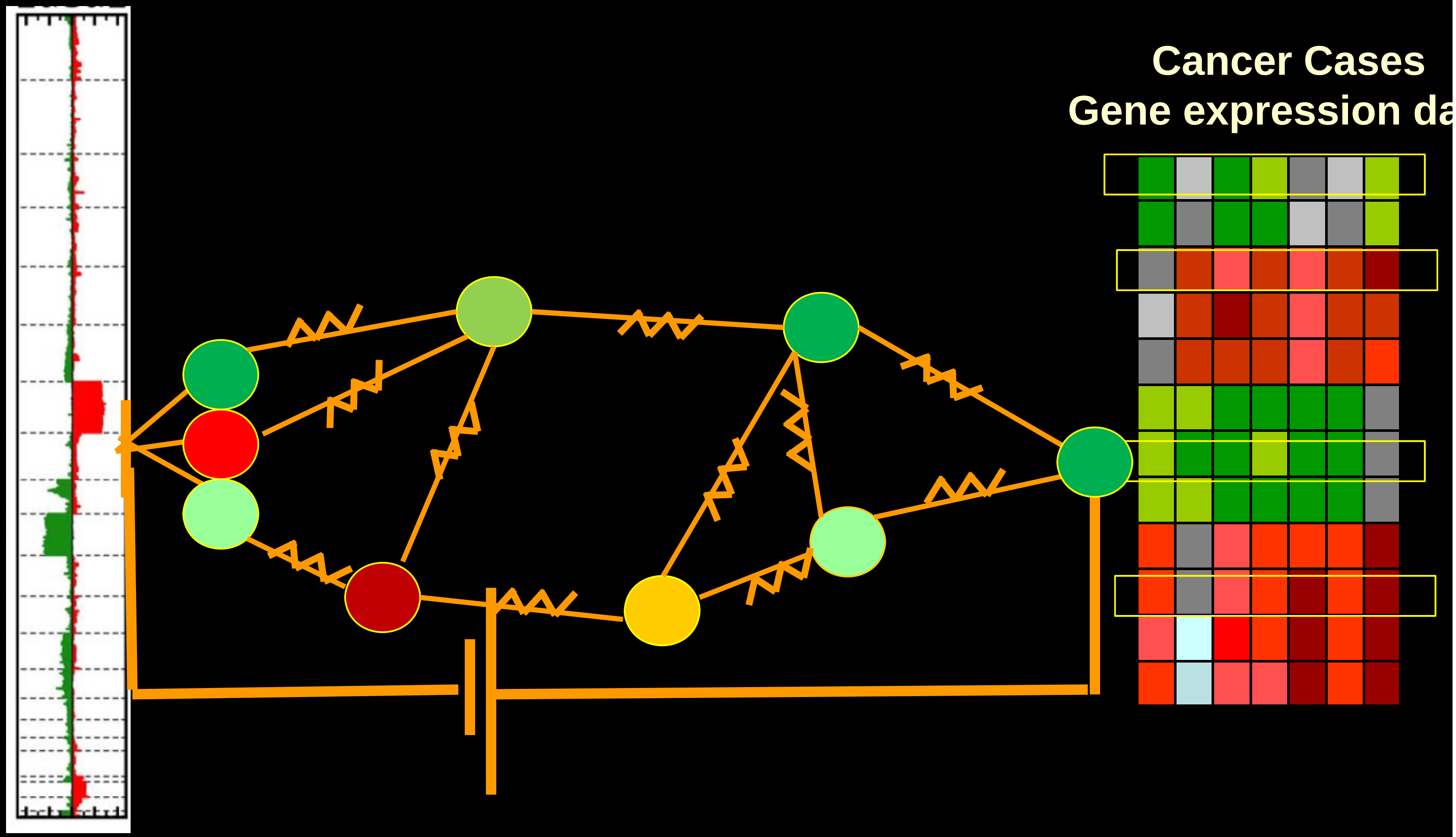
Cancer Cases

Gene expression da

Using expression to guide path discovery

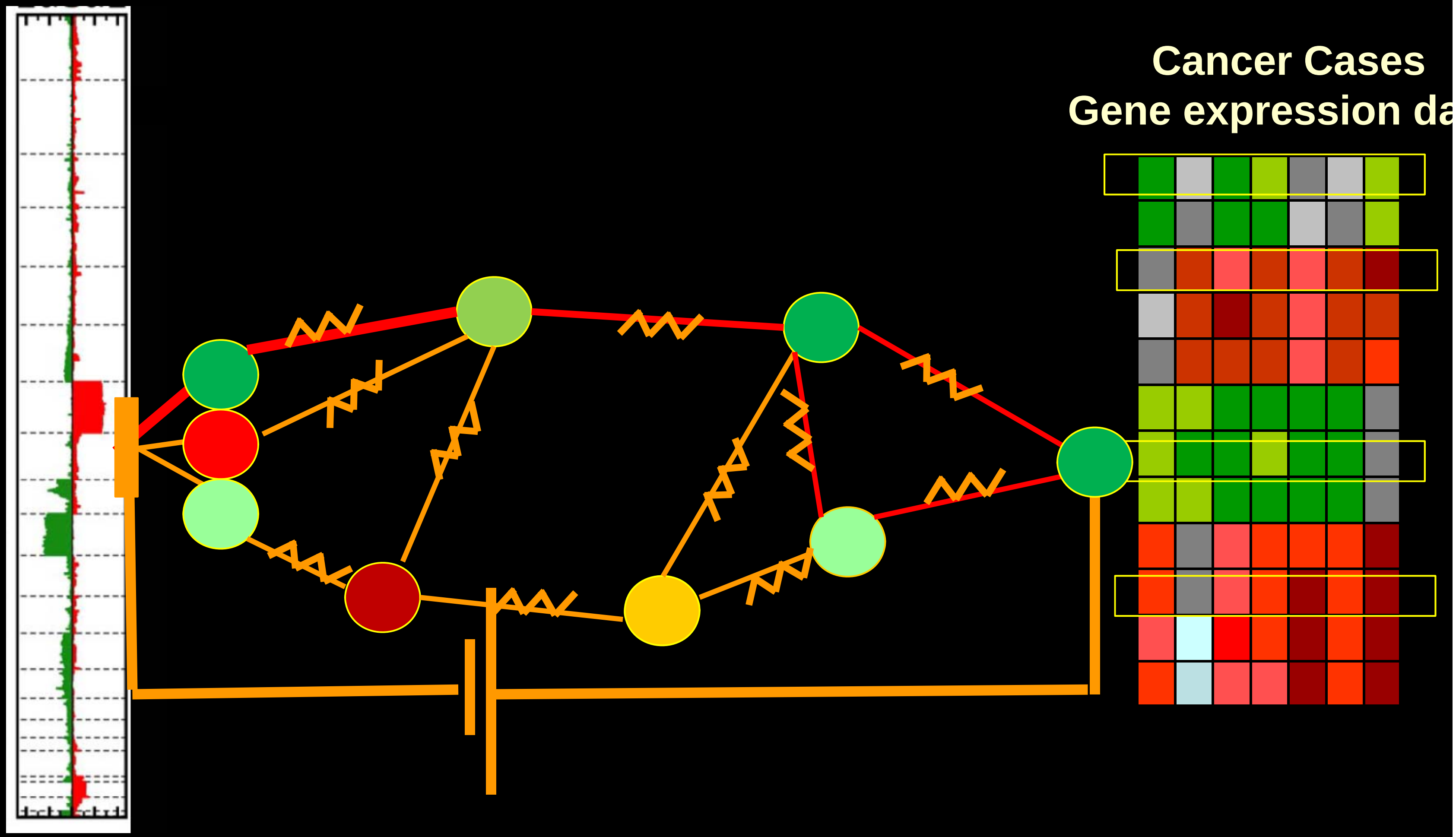


Translating probabilities it resistances



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



