Combinatorics of *cis*-regulatory elements in osmotic stress response of Ascomycetes.

Aurian García-González
Mentors: Sushmita Roy, Jay Konieczka, Dawn Thompson
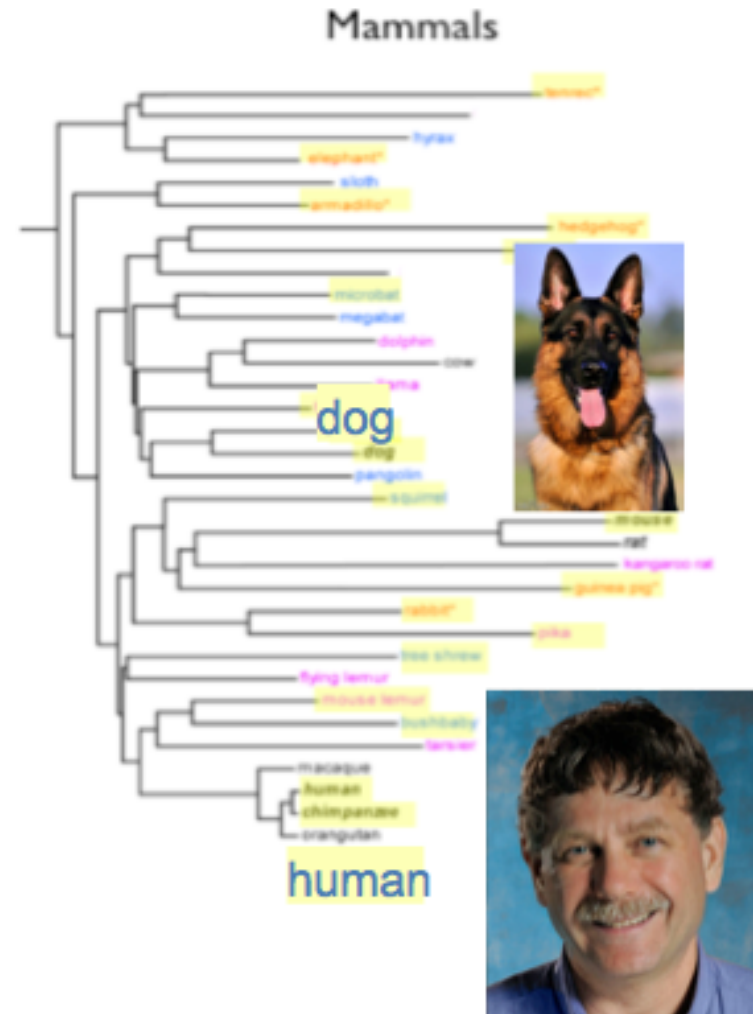Broad Institute- Summer Research Program 2011
August 3, 2011

Question:

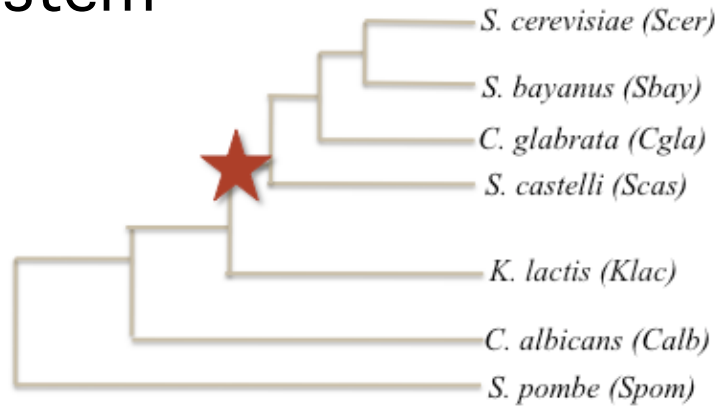- Where does phenotypic diversity come from?

Question:

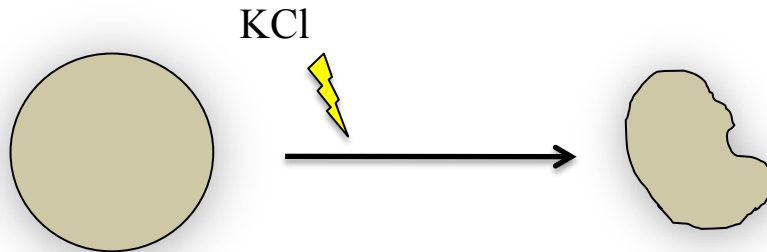- If their gene content is so similar, why are they so different?
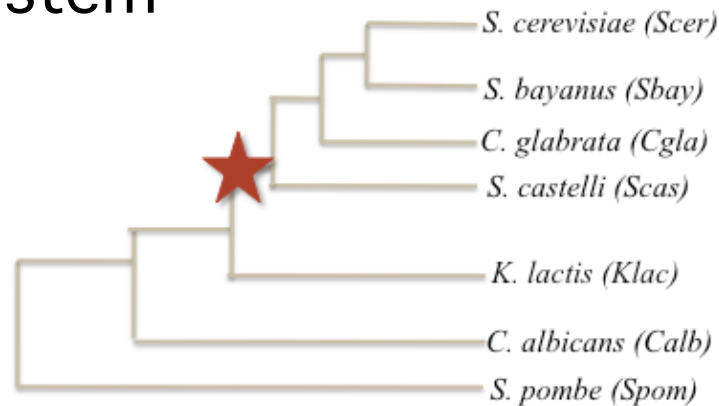
How do we study evolution of gene regulation?

Model system

Ascomycota Fungi

~ 300 MYA

Mammals

dog

human

~ 95 MYA

How do we study evolution of gene regulation?

