DEPARTMENT OF INFORMATION ENGINEERING

From genetic variants to RNAs and protein signaling networks: towards a system level understanding of biology of Type 2 Diabetes

Barbara Di Camillo
Dipartimento di Ingegneria dell'Informazione Università di Padova

STATSEQ meeting on Gene Network Inference with Systems genetic data and beyond.
28-29th March 2013, Paris, France

Layout

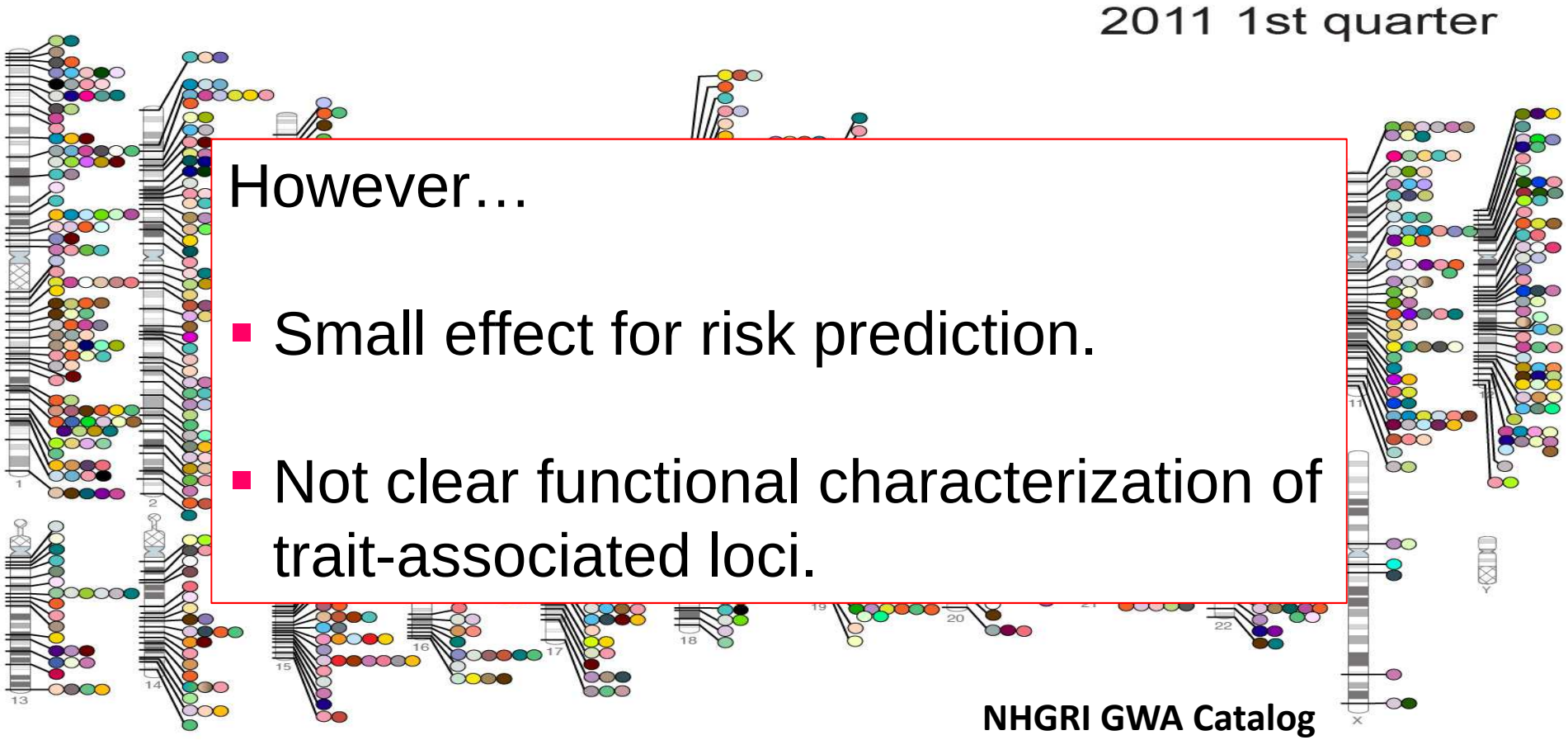
- Using regulatory pathway concept to identify SNP variants associated with T2D in GWAS

- Transcriptional response to SNP variants: Gene Network Inference

- Quantitative modeling of insulin signaling: possible effects of mutations on signaling pathways

GWAS studies have successfully identified a number of significant SNP-disease associations

Published Genome-Wide Associations through 03/2011,
1,319 published GWA at $p \leq 5 \times 10^{-8}$ for 221 traits



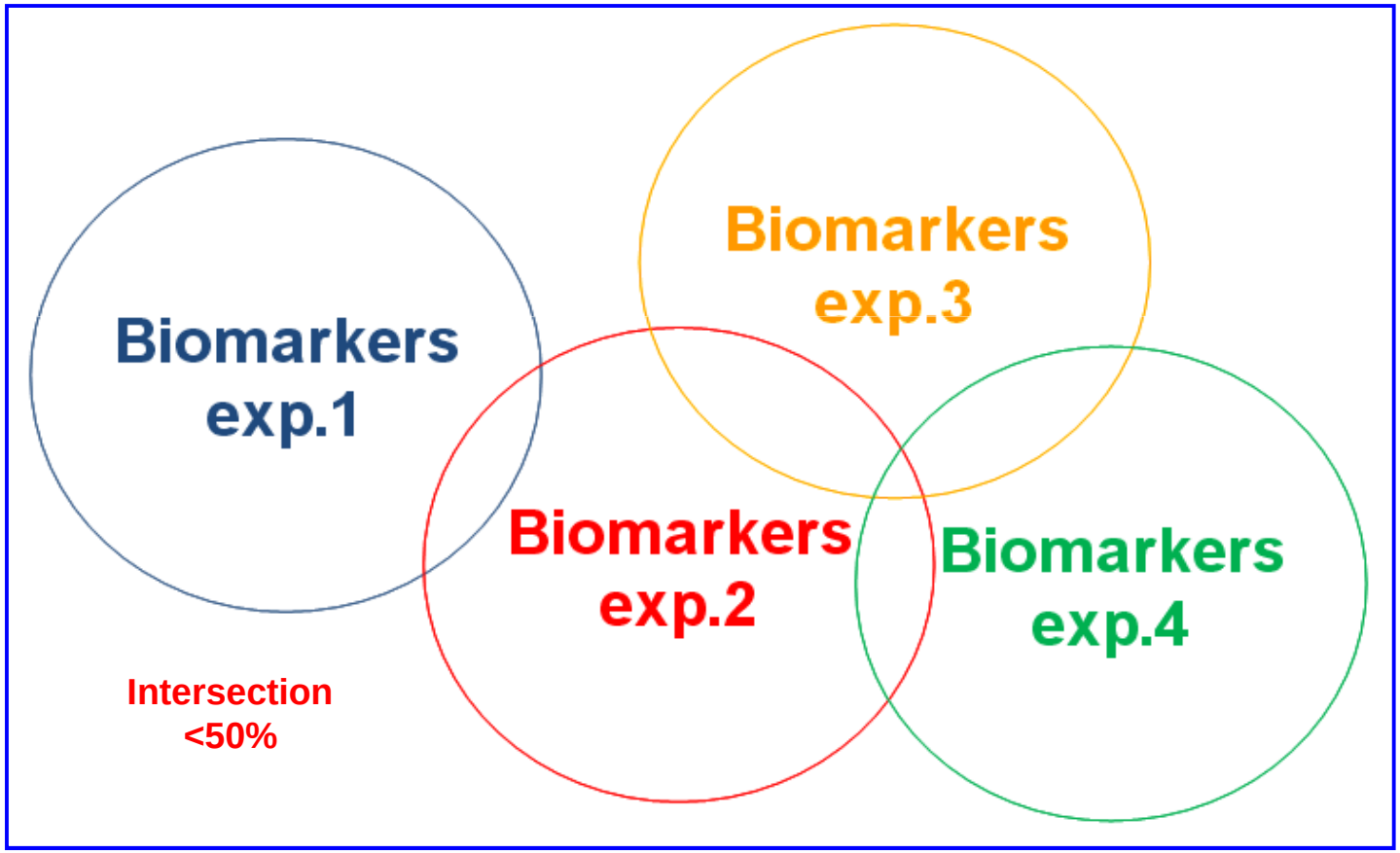
However…

- Small effect for risk prediction.
- Not clear functional characterization of trait-associated loci.