Cancer Cases
Gene expression da

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

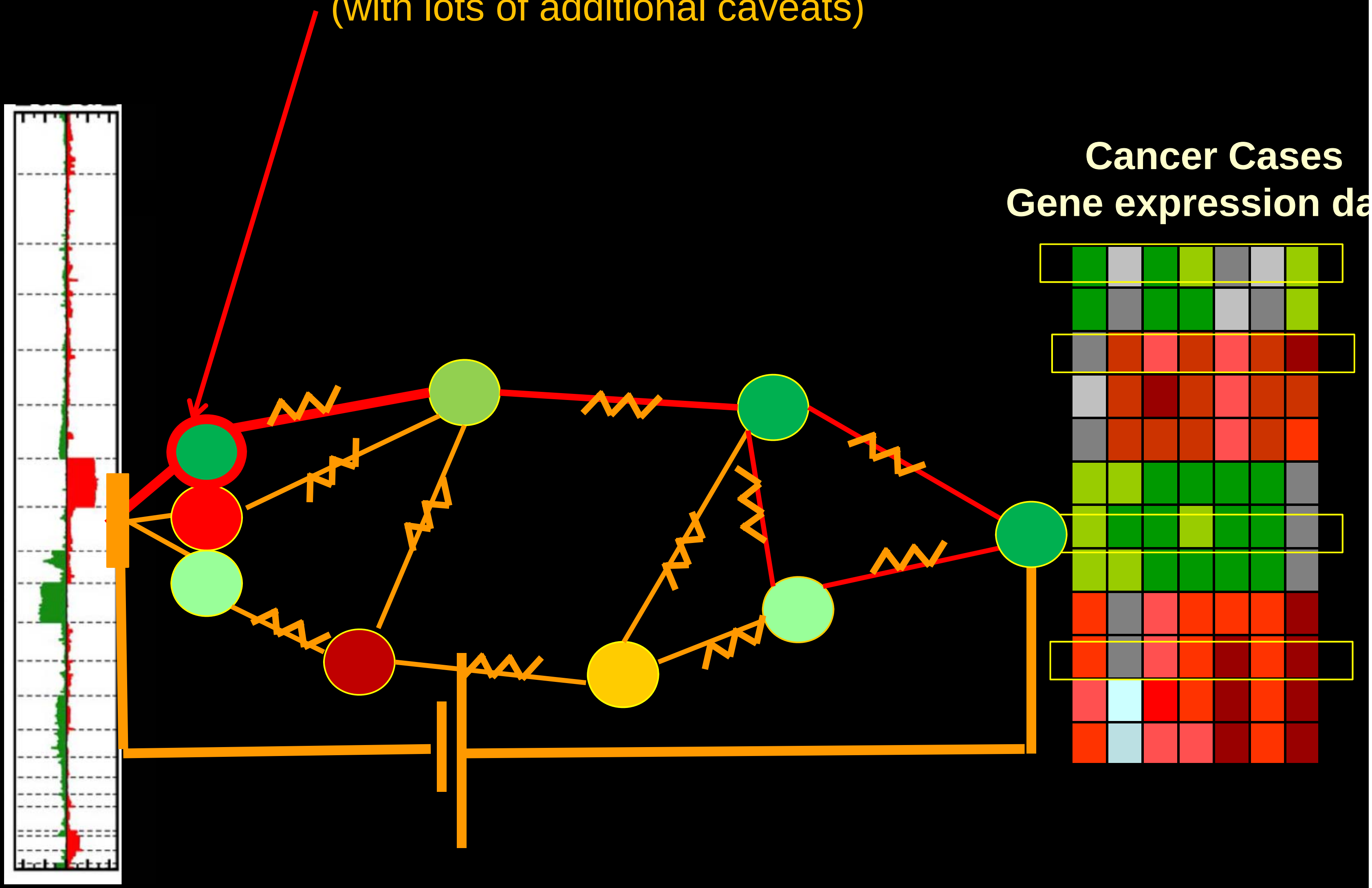
(reversely proportional to the average correlation of the expression of the adjacent genes with expression of the target gene)