Model system

FEBS Letters 583 pp. 3959-3965, 2009.
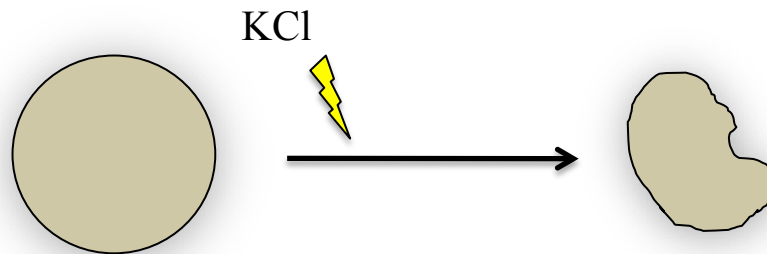
Phenotype: *osmotic stress*

How do we study evolution of gene regulation?

Model system

FEBS Letters 583 pp. 3959-3965, 2009.

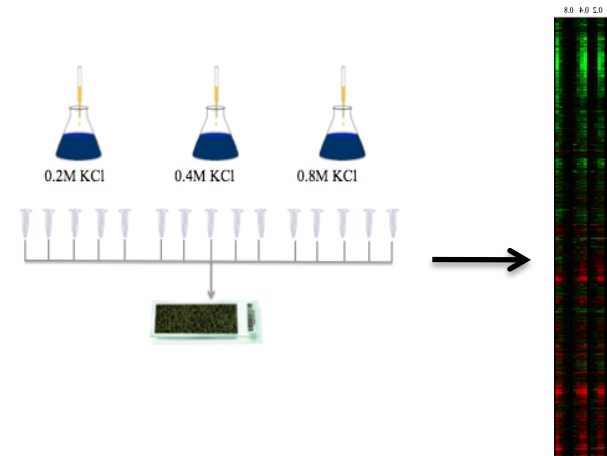## Phenotype: *osmotic stress*

- Readily tractable experimentally.
- Thoroughly described in *S. cerevisiae* as Hog1-directed.
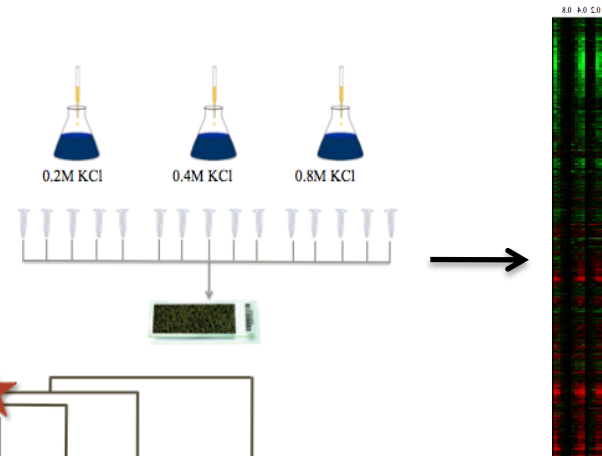
Nat Genet. 2008 Nov; 40(11):1300-6.

Characterizing osmotic stress response

1) Perform **stress experiments** to obtain expression profiles

Characterizing osmotic stress response

1) Perform **stress experiments** to obtain expression profiles

2) Cluster genes to identify **phenomenologs**

Clusters with similar expression – regardless of gene content.

0.2M KCl   0.4M KCl   0.8M KCl

Scer Sbay Cgla Scas Klac Calb Spom
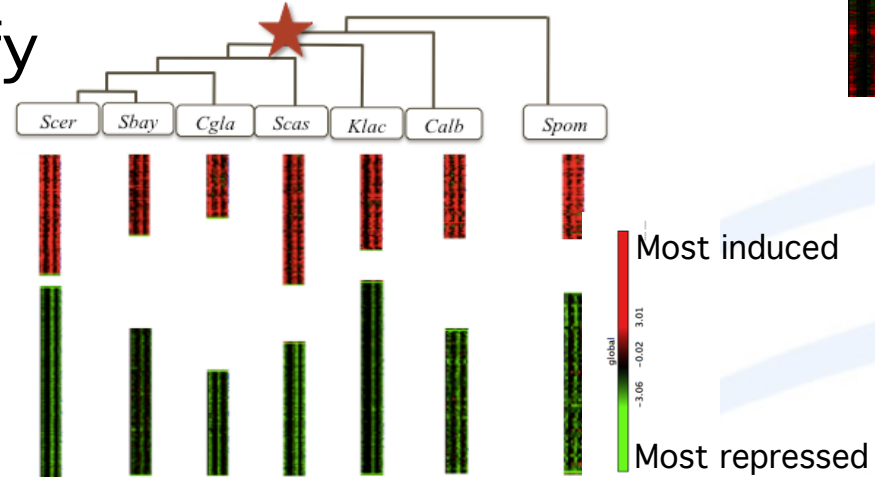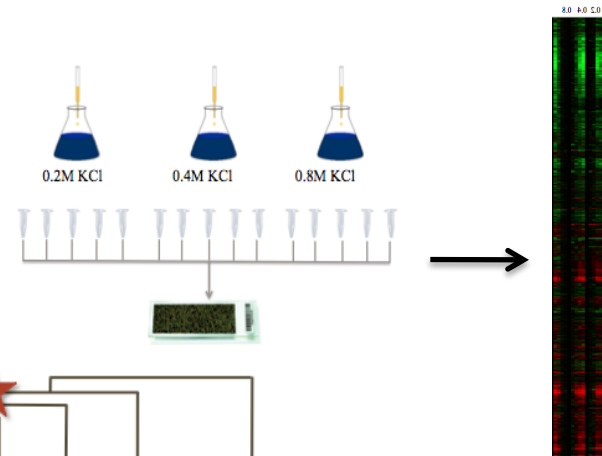
Most induced

Most repressed

Characterizing osmotic stress response

1) Perform **stress experiments** to obtain expression profiles

2) Cluster genes to identify **phenomenologs**

Clusters with similar expression – regardless of gene content.