NHGRI GWA Catalog
www.genome.gov/GWAStudies

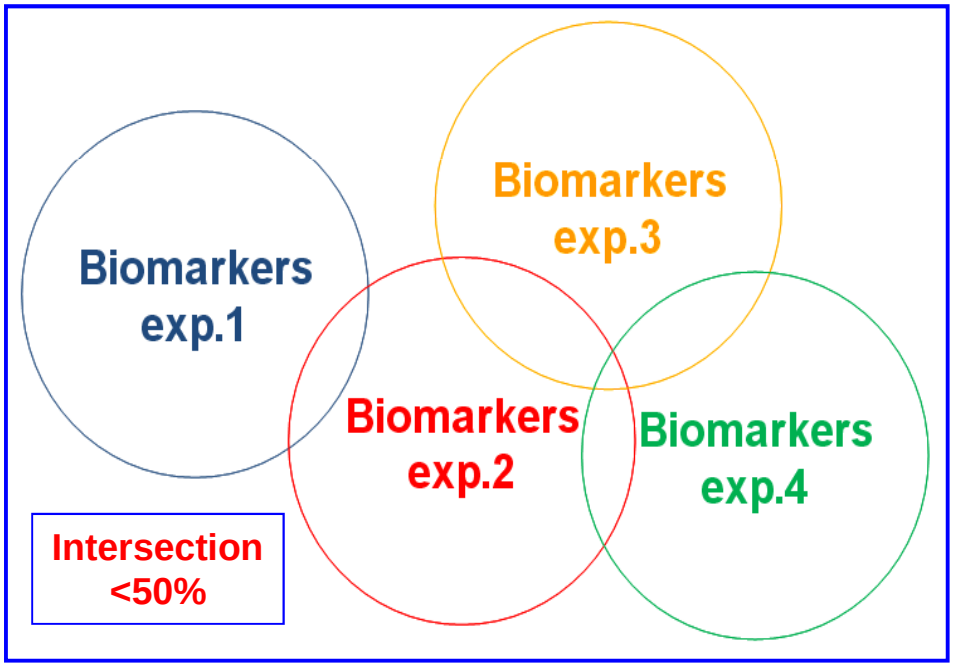
Limited overlap of biomarker lists

A common experience with high-throughput data



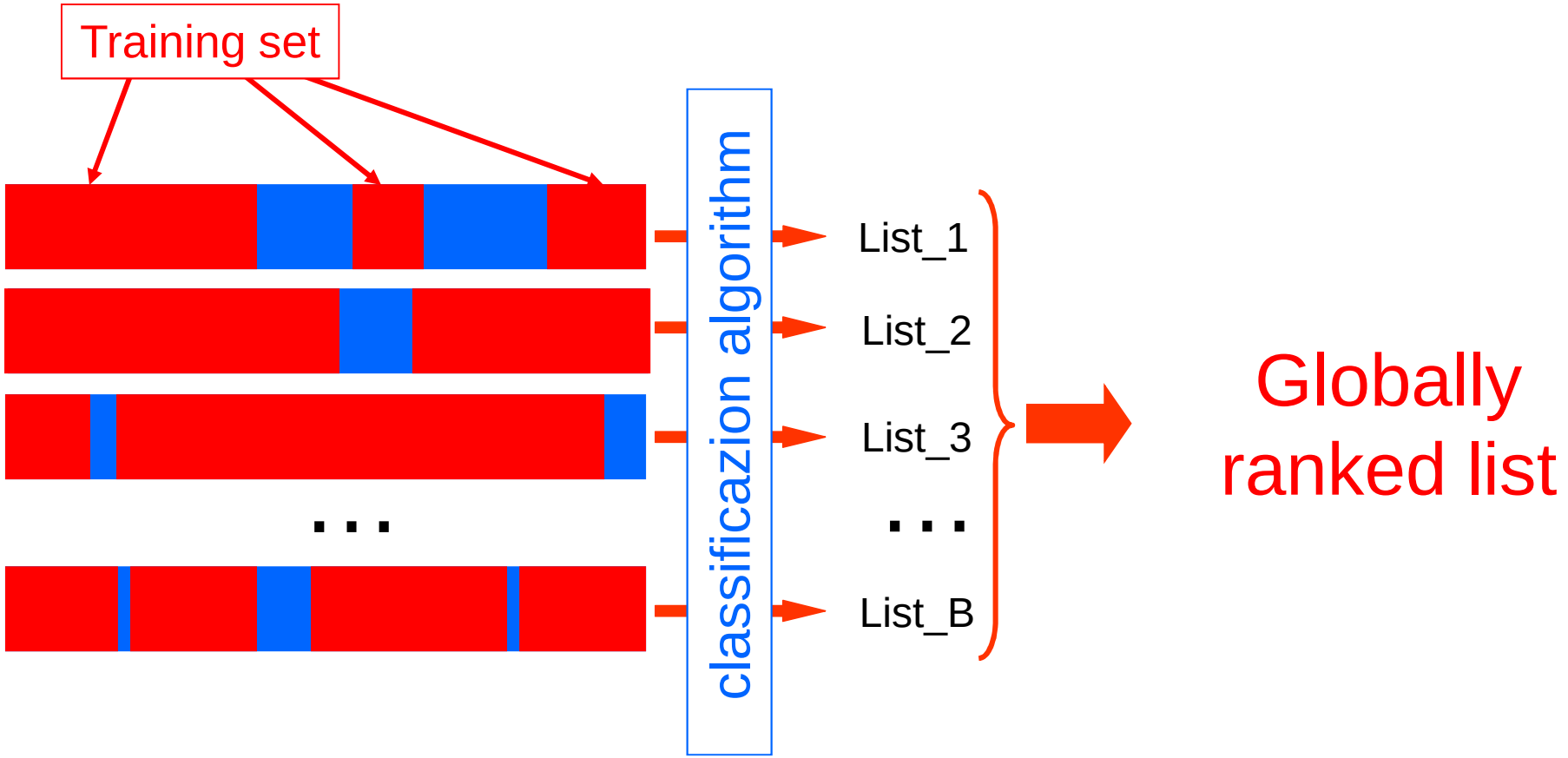
Possible reasons of list instability (1)

- Datasets with a low number of subjects vs. high number of variables
- Data are highly correlated (co-regulated)



Accounting for data dimension

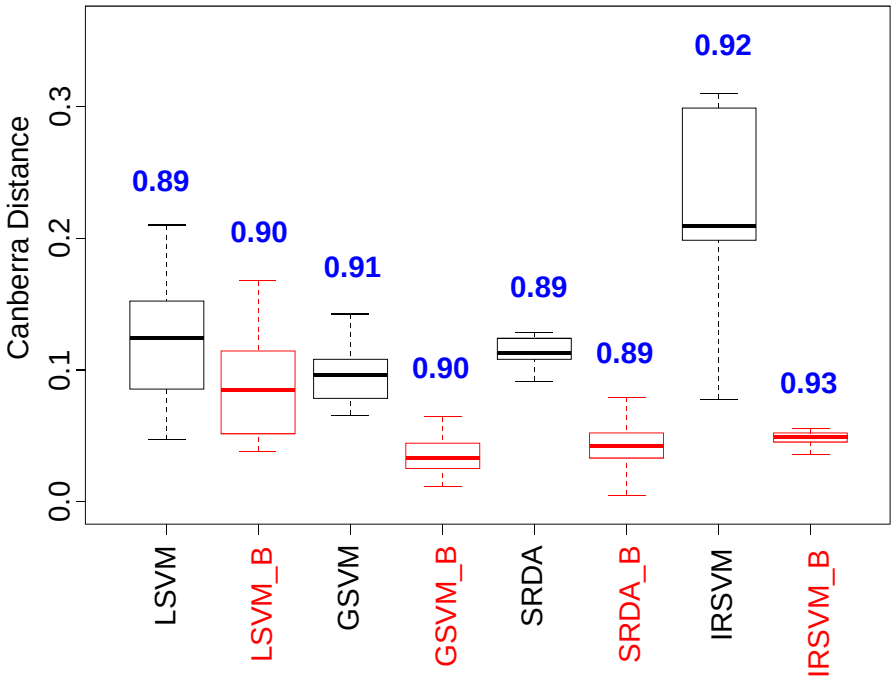
Bootstrap: Sample (with replacement) N subjects from the original dataset. Run the algorithm and extract the biomarker list. Repeat B times (e.g. B=100)

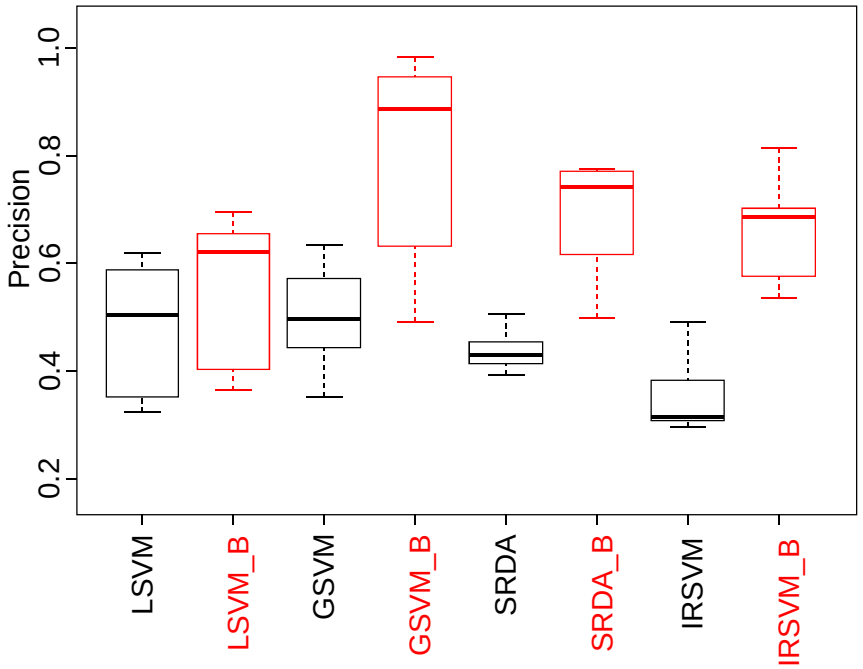


Resampling results

Resampling improves biomarker list reproducibility and precision.

Simulated data: 10 datasets with 2 classes, 20 subjects per class, 10000 features and 160 biomarkers



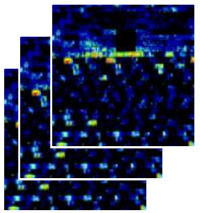


Di Camillo B. et al., PLoS ONE, 2012
Sambo F. et al., BMC Bioinformatics, 2012

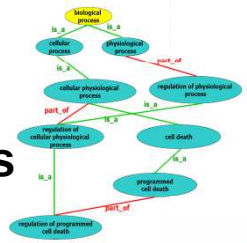
Accounting for data co-regulation

Using prior biological information to define to regularize the discriminant functions of the classification algorithms.

