

From Transcription factor binding sites to metabolic phenotype: A tale of MORE gene regulation

Workshop on BIOINFORMATICS OF GENE REGULATION

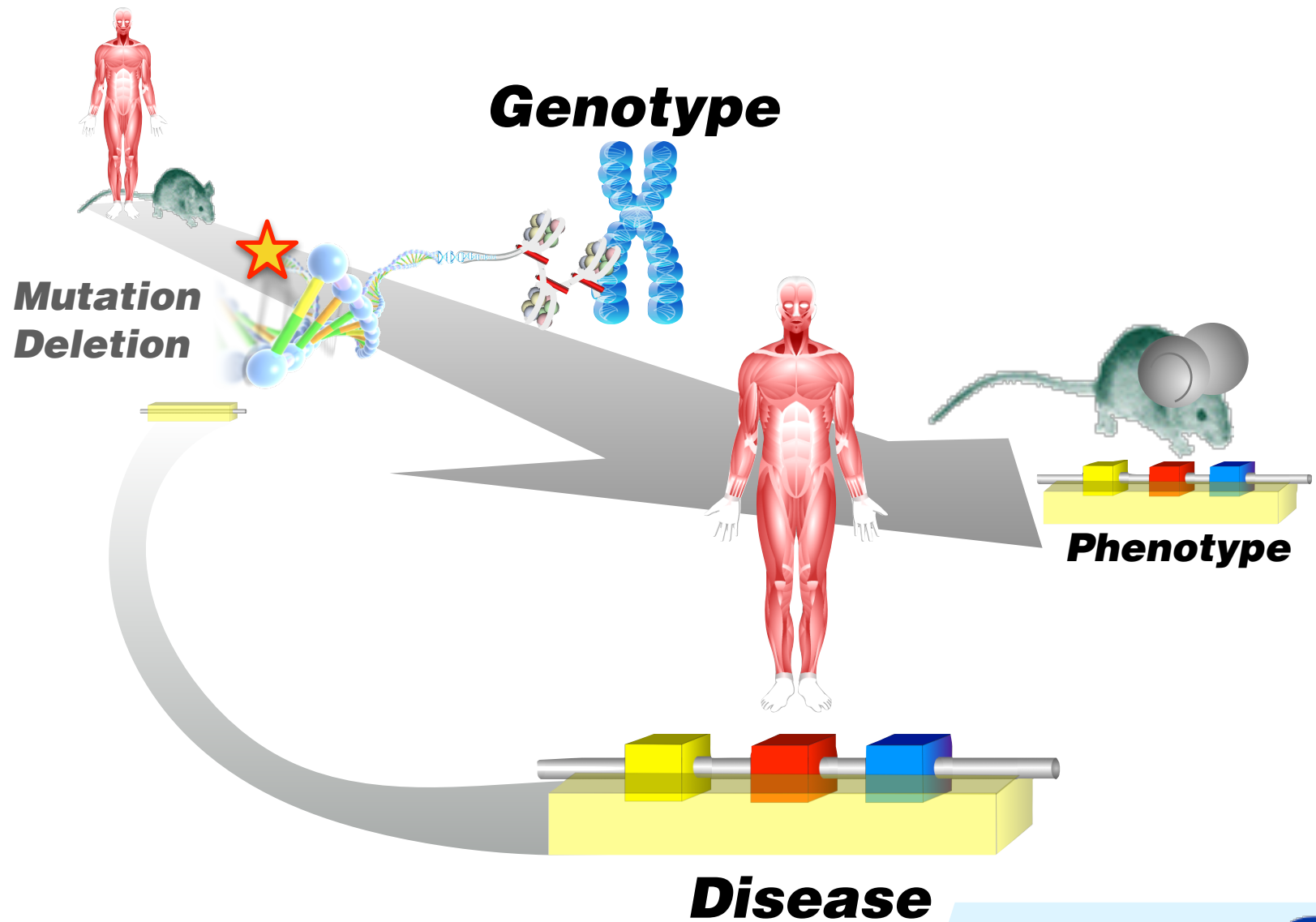
on the occasion of

30 Years TRANSFAC

Göttingen, 7.-9. March 2018

by the Institute of Bioinformatics, UMG Göttingen

From Transcription factor binding sites to metabolic phenotype: A tale of MORE gene regulation



18 years after the human genome had been sequenced

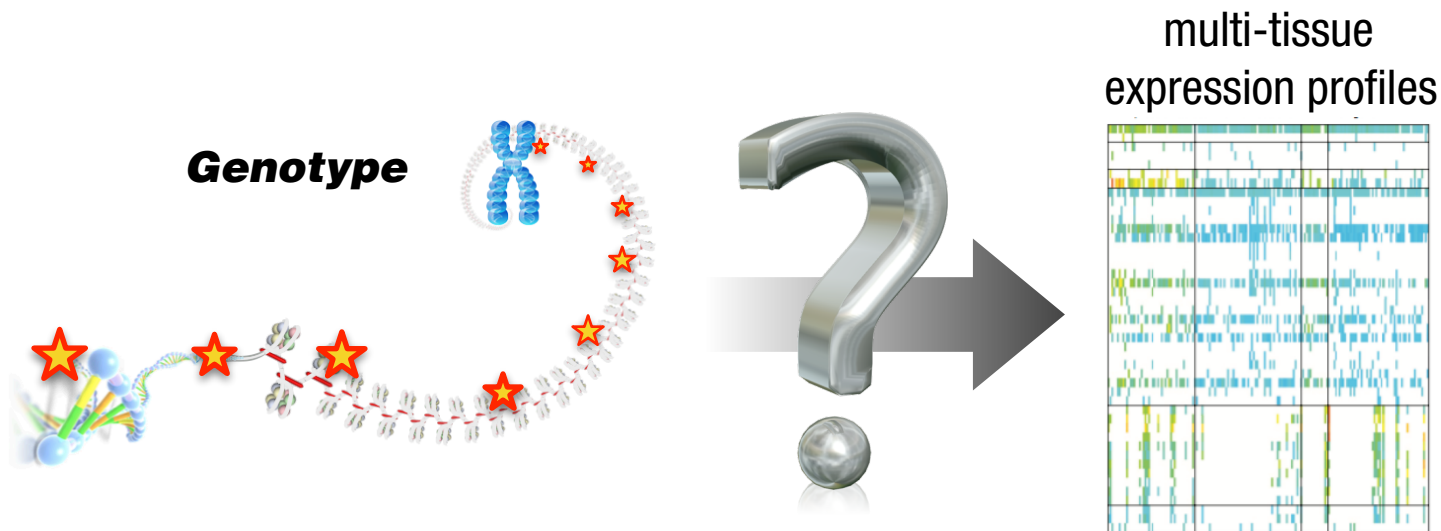
NEWS & VIEWS

For News & Views online, go to
nature.com/newsandviews

HUMAN GENOMICS

Cracking the regulatory code

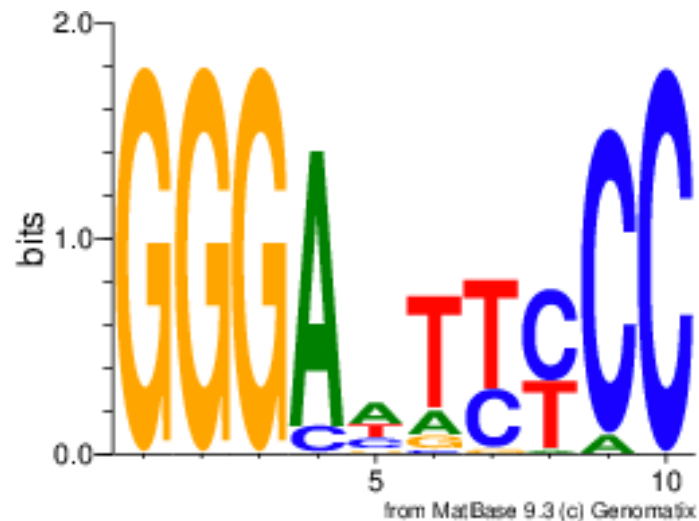
A collection of papers catalogues the associations between genetic variation and gene expression in healthy tissues — the largest analysis of this kind so far. **SEE ARTICLE P.204 & LETTERS P.239, P.244 & P.249**



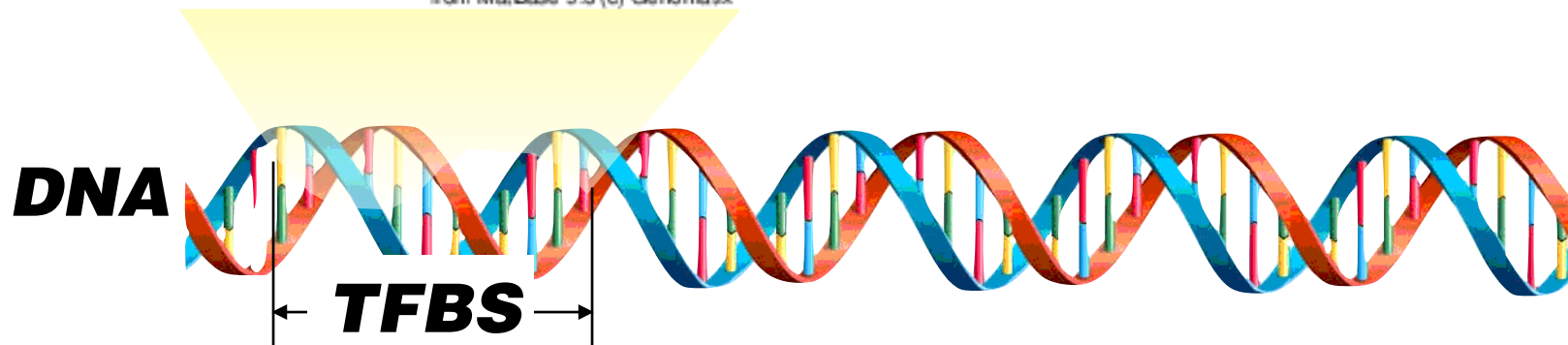
We need to know MORE than just a SNP - expression correlation

Which code? A code can be read from the genome...

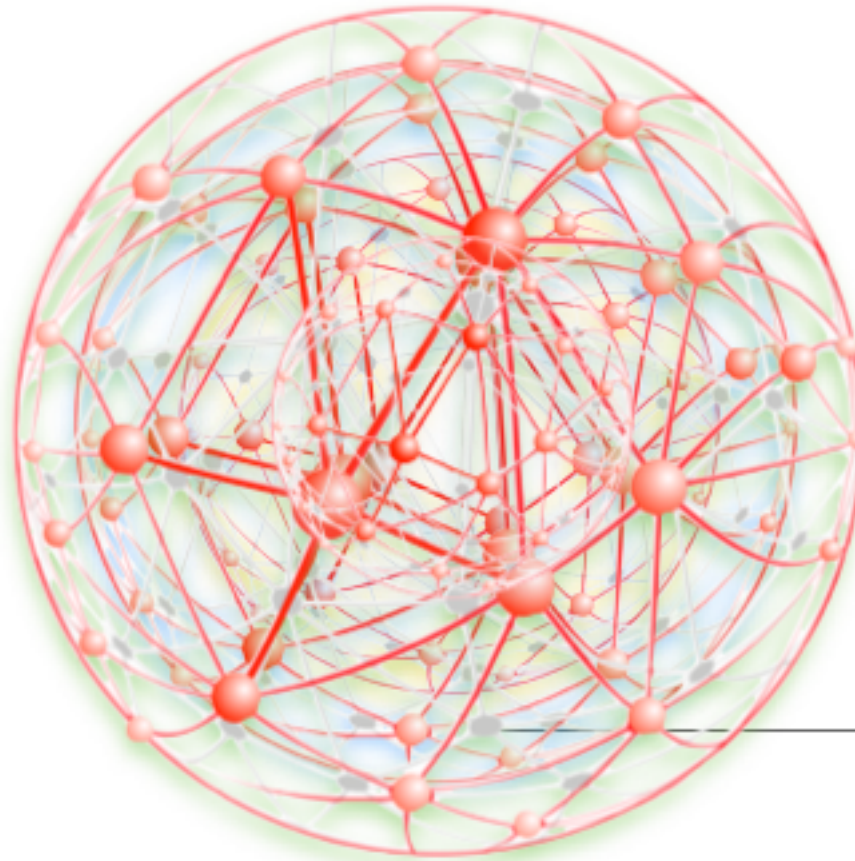
It all starts with Transcription Factor Binding Sites (TFBSs)...



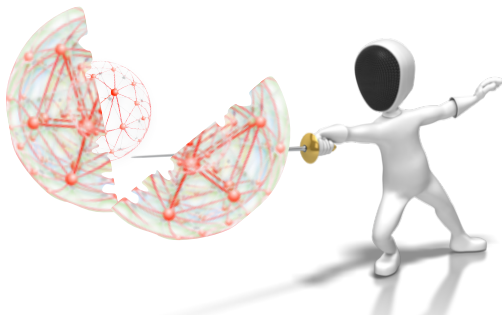
TFBSs are like the words in a dictionary...



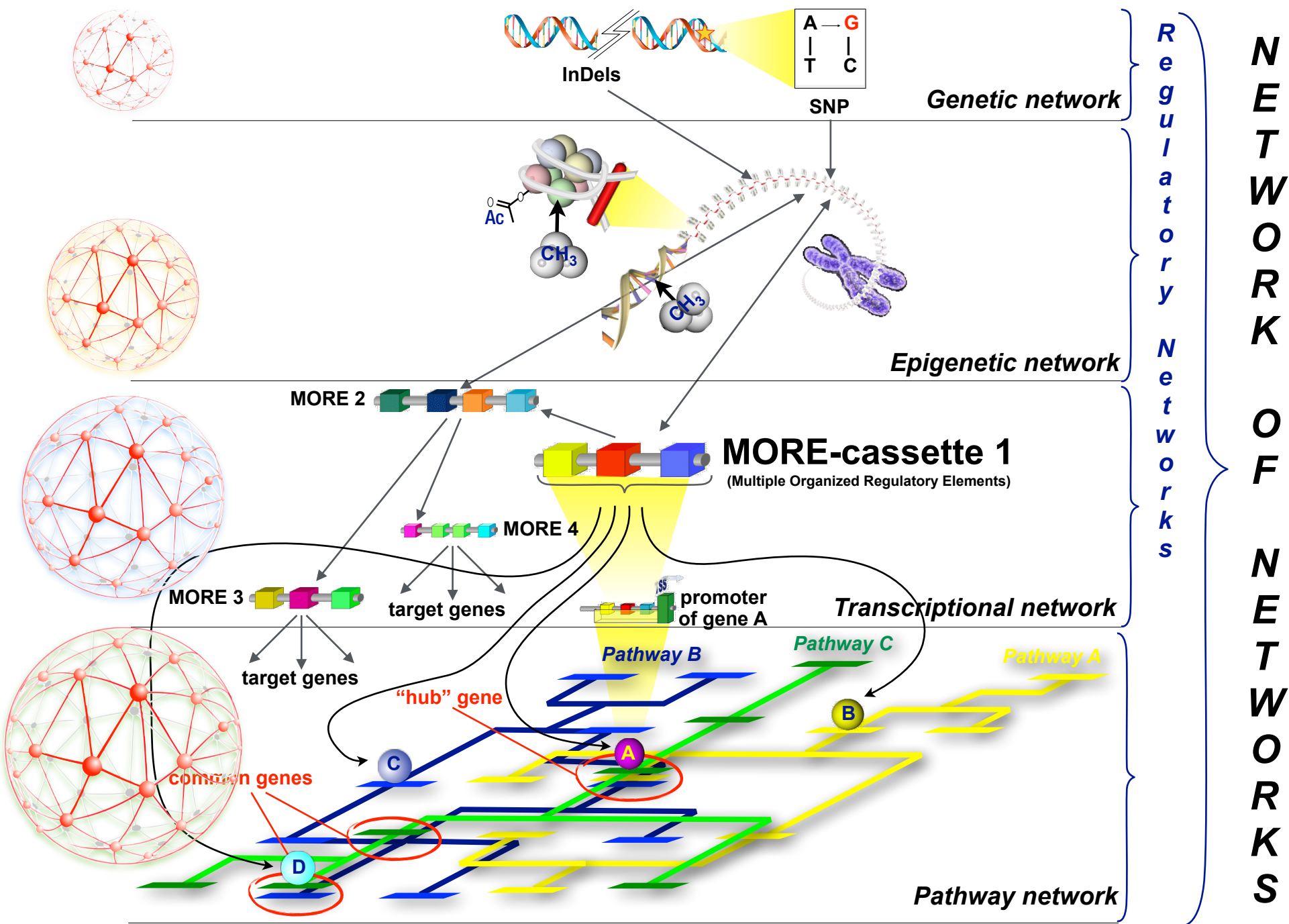
...and continues with networks!