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

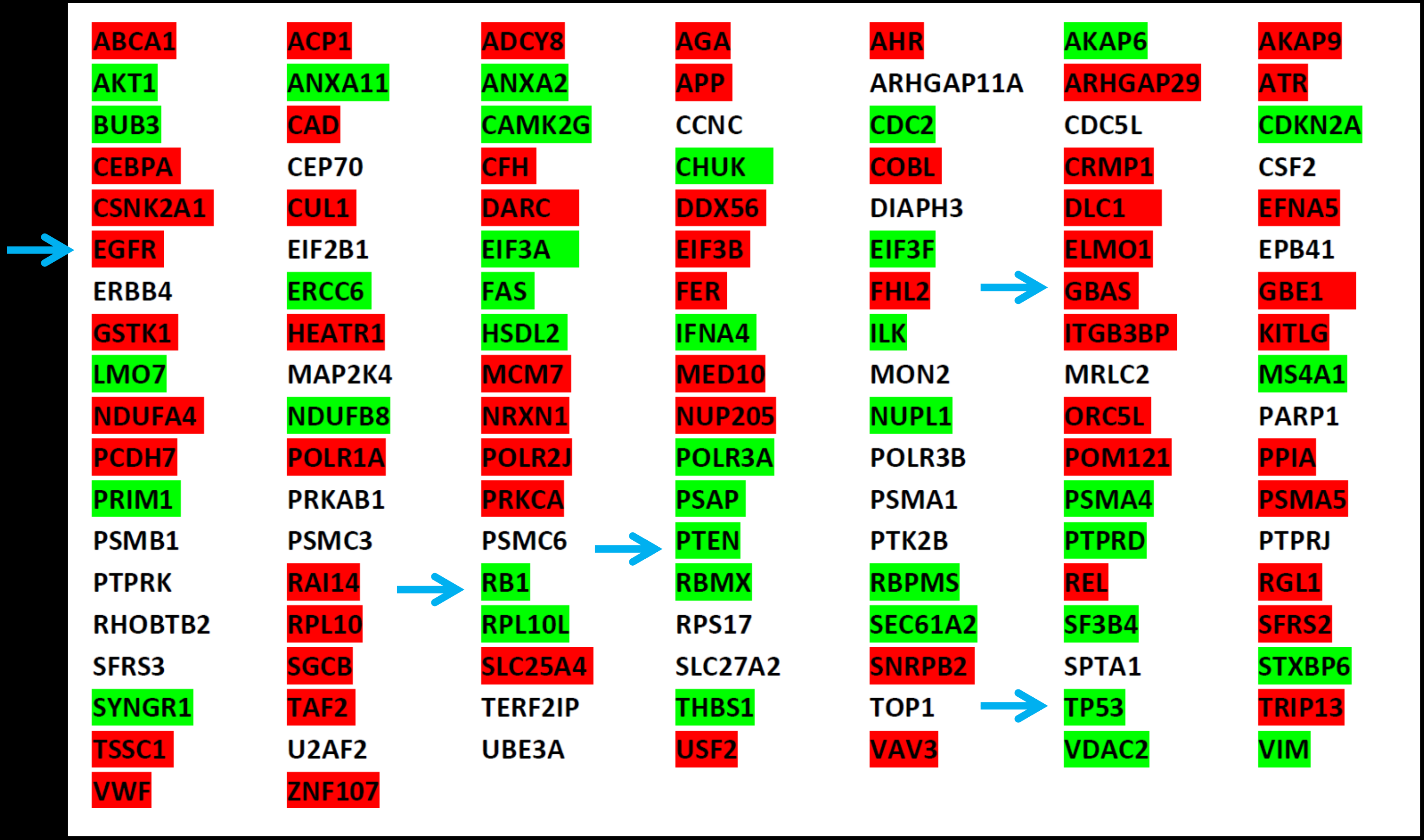
(with lots of additional caveats)



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)

Causal copy number aberrations



ABCA1
AKT1
BUB3
CEBPA
CSNK2A1
EGFR
ERBB4
GSTK1
LMO7
NDUFA4
PCDH7
PRIM1
PSMB1
PTPRK
RHOBTB2
SFRS3
SYNGR1
TSSC1
VWF

ACP1
ANXA11
CAD
CEP70
CUL1
EIF2B1
ERCC6
HEATR1
MAP2K4
NDUFB8
POLR1A
PRKAB1
PSMC3
RAI14
RPL10
SGCB
TAF2
U2AF2
ZNF107

ADCY8
ANXA2
CAMK2G
CFH
DARC
EIF3A
FAS
HSDL2
MCM7
NRXN1
POLR2J
PRKCA
PSMC6
RB1
RPL10L
SLC25A4
TERF2IP
UBE3A

AGA
APP
CCNC
CHUK
DDX56
EIF3B
FER
IFNA4
MED10
NUP205
POLR3A
PSAP
PTEN
RBMX
RPS17
SLC27A2
THBS1
USF2

AHR
ARHGAP11A
CDC2
COBL
DIAPH3
EIF3F
FHL2
ILK
MON2
NUPL1
POLR3B
PSMA1
PTK2B
RBPMS
SEC61A2
SNRPB2
TOP1
VAV3

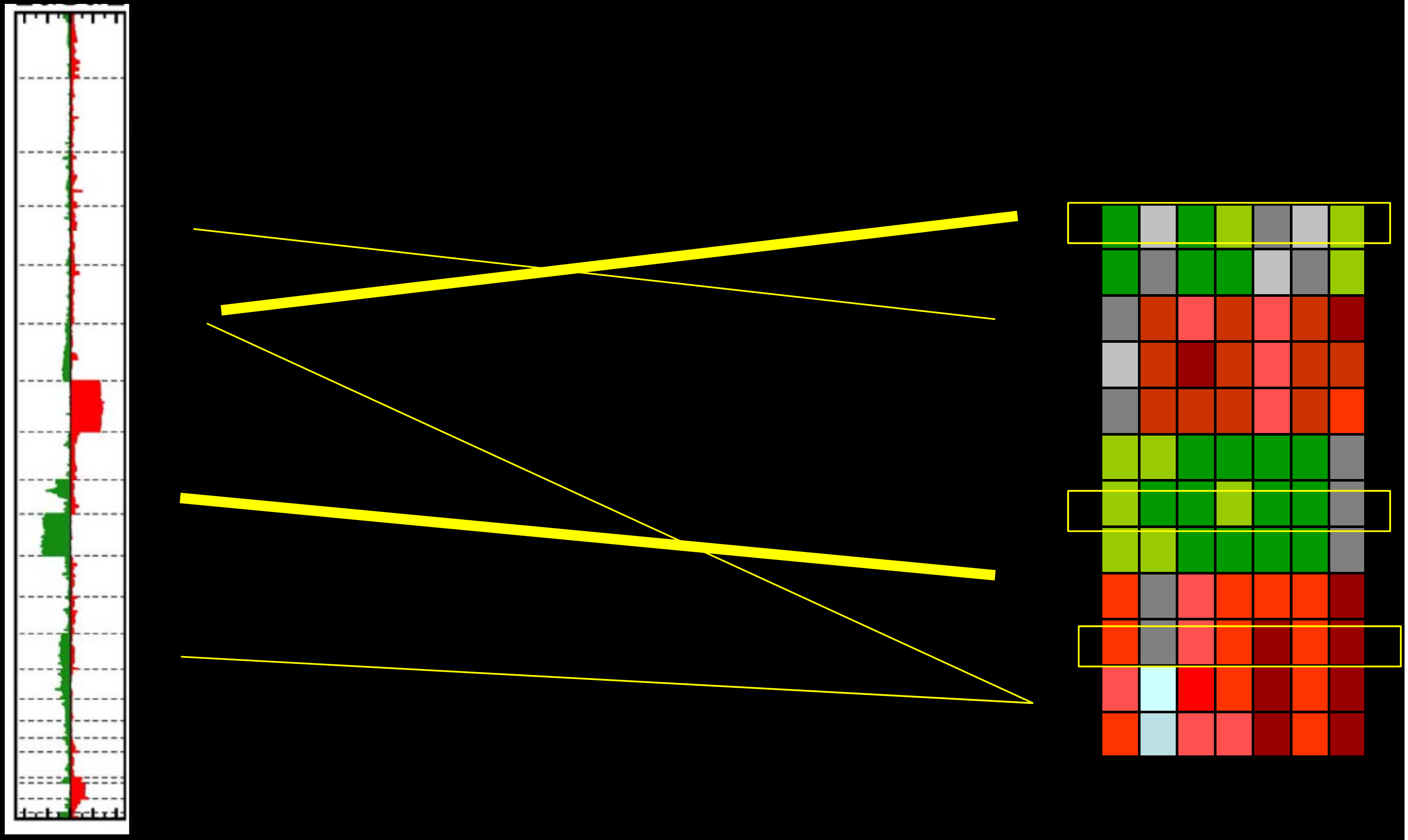
AKAP6
ARHGAP29
CDC5L
CRMP1
DLC1
ELMO1
GBAS
ITGB3BP
MRLC2
ORC5L
POM121
PSMA4
PTPRD
REL
SF3B4
SPTA1
TP53
VDAC2

