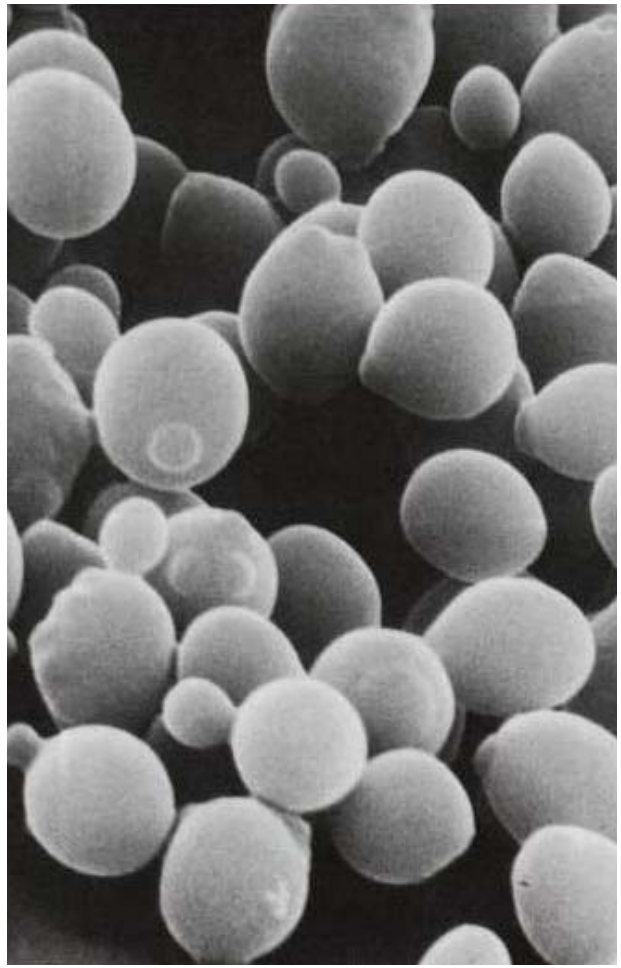
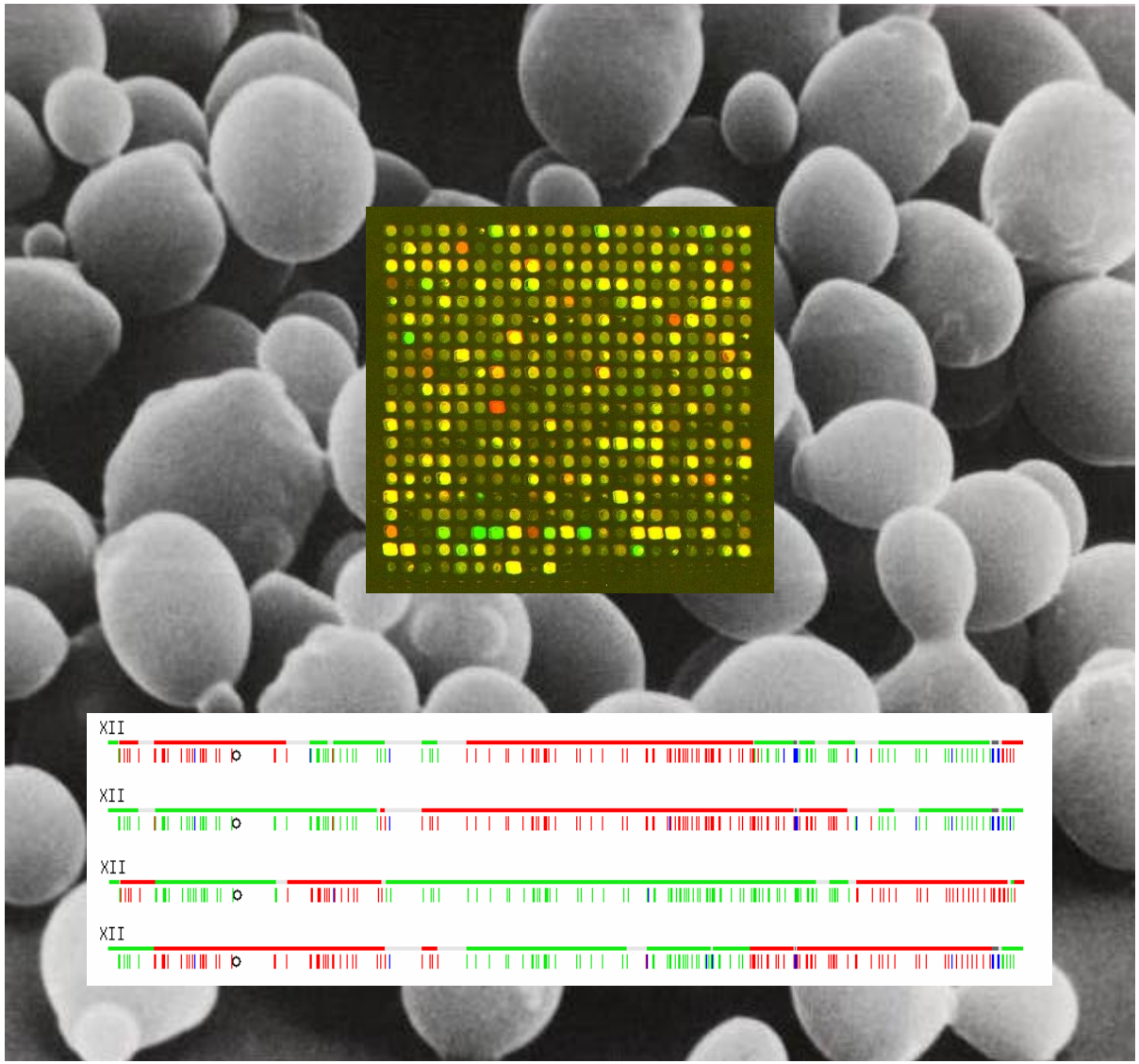
What is the genetic basis of phenotypic variation?



Genetics of gene expression



Why study genetics of gene expression?

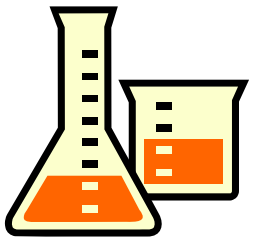
- Regulatory variation is source of intraspecific differences and evolutionary change between species

- Connection between genotypes and organismal phenotypes

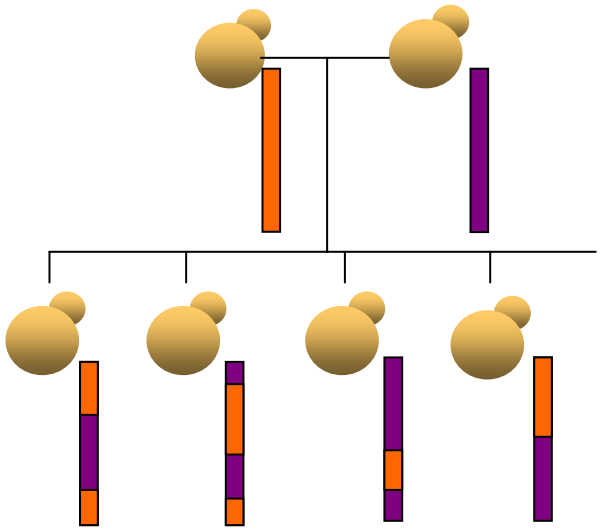
DNA → mRNA → Protein → Phenotypes

- Genetic architecture as model for other quantitative traits
- Genetics as tool to probe regulatory networks

Transcript variation in a yeast cross



Lab (BY)

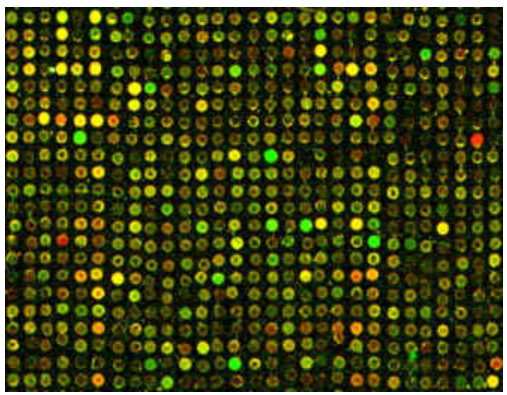


...