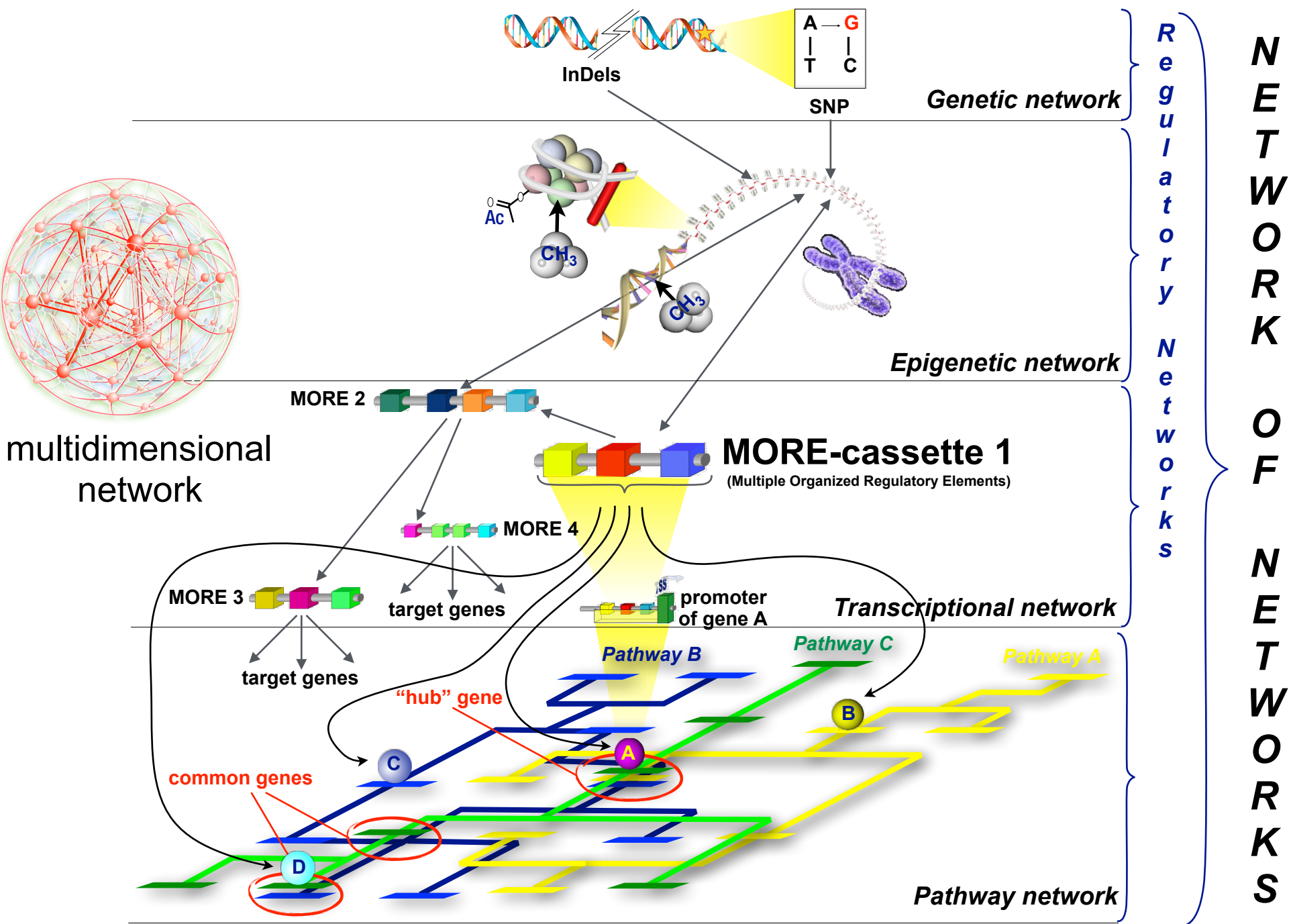
Life in a cell is a single network



Fortunately, divide and conquer does work (to some extent)



Gene expression

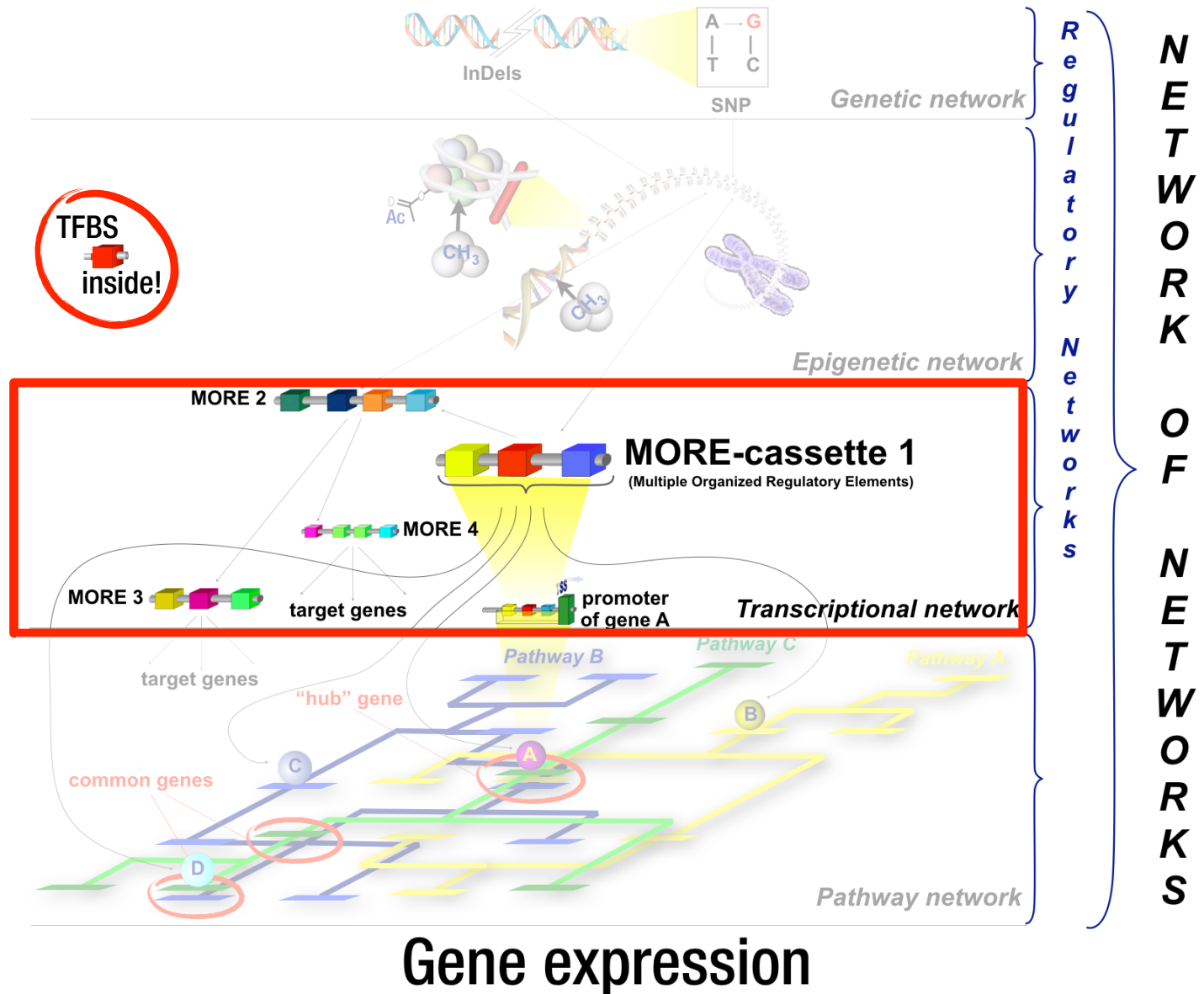


**N
E
T
W
O
R
K

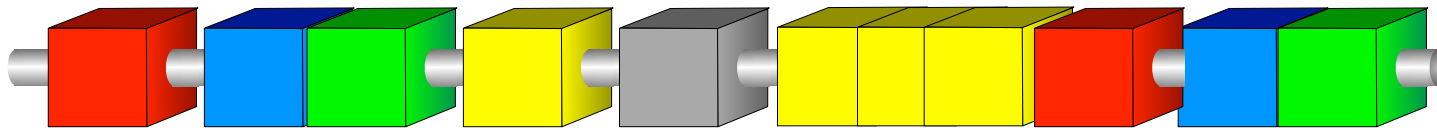
O
F

N
E
T
W
O
R
K
S**

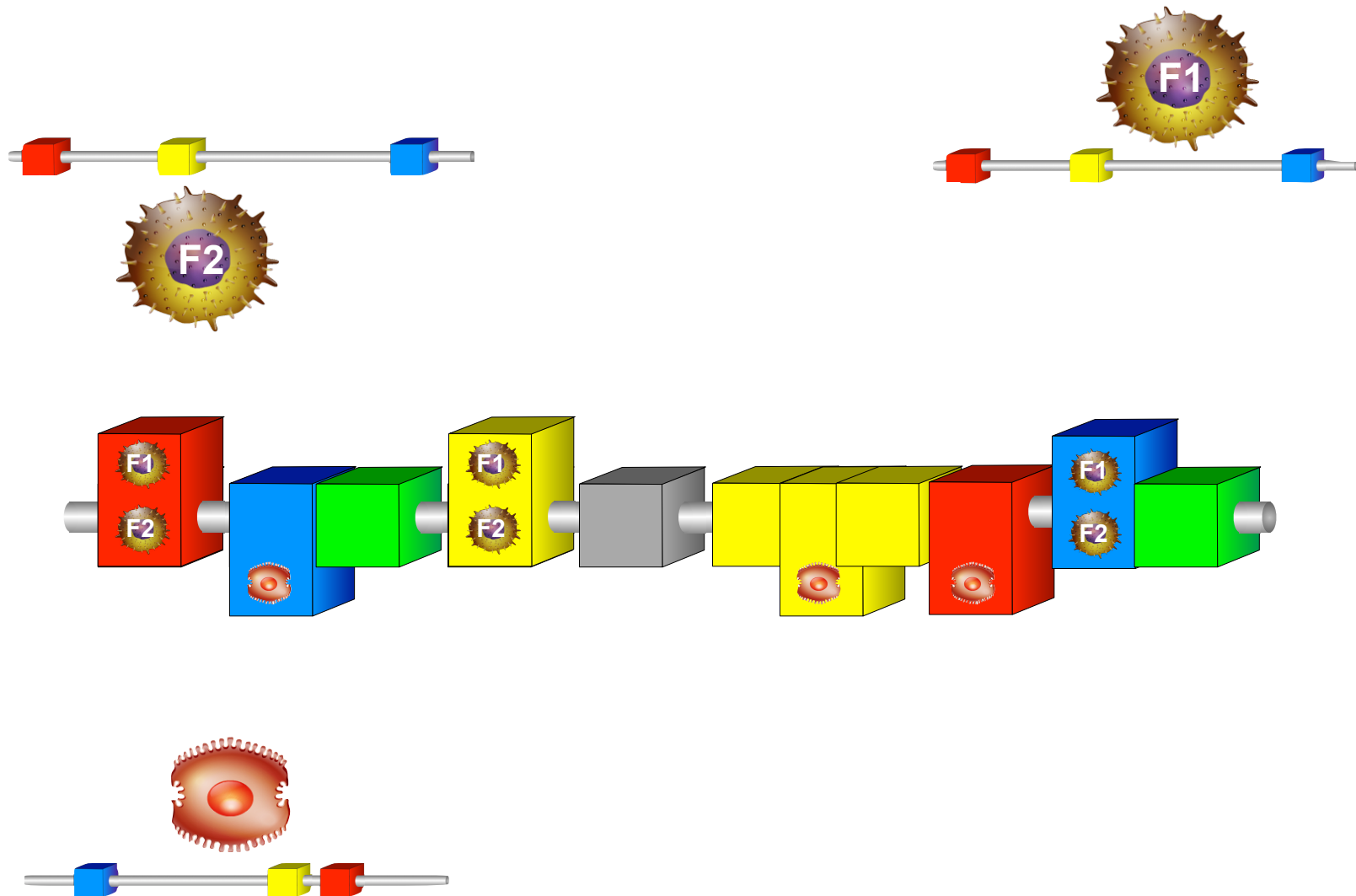
It is time to tell you MORE about gene regulation



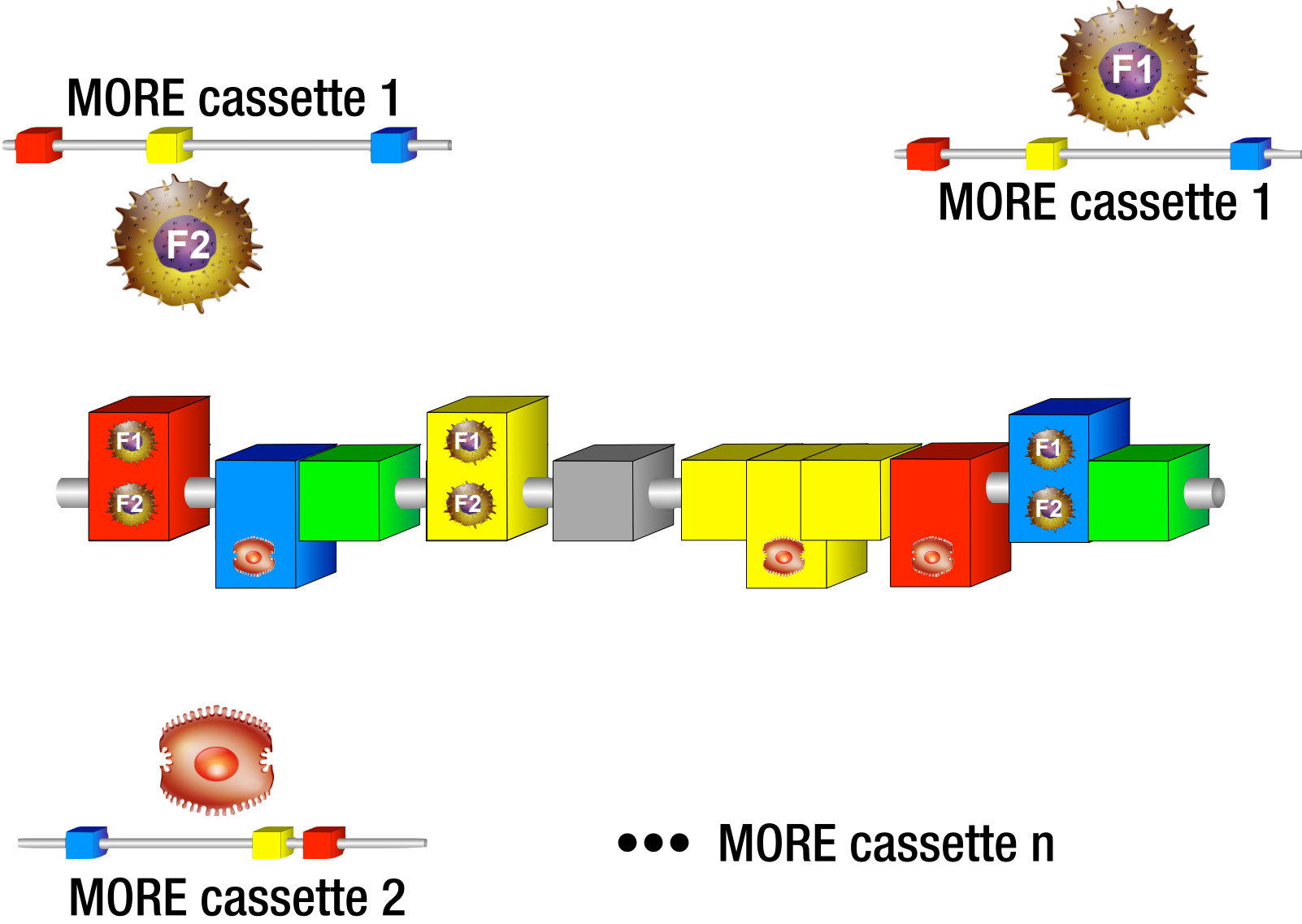
TFBSs are ubiquitous - present everywhere and every time



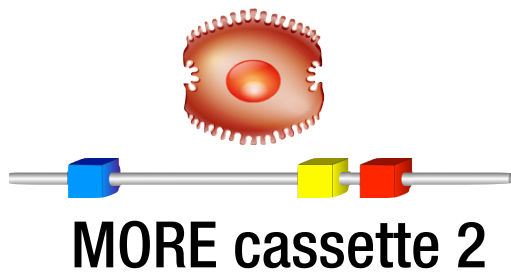
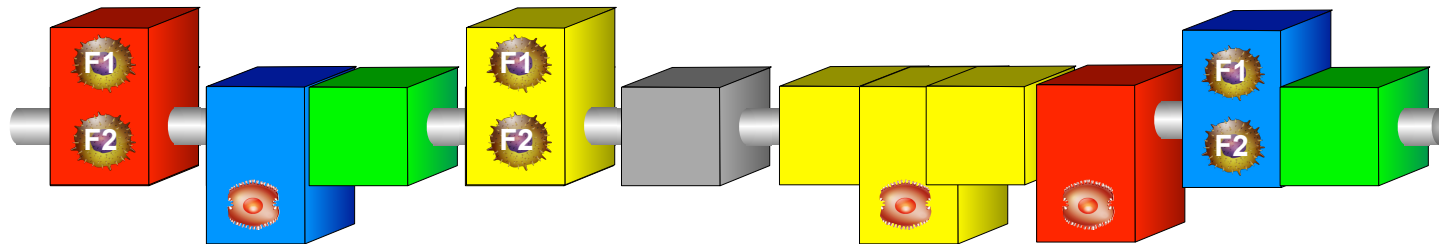
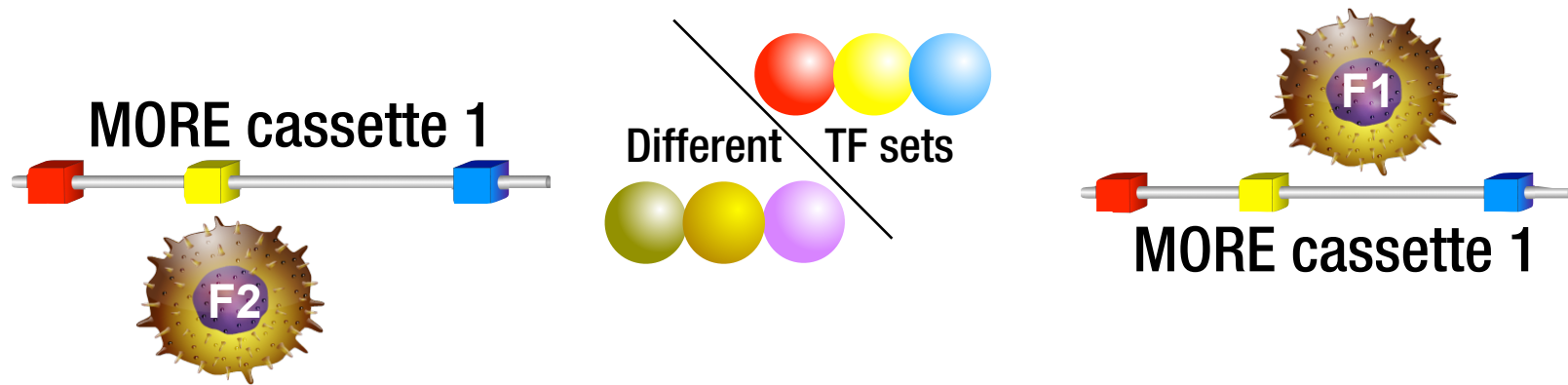
TFBSs are ubiquitous - present everywhere and every time