AKAP9
ATR
CDKN2A
CSF2
EFNA5
EPB41
GBE1
KITLG
MS4A1
PARP1
PPIA
PSMA5
PTPRJ
RGL1
SFRS2
STXBP6
TRIP13
VIM

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

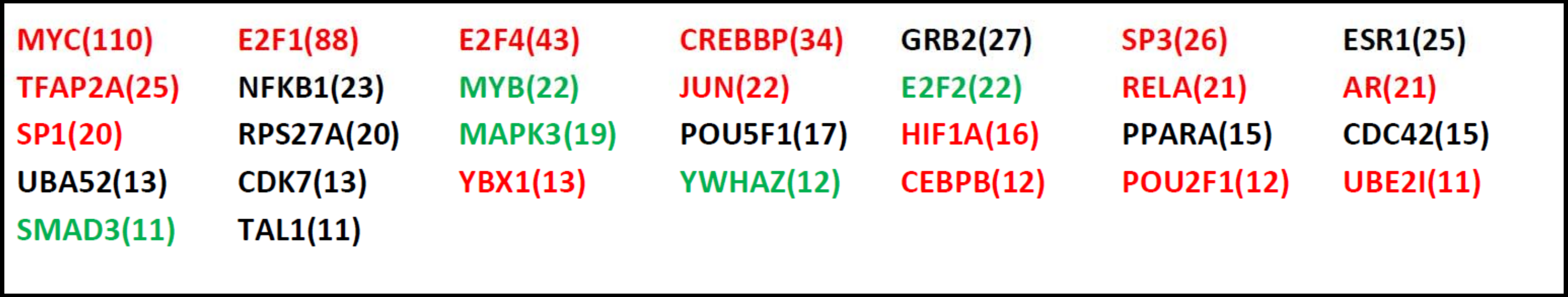
Solve current flow for all pairs and find nodes belonging to many paths



Cancer Cases
CNV data

Cancer Cases
Gene expression data

Hubs



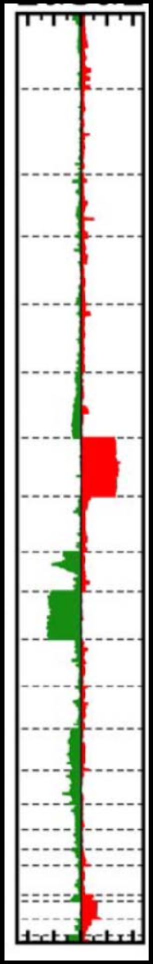
MYC(110)
E2F1(88)
E2F4(43)
CREBBP(34)
GRB2(27)
SP3(26)
ESR1(25)
TFAP2A(25)
NFKB1(23)
MYB(22)
JUN(22)
E2F2(22)
RELA(21)
AR(21)
SP1(20)
RPS27A(20)
MAPK3(19)
POU5F1(17)
HIF1A(16)
PPARA(15)
CDC42(15)
UBA52(13)
CDK7(13)
YBX1(13)
YWHAZ(12)
CEBPB(12)
POU2F1(12)
UBE2I(11)
SMAD3(11)
TAL1(11)

Goals :

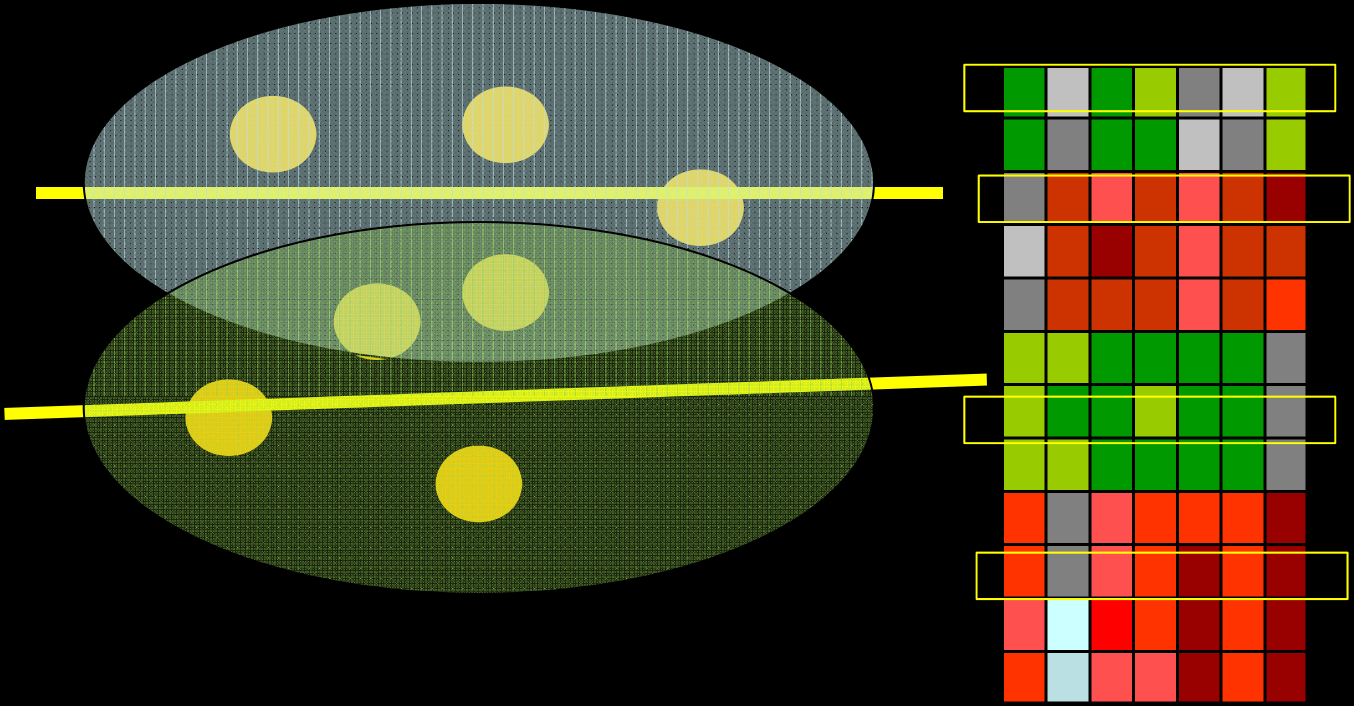
- ✓ A method for system level analysis of propagation of such perturbation in the network
- ✓ Prediction of “causal” mutations
- ✓ Prediction of “master regulators” (network hubs) involved in disease
- Prediction of pathways dys-regulated in disease