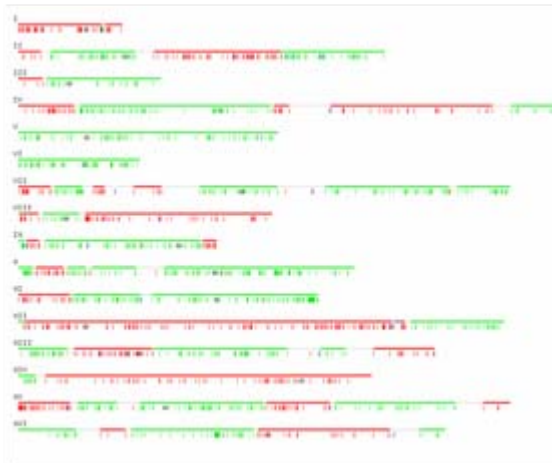
Wine (RM)

profile
gene
expression

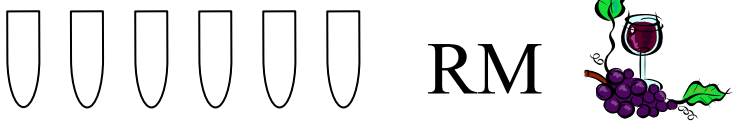
determine
genome
segregation

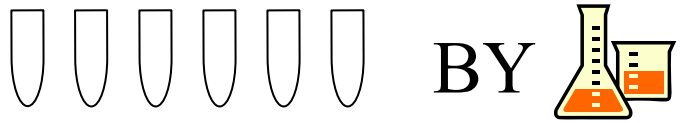


correlate genotype
with
transcript abundance



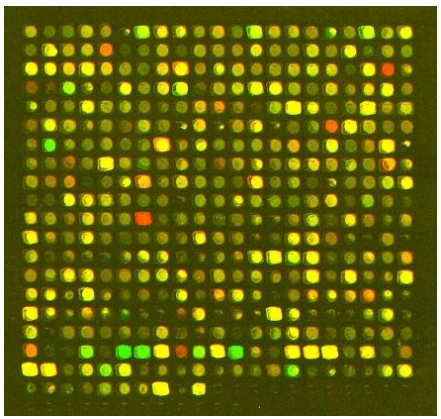
Comparison of gene expression in two strains



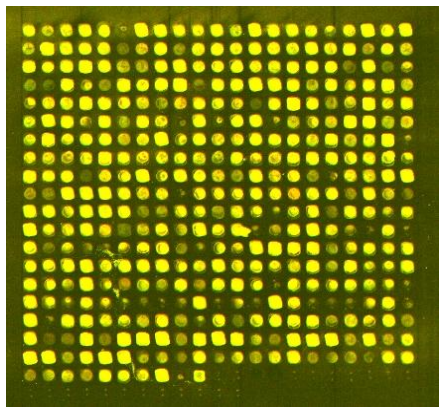


Isolate RNA
Label
Hyb vs. reference

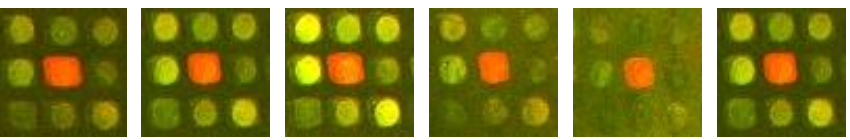
Isolate RNA
Label
Hyb vs. reference

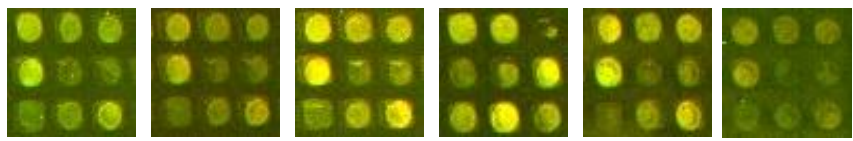


arrays cover
all yeast genes
~6000



compare using Wilcoxon-Mann-Whitney rank test



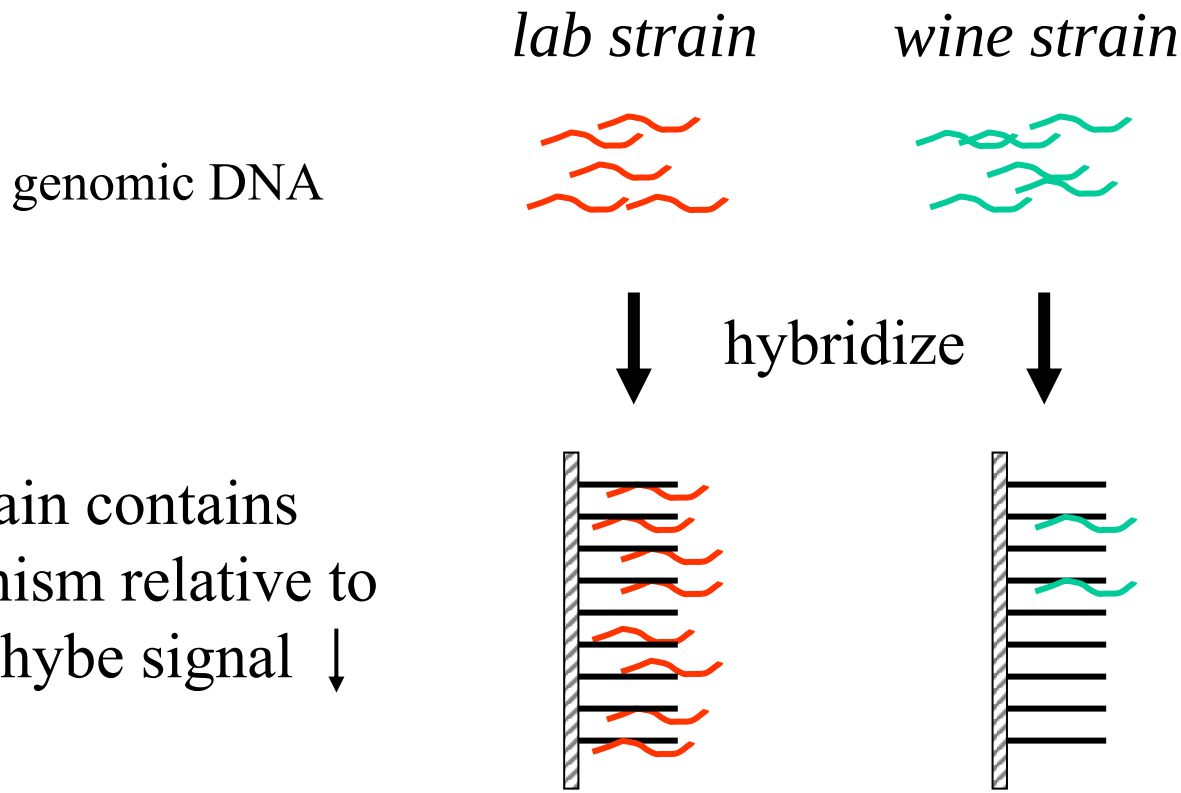


2932 genes differ at FDR = 0.05

Science 2002

Genotyping by hybridization to oligo arrays

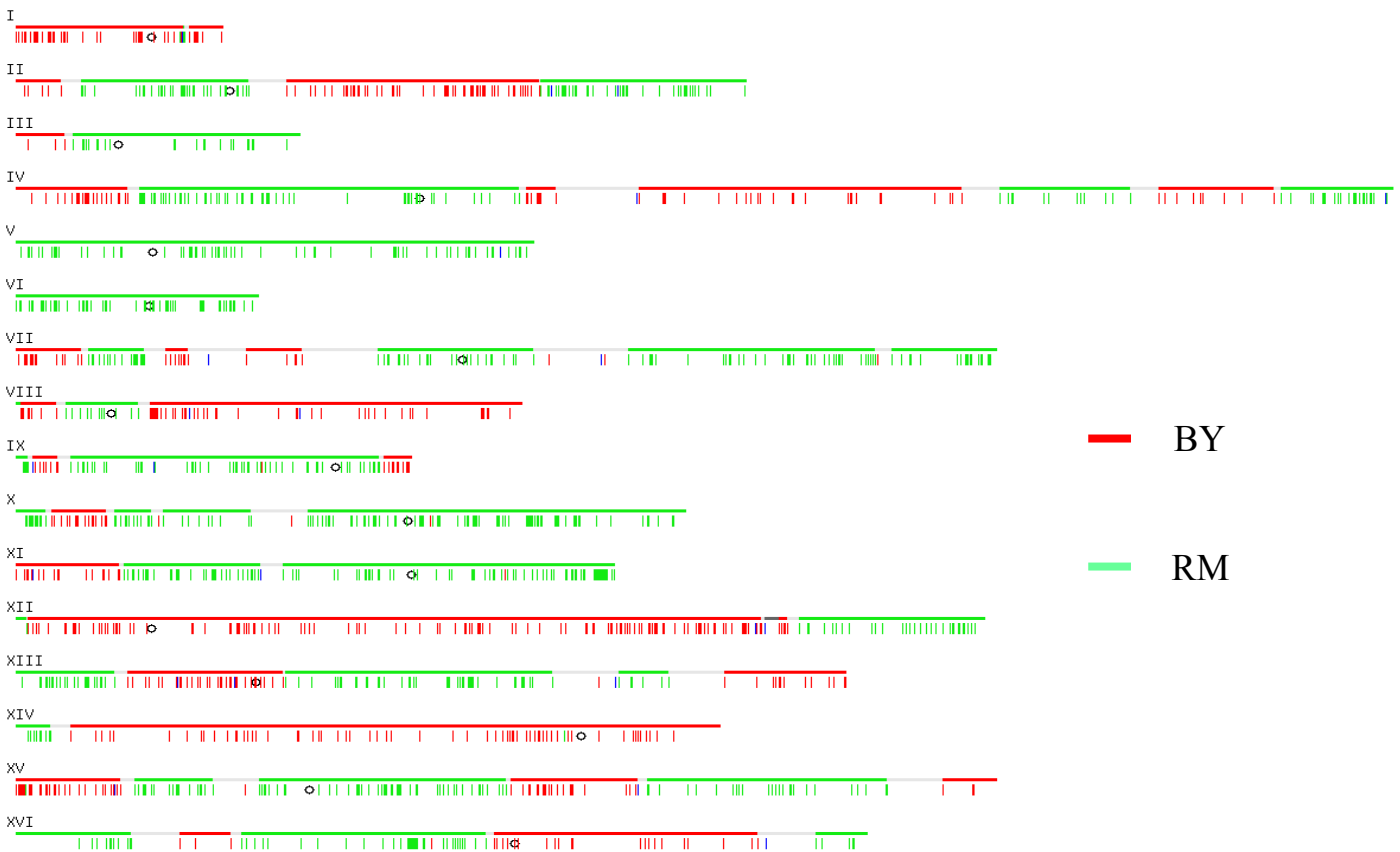
Affymetrix yeast chips contain `134,175` oligos (25 bp) from lab strain