TFBSs are ubiquitous - present everywhere and every time




TFBSs are ubiquitous - present everywhere and every time



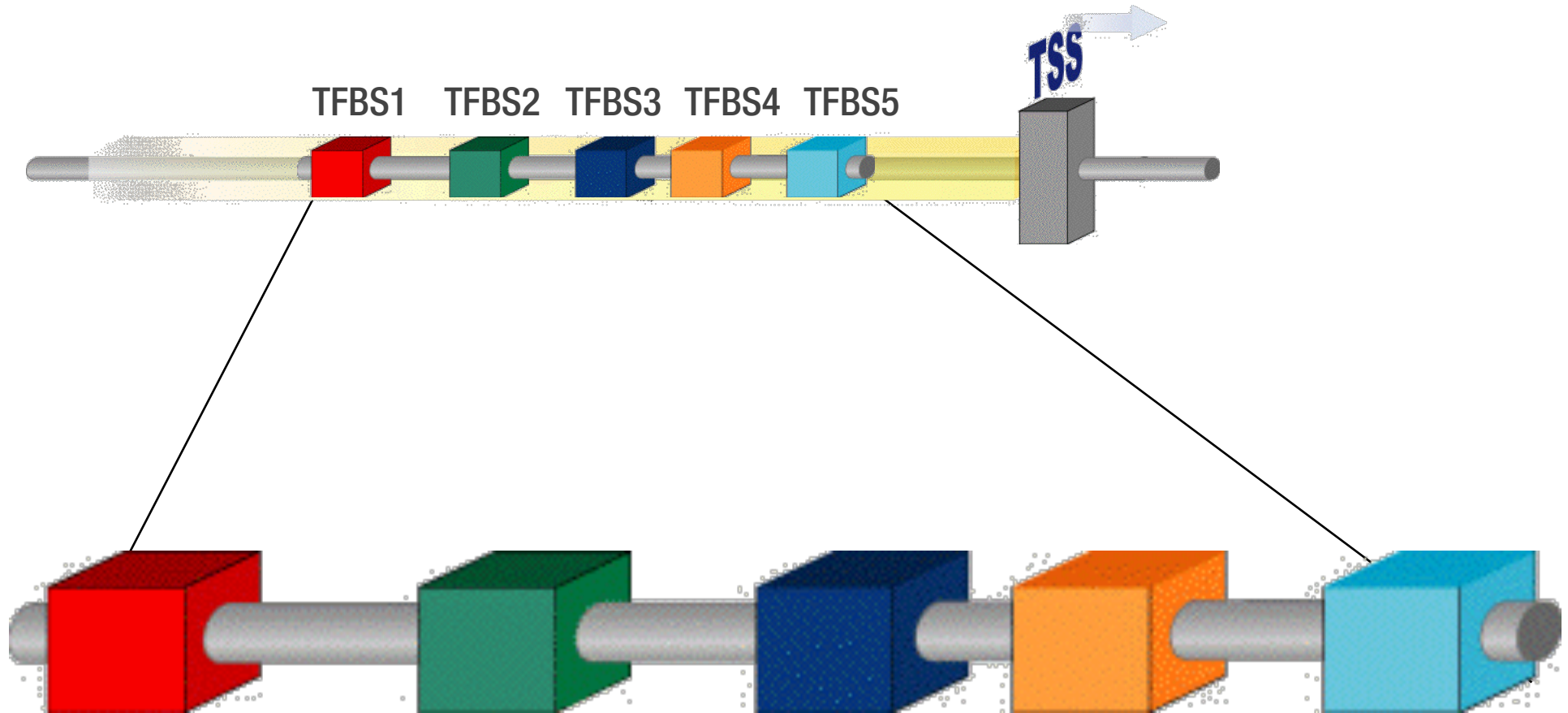
●●● MORE cassette n

A MORE cassette is a regulatory program written in Transcription Factor Binding Sites (TFBSs)

A MORE-cassette
has a clear structure




TFBSs order (1-n)
TFBSs distance ranges
Distance range variation
MORE-strand orientation



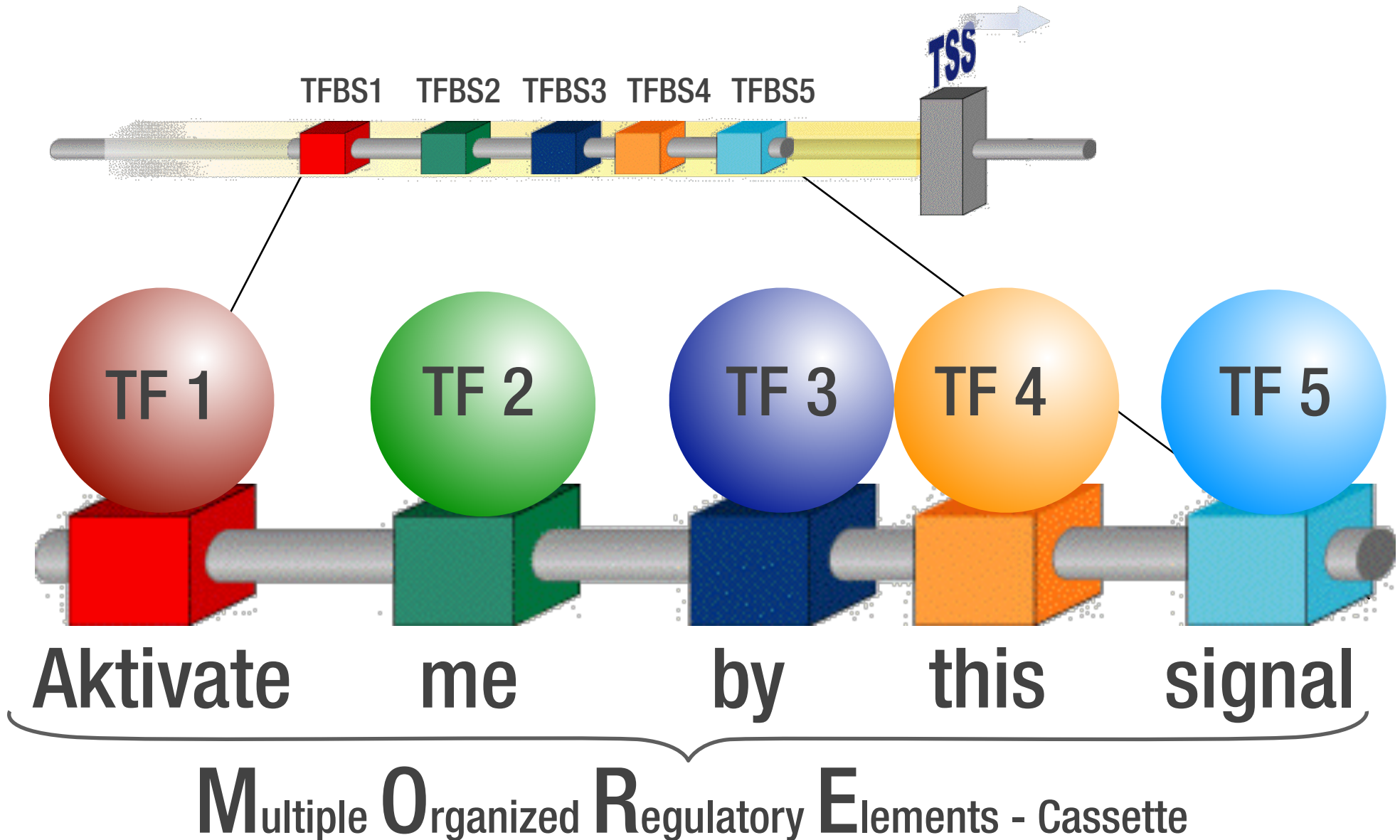
Multiple **O**rganized **R**egulatory **E**lements - Cassette

A MORE cassette is a regulatory program written in Transcription Factor Binding Sites (TFBSs)

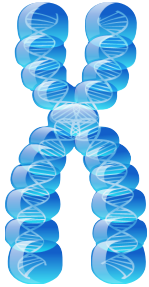
A MORE-cassette
has a clear structure



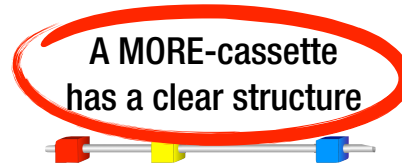
TFBSs order (1-n)
TFBSs distance ranges
Distance range variation
MORE-strand orientation



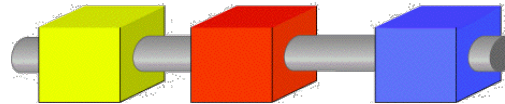
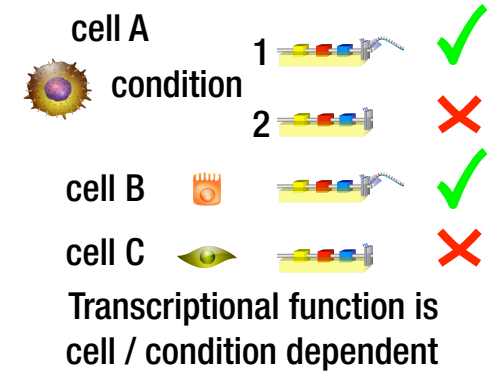
Basic properties of MORE-cassettes



Invariant part of genomic sequence
in every cell and all the time



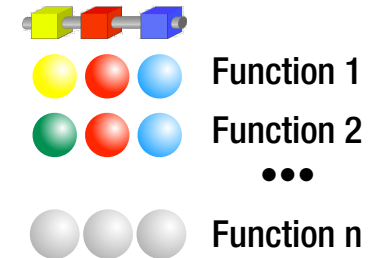
TFBSs order (1-n)
TFBSs distance ranges
Distance range variation
MORE-strand orientation



A MORE-cassette
represents context!



Individual TFBSs can be part
of multiple MORE-cassettes

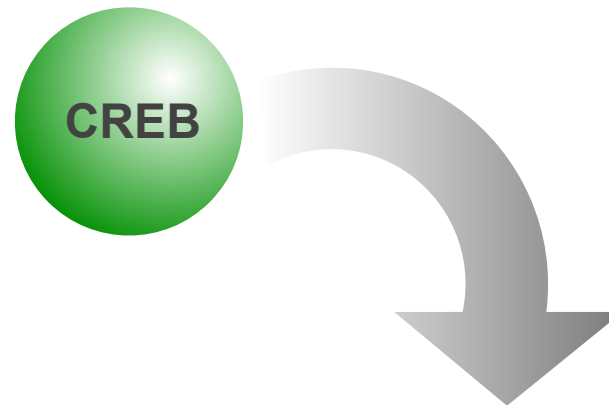


Function emerges upon binding
of TF-sets NOT individual TFs

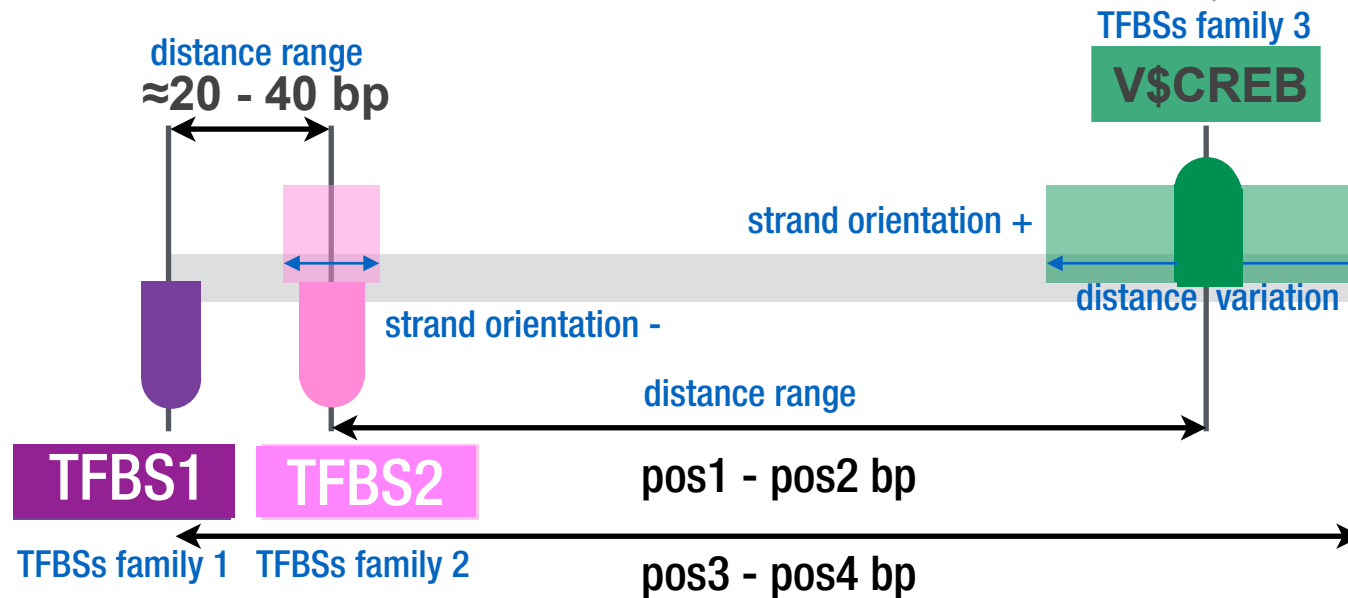
Neither TFs nor TFBSs are functional without context!

A tale of a transcription factor (CREB)

CREB is involved in hundreds of different biological processes, among them:
circadian rhythm



CREB binding sites are present in 96.2 % of all vertebrate promoters!

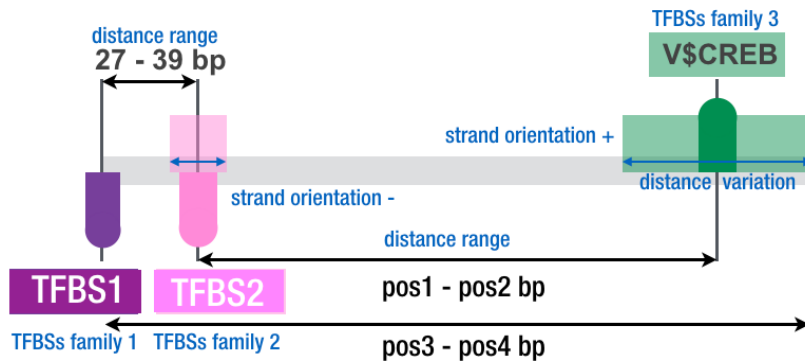


Associated with the circadian regulation of TF1 by CREB

Multiple Organized Regulatory Elements (MORE) - cassette

(formerly known as TFBSs framework)

The structure of a MORE cassette is the basis of the function

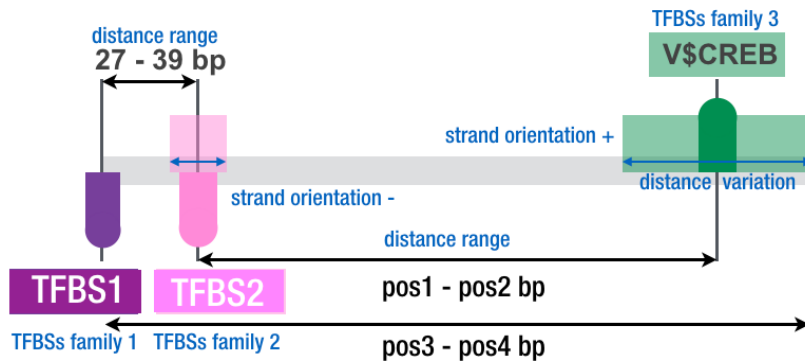


Selectivity: 1/ 2360

- 24 matches in 56656 rat promoters
- 5 are CREB regulated in circadian rhythm (in the pineal gland, 2019 promoters)
- 5.8 fold overrepresented

simple

The structure of a MORE cassette is the basis of the function

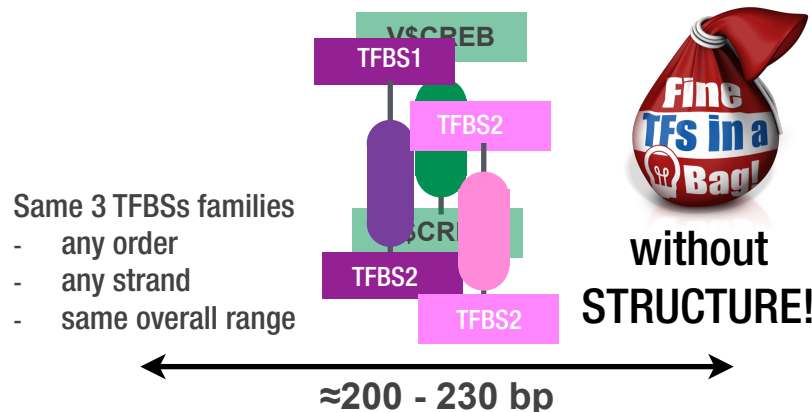


Selectivity: 1/ 2360

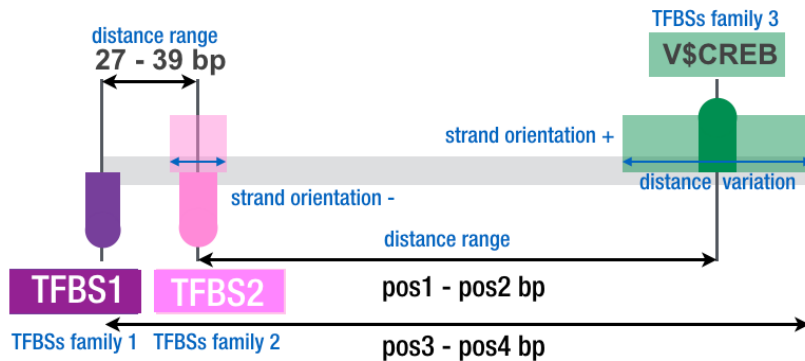
- 24 matches in 56656 rat promoters
- 5 are CREB regulated in circadian rhythm (in the pineal gland, 2019 promoters)
- 5.8 fold overrepresented

simple

no order, any distance!