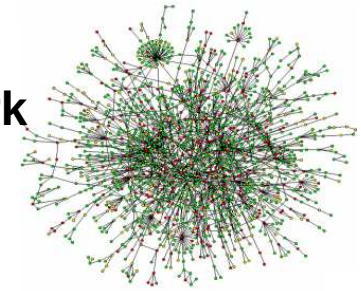
Correlation between genes



Functional annotations



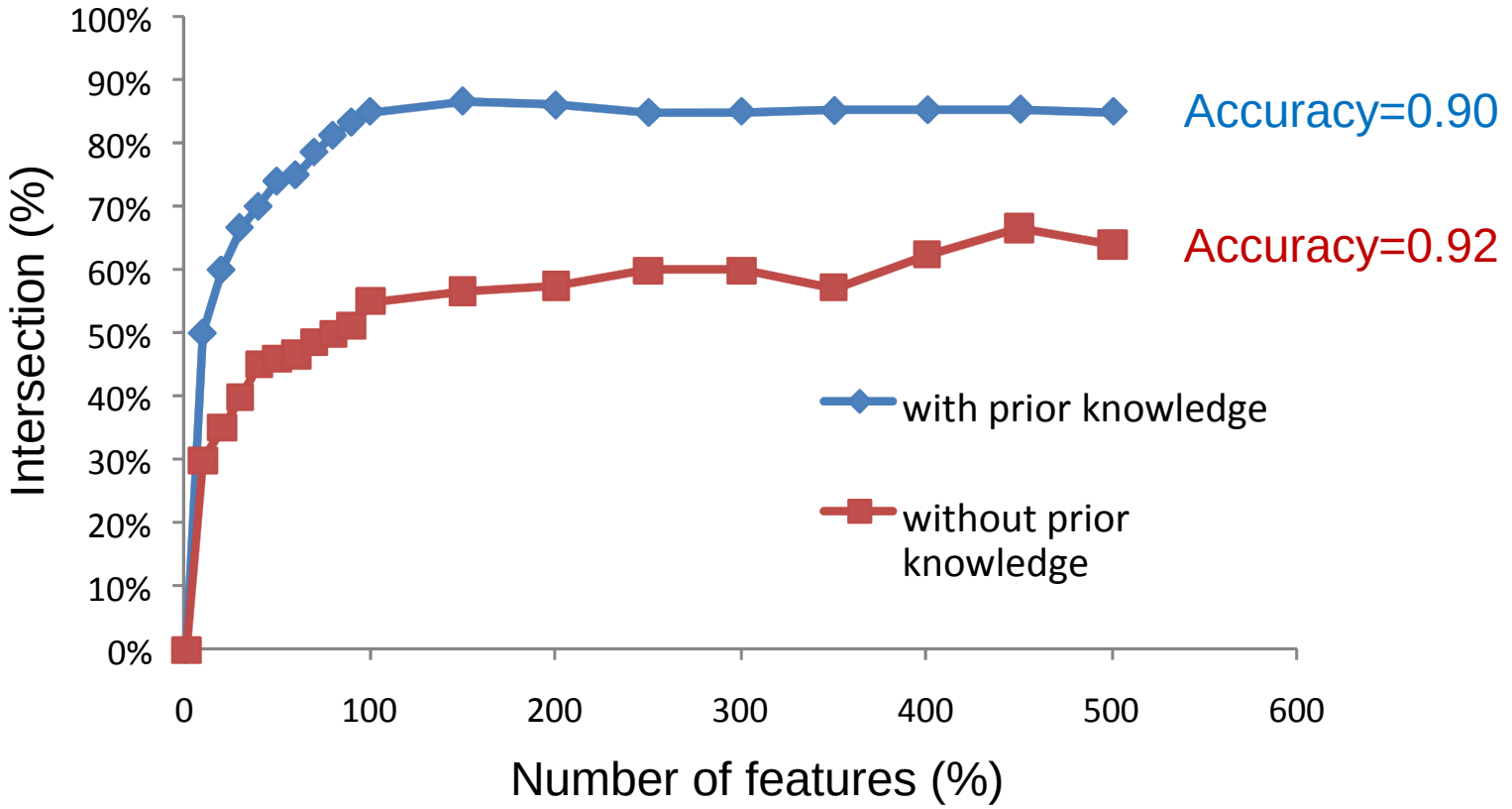
PPI Network topology



Similarity Matrix **S**

Constrain the kernel to place “similar” features close to each other in the biomarker list, if they are also discriminant

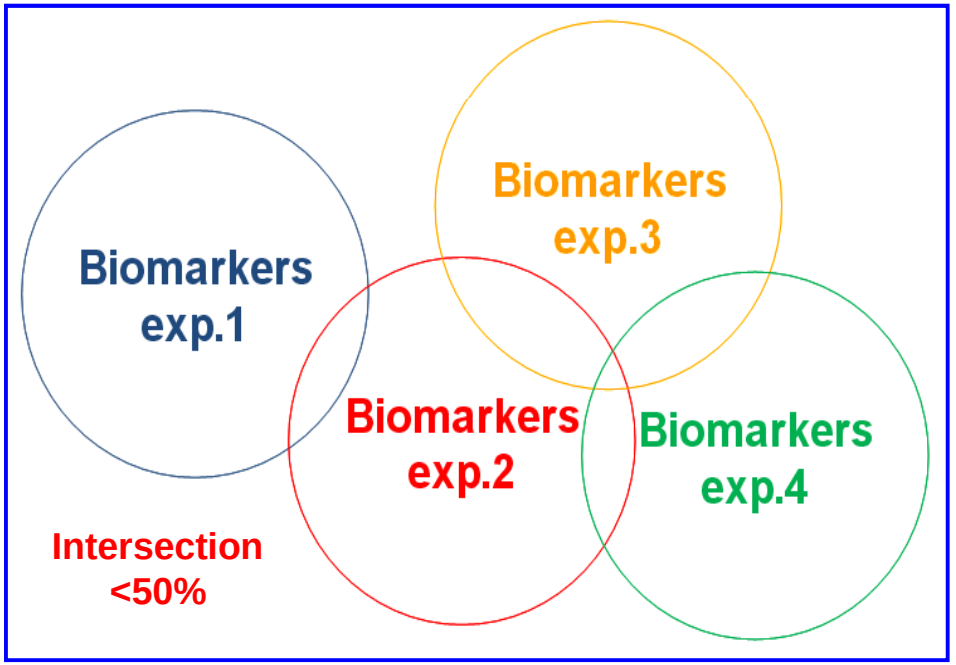
Prior-knowledge Integration Results



Sanavia et al., BMC Bioinformatics 2012

Possible reasons of list instability (2)

- Diseases are heterogeneous and multicausal: *each patient may exhibit a different combination of gene alterations that are sufficient to perturb some specific pathways.*



Using Pathways to analyze SNP data

IDEA: studying the cumulative variation of SNPs in genes mapping on the same pathway (interacting genes)

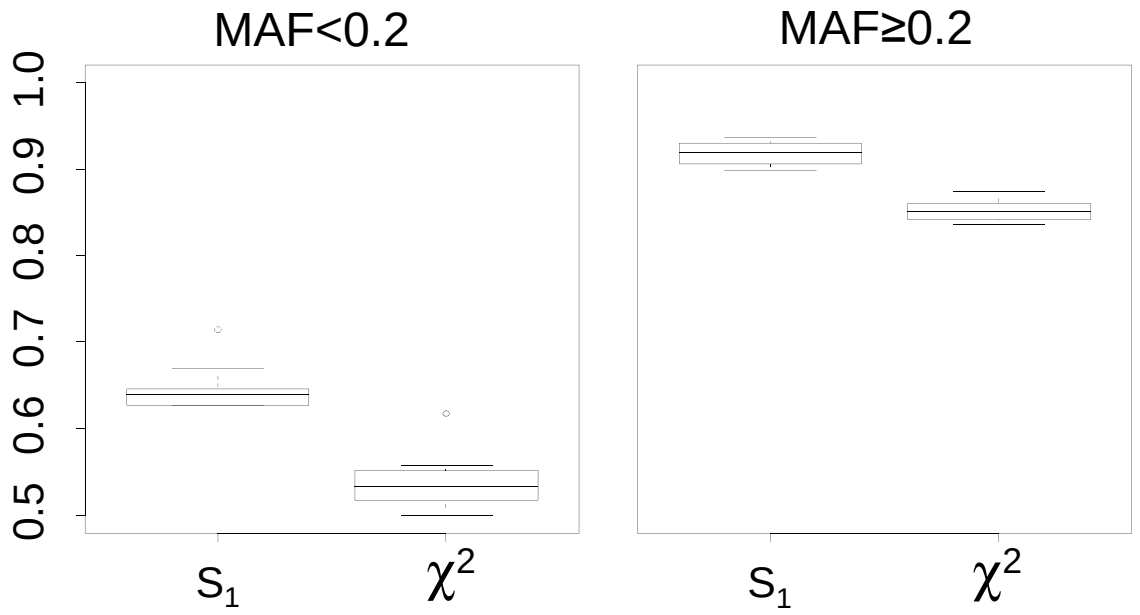
SP-ABACUS

*SNP & Pathway Analysis b*ased on a *B*iv*A*riate *CU*mulative *S*tatistics

- SNPs mapped on genes and genes on pathways
- Each pathway is tested independently applying an entropy based bivariate test of association for each pair of SNPs
  - ✓ VS. univariate, easier to detect TP (multiple evidence)
  - ✓ VS. multivariate, avoid massive optimization problems and keep low associations
- Cumulative statistic calculated for each pathway

Performance on simulated data

SP-ABACUS Recall at FPrate=5% on single test



✓More sensitive than Chi-square test, especially at low MAF

SP-ABACUS characteristics

- ***SNP p-values depend on the pathway being analyzed!***

SNP significance increases with:

- ✓*Strength of association*
- ✓*Number of SNPs associated to the disease in the pathway*
- ✓*Possible epistatic interactions*

- *Significance does not depend on number of SNPs mapping on genes and on the pathway*

Di Camillo et al., 2013 submitted

Application to WTCCC T2D data

1924 Type 2 diabetes/2938 healthy/ 500K SNPs (Affymetrix)

SNPs mapped on genes (Affymetrix annotation files) and genes on 860 KEGG + REACTOME pathways (MSigDB)

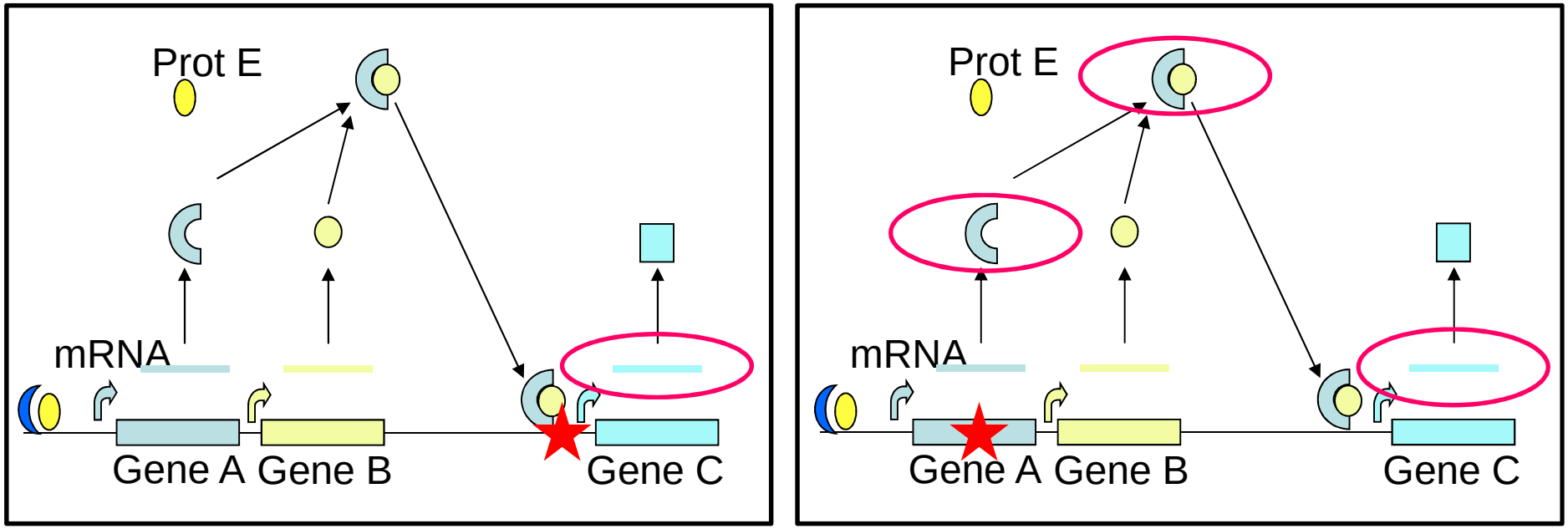
- **PATHWAYS IN CANCER**: TCF7L2, PPARG, MYC, …
- **TRYPTOPHAN_METABOLISM:** ASMT
- **INSULIN SIGNALING**: IRS1, PI3K
- **CELL CYCLE:** MYC, PPM1A
- **REGULATION_OF_INSULIN_SECRETION:** ADCY8, KCNC2
- **CLASS_B_2_SECRETIN_FAMILY_RECEPTORS:** GNG2, PTH2R,WNT4
- **NA_INDEPENDENT_GLUCOSE_TRANSPORTERS:** SLC2A9
- **...**