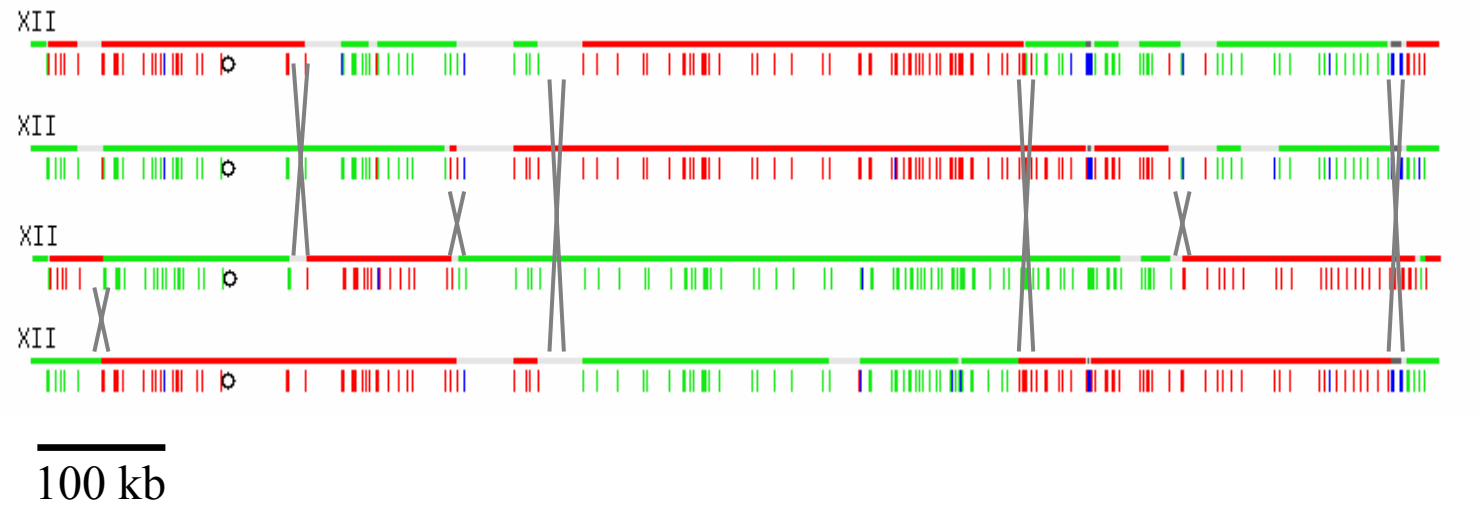
if wine strain contains polymorphism relative to lab strain, hybe signal ↓

~3000 reproducible differences: lower bound one change per 1000 bp

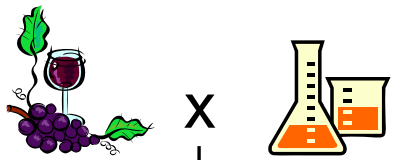
Genetic map: 3312 markers; >99% genome coverage



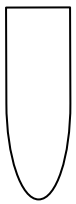
Genotypes of 4 segregants from one tetrad