The structure of a MORE cassette is the basis of the function



Selectivity: 1/ 2360

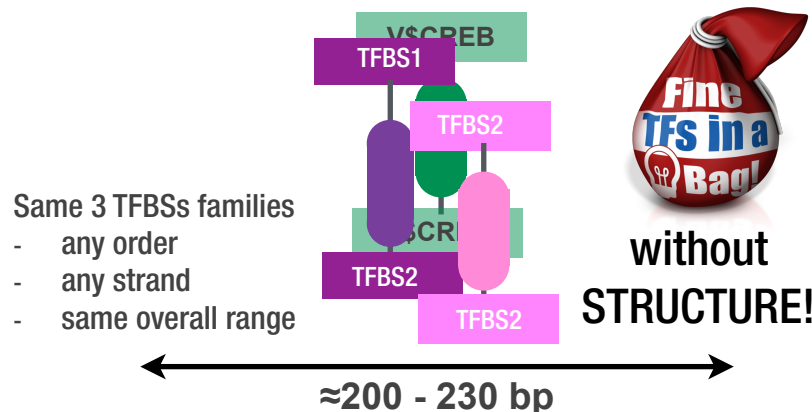
- 24 matches in 56656 rat promoters
- 5 are CREB regulated in circadian rhythm (in the pineal gland, 2019 promoters)
- 5.8 fold overrepresented

simple
including structure!

context

emergence of function

no order, any distance!



too simple

~~Selectivity: 1/ 3~~

- > 18,000 matches in 56656 rat promoters
- 714 are CREB regulated in circadian rhythm (in the pineal gland, 2019 promoters)
- not overrepresented (1.1 fold)

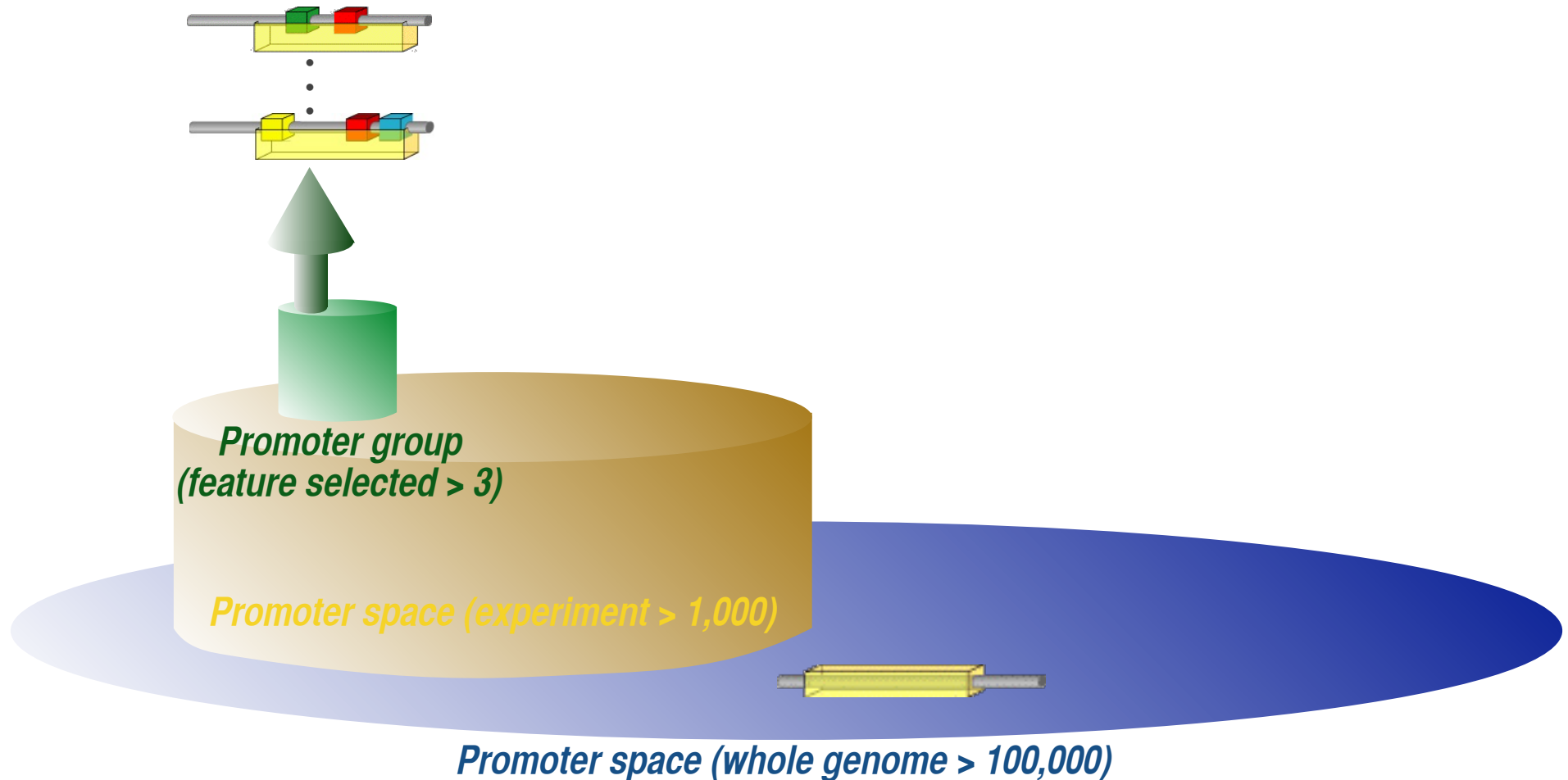
MORE cassettes are part of the genomic sequence!

All other promoters or sequences containing the same MORE-cassette can be found



Invariant part of genomic sequence
in every cell and all the time

Definition of MORE-cassettes



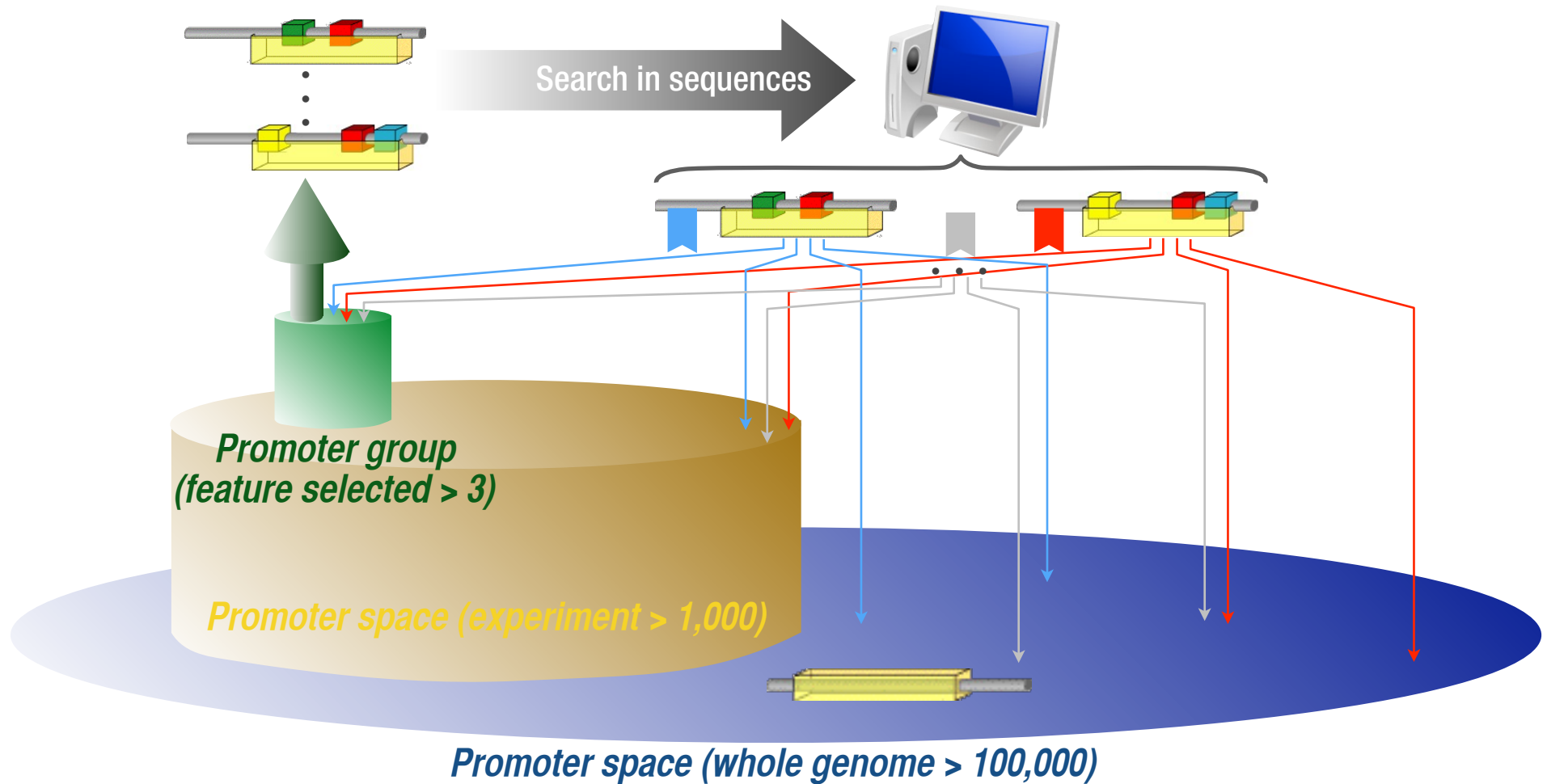
MORE cassettes are part of the genomic sequence!

All other promoters or sequences containing the same MORE-cassette can be found



Invariant part of genomic sequence
in every cell and all the time

Definition of MORE-cassettes



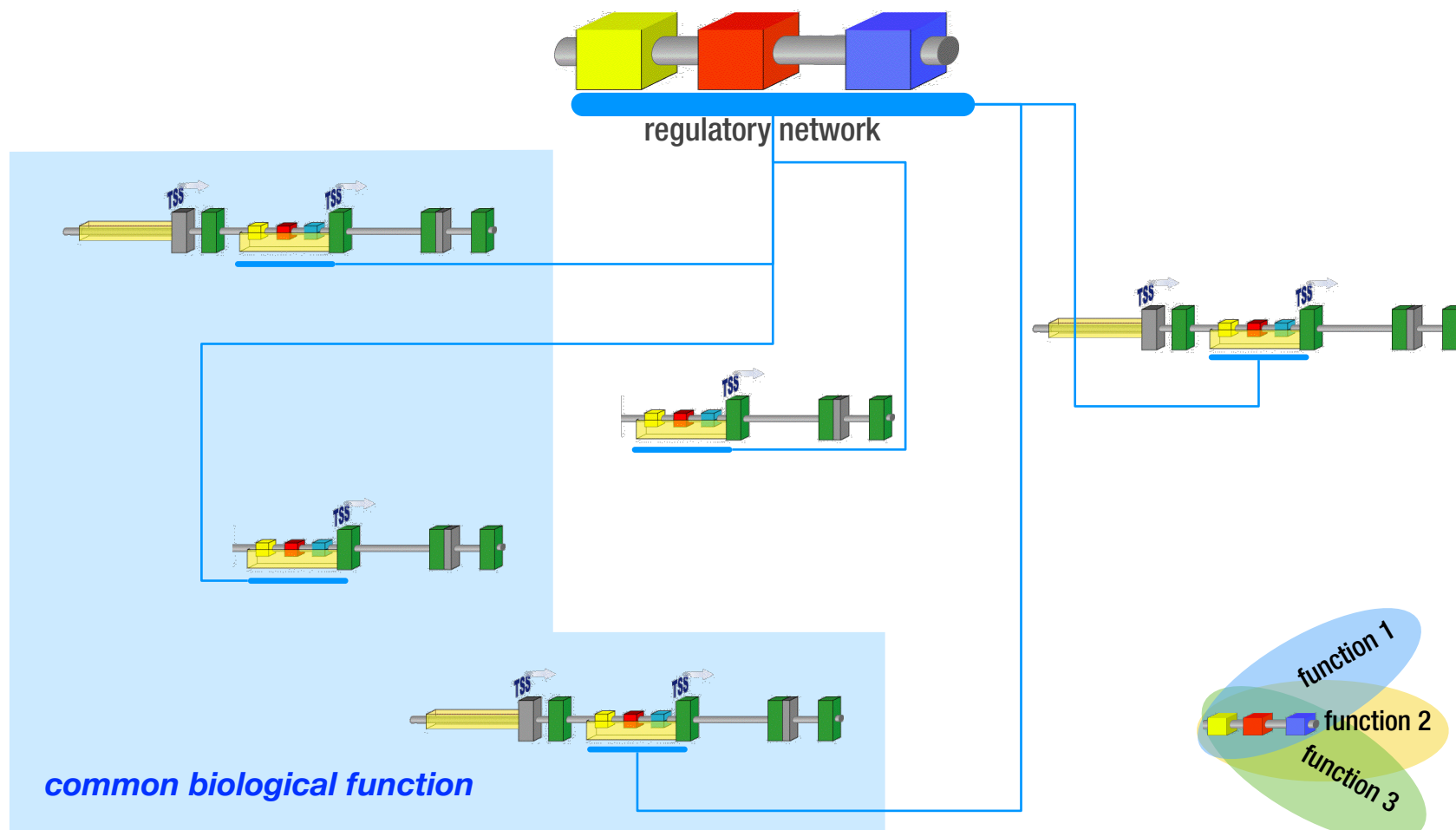
MORE-cassettes find ALL matching promoters regardless of the initial set



Individual TFBSs can be part of multiple MORE-cassettes

MORE cassettes represent regulatory networks!

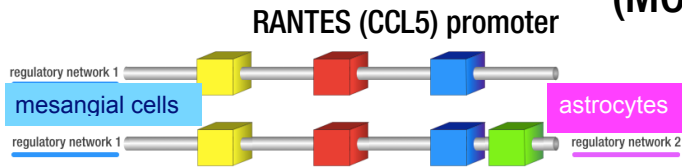
Promoters with the same MORE cassettes can be part of the same functional network



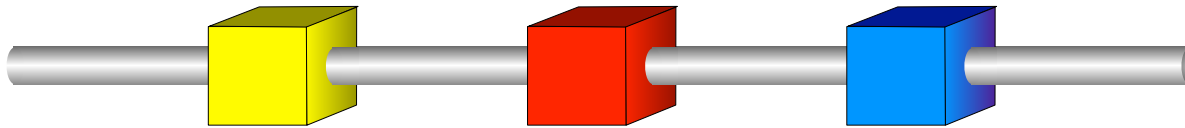
MORE-cassettes are function-associated but usually not function-specific!