Most induced

Most repressed

3) What are the *cis*-regulatory elements present in these phenomenologs?

5'   promoter   coding region   3'
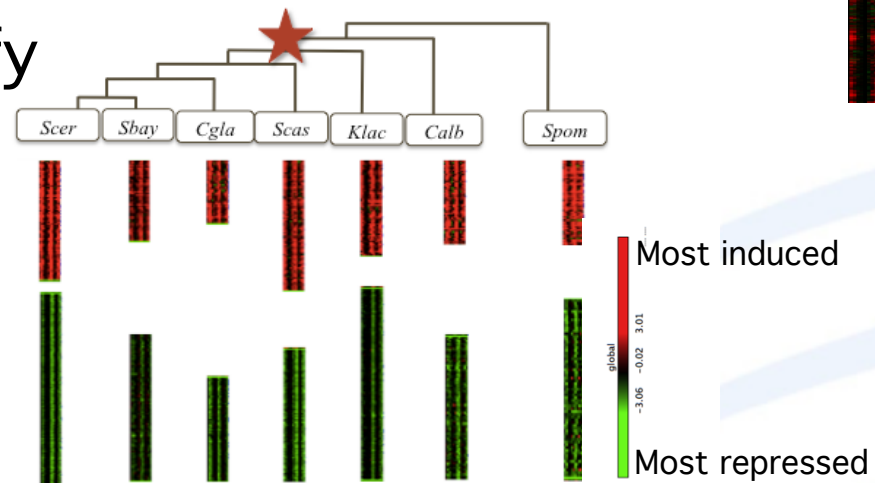
BROAD INSTITUTE

Characterizing osmotic stress response

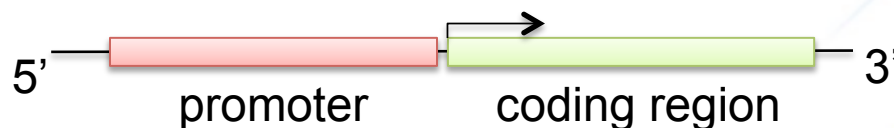1) Perform **stress experiments** to obtain expression profiles

2) Cluster genes to identify **phenomenologs**

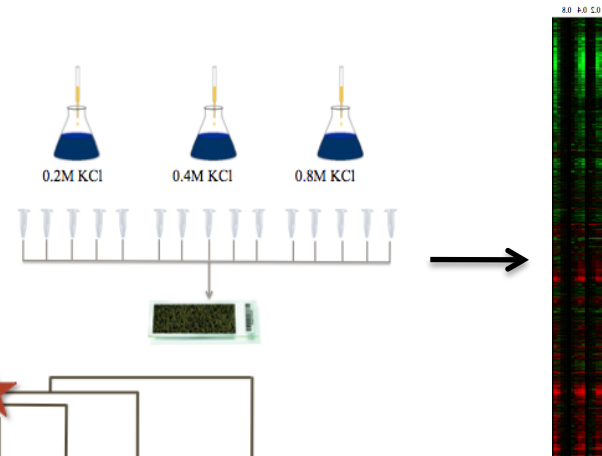Clusters with similar expression – regardless of gene content.

3) What are the *cis*-regulatory elements present in these phenomenologs?
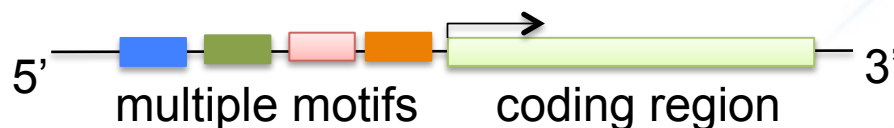
0.2M KCl   0.4M KCl   0.8M KCl

Scer   Sbay   Cgla   Scas   Klac   Calb   Spom

Most induced

Most repressed

5'   single motif   coding region   3'
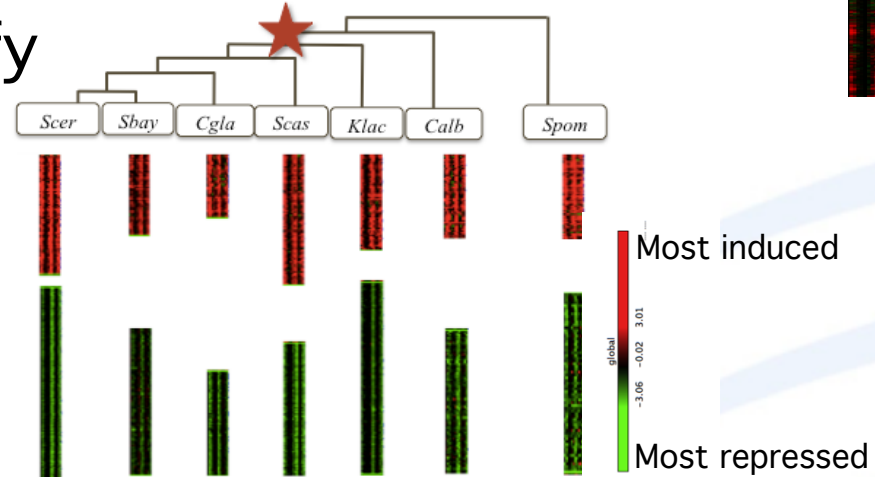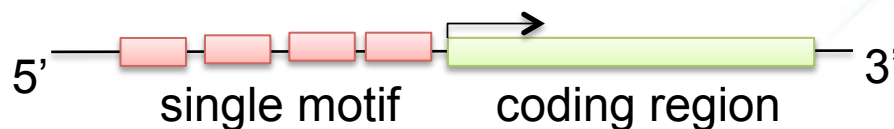
BROAD
INSTITUTE