Di Camillo et al., 2013 submitted

Layout

- Using regulatory pathway concept to identify SNP variants associated with T2D in GWAS
- Transcriptional response to SNP variants: Gene Network Inference
- Quantitative modeling of insulin signaling: possible effects of mutations on signaling pathways

Transcriptional response to SNP variants



Strategy based on differential expression analysis plus quality reasoning to disambiguate directed/undirected interactions

(SPQR, Badaloni et al., IEEE/ACM Trans Comput Biol Bioinform. 2012)

"Honorable Mention for Best Performer (fourth place) in the Systems Genetics Challenge" in DREAM 5"

Differential wiring

- GSE18732, human muscle tissue, 45 T2D, 47 controls (no eQTL)
- 14 probes (12 genes) differentially expressed at FDR 10%

Regulators: the genes annotated with SNPs associated to T2D

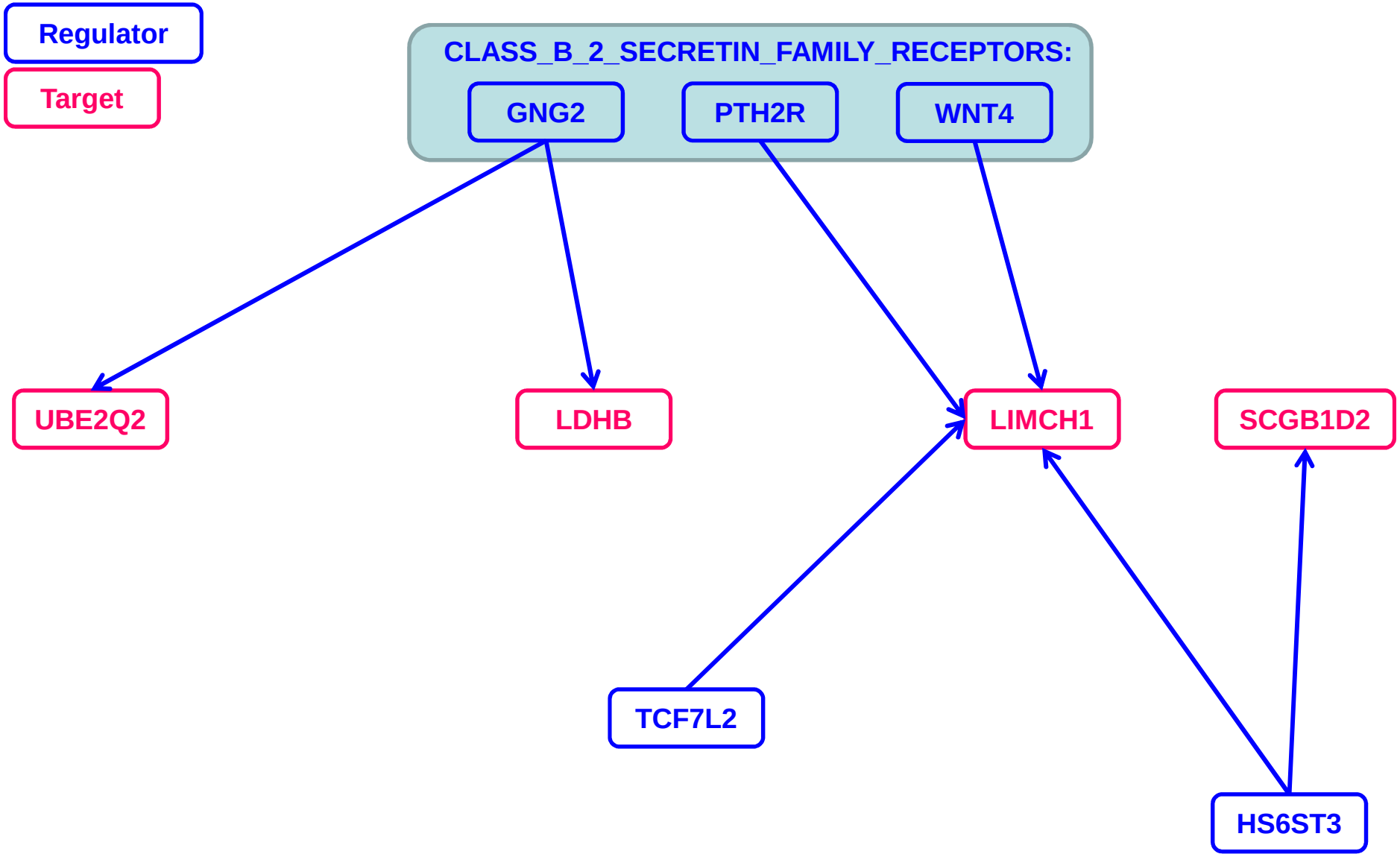
Targets: the differentially expressed genes in T2D vs. NORMAL

For each pair (regulator, target) we calculate:

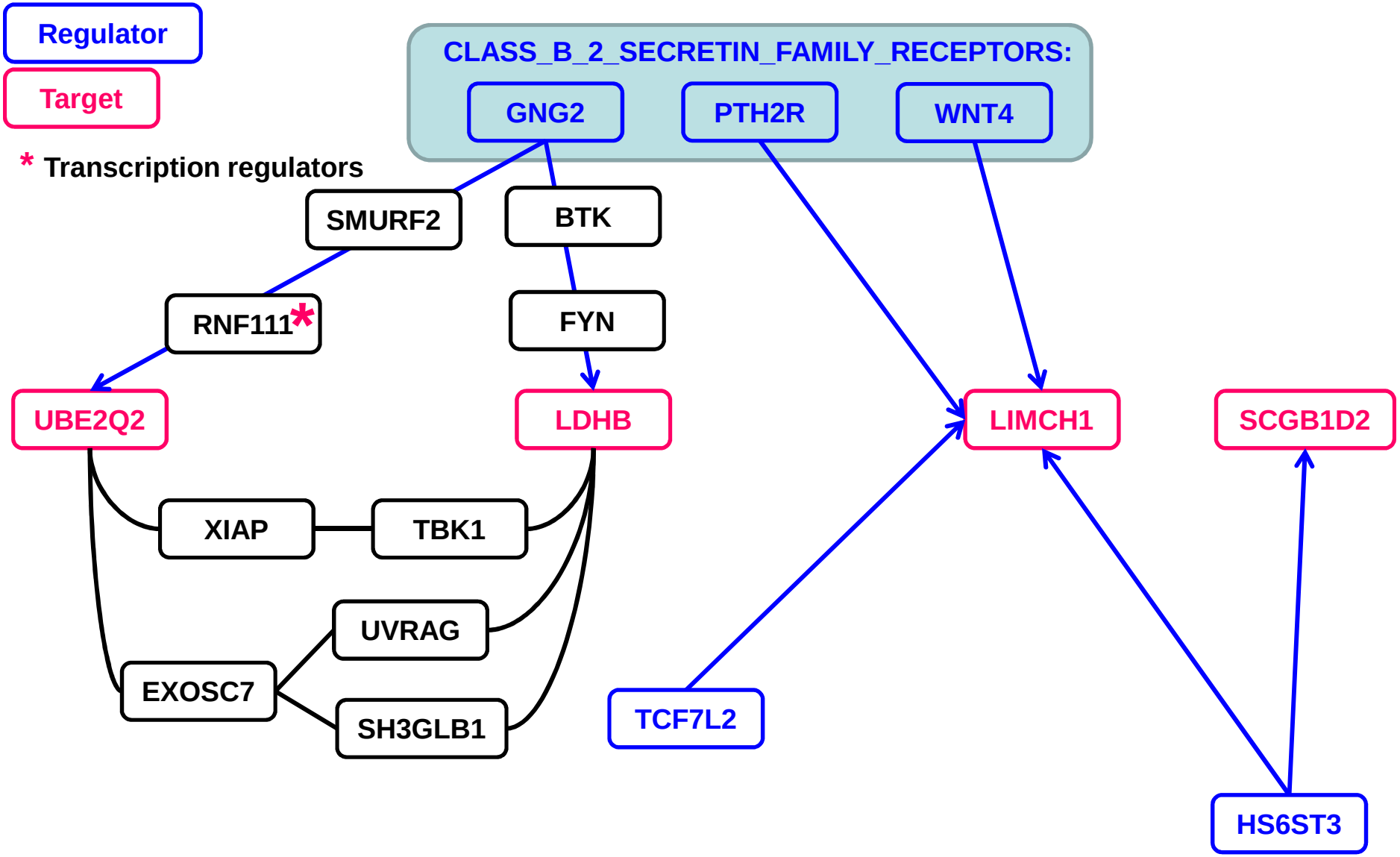
$$\Delta r(\text{regulator, target}) = r(\text{regulator, target})_{\text{NORMAL}} - r(\text{regulator, target})_{\text{T2D}}$$

Significant differences are hints of downstream effect of the mutations on the targets