The most universal feature of languages is redundancy

(MORE-cassettes are reused and extended)



See that ridge



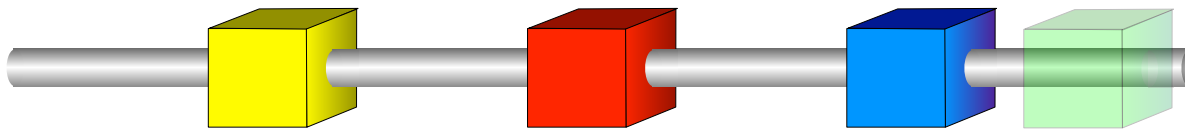
Individual TFBSs can be part of multiple MORE-cassettes

The most universal feature of languages is redundancy

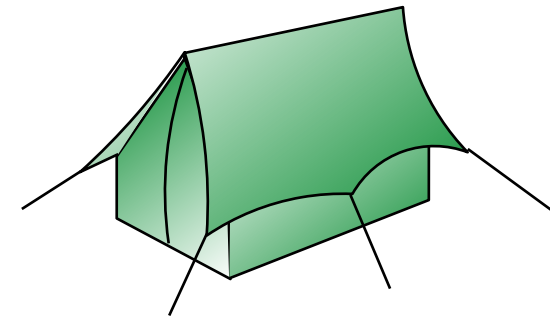
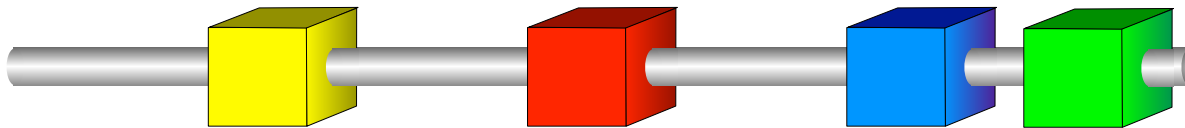
(MORE-cassettes are reused and extended)



See that ridge



See that ridge tent



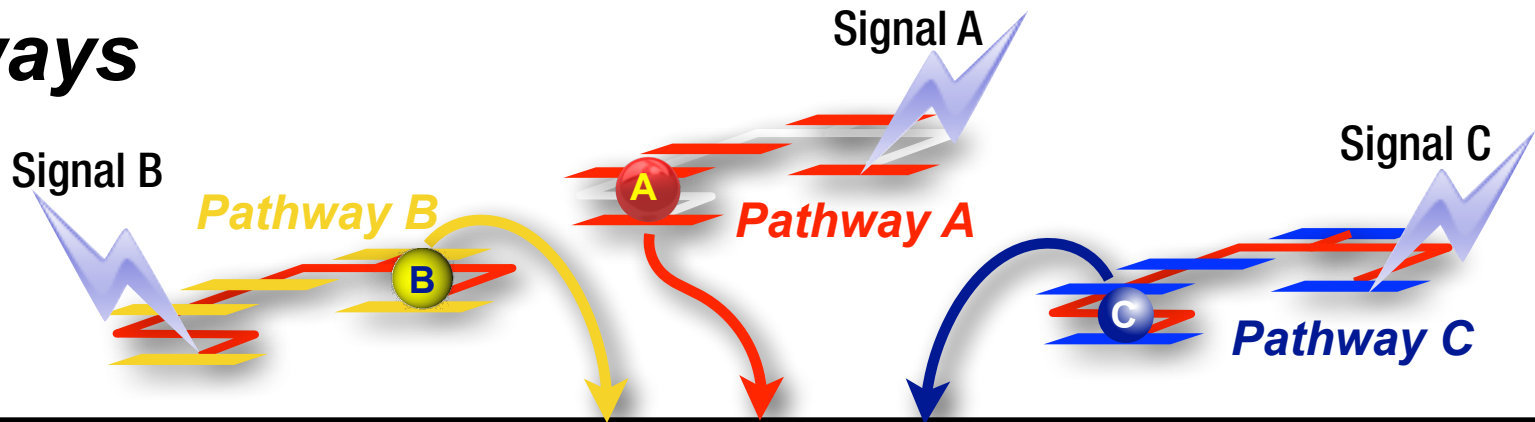
Emergence of function is context-dependent!



Individual TFBSs can be part of multiple MORE-cassettes

MORE-cassettes and signaling pathway form networks

pathways



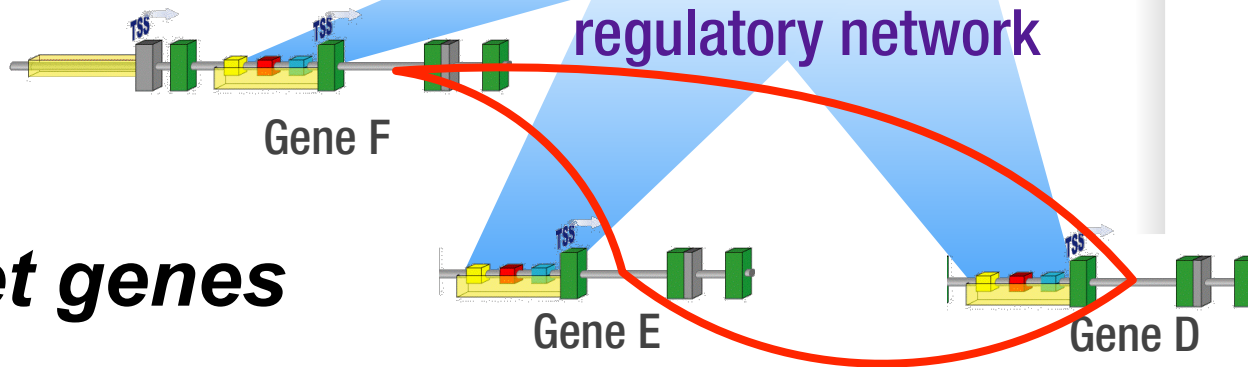
MOREs



regulatory network

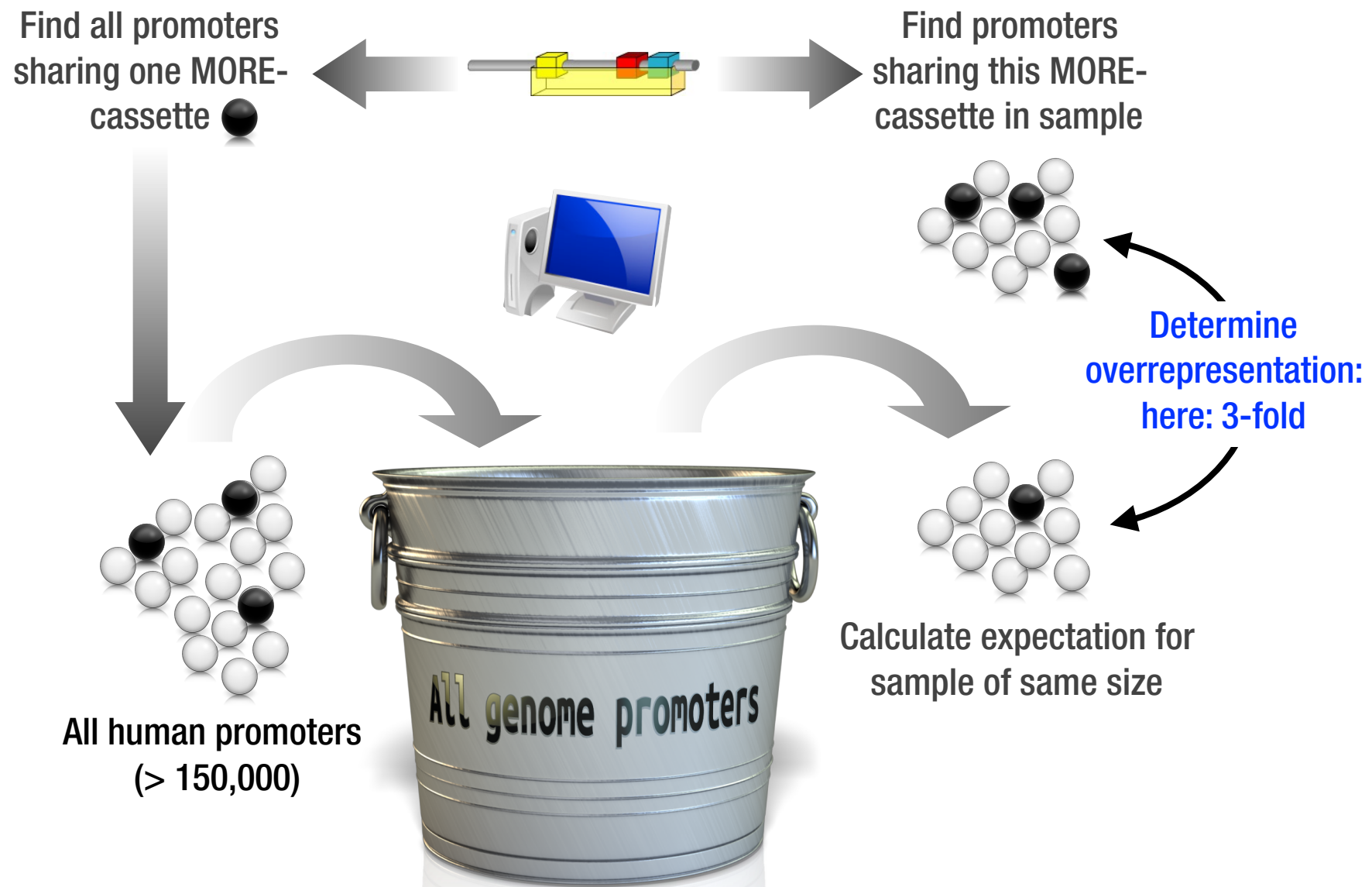
metabolic network

target genes



Relevant MORE cassettes are associated with gene sets

Gene promoters can contain MOREs ● or not ○





OK, so MORE-cassettes finally explain regulation of gene transcription?

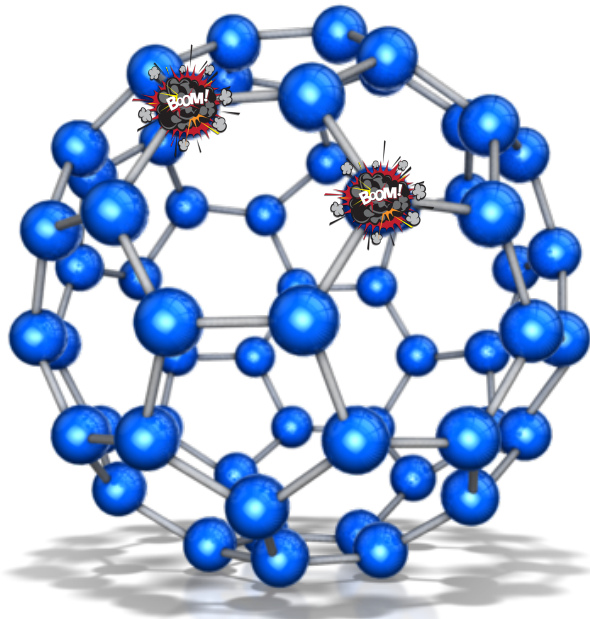
They are definitely part of the answer, but only part!

*The belief that there is only one truth,
and that oneself is in possession of it,
is the root of all evil in the world.*

Max Born

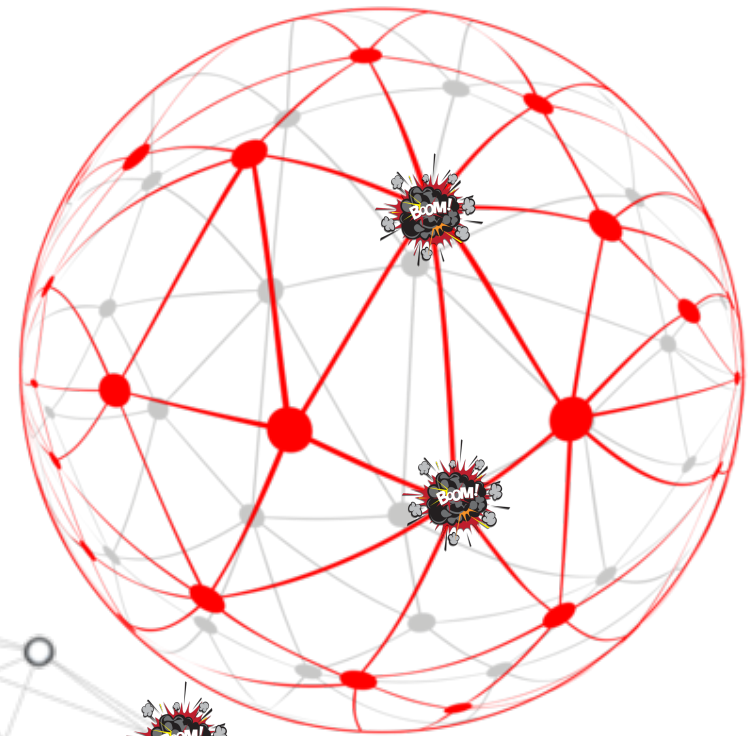
What happens with genomic mutations / deletions?

regulatory network

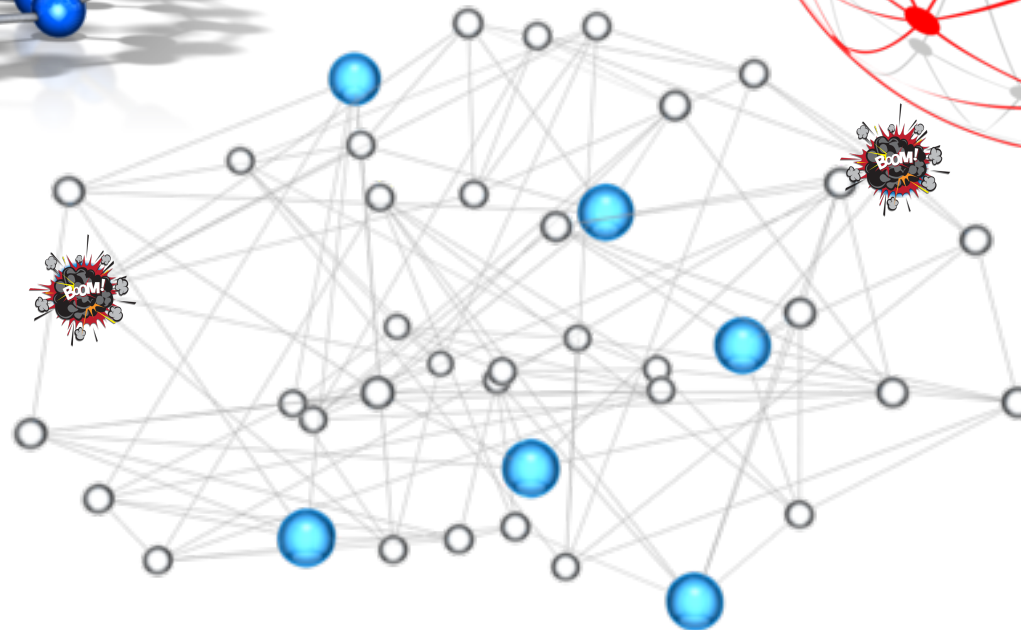


BOOM!
knockout

metabolic network



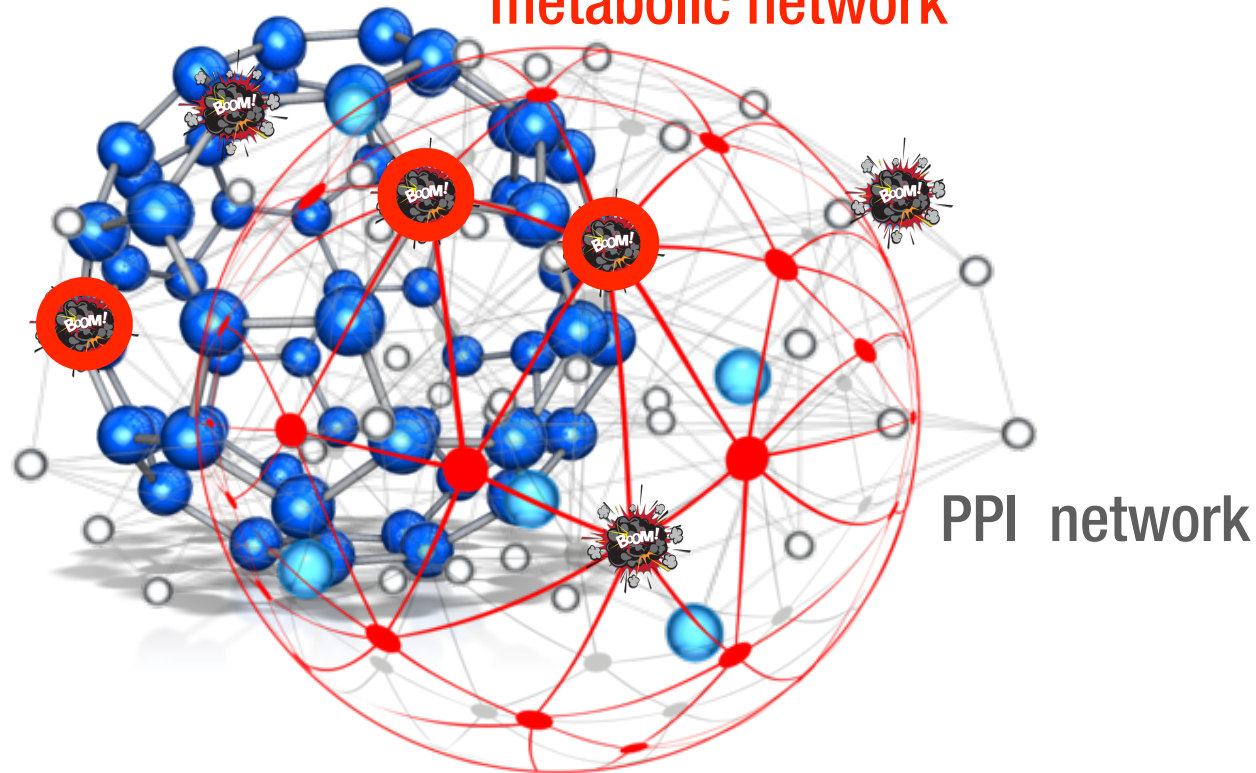
PPI network



What happens with genomic mutations / deletions?

regulatory network

metabolic network



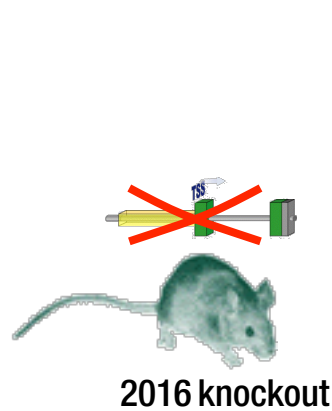
some genes are connected in one network



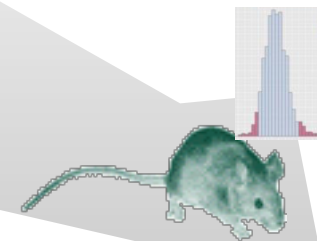
some genes are connected in multiple networks

Use the knockout genes to find the networks involved!