Characterizing osmotic stress response

1) Perform **stress experiments** to obtain expression profiles

2) Cluster genes to identify **phenomenologs**

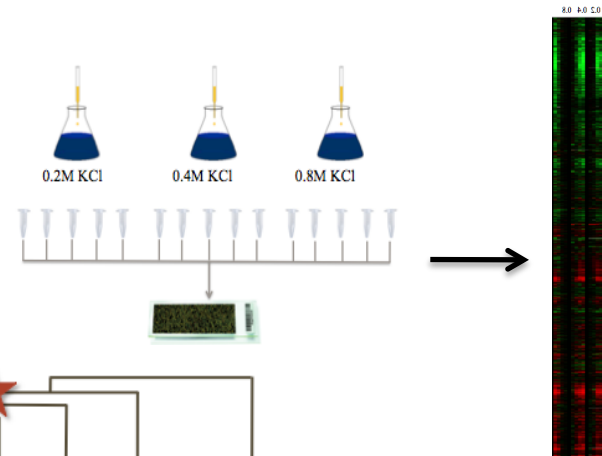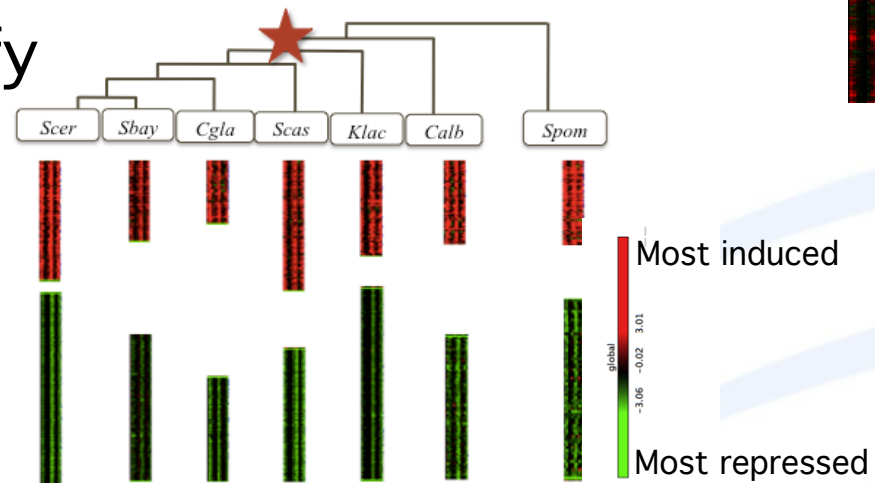   Clusters with similar expression – regardless of gene content.

   Most induced

   Most repressed

3) What are the *cis*-regulatory elements present in these phenomenologs?

   5' multiple motifs coding region 3'

Use ModuleDigger to find CRMs

ModuleDigger:

Uses motif data for genes in a cluster and computes hierarchical scores for modules.

S. cerevisiae

BROAD
INSTITUTE

Use ModuleDigger to find CRMs

ModuleDigger:

Uses motif data for genes in a cluster and computes hierarchical scores for modules.

Geneset specificity score (f_g) for the motifs in each gene

BMC Bioinformatics 2009, 10(Suppl I):S30.

BROAD
INSTITUTE

Use ModuleDigger to find CRMs

ModuleDigger:

Uses motif data for genes in a cluster and computes hierarchical scores for modules.