Are there common functional pathways?

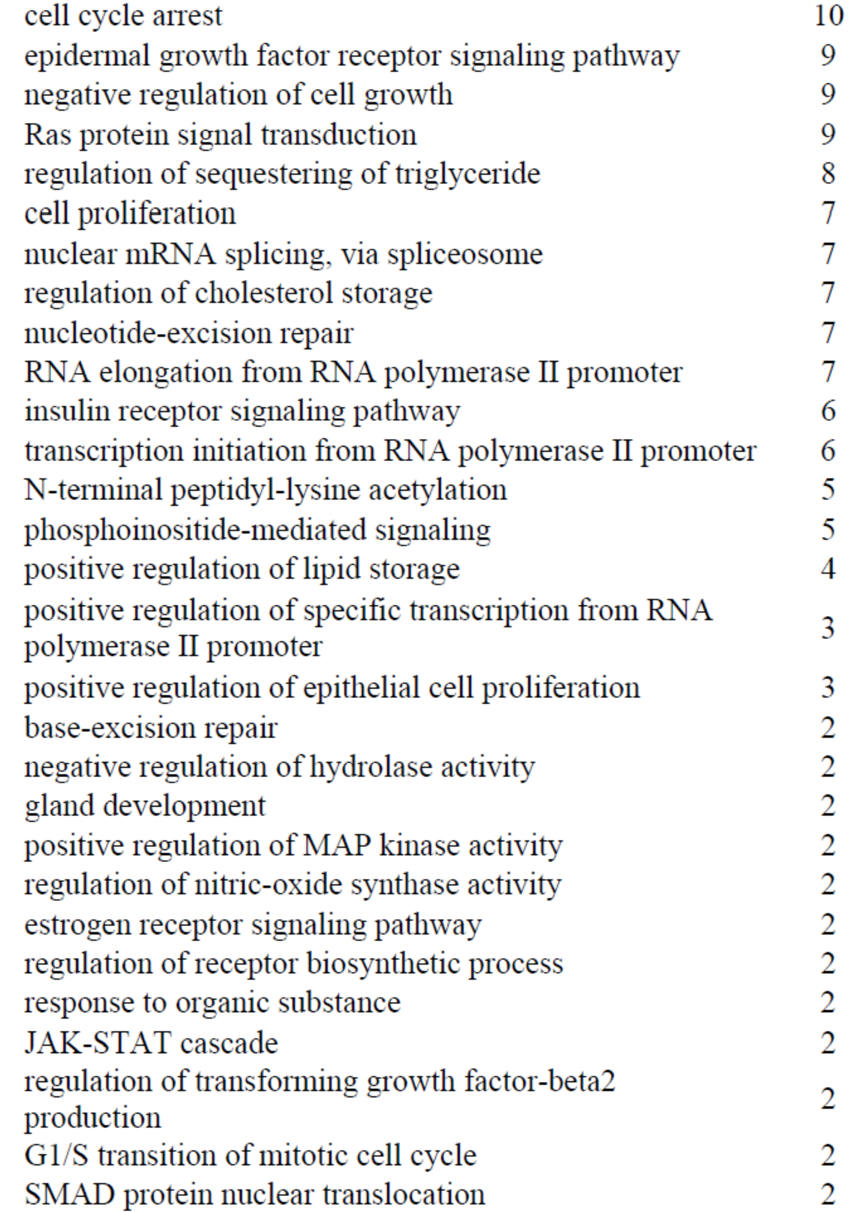
Cancer Cases
CNV data



Cancer Cases
Gene expression dat



Common GO pathways



cell cycle arrest	10
epidermal growth factor receptor signaling pathway	9
negative regulation of cell growth	9
Ras protein signal transduction	9
regulation of sequestering of triglyceride	8
cell proliferation	7
nuclear mRNA splicing, via spliceosome	7
regulation of cholesterol storage	7
nucleotide-excision repair	7
RNA elongation from RNA polymerase II promoter	7
insulin receptor signaling pathway	6
transcription initiation from RNA polymerase II promoter	6
N-terminal peptidyl-lysine acetylation	5
phosphoinositide-mediated signaling	5
positive regulation of lipid storage	4
positive regulation of specific transcription from RNA polymerase II promoter	3
positive regulation of epithelial cell proliferation	3
base-excision repair	2
negative regulation of hydrolase activity	2
gland development	2
positive regulation of MAP kinase activity	2
regulation of nitric-oxide synthase activity	2
estrogen receptor signaling pathway	2
regulation of receptor biosynthetic process	2
response to organic substance	2
JAK-STAT cascade	2
regulation of transforming growth factor-beta2 production	2
G1/S transition of mitotic cell cycle	2
SMAD protein nuclear translocation	2

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

Design details under the hood

- Current flow reduces to solving a set of linear equations (Kirchhoff's laws)

  Caveat: We had to solving a linear system with 20,000 variables thousands of times for permutation test required new methodology

- Many biological interactions are directional. This can be taken care by solving linear program with corresponding constraints - Caveat: the network is to big for solving thousands of linear programs