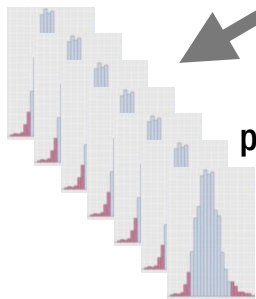
The regulatory mechanisms behind phenotypes



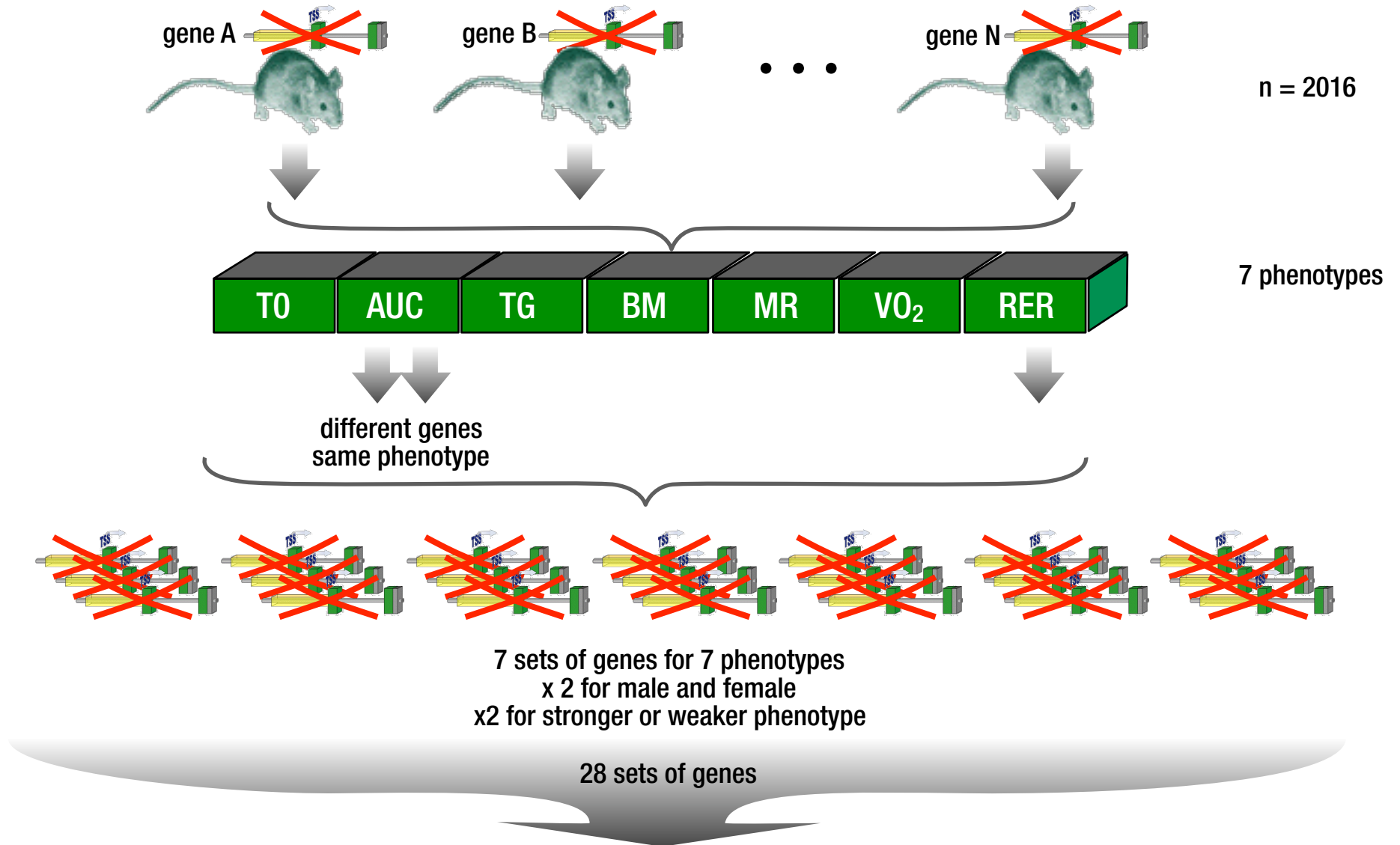
Phenotyping



Is there a MORE-phenotype correlation?



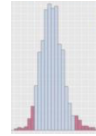
IMPC experimental strategy



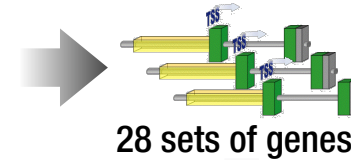
Regulatory network analysis

Helmholtz / Werner analytical strategy - MORE cassettes

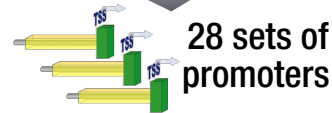
strong phenotype
(974 genes)



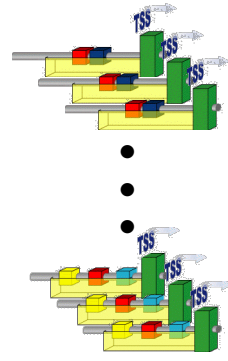
Ranking/Determination
of 28 outlier lists
0%-5%, 95%-100%



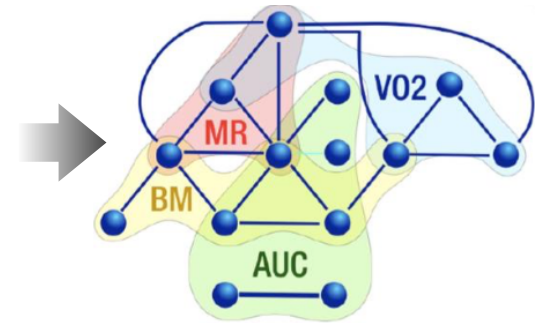
28 sets of genes



28 sets of
promoters

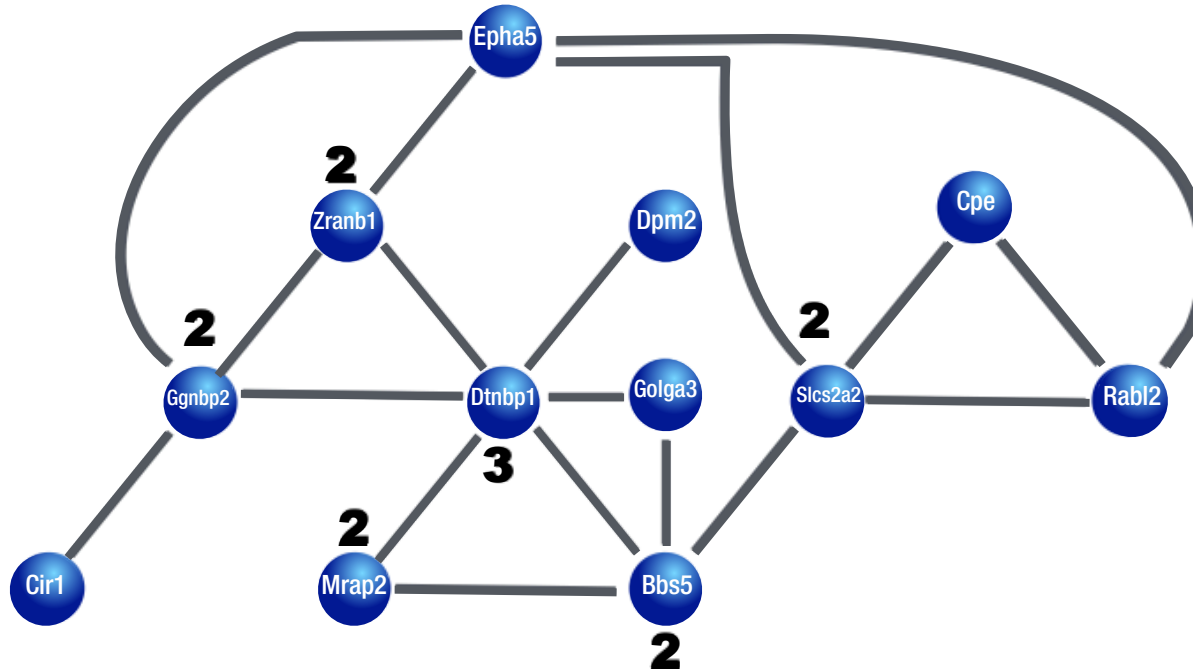


Promoter analysis
(225 MORE-sets
428 genes, 44%)

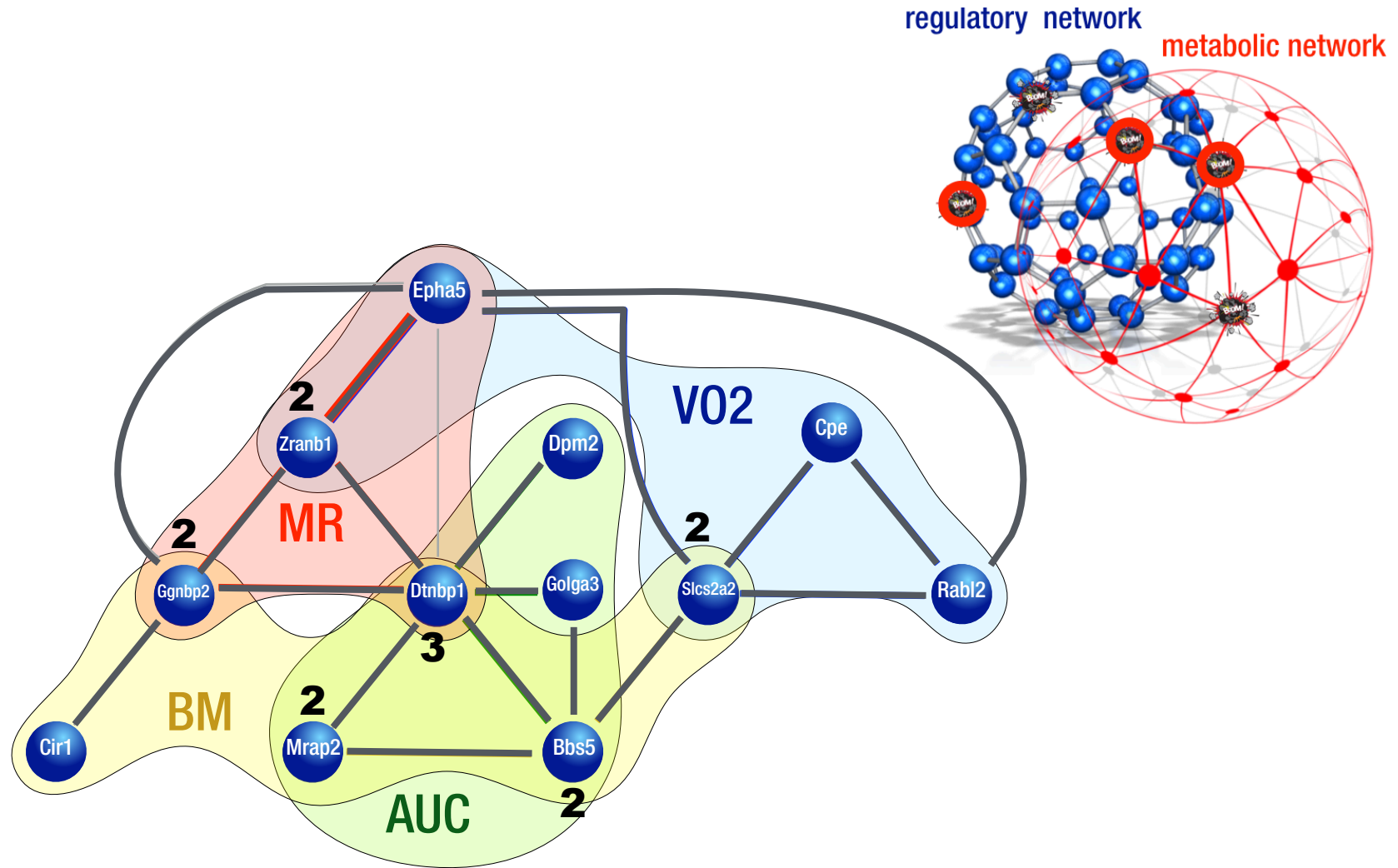


MORE-set /
phenotype networks

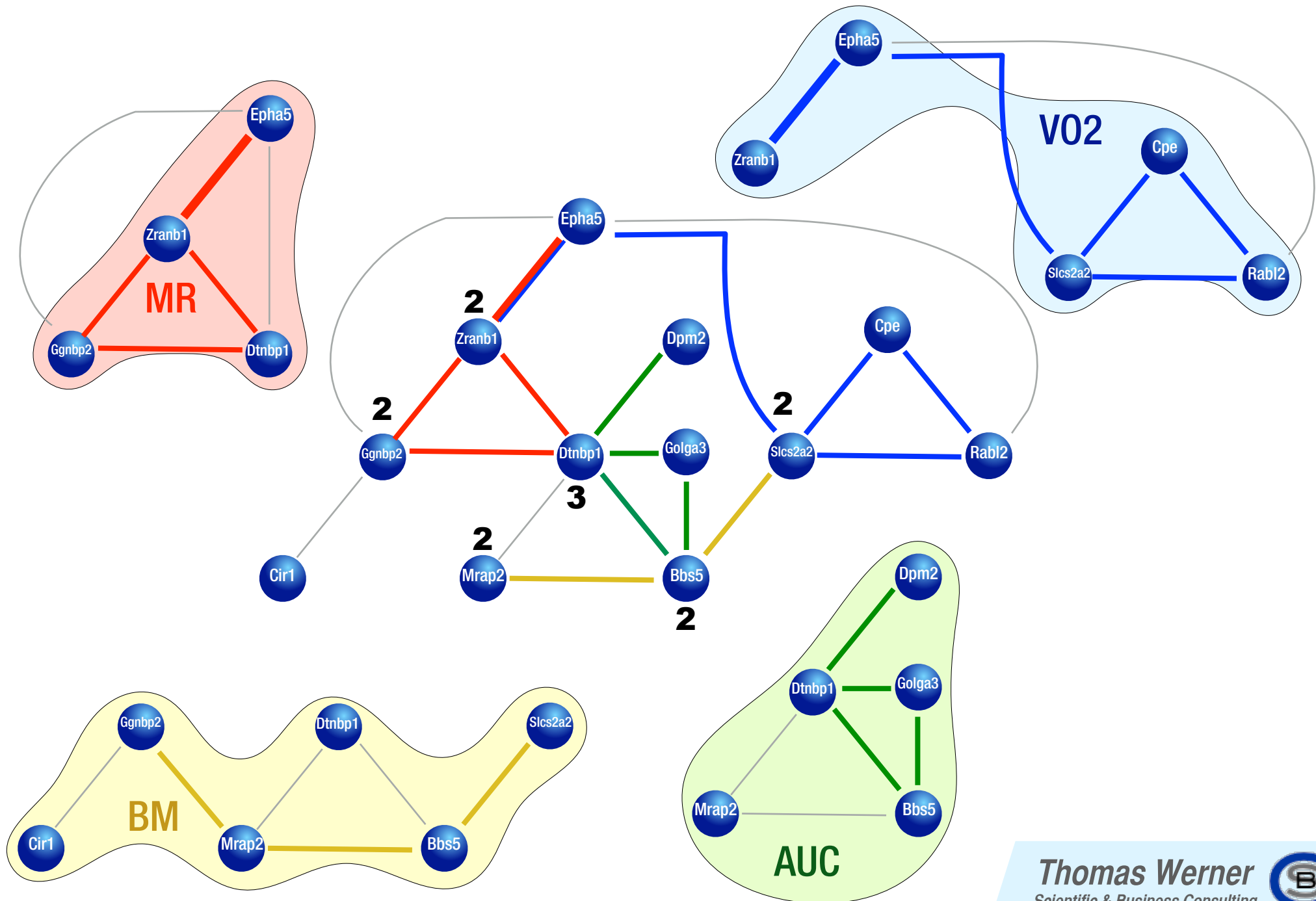
Knockout genes are connected by MOREs



Knockout genes are connected by MOREs



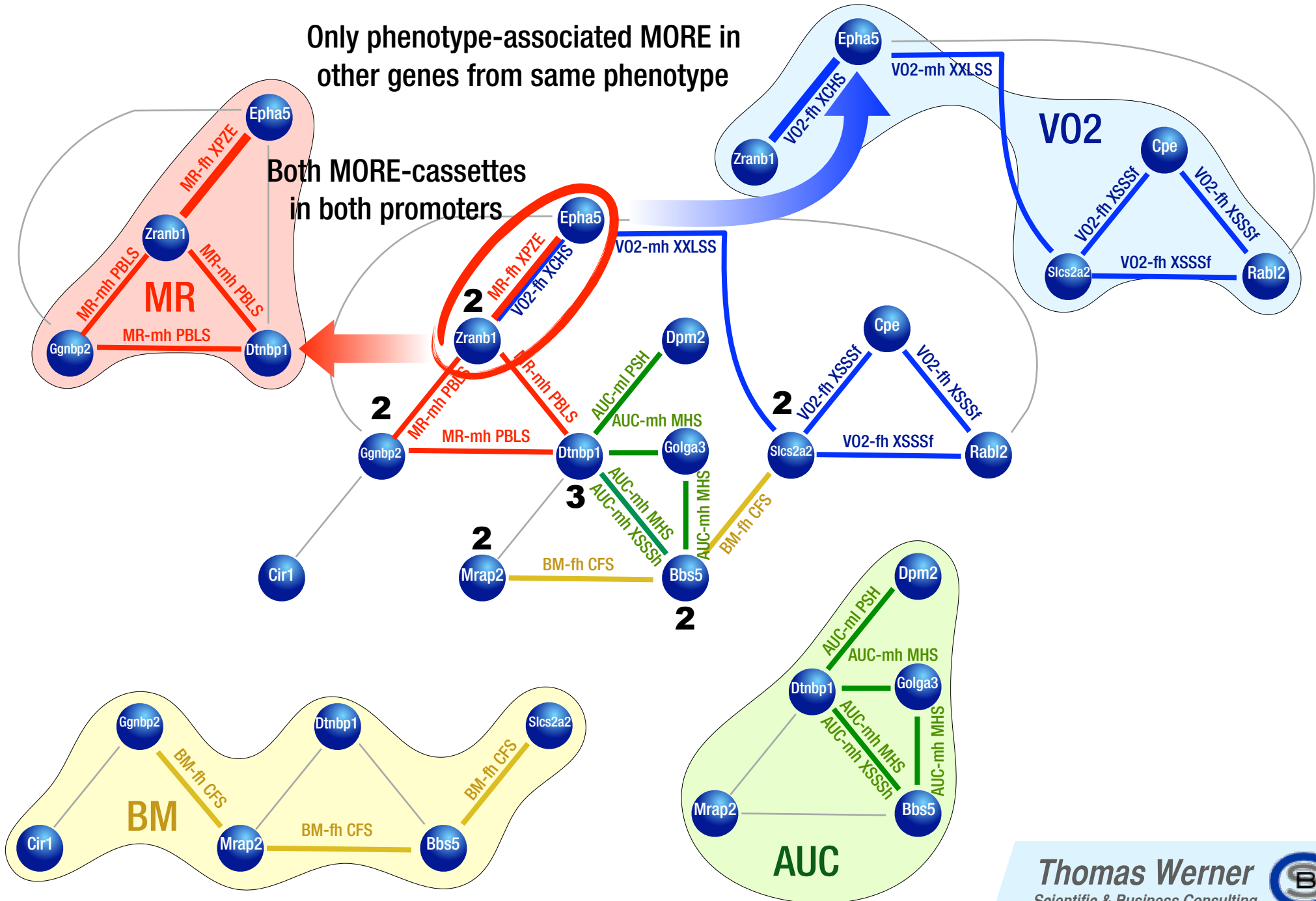
Knockout genes are connected by phenotype-related MOREs