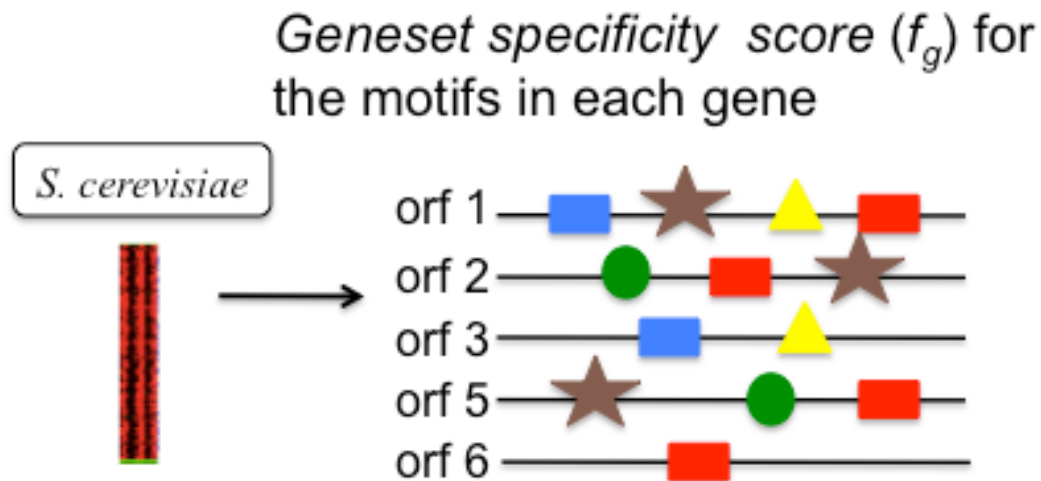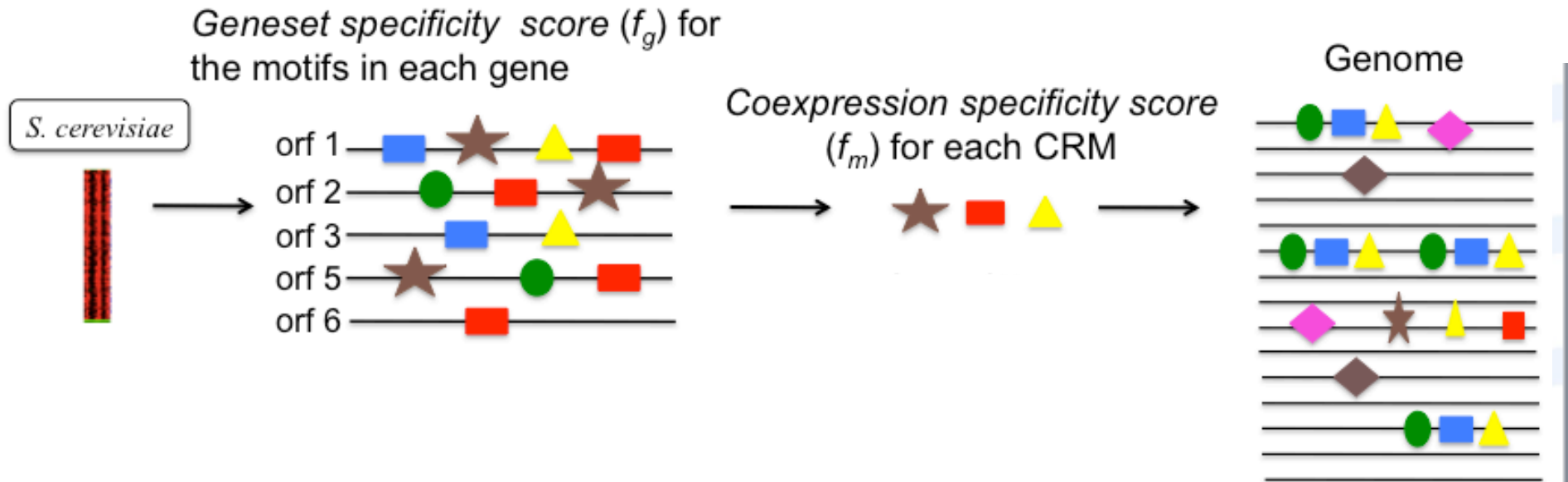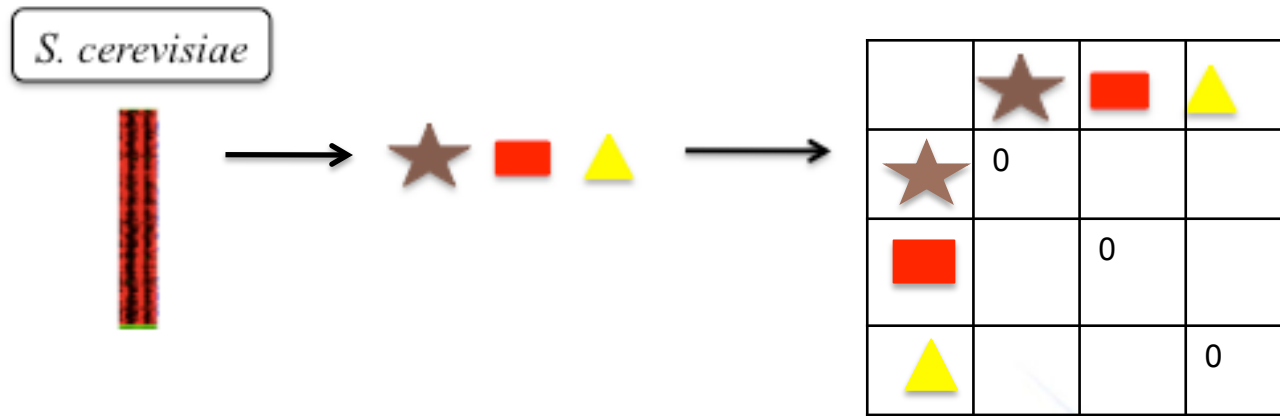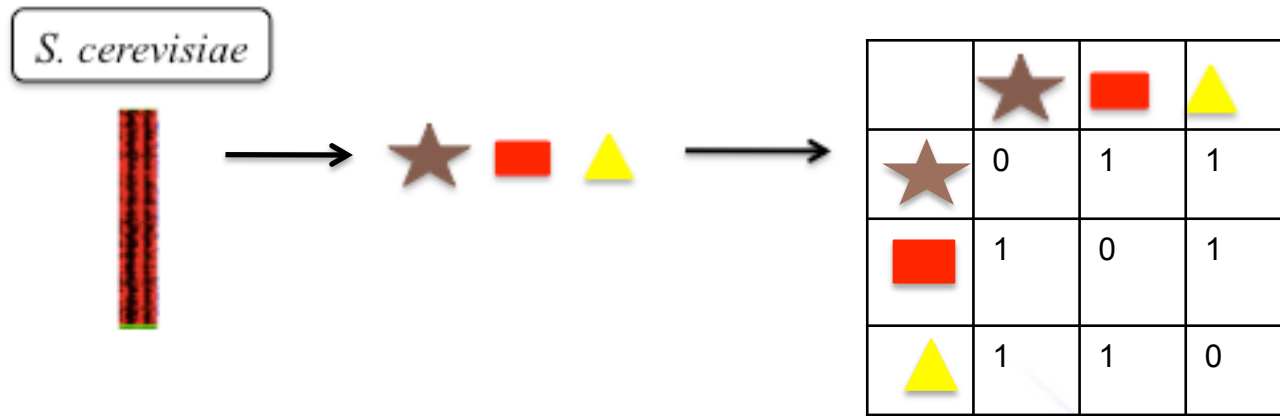# Studying combinations of motifs in CRMs