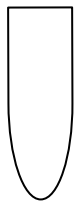


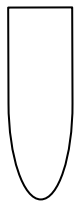
Genotyping and phenotyping segregants



x







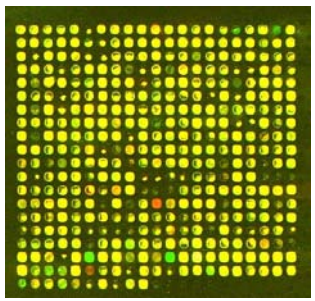
Culture

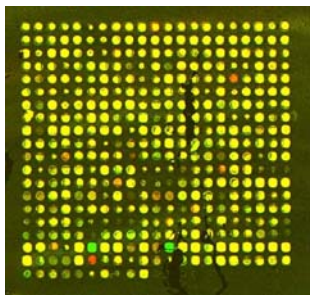
Seg 1

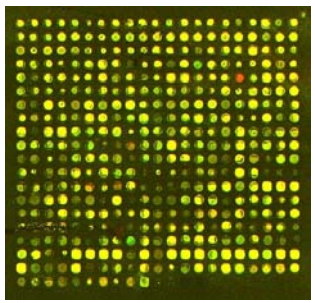
Seg 2

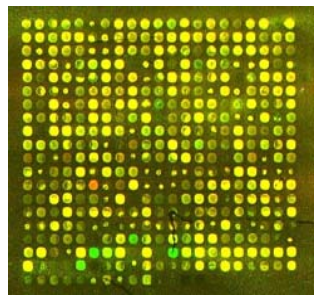
…

Seg 112

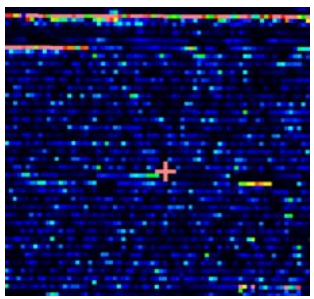


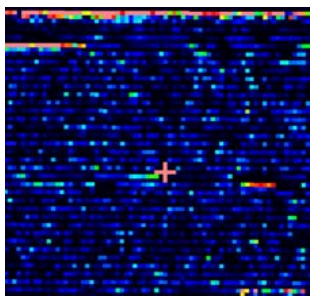


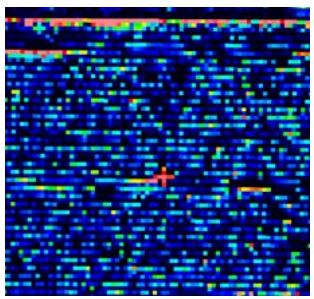


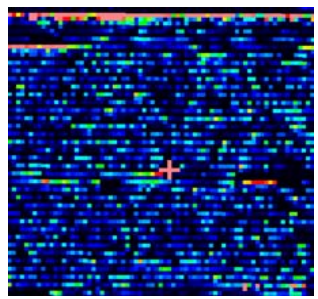


mRNA expression



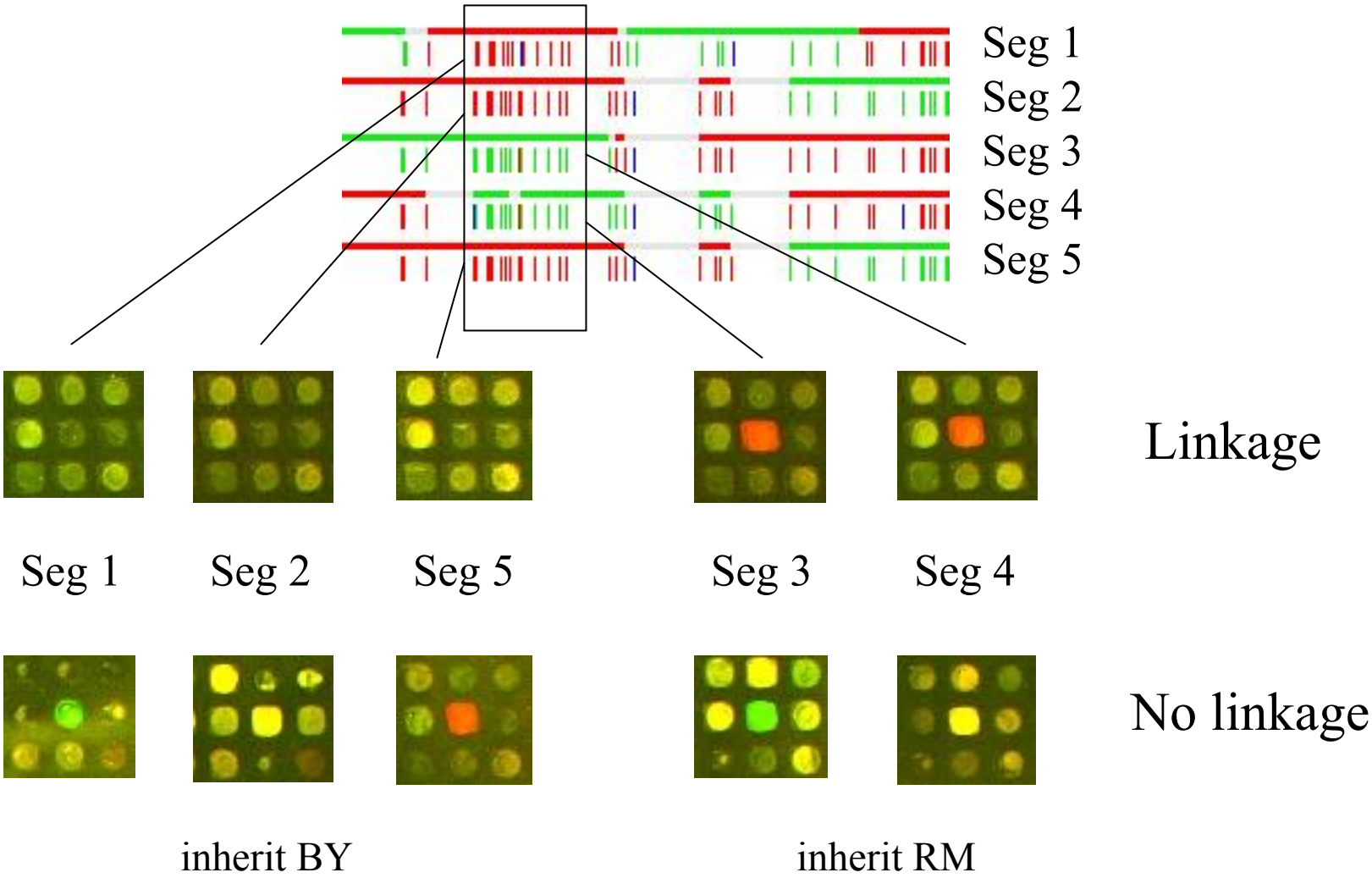






Oligo Genotyping

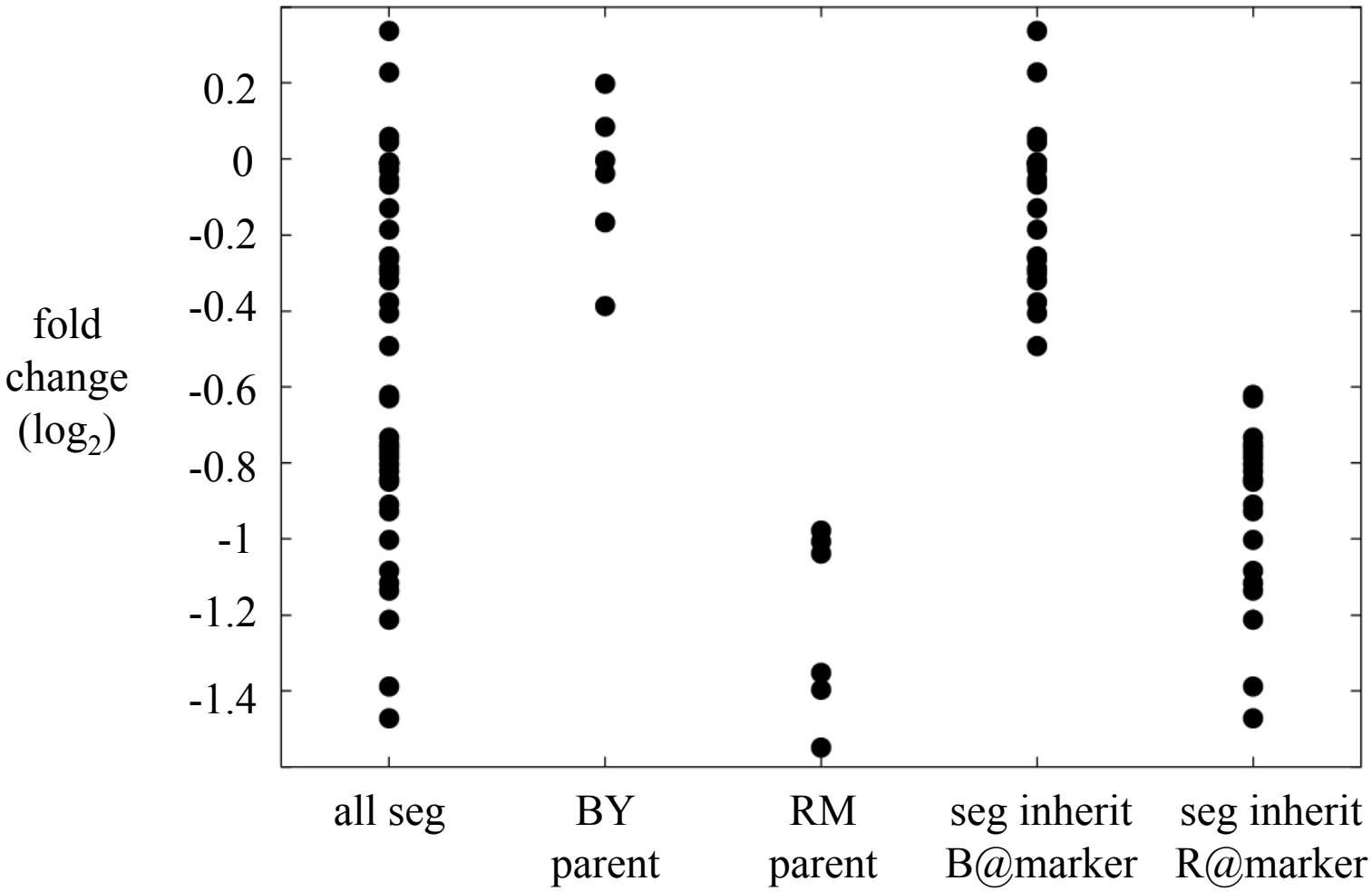
Testing for linkage in segregants



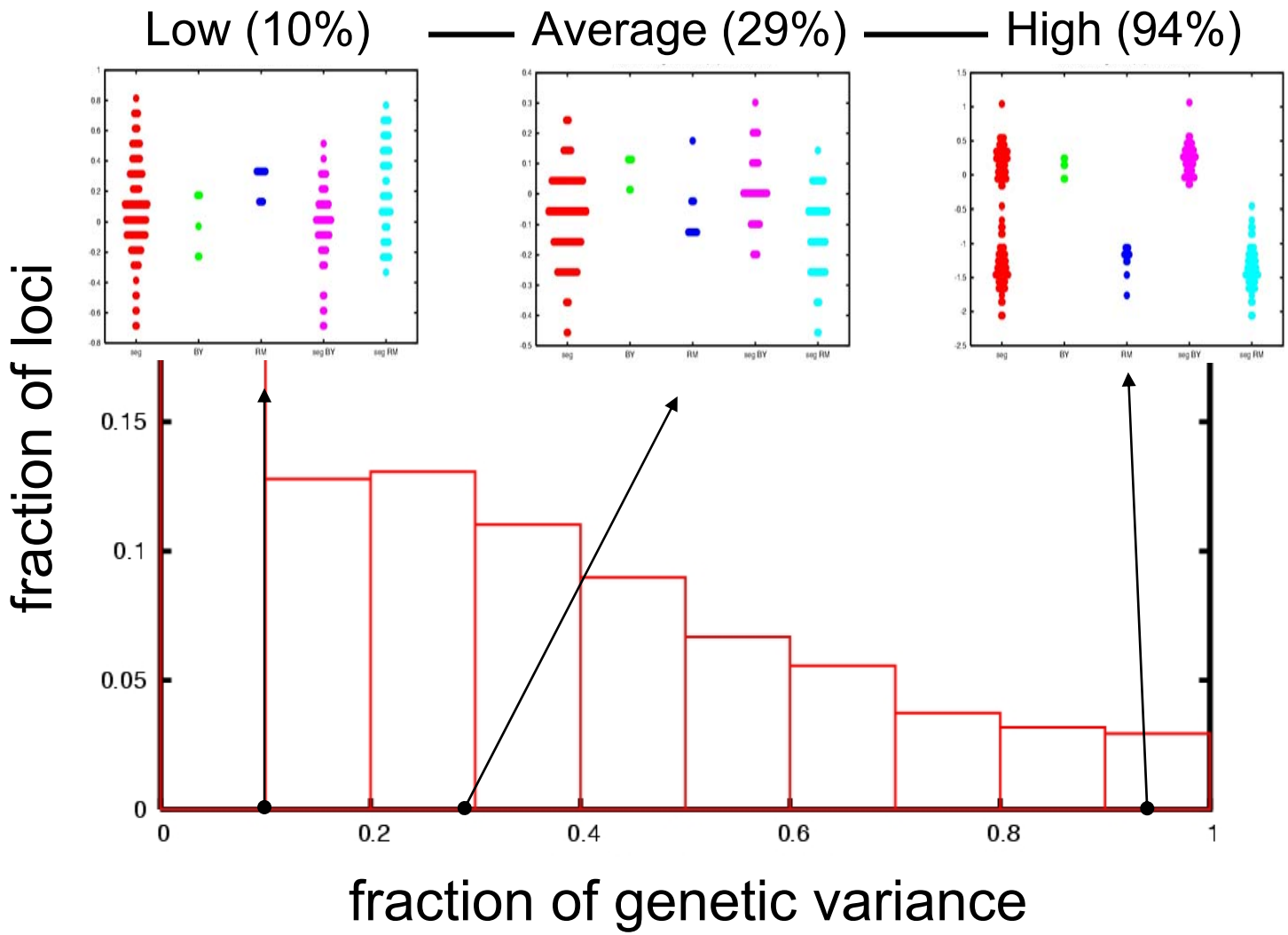
2984 genes link at FDR = 0.05

Science 2002

Example of mRNA linkage



Variance explained by detected loci