Knockout genes are connected by phenotype-related MOREs

Only phenotype-associated MORE in other genes from same phenotype

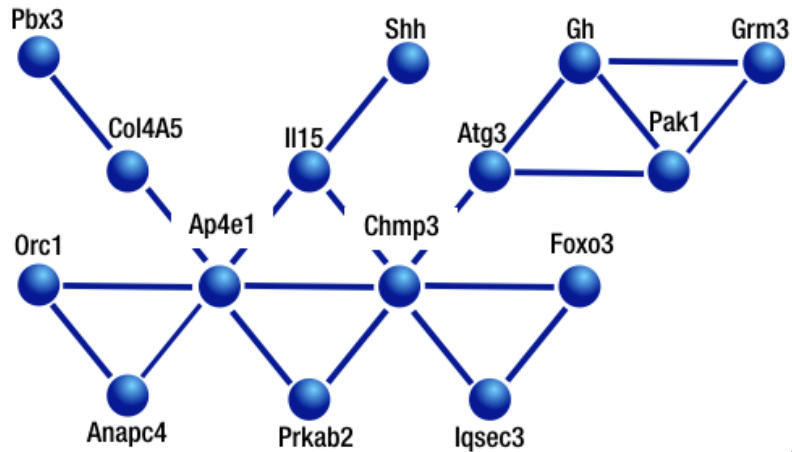
Both MORE-cassettes in both promoters



Regulatory and metabolic networks of knockout genes overlap

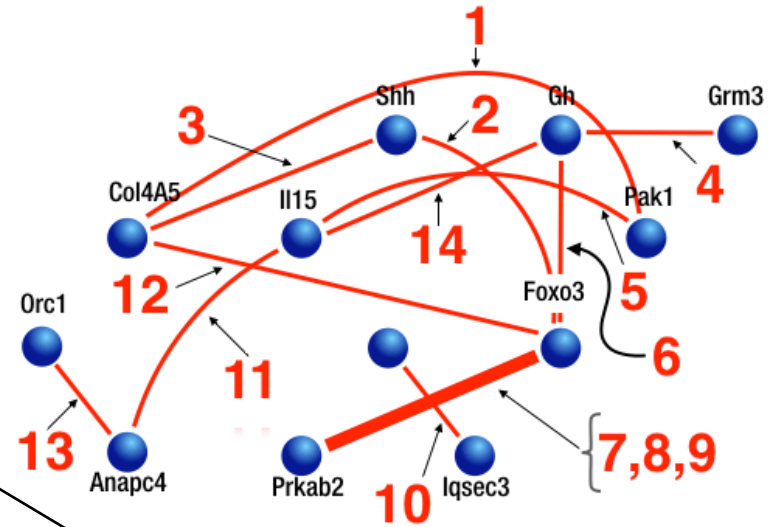
Regulatory network

MORE-set / phenotype network (AUC)

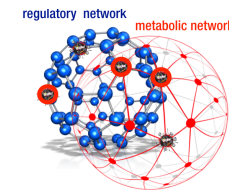
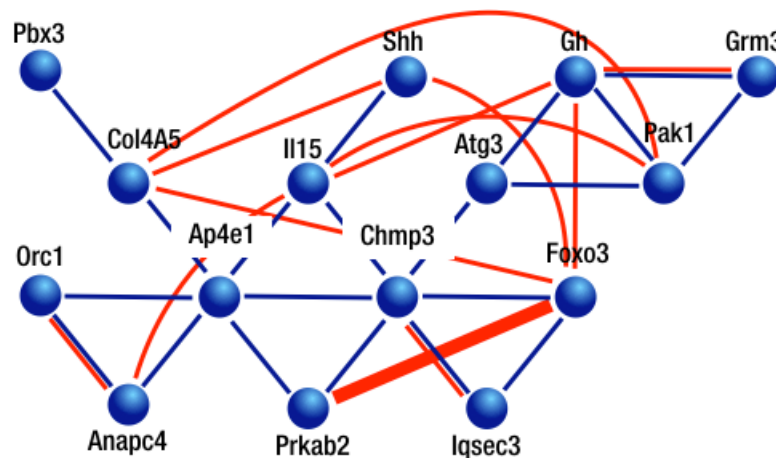


Metabolic network

KEGG pathways / phenotype network



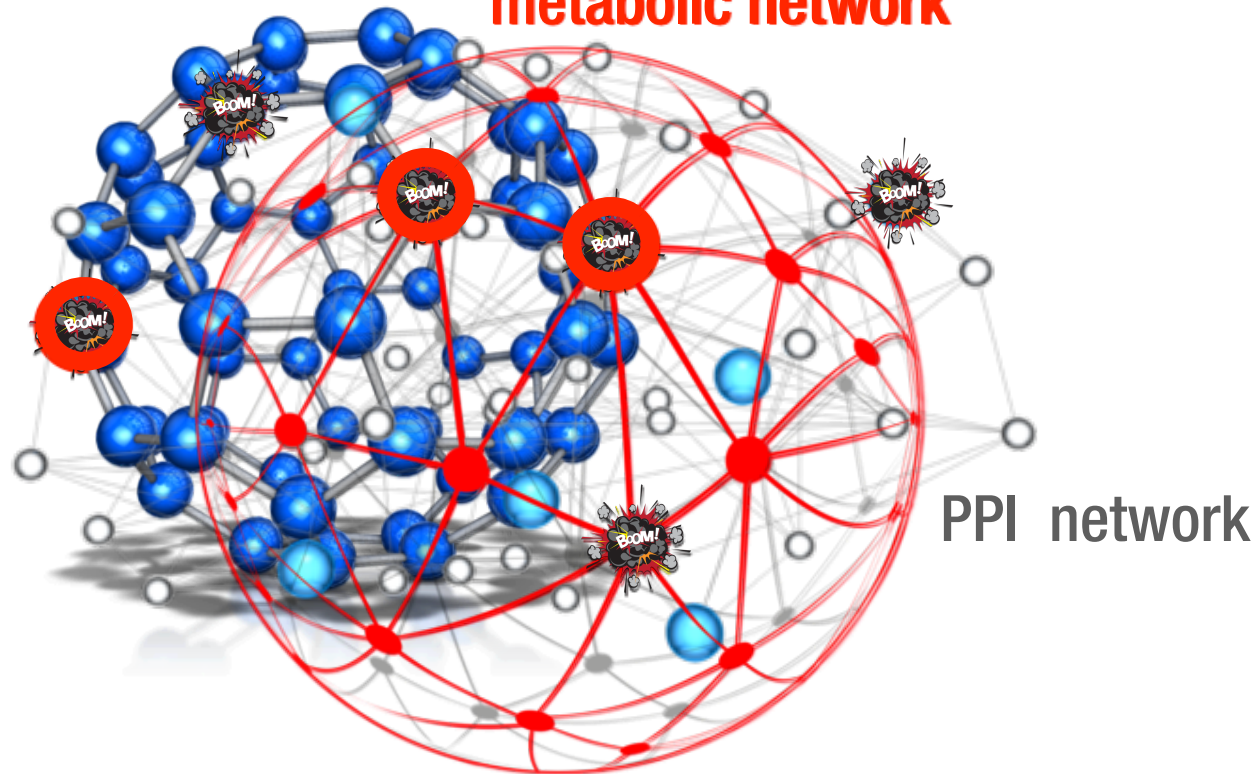
Composite network



What happens with genomic mutations / deletions?

regulatory network

metabolic network



some genes are connected in one network



some genes are connected in multiple networks

Use the knockout genes to find the networks involved...

...and use the networks to find genes!

What about genes we had not seen before?

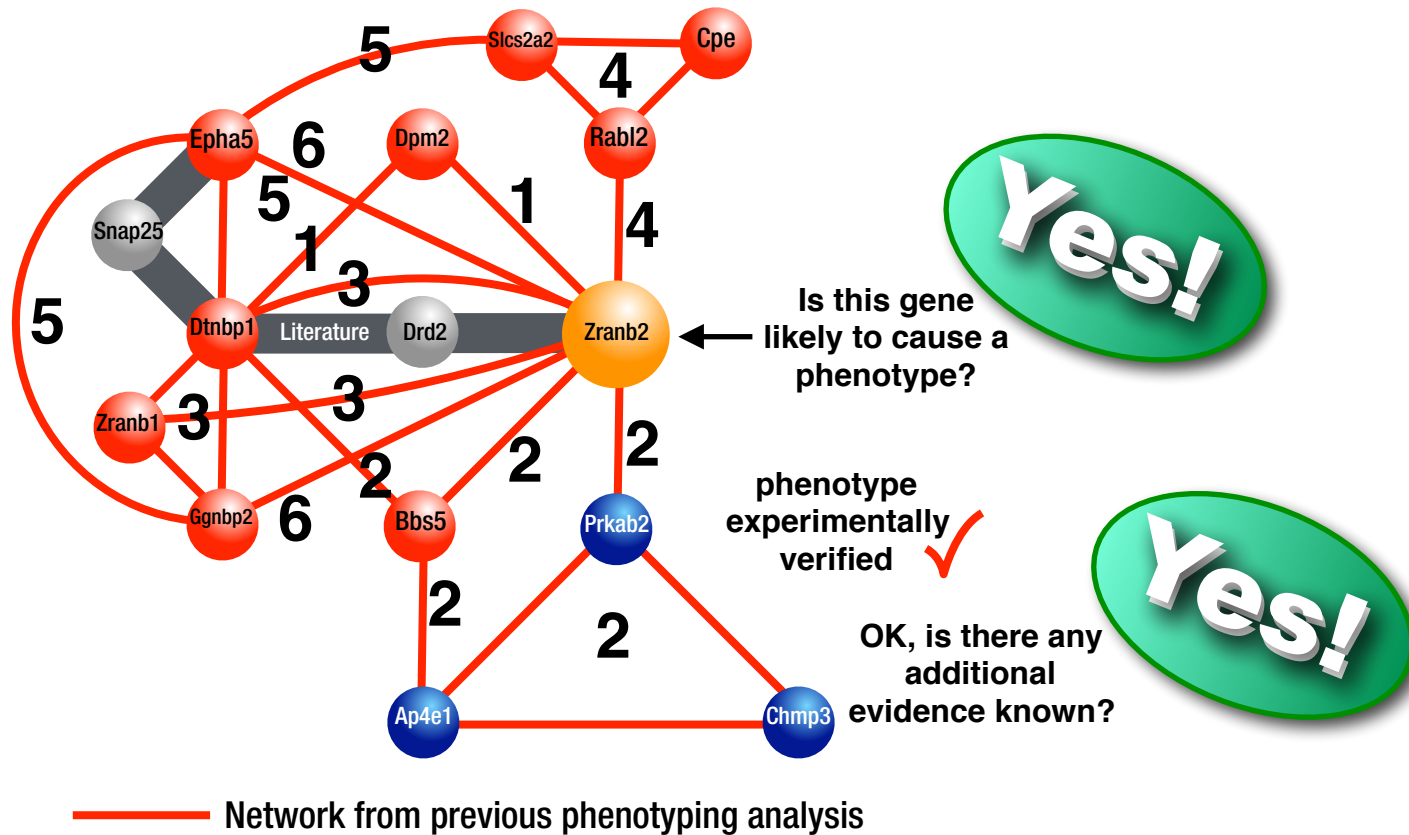


Zranb2

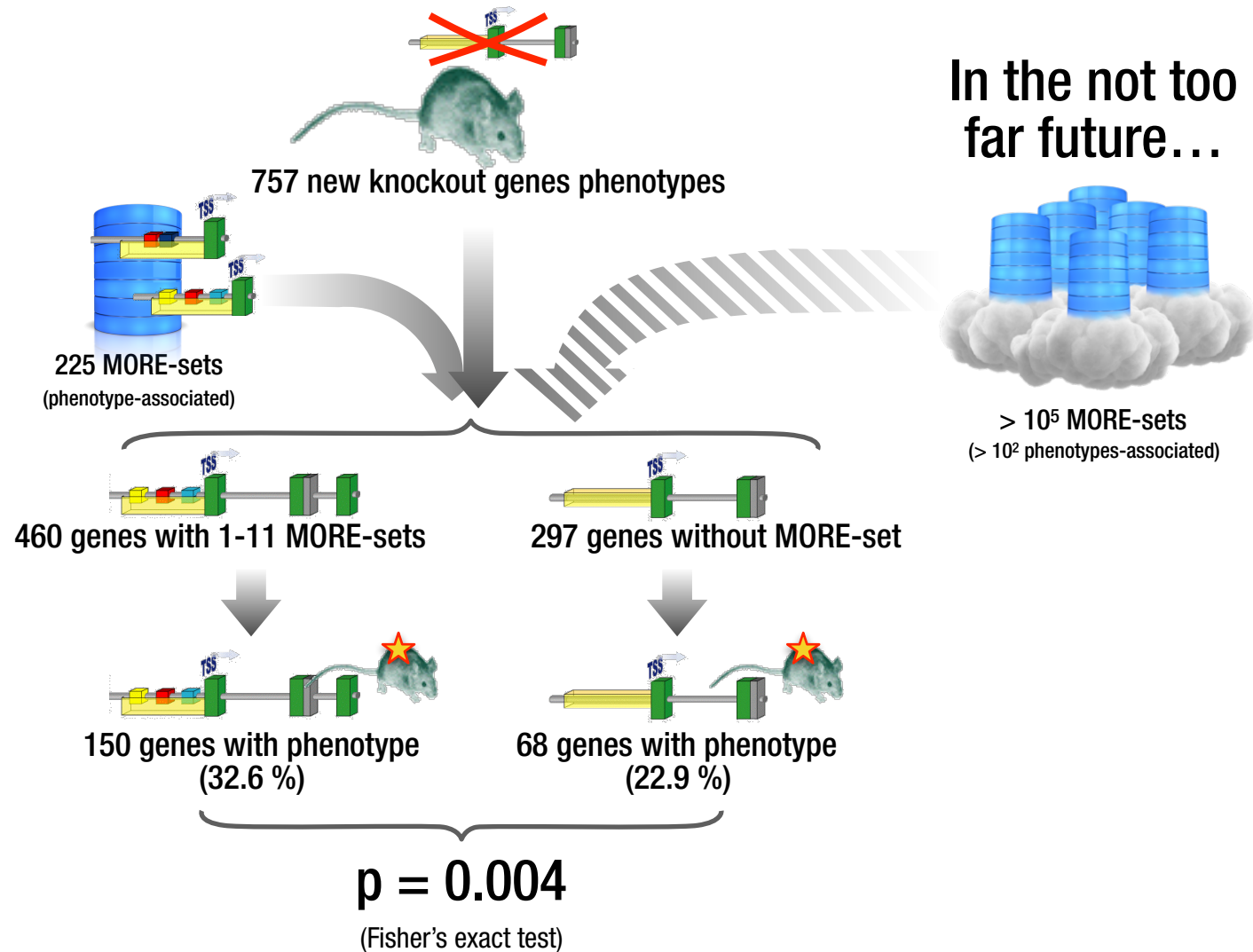


Is this gene
likely to cause a
phenotype?

What about genes we had not seen before?



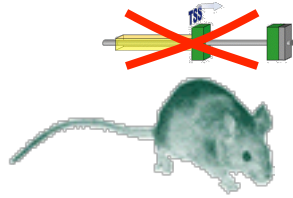
Phenotype prediction of knockout - by MORE cassettes!



In the not too far future...

MOREs connect more than a few genes, they can indicate a phenotype!