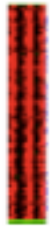# Studying combinations of motifs in CRMs
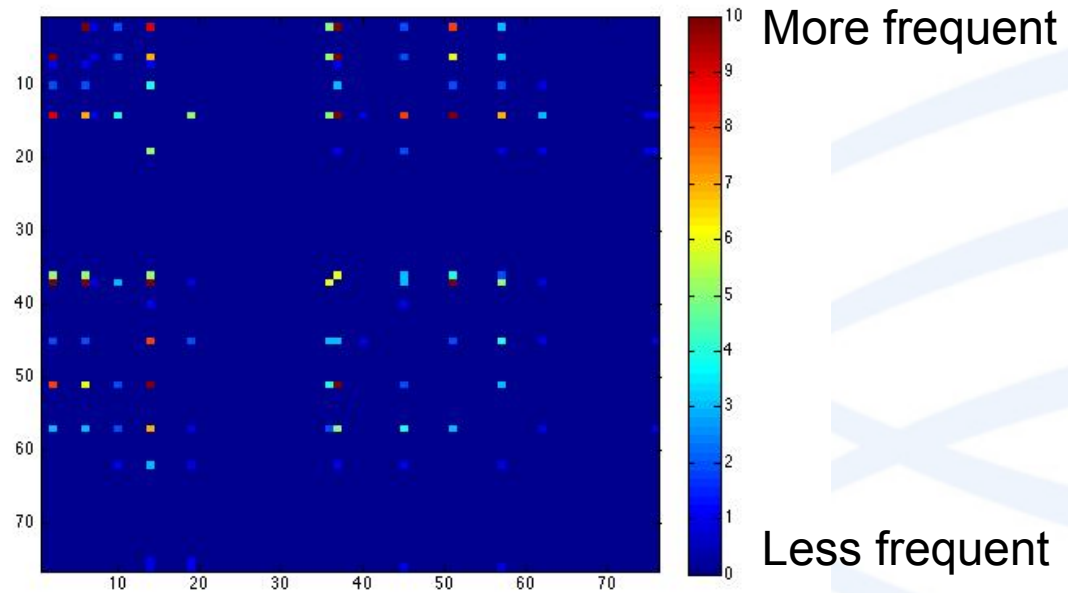
Studying combinations of motifs in CRMs

Studying combinations of motifs in CRMs
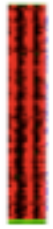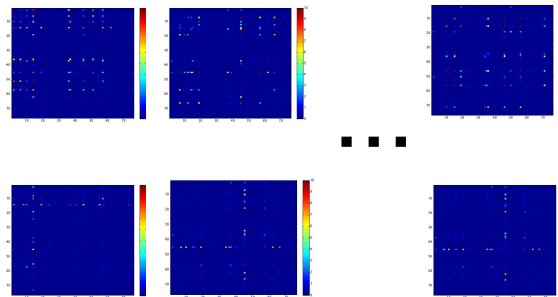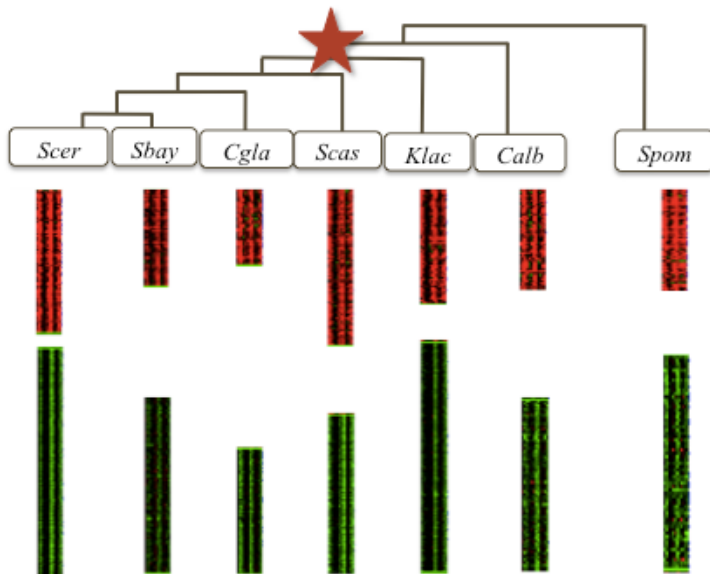
Studying combinations of motifs in CRMs