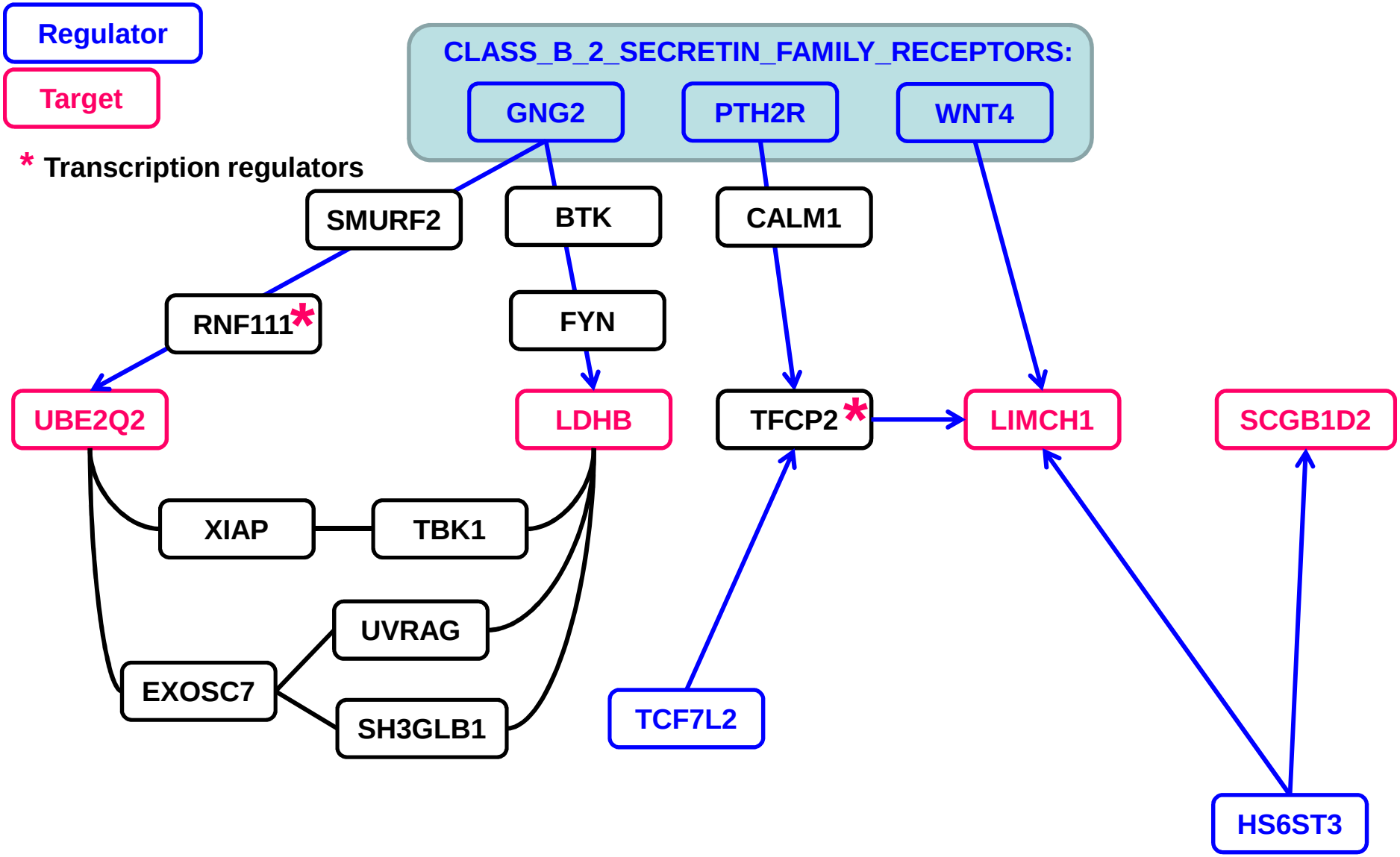
The inferred network



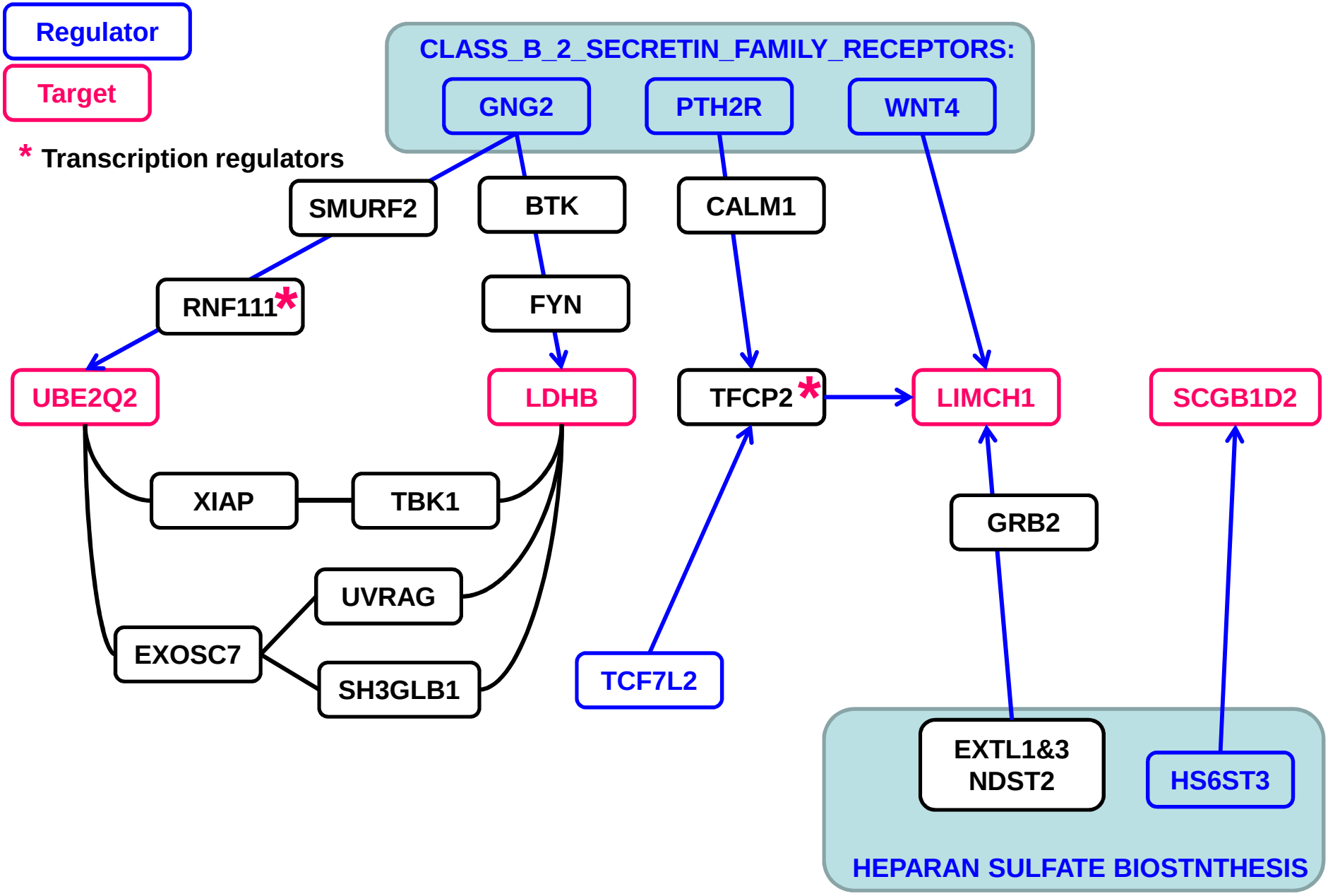
Validation


Regulator
Target
* Transcription regulators
CLASS_B_2_SECRETIN_FAMILY_RECEPTORS:
GNG2
PTH2R
WNT4
SMURF2
BTK
RNF111*
FYN
UBE2Q2
LDHB
LIMCH1
SCGB1D2
XIAP
TBK1
UVRAG
EXOSC7
SH3GLB1
TCF7L2
HS6ST3

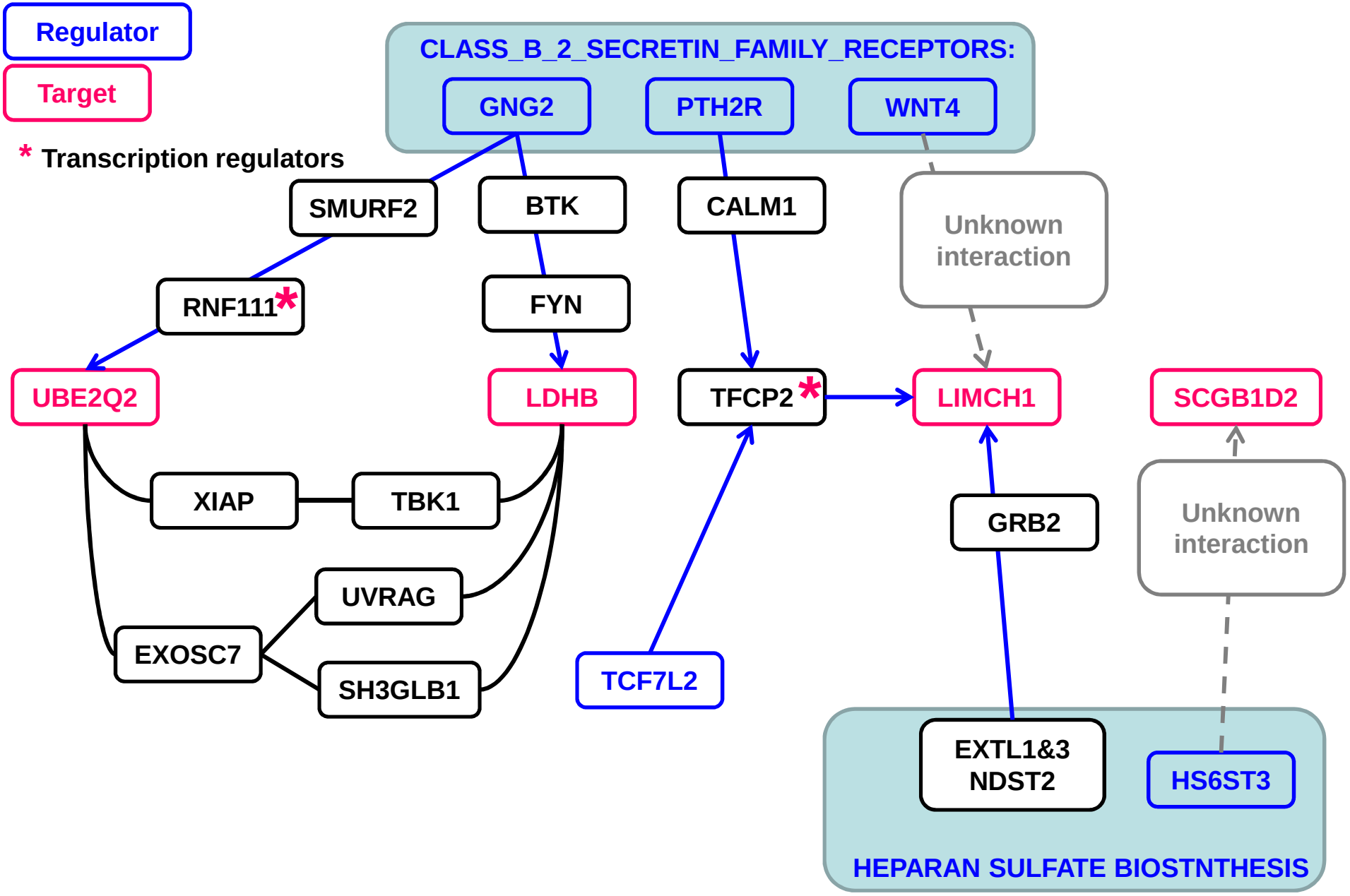
Validation


Regulator
Target
* Transcription regulators
CLASS_B_2_SECRETIN_FAMILY_RECEPTORS:
GNG2
PTH2R
WNT4
SMURF2
BTK
CALM1
RNF111*
FYN
UBE2Q2
LDHB
TFCP2*
LIMCH1
SCGB1D2
XIAP
TBK1
UVRAG
EXOSC7
SH3GLB1
TCF7L2
HS6ST3