Predicting vs verification of phenotypes



> 10³ knockout

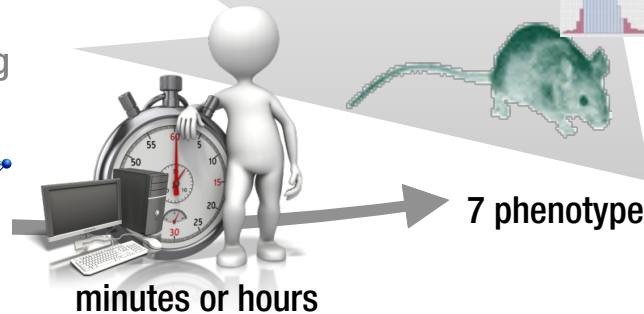
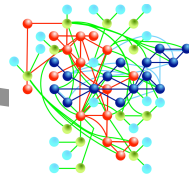


MOREs



MORE-phenotype correlation

Phenotyping



7 phenotypes

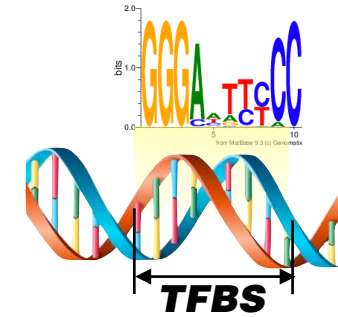
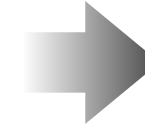
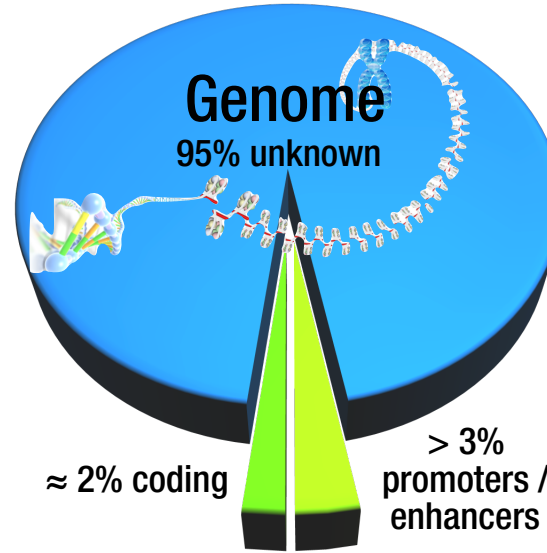
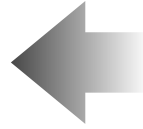
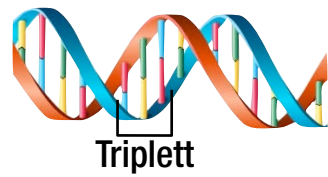
minutes or hours

Predict phenotype from regulation using the Regulatory Code

Genetic Code

versus

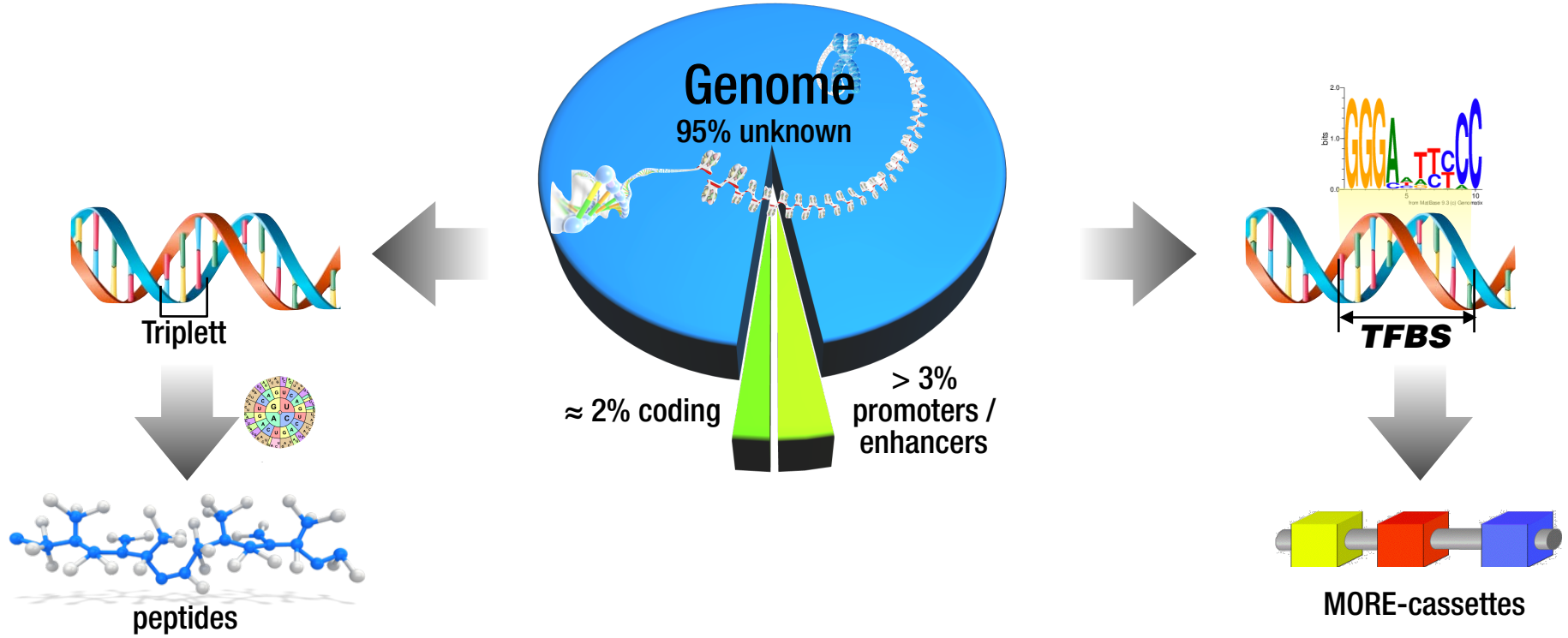
Regulatory Code



Genetic Code

versus

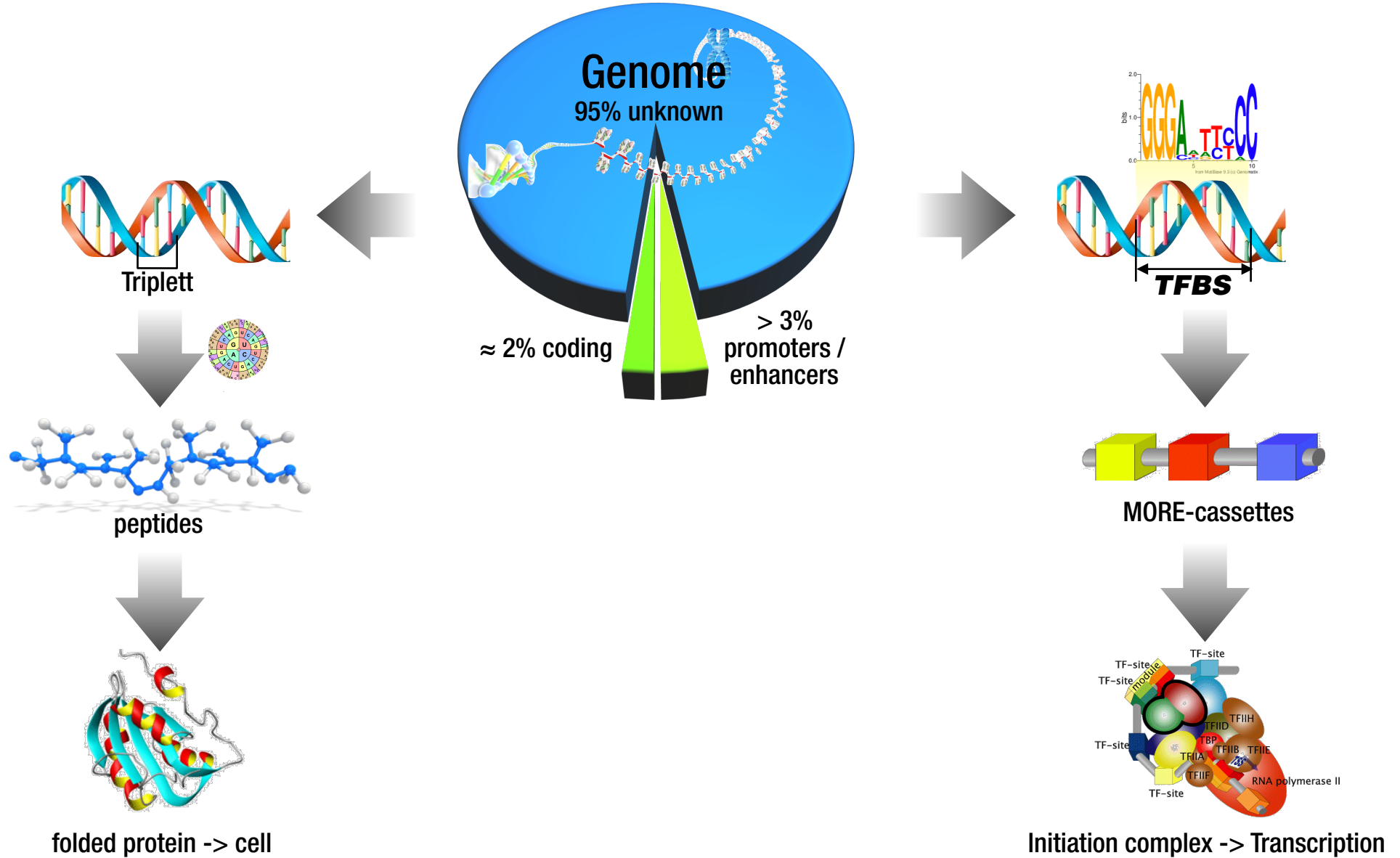
Regulatory Code



Genetic Code

versus

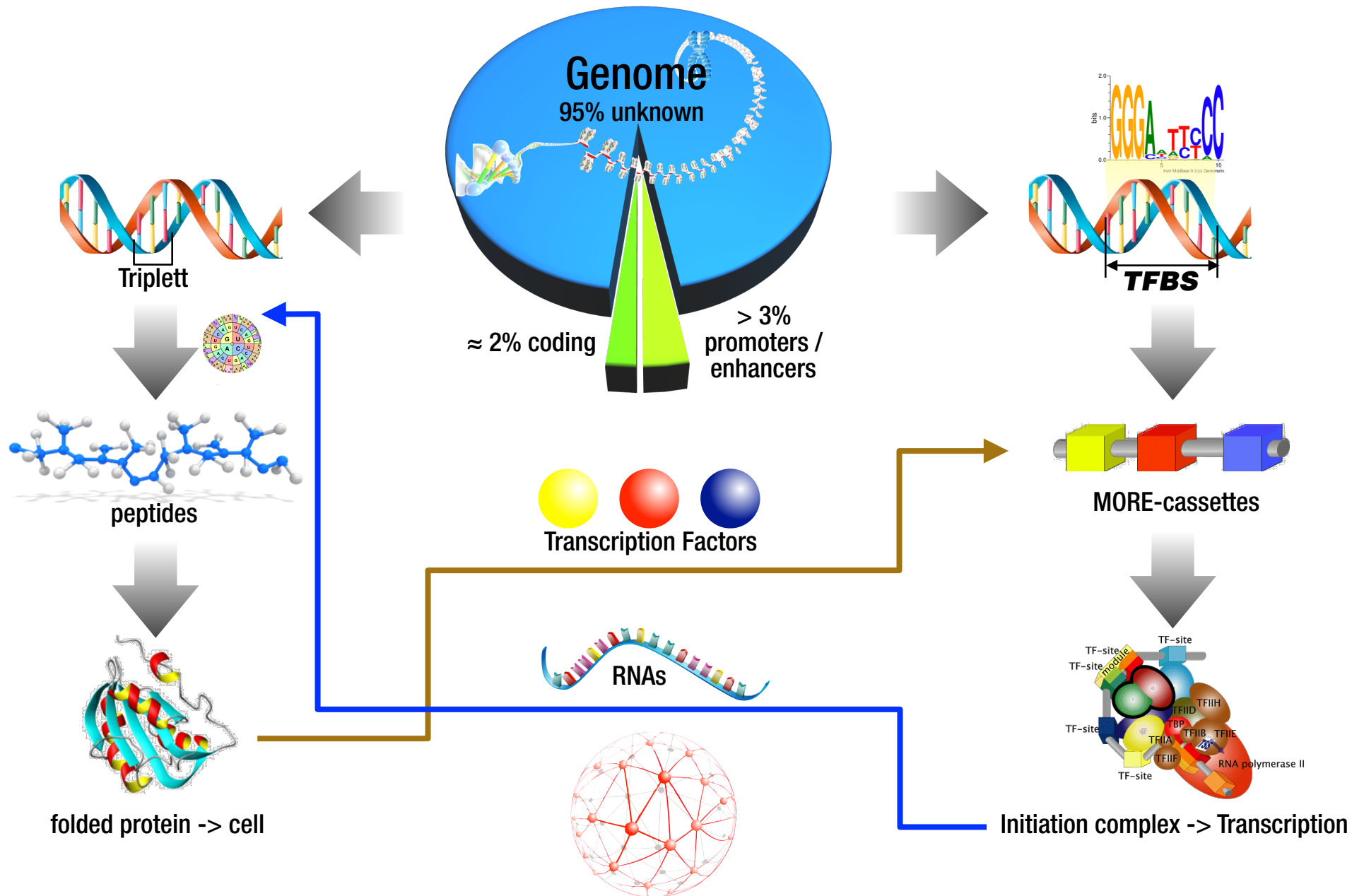
Regulatory Code



Genetic Code

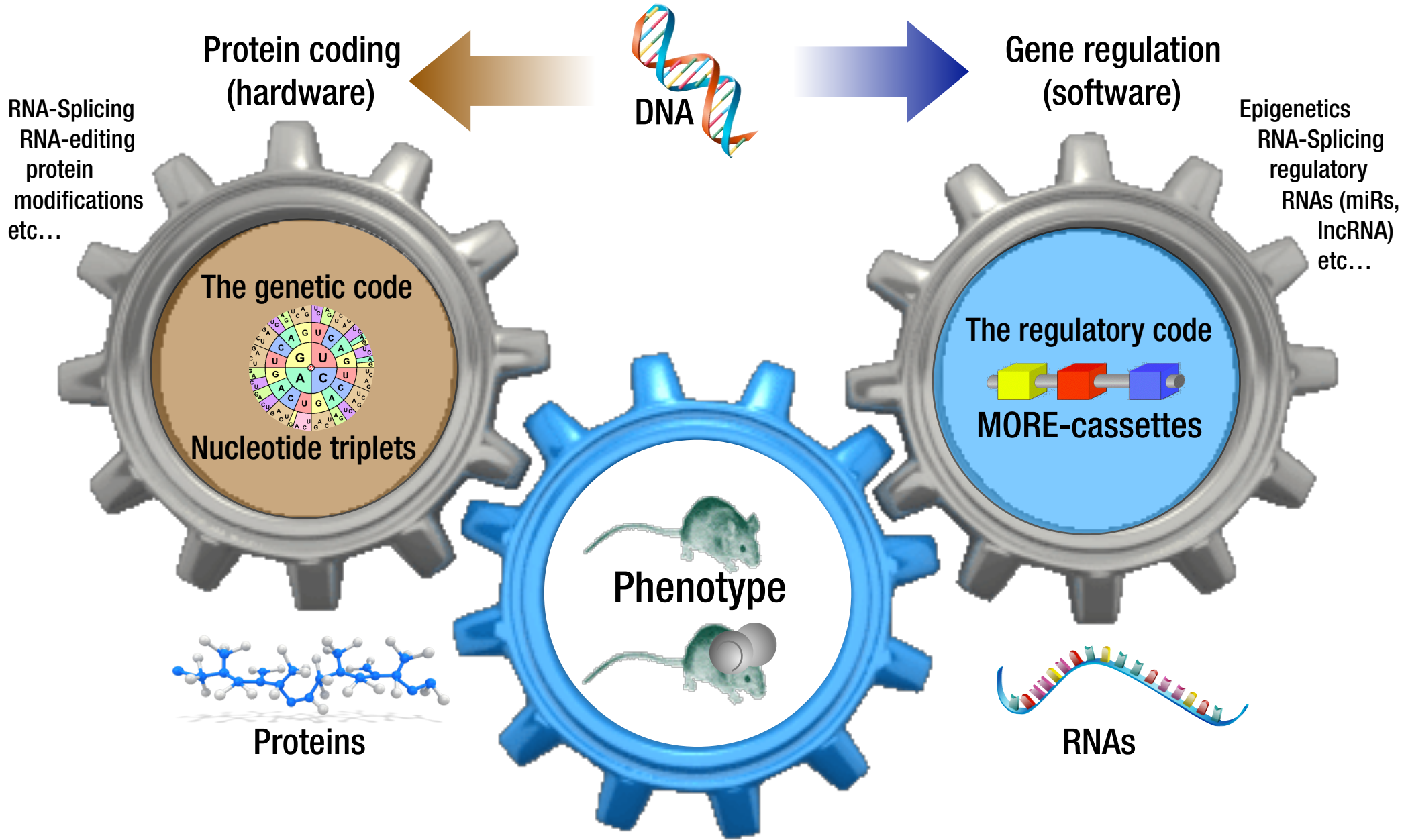
versus

Regulatory Code



It is all one big network!

Mechanisms of life



Acknowledgments

of the many people at these institutions without whom this would have been impossible...

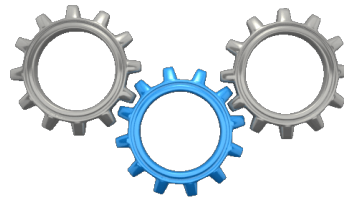
EURen  Omics

 **genomatix**

 **MICHIGAN MEDICINE**
UNIVERSITY OF MICHIGAN

HelmholtzZentrum münchen
German Research Center for Environmental Health

**Institute of
Experimental Genetics**



**THANK
YOU**

...and who made a number of
break-throughs possible!

Thomas Werner
Scientific & Business Consulting