Are they conserved across all species?
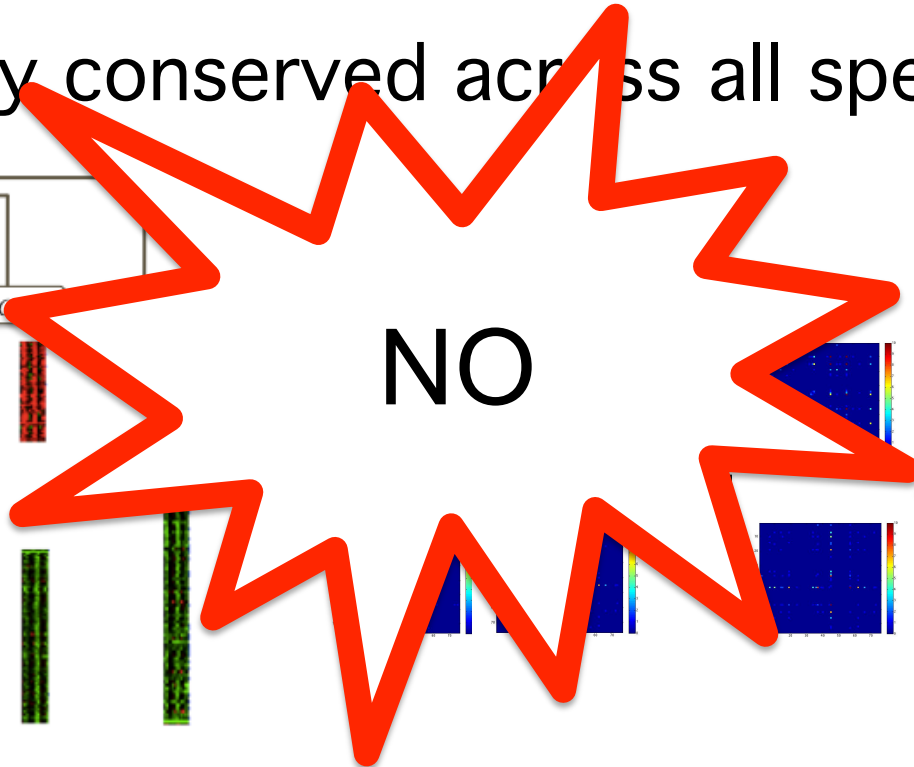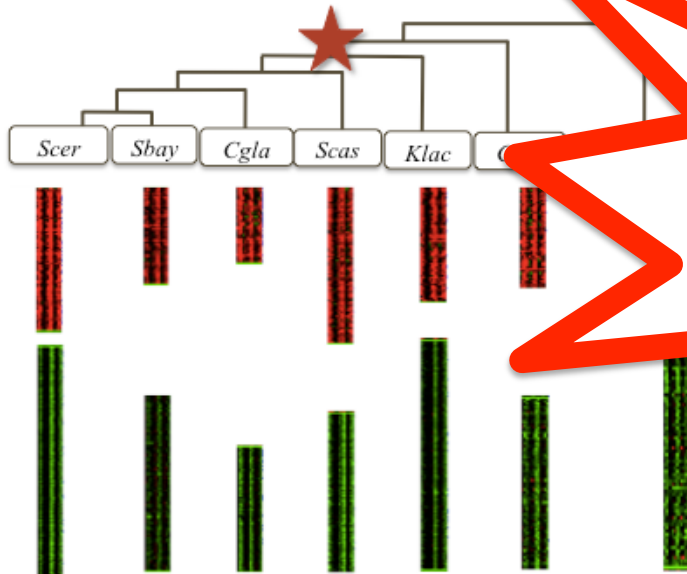
Studying combinations of motifs in CRMs

Motif combinations are conserved in clades

WGD: whole genome duplication

Pre-WGD

S. cerevisiae
S. bayanus
C. glabrata
S. castelli
K. lactis
C. albicans
S. pombe
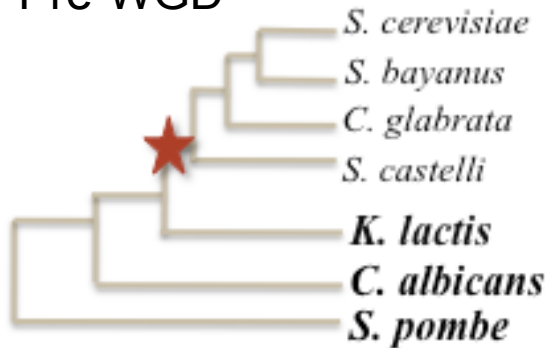
Most induced

RCS1, SIP4

Most repressed

RCS1, SIP4
MBP1, TEC1
MBP1, PDR3

Motif combinations are conserved in clades

WGD: whole genome duplication

Pre-WGD

- *S. cerevisiae*
- *S. bayanus*
- *C. glabrata*
- *S. castelli*
- **K. lactis**
- **C. albicans**
- **S. pombe**

Post-WGD

- **S. cerevisiae**
- **S. bayanus**
- **C. glabrata**
- **S. castelli**
- *K. lactis*
- *C. albicans*
- *S. pombe*
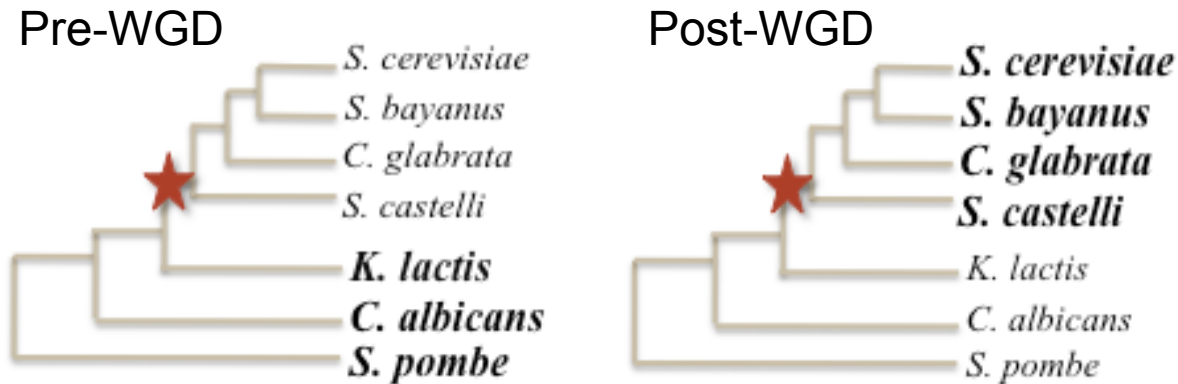
Most induced