PNAS 2005

Estimating genetic complexity

Assumption: genetics controlled by additive loci of equal effect

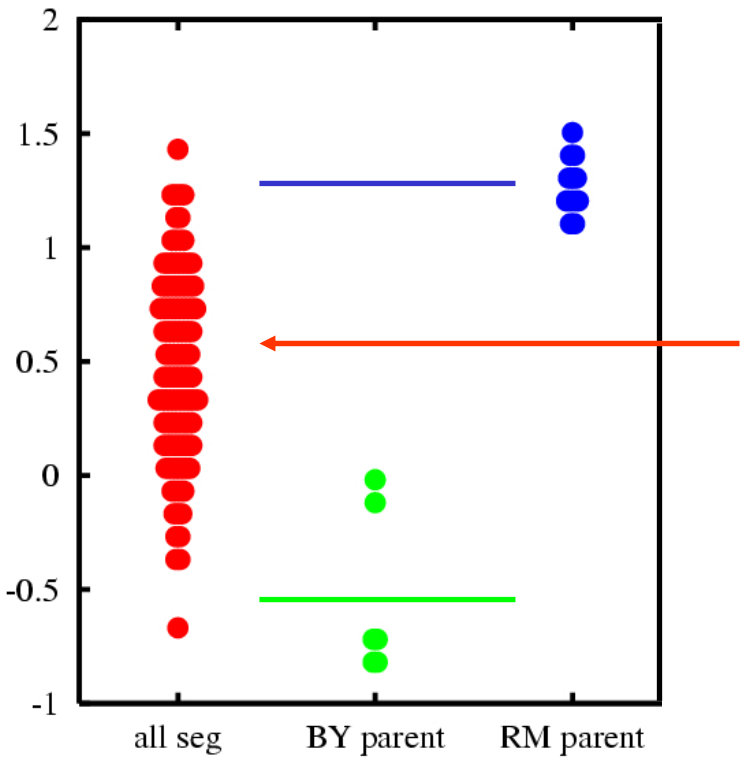
Single-locus genetics:	4-5% of transcripts
More than 2 loci:	75% or more
More than 4 loci:	50% or more
More than 8 loci:	20% or more

Best-fit model: 60% of transcripts controlled by 1-10 loci

40% of transcripts more complex

PNAS 2005

Complexity I: multiple additive loci of same sign

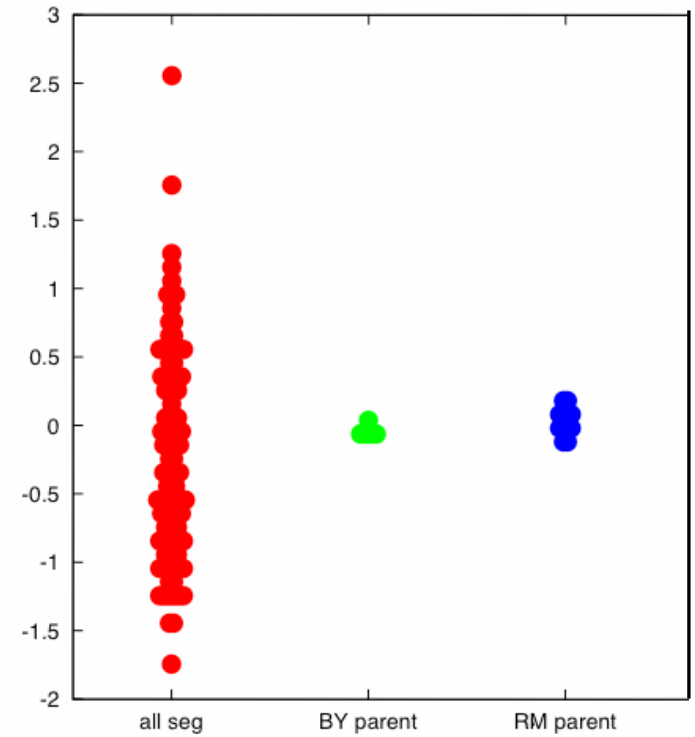


most segregants fall between parent means

729 traits pass at FDR = 0.05

PNAS 2005

Complexity II: transgressive segregation



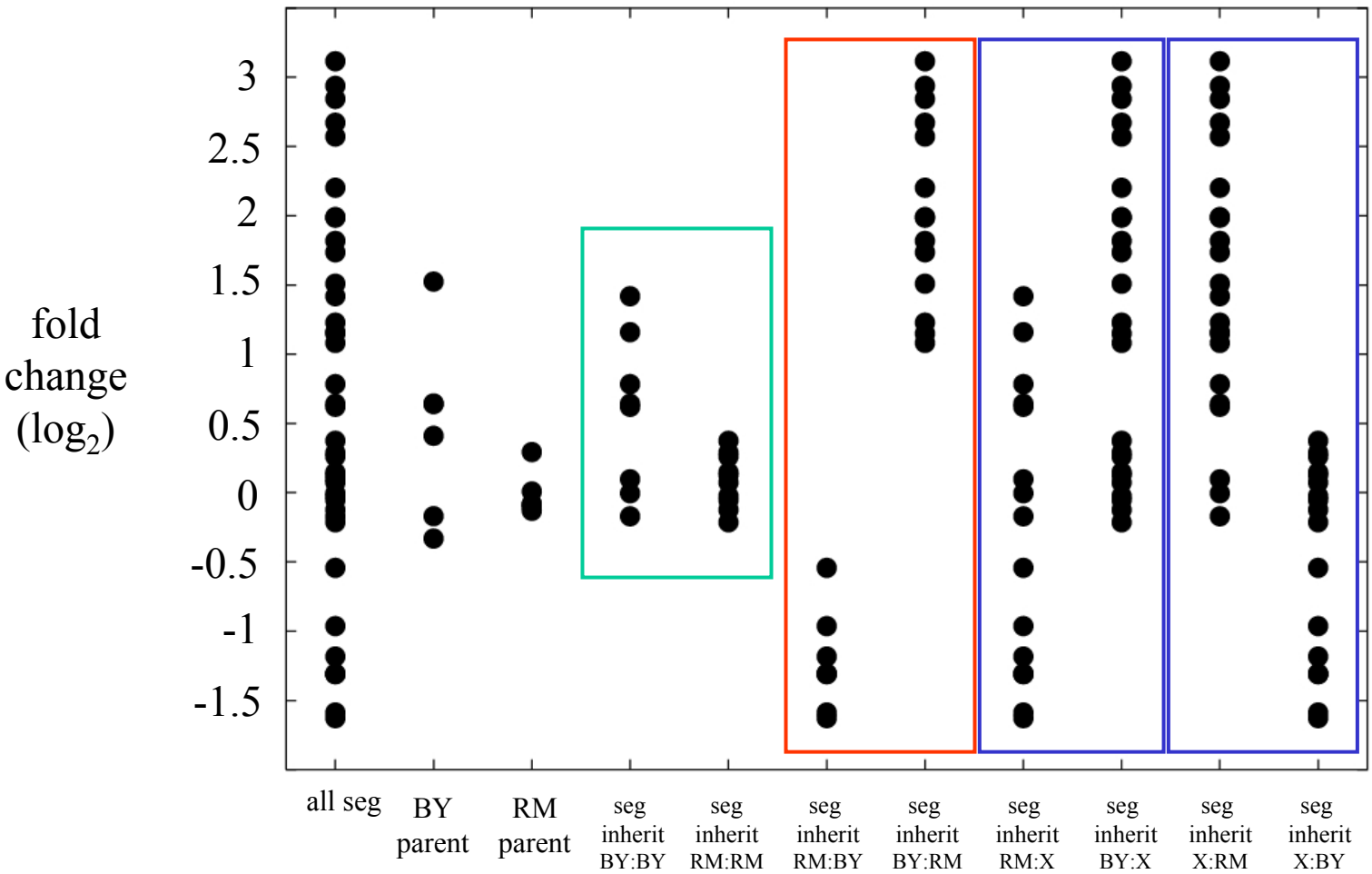
most segregants fall outside parent means

developed formal test with high power to detect transgression (21 seg ≥ 2 SD outside)