Validation


Regulator
Target
* Transcription regulators
CLASS_B_2_SECRETIN_FAMILY_RECEPTORS:
GNG2
PTH2R
WNT4
SMURF2
BTK
CALM1
RNF111*
FYN
UBE2Q2
LDHB
TFCP2*
LIMCH1
SCGB1D2
XIAP
TBK1
GRB2
UVRAG
EXOSC7
SH3GLB1
TCF7L2
EXTL1&3
NDST2
HS6ST3
HEPARAN SULFATE BIOSTNTHESIS

Validation


Regulator
Target
* Transcription regulators
CLASS_B_2_SECRETIN_FAMILY_RECEPTORS:
GNG2
PTH2R
WNT4
SMURF2
BTK
CALM1
Unknown interaction
RNF111*
FYN
UBE2Q2
LDHB
TFCP2*
LIMCH1
SCGB1D2
XIAP
TBK1
GRB2
Unknown interaction
UVRAG
EXOSC7
SH3GLB1
TCF7L2
EXTL1&3
NDST2
HS6ST3
HEPARAN SULFATE BIOSTNTHESIS

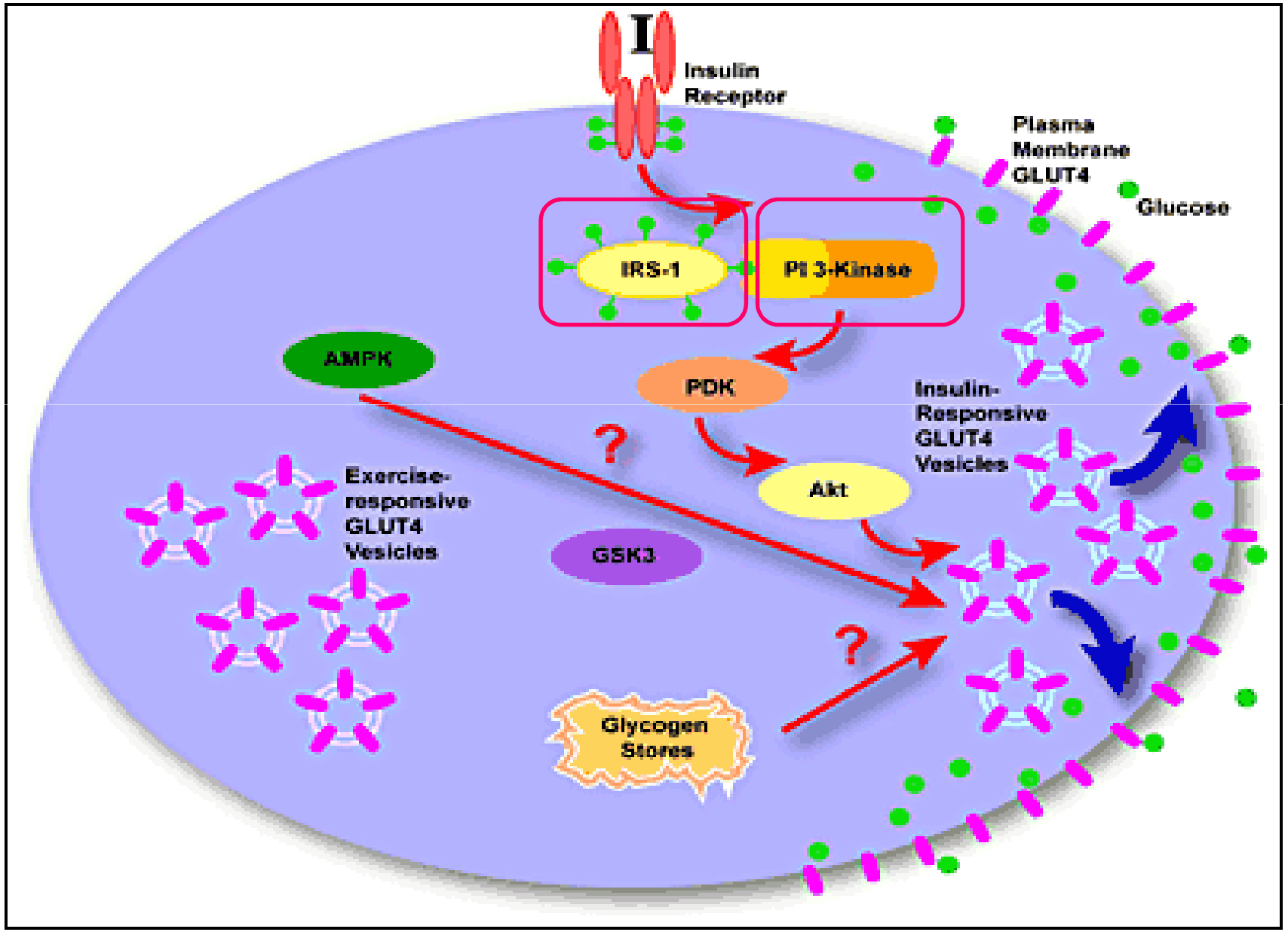
Layout

- Using regulatory pathway concept to identify SNP variants associated with T2D in GWAS

- Transcriptional response to SNP variants: Gene Network Inference

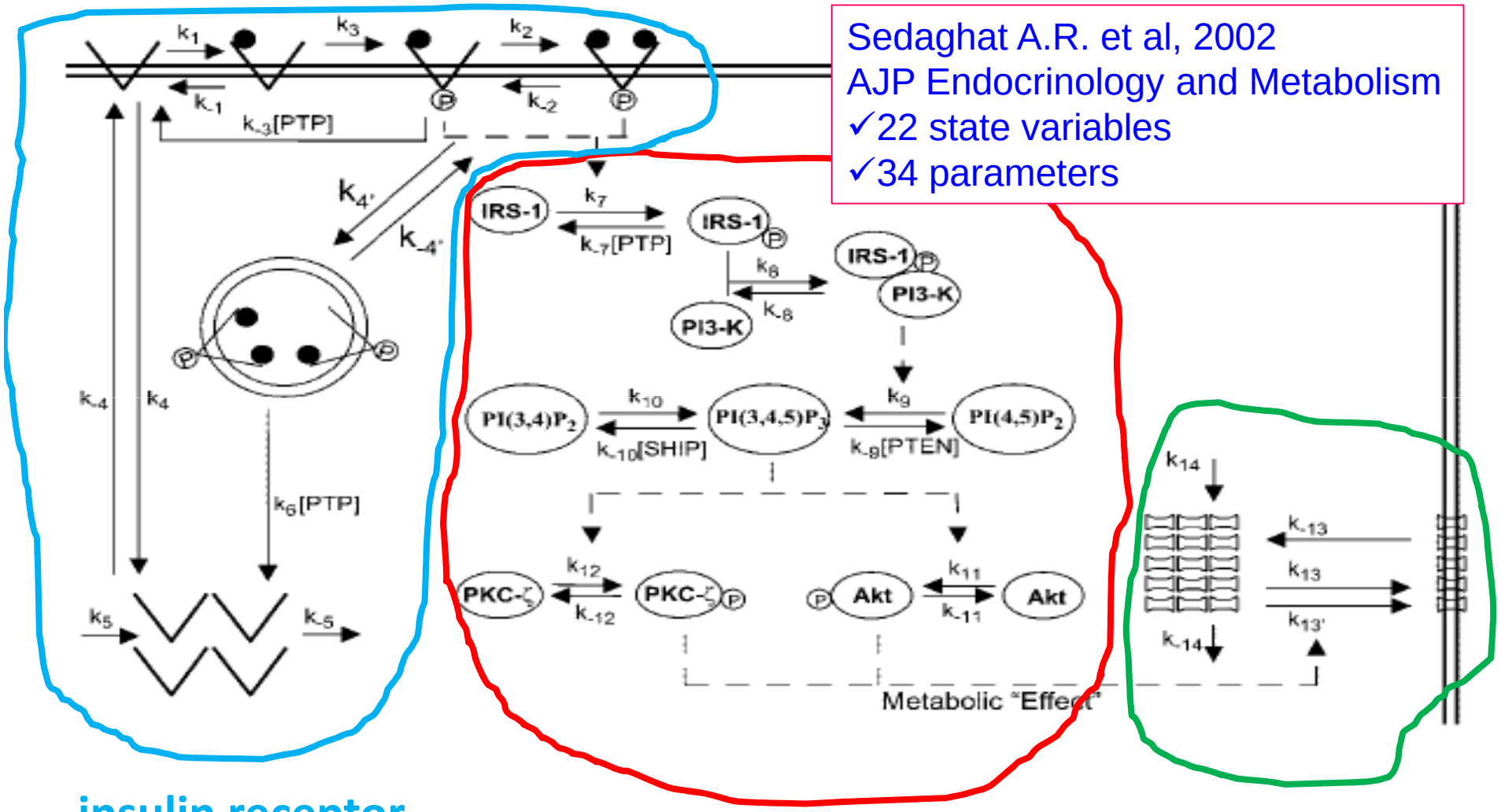
- Quantitative modeling of insulin signaling: possible effects of mutations on signaling pathways

Insulin Signaling Pathway (ISP)



Mass Action Kinetics model of ISP

Sedaghat A.R. et al, 2002
AJP Endocrinology and Metabolism
✓22 state variables
✓34 parameters