RCS1, SIP4

Most induced

MIG1, MSN2/MSN4
MIG1, ADR1
MSN2/MSN4, ADR1

Most repressed

RCS1, SIP4
MBP1, TEC1
MBP1, PDR3

Most repressed

Motif combinations are conserved in clades

WGD: whole genome duplication

Pre-WGD

S. cerevisiae
S. bayanus
C. glabrata
S. castelli
K. lactis
C. albicans
S. pombe

Most induced

RCS1, SIP4

Most repressed

RCS1, SIP4
MBP1, TEC1
MBP1, PDR3

Post-WGD

S. cerevisiae
S. bayanus
C. glabrata
S. castelli
K. lactis
C. albicans
S. pombe

Most induced

MIG1, MSN2/MSN4
MIG1, ADR1
MSN2/MSN4, ADR1

Most repressed

Pathogens

S. cerevisiae
S. bayanus
C. glabrata
S. castelli
K. lactis
C. albicans
S. pombe

Most induced

RCS1, SIP4
CRZ1, SIP4

Most repressed

RCS1, SIP4

Stress-related regulatory elements co-occur with glucose-related elements
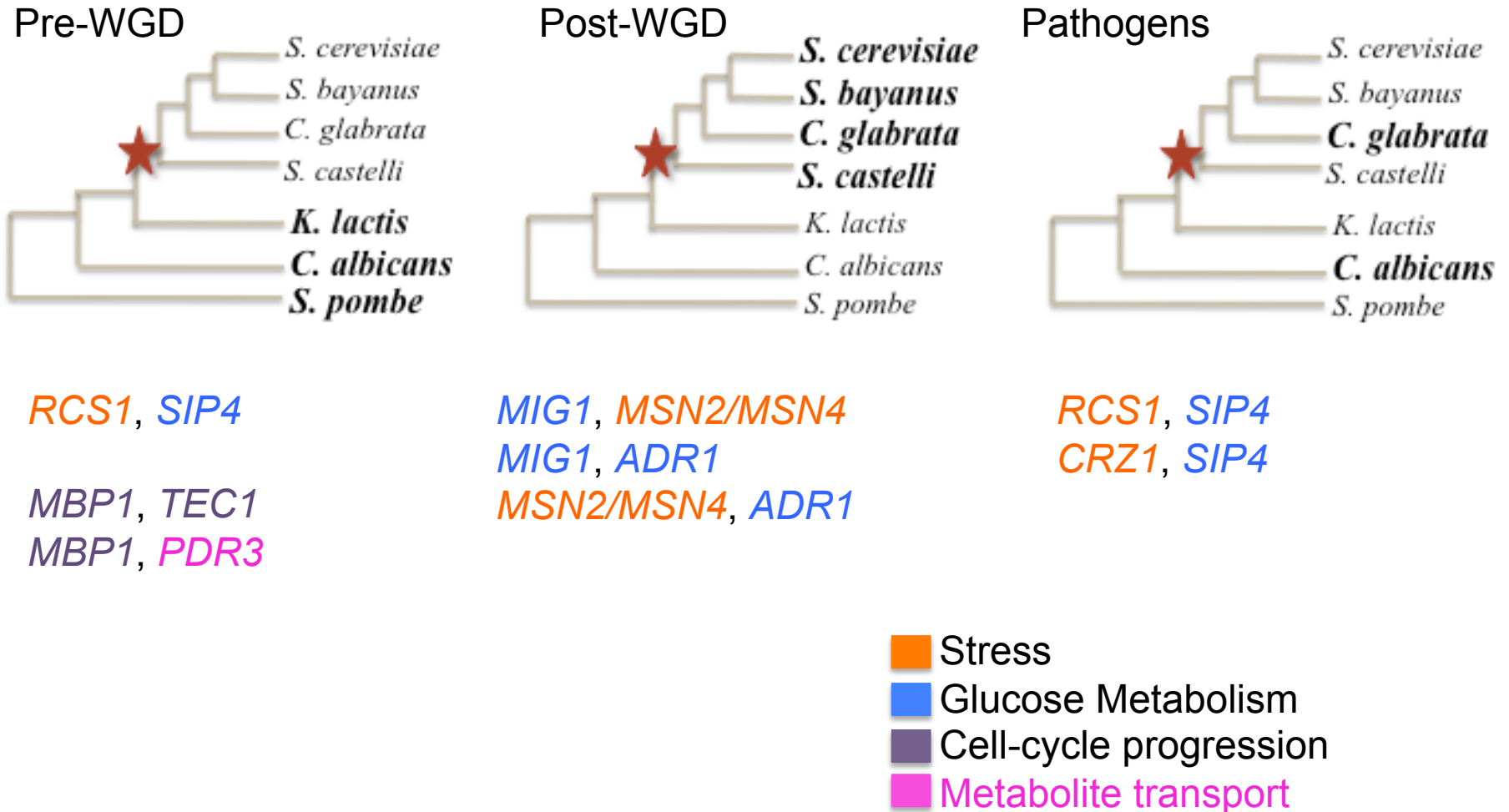
Conclusions

- Very few motif pairs were species-specific. Combinatorics are conserved in at least two species.

BROAD
INSTITUTE

Conclusions

- Very few motif pairs were species-specific. Combinatorics are conserved in at least two species.

- The most induced clusters showed more motif pairs than the repressed clusters.

BROAD
INSTITUTE

Conclusions

- Very few motif pairs were species-specific. Combinatorics are conserved in at least two species.

- The most induced clusters showed more motif pairs than the repressed clusters.

- Glucose responsive elements occur with stress elements.

BROAD
INSTITUTE

Future work

- Explore higher order combinations.

BROAD
INSTITUTE

Future work

- Explore higher order combinations.
- Develop more sophisticated algorithms to discover significant CRMs.

BROAD
INSTITUTE

Future work

- Explore higher order combinations.
- Develop more sophisticated algorithms to discover significant CRMs.
- Correlate *trans* interactions with observed phenotype.

BROAD
INSTITUTE

Acknowledgements

Mentors:
Sushmita Roy
Jay Konieczka
Dawn Thompson
Aviv Regev
Regev Lab

SRPG:
Eboney Smith
Bruce Birren
Lori Thomae
Nicole Edmonds
Broad staff

BROAD INSTITUTE