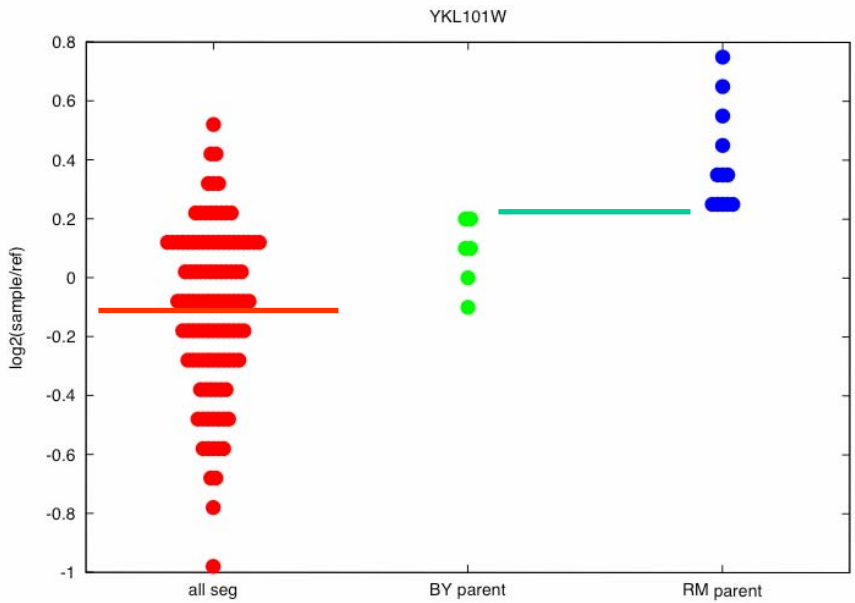
1716 traits pass at FDR = 0.05

PNAS 2005

Two-locus inheritance with transgression



Complexity III: interacting loci



segregant mean differs from mid-parent mean

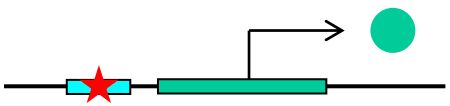
835 traits pass at FDR = 0.05

PNAS 2005

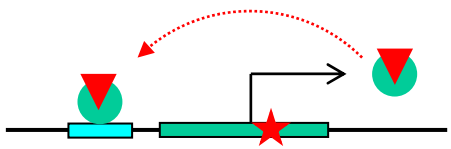
Local regulatory variation

1431 transcripts (25%) link to marker closest to encoding gene

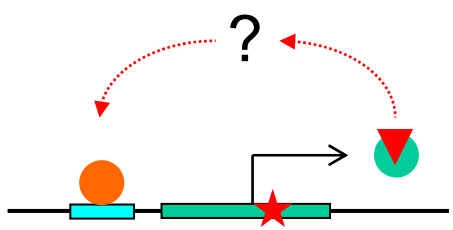
cis-regulatory
(allele-specific)