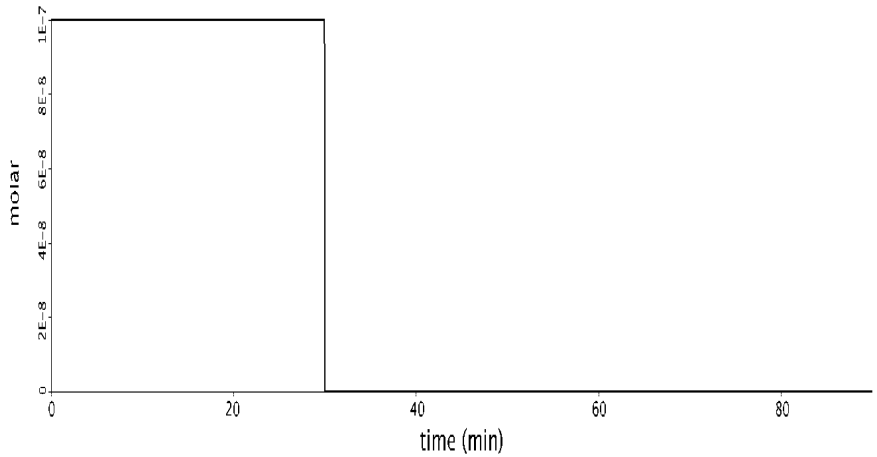
insulin receptor binding and receptor recycling

postreceptor signaling cascade

GLUT4 translocation

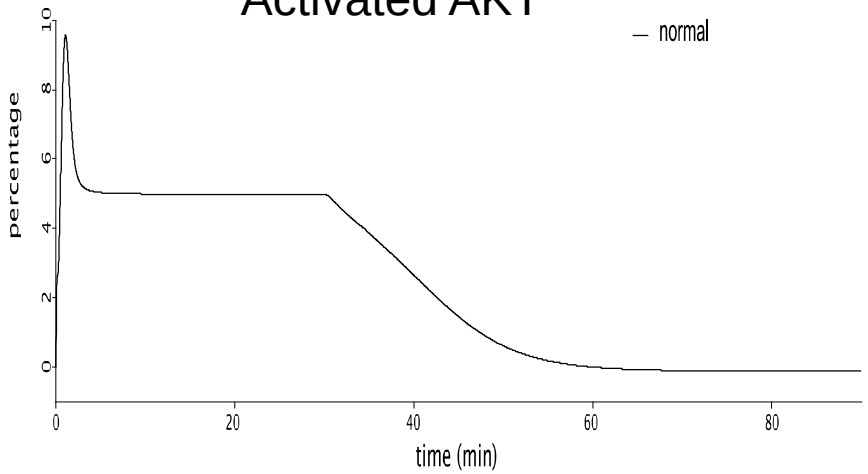
Simulation (Normal)

insulin

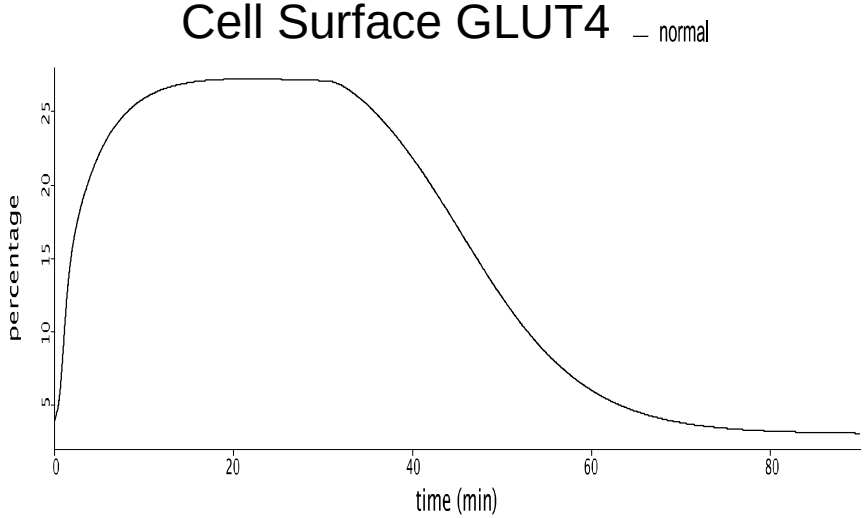


Step input ranging from 0 to 10^{-7} M for 30 minutes

Activated AKT



Cell Surface GLUT4



Local Sensitivity Analysis

- Output *O*: GLUT4 activation

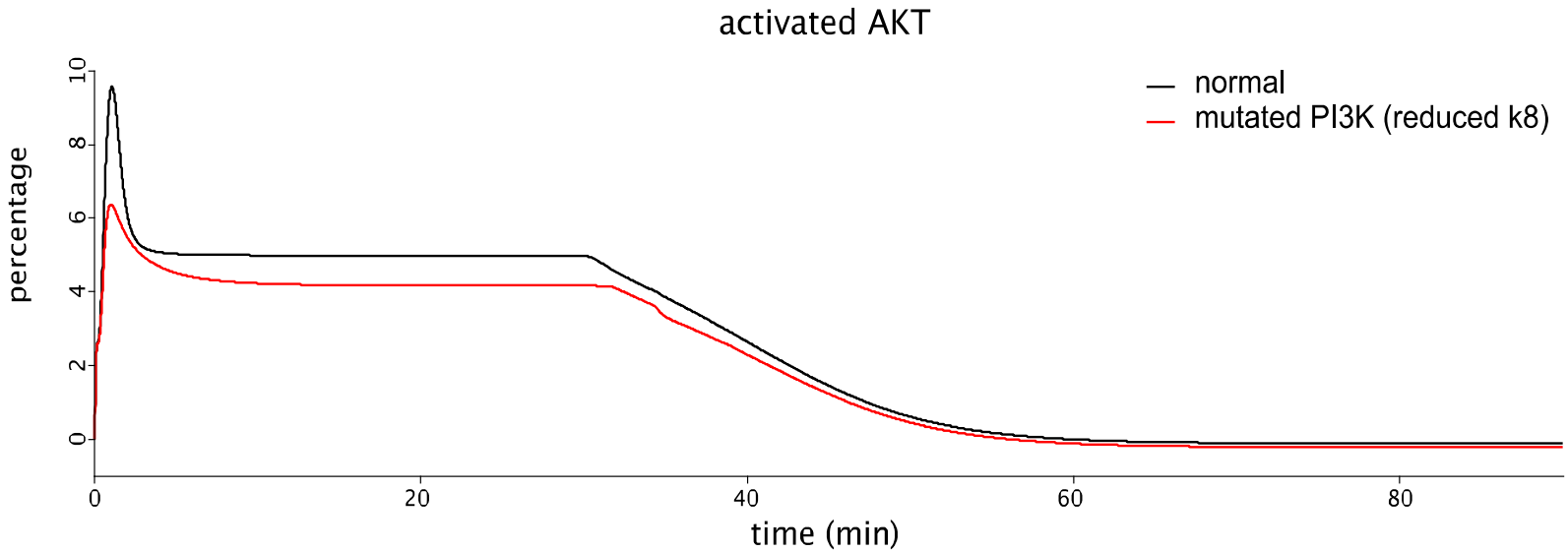
$$S_j = \frac{\partial O / O}{\partial p_j / p_j} = \frac{\left(O\left(p_j + \Delta p_j\right) - O\left(p_j - \Delta p_j\right)\right)/O}{2\Delta p_j / p_j}$$

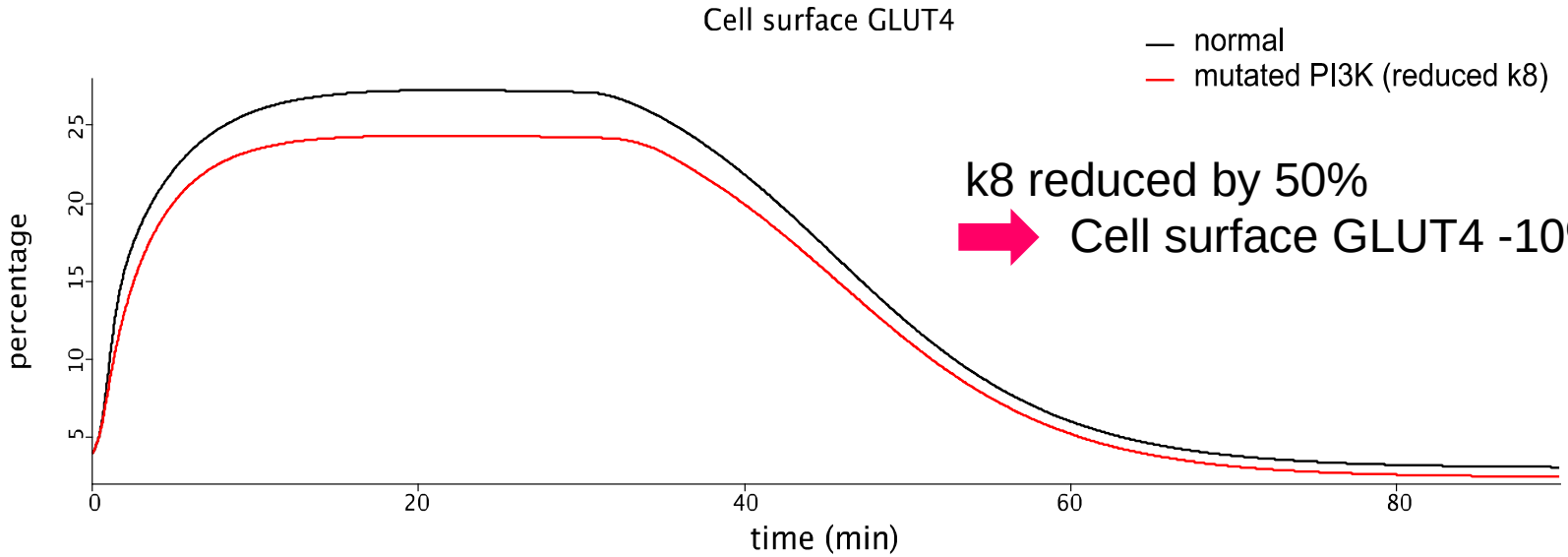
$\Delta p_j = 0.001 \cdot p_j$

parameter	S
k3	-2.40
k9	0.24
k8	0.21
k7b	-0.13
k-7b	0.13
k-13	-0.07
k13	0.07
k-9	-0.06
…	…

http://www.ebi.ac.uk/biomodels-main/BIOMD0000000137
http://physiome.org/jsim/

Simulation (Mutated PI3K)