- Null model and p-value estimations

Kim, Wuchty, Przytycka – *PloS Comp Bio 2011*
Kim, Przytycki, Wuchty, Przytycka – *Phys. Bio.* 2011

Acknowledgments

Group members:

Yoo-Ah Kim

DongYeon Cho

Xiangjun Du

Jan Hoinka

Yang Huang

Raheleh Salari

Damian Wojtowicz

Collaborators:

Stefan Wuchty (NCBI)

Jozef Przytycki (GWU)

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

my great-great uncle (the "Giant")

[redacted]

Acknowledgments

Group members:

Yoo-Ah Kim

DongYeon Cho

Xiangjun Du

Jan Hoinka

Yang Huang

Raheleh Salari

Damian Wojtowicz

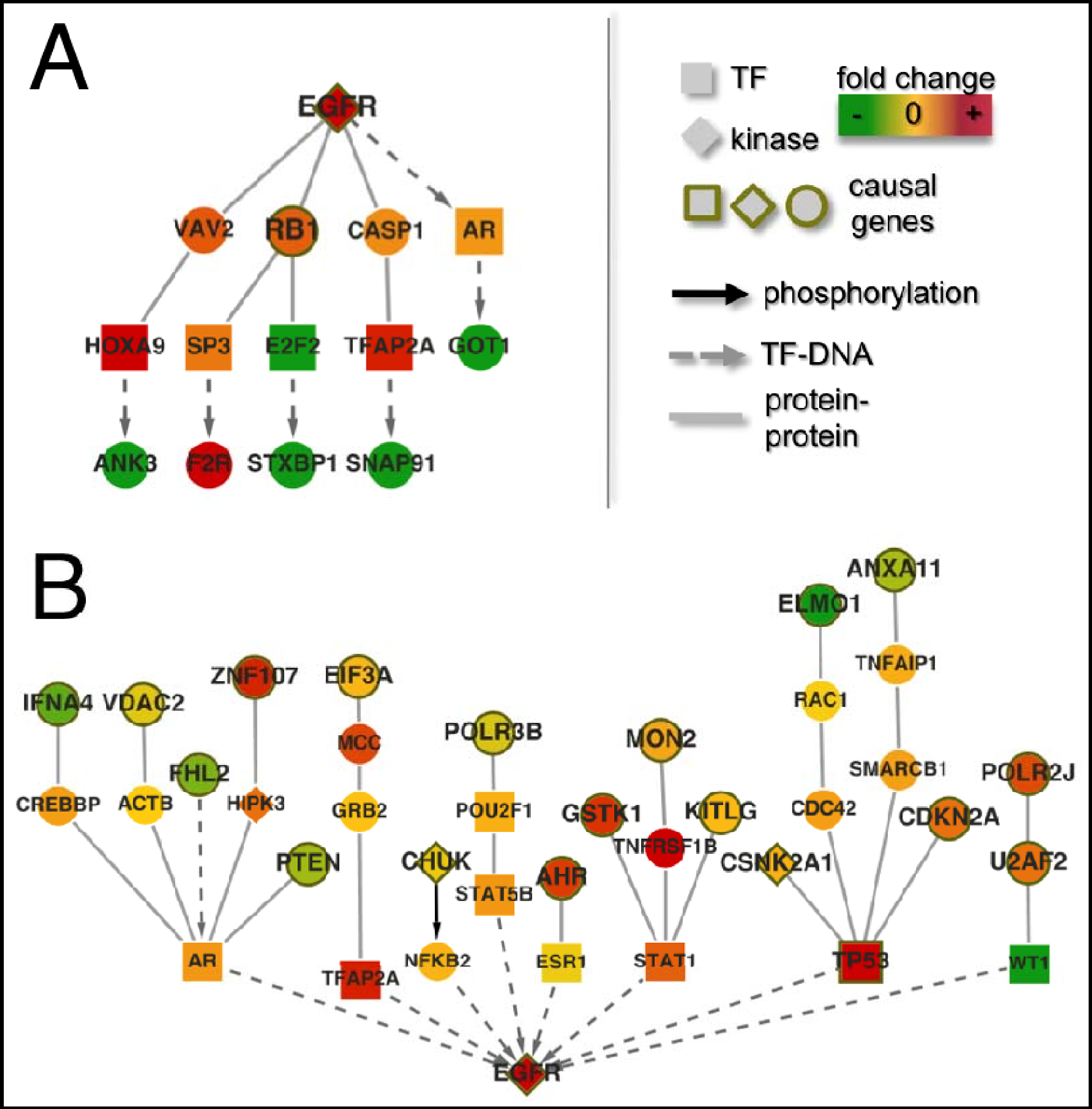
Collaborators:

Stefan Wuchty (NCBI)

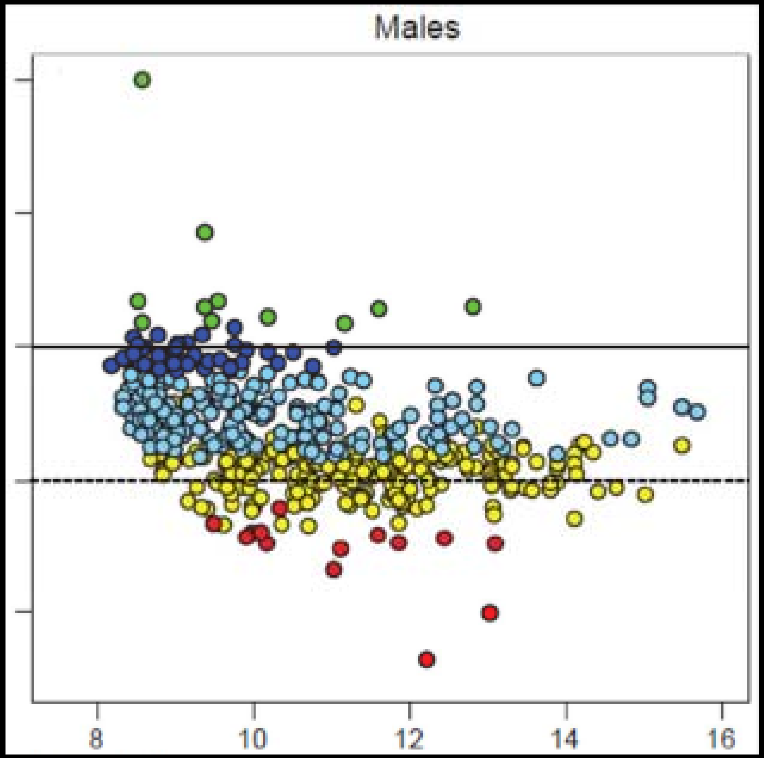
Brian Oliver (NIDDK)
John Malone
Nicolas Mattiuzzo