auto-regulatory



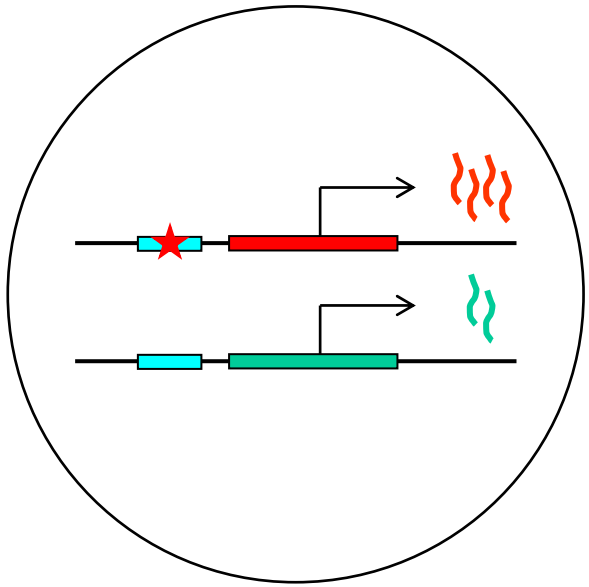
feedback



PLoS Genet 2005

Testing allele-specific expression

diploid hybrid

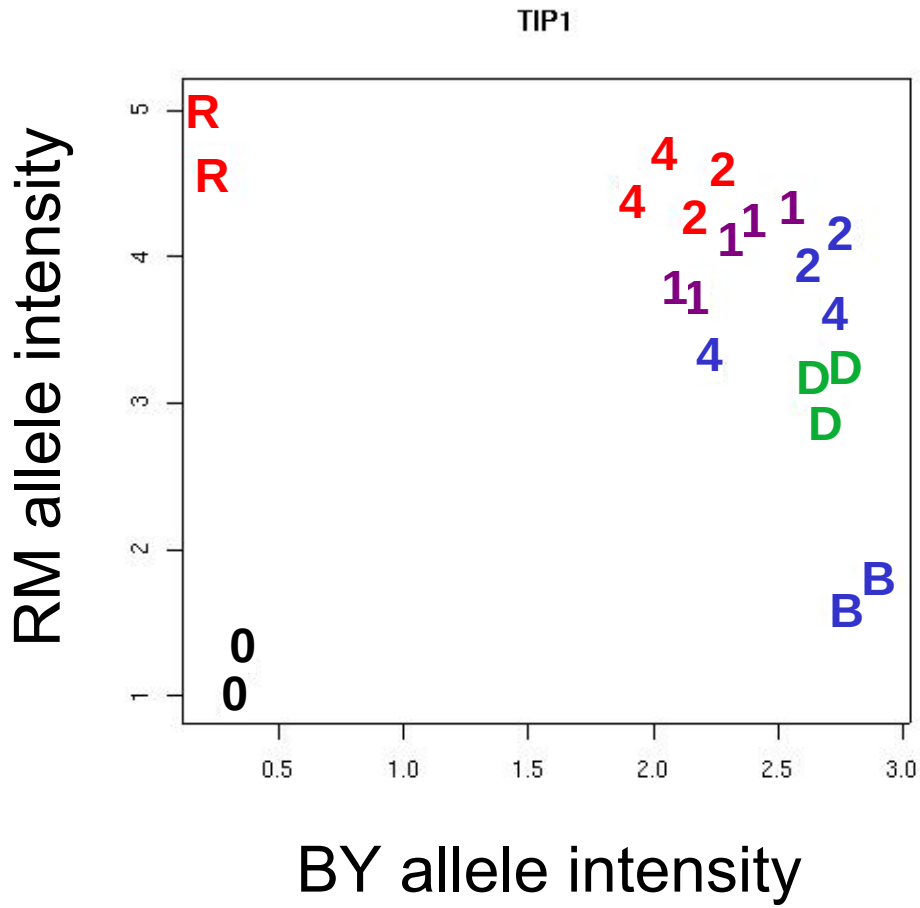


44 of 77 self-linkers show ASE at $p < .05$

2 of 16 trans-linkers show ASE at $p < .05$

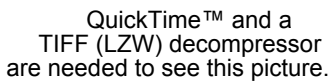
PLoS Genet 2005

Allele-specific expression of *TIP1*



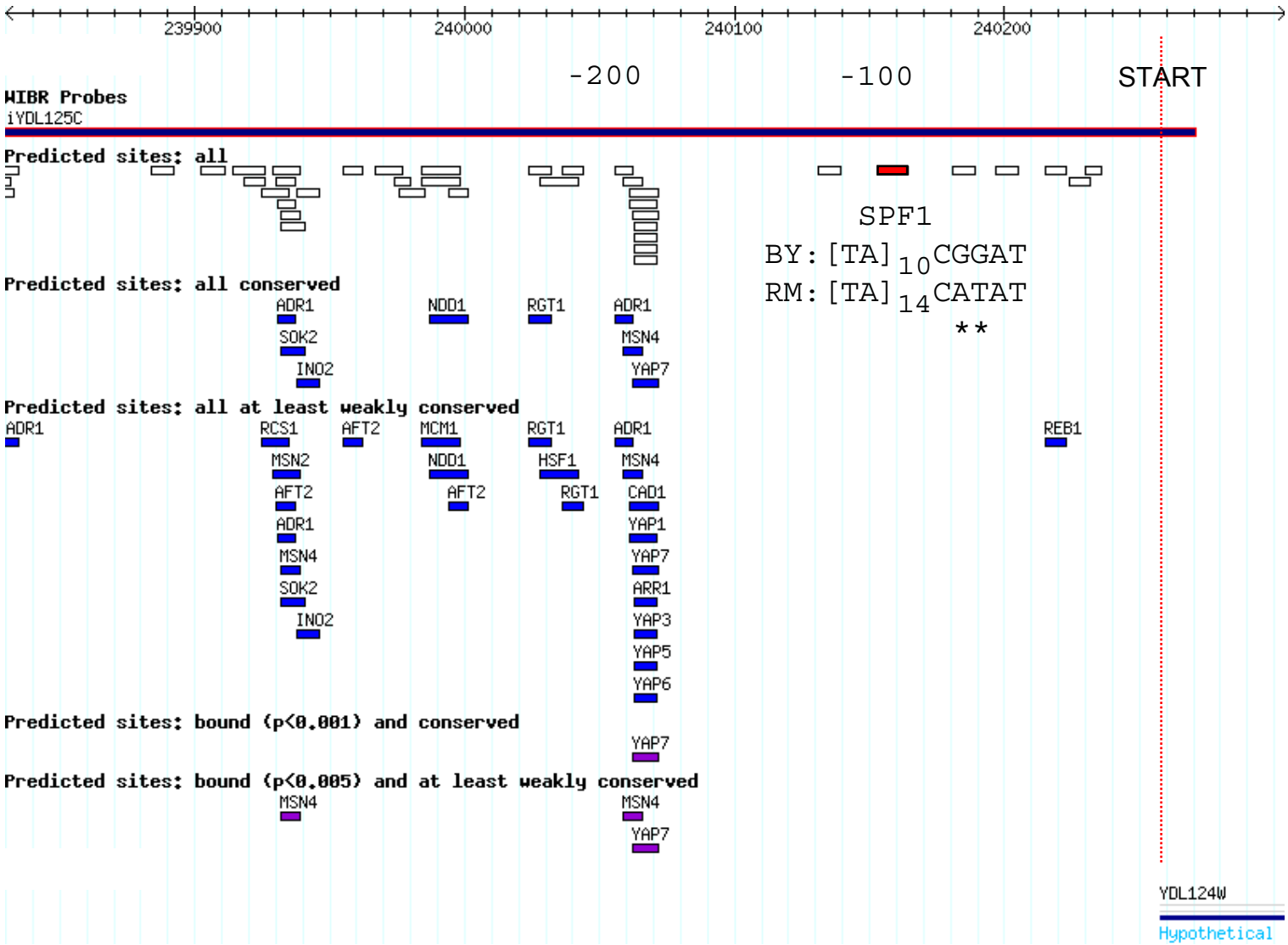
Genome Res 2005

More SNPs in regulatory regions of self-linkers



PLoS Genet 2005

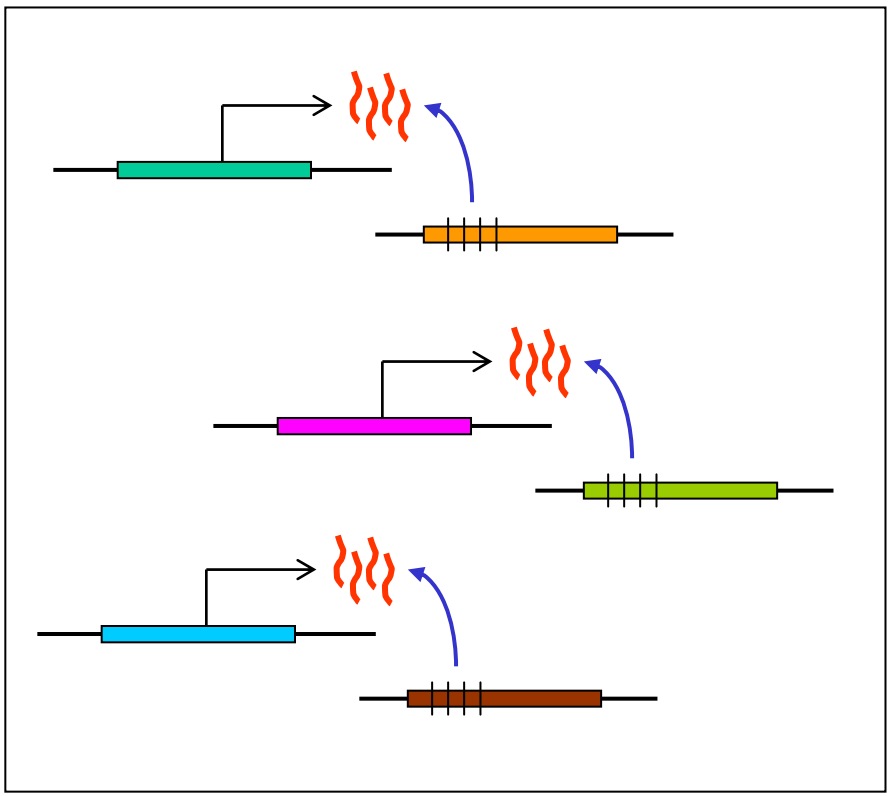
What are the regulatory polymorphisms?