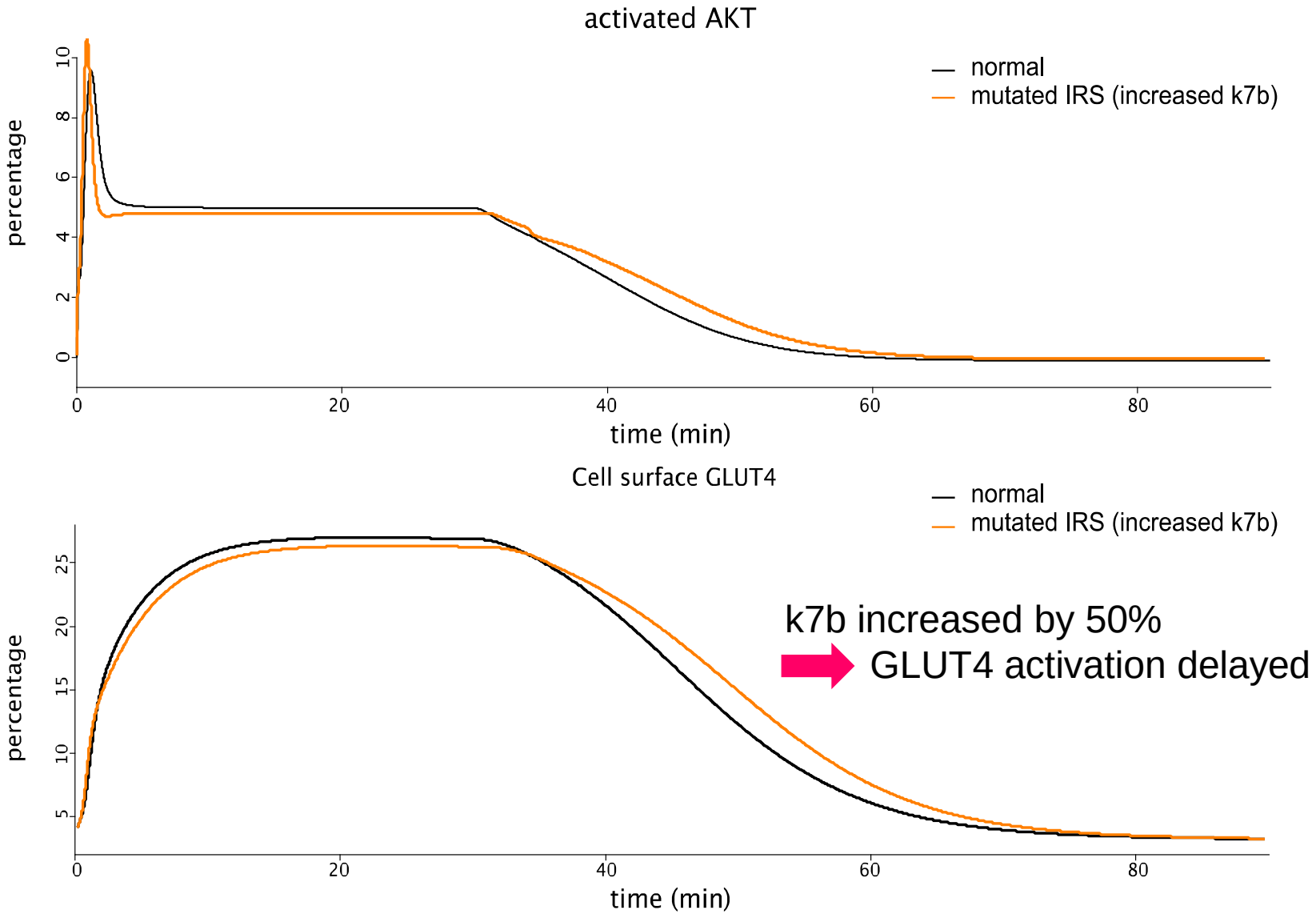




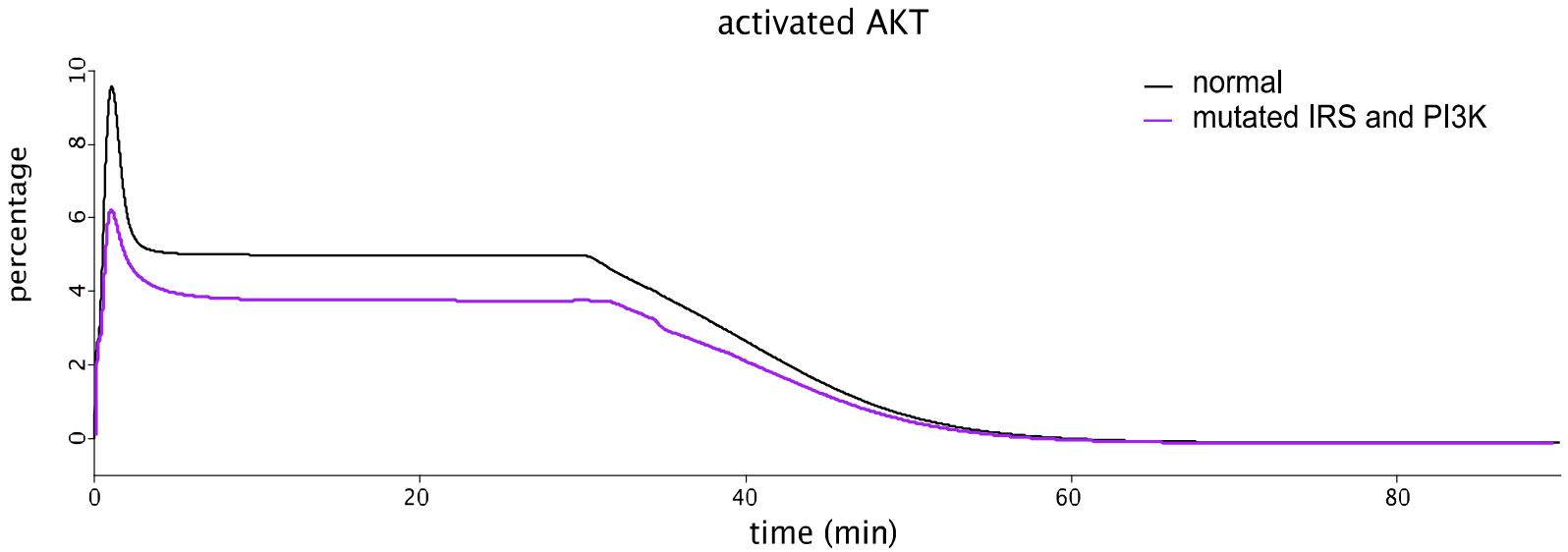
k8 reduced by 50%

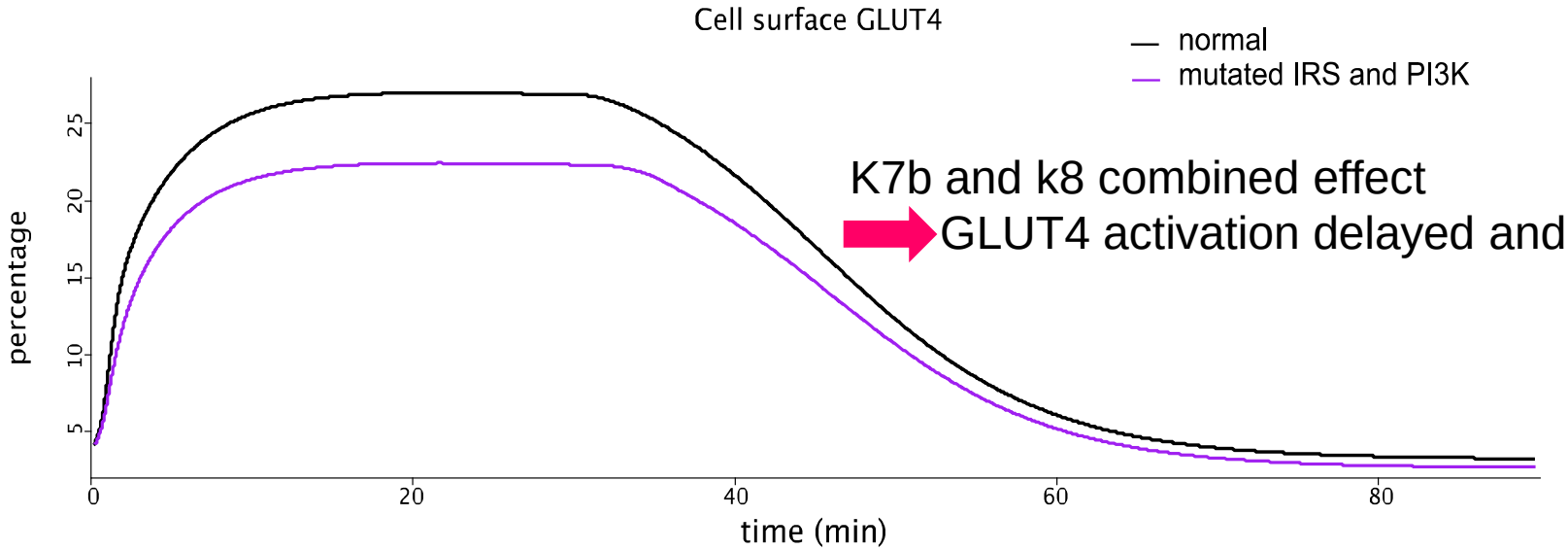
➡ Cell surface GLUT4 -10%

Simulation (Mutated IRS)



Simulation (Mutated PI3K and IRS)





K7b and k8 combined effect

➡ GLUT4 activation delayed and -20%

In Summary

- SP-ABACUS

Using SP-ABACUS allowed gaining sensitivity and formulating new hypothesis on involvement of tryptophan metabolism on T2D etiology

- Differential Wiring

Significant differences of inferred regulatory links (T2D vs. Normal) revealed downstream targets of the mutations

- Quantitative model of insulin signaling

The modeling approach yielded important insights into reciprocal relationships between insulin resistance and mutations in IRS/PI3K/Akt pathway that might be relevant for generating novel therapeutic approaches.

Conclusions

- The most critical aspect is to explain the data from a clinical/biological point of view

- To this purpose, a key challenge we have to face is the integration of data within and across domains and levels of granularity in a multiscale approach.

Bioengineering - Genomic Group

DEPARTMENT OF INFORMATION ENGINEERING

Angela Grassi, Tiziana Sanavia, Federica Eduati

Francesco Sambo, Emanuele Trifoglio, Francesca Finotello

Gianna Toffolo, Alberto Corradin, Barbara Di Camillo