Justin Andrews (Indiana University)

Jozef Przytycki (GWU)


A
EGFR
VAV2
RB1
CASP1
AR
HOXA9
SP3
E2F2
TFAP2A
GOT1
ANK3
F2R
STXBP1
SNAP91
TF
kinase
fold change
-
0
+
causal genes
phosphorylation
TF-DNA
protein-protein
B
ANXA11
ELMO1
TNFAIP1
RAC1
ZNF107
EIF3A
IFNA4
VDAC2
MCC
POLR3B
MON2
SMARCB1
POLR2J
FHL2
CREBBP
ACTB
HIPK3
GRB2
POU2F1
GSTK1
KITLG
CDC42
CDKN2A
TNFRSF1B
PTEN
CHUK
AHR
CSNK2A1
U2AF2
STAT5B
AR
NFKB2
ESR1
STAT1
TP53
WT1
TFAP2A
EGFR

Impact of gene copy number on gene expression in Drosophila melanogaster



Expression fold change ($\log_2$)

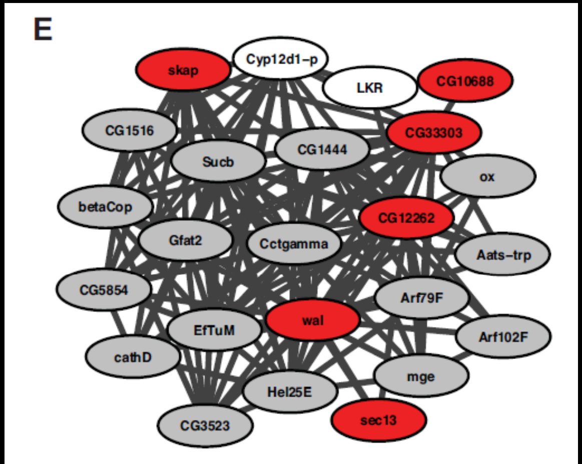
0

-1

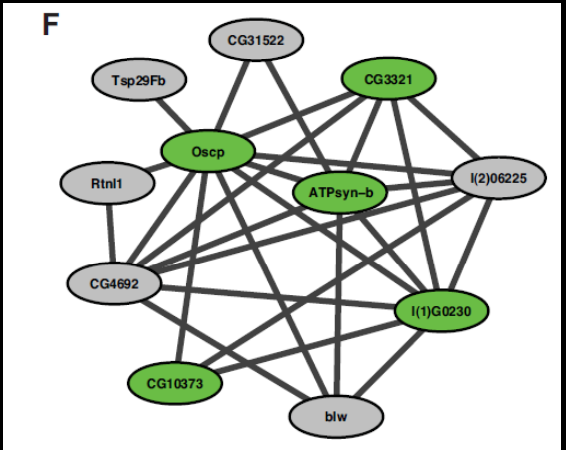
Expression (wild type)

collaboration with *Brian Oliver* group (NIDDK)

CNV-related perturbations propagate trough interaction network



DPiM#8 in *Df(2L)ED60/+* female

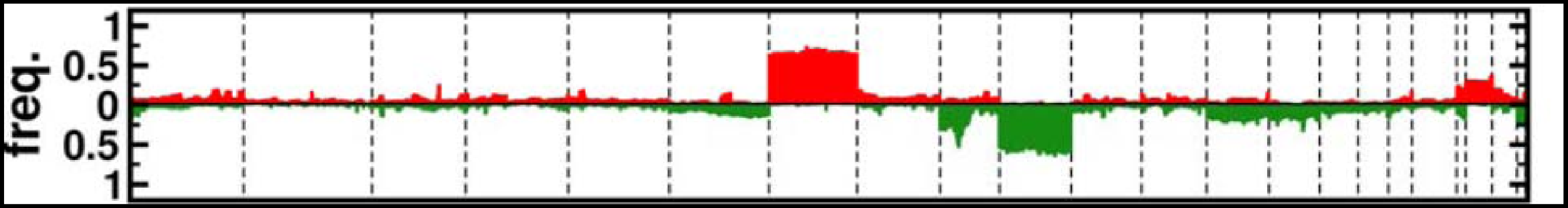


DPiM#25 in *Df(2L)ED2809/+* male

Co-complex network from Artavanis-Tsakonas group (unpublished)

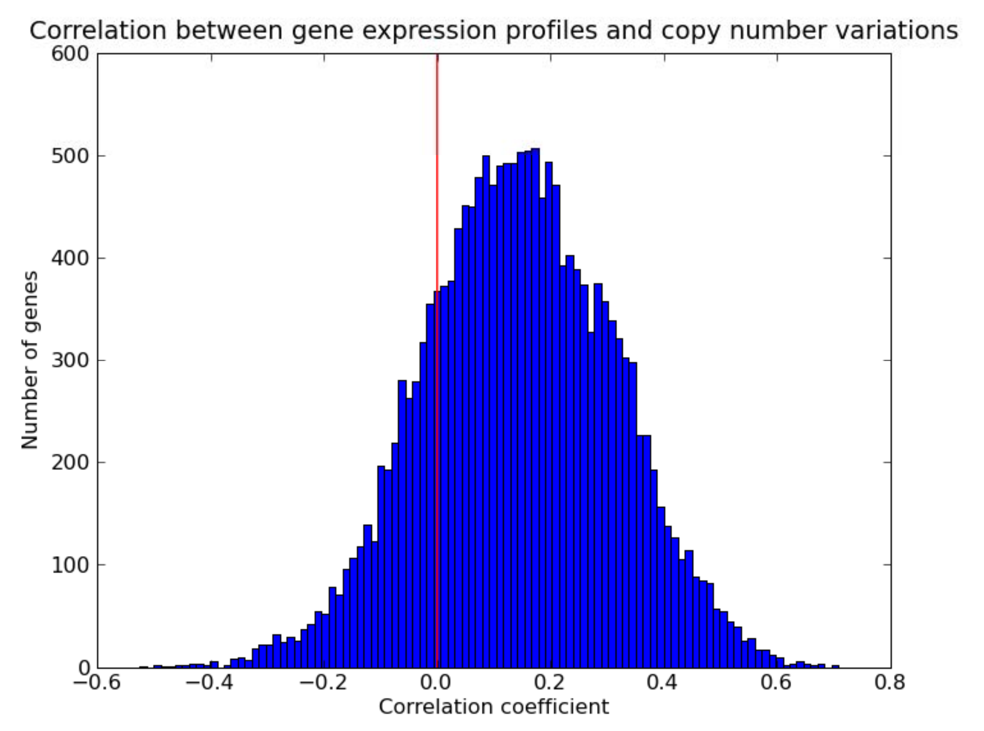
Impact on copy number on gene expression in glioma

CNV

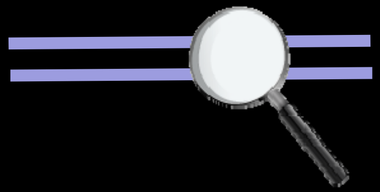


Chromosomes

Correlation between CNV and expression



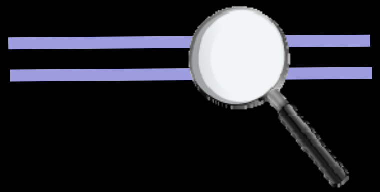
Genotype:



effects of genotypic variation:

- change in amino acid
- change in gene structure
- copy number variations ….

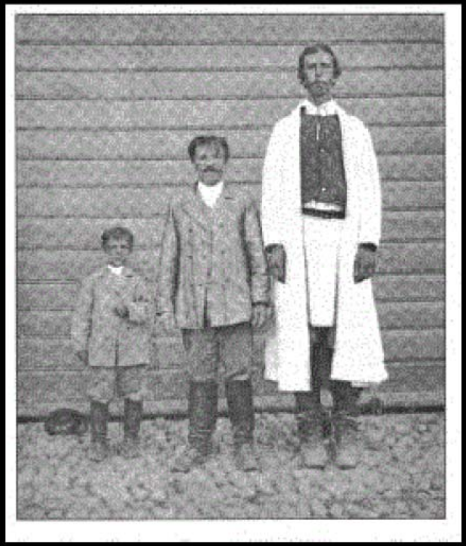
Genotype:



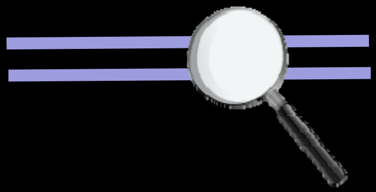
effects of genotypic variation:

- change in amino acid
- change in gene structure
- copy number variations ….

Phenotype



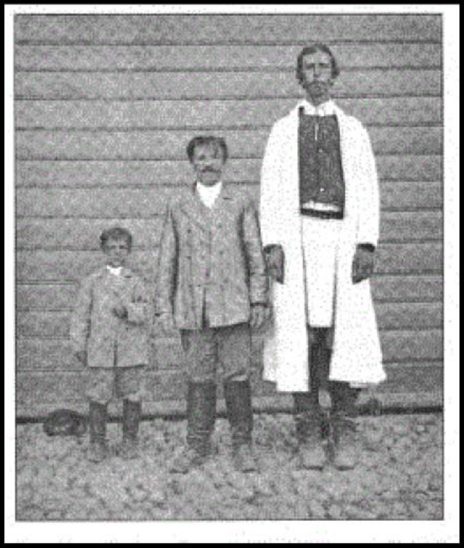
Genotype:



effects of genotypic variation:

- change in amino acid
- change in gene structure
- copy number variations ….

Phenotype



Molecular phenotypes

- gene expression
- Metabolite level

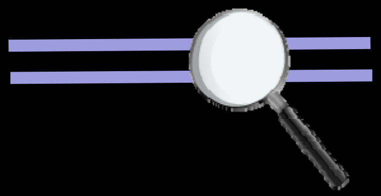
Copy number variations (CNV)
(gene dosage)

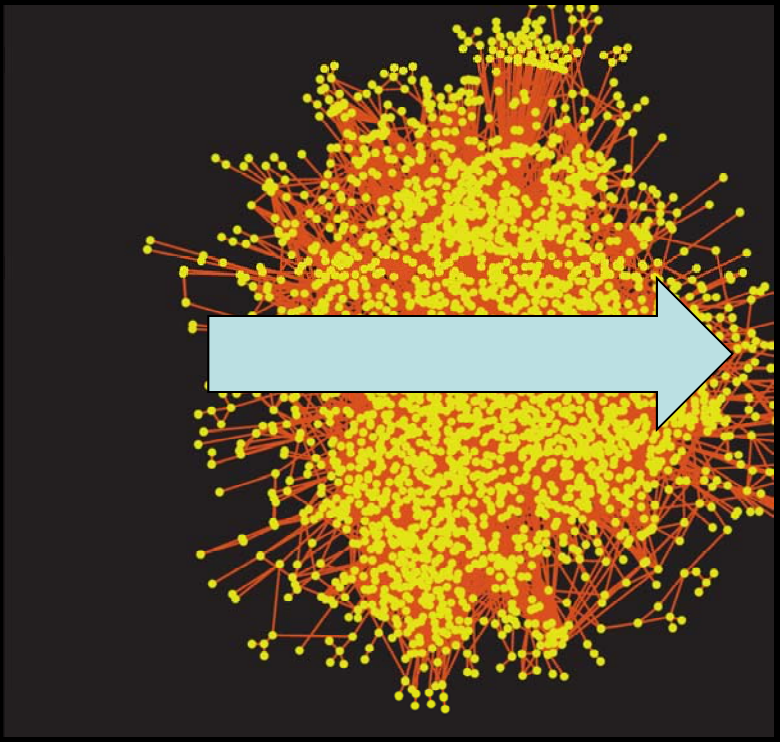
- implicated in large number of human diseases (cancer, Crohn's disease, autism)

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

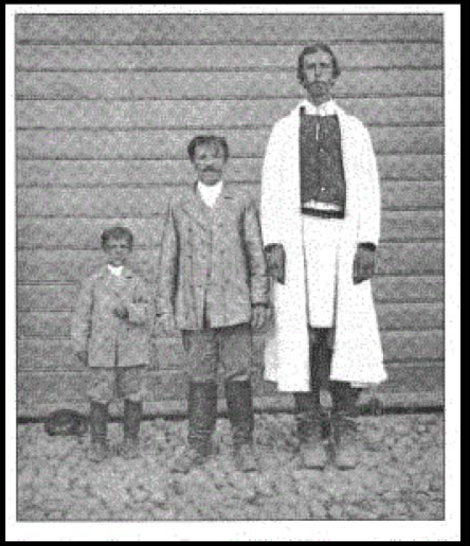
- Frequent type of somatic mutations in cancer

Genotype:





Phenotype



Molecular phenotypes

- gene expression
- Metabolite level