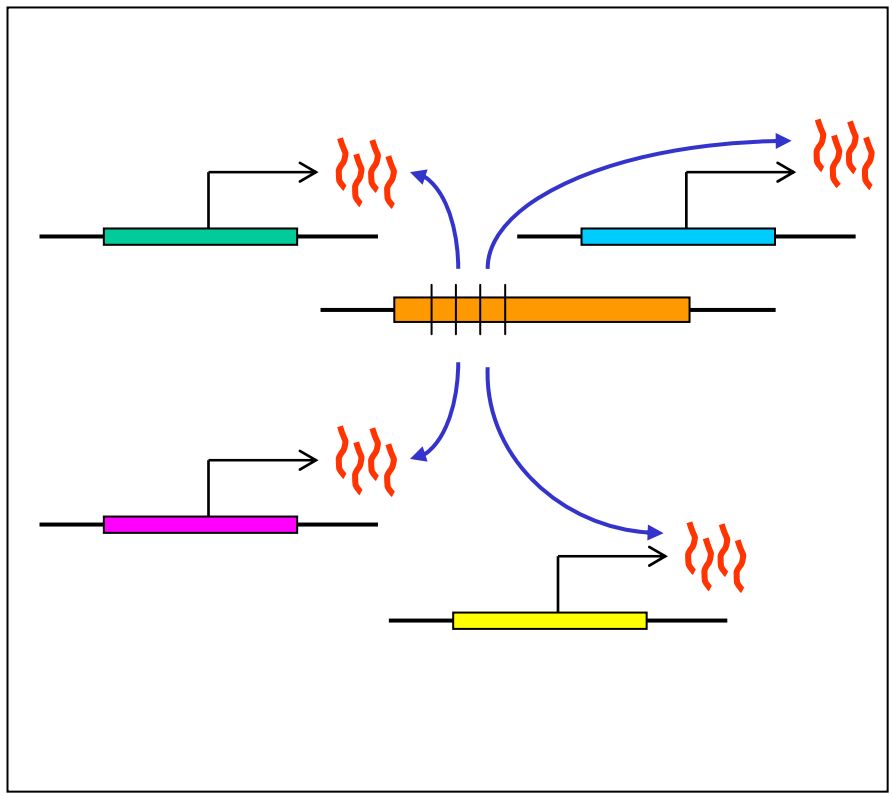


Trans-acting regulatory variation

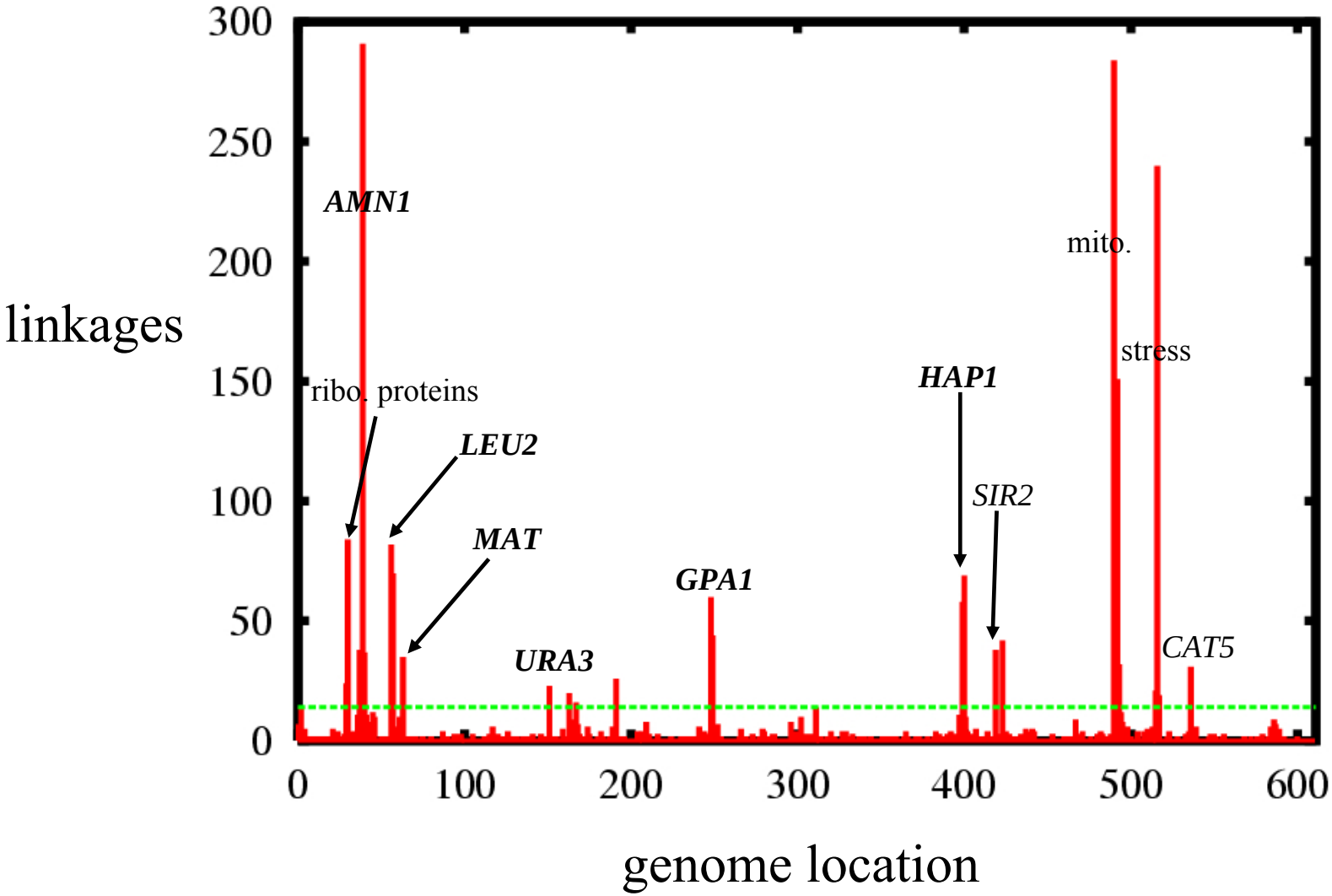
3055 messages link to 100-250 distinct loci

Is the distribution of linkages random?





Genomewide distribution of *trans* linkages



1840 messages link to 15 loci (59% of all linking traits)

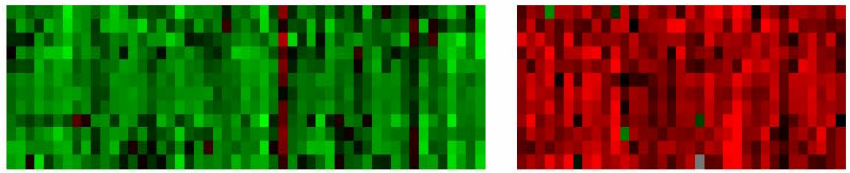
Science 2002

Positional cloning of *AMN1*

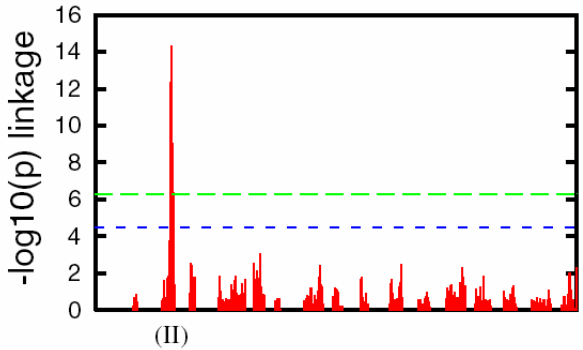
BUD9
DSE3
CST13
ISR1
YKL132C
DSE2
DSE1
SCW11
PRY3
SUN4
EGT2
DSE4

chr 2 locus:
inherit RM

chr 2 locus:
inherit BY



Daughter cell specificity



expression of cluster
shows linkage

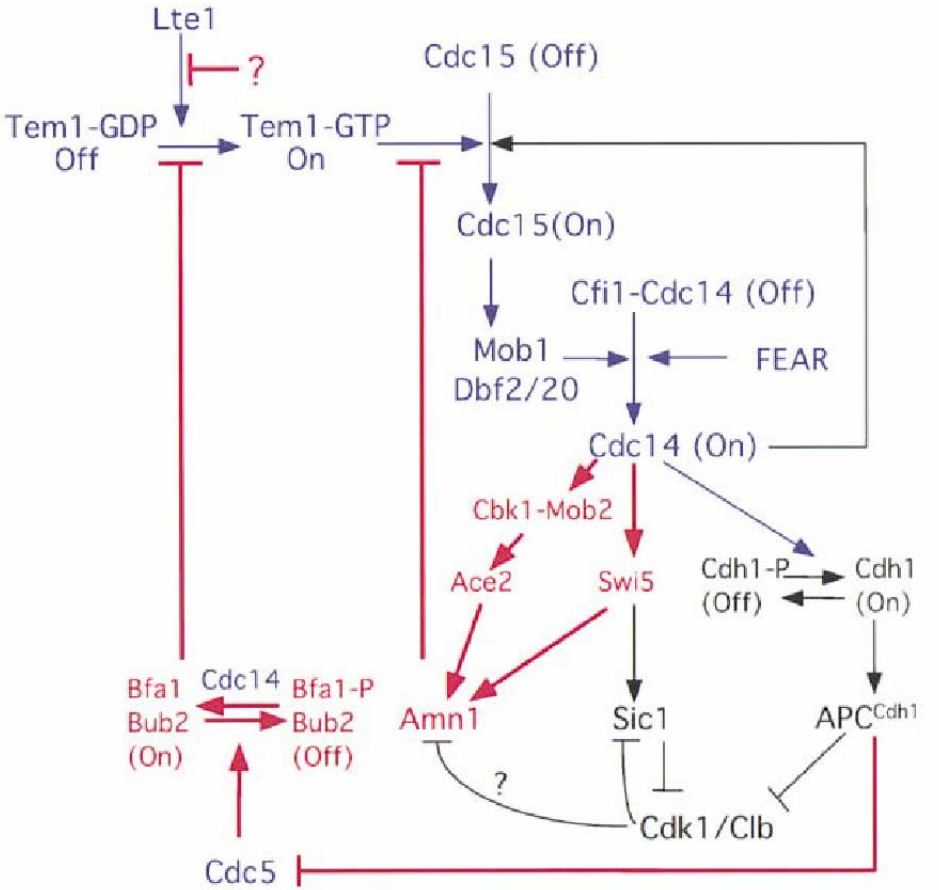
recombinants narrow region to a single gene: *AMN1*

BY allele of AMN1 carries missense mutation at conserved site

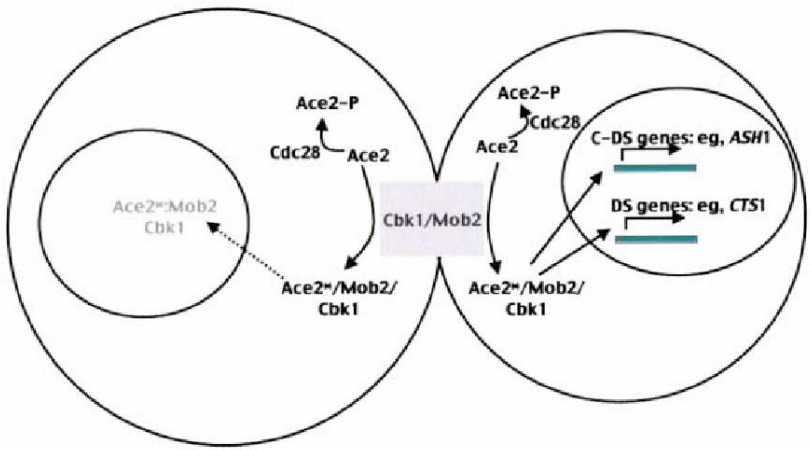
```
        RM  DNFLLRLSQS IPNLKHLDLR ACDNVSDSGV VCIALNCPKL KTFNIGRHRR
        BY  DNFLLRLSQS IPNLKHLVLR ACDNVSDSGV VCIALNCPKL KTFNIGRHRR
S. paradoxus                   D
S. mikatae                     D
S. bayanus                     D
S. servazzii                   D
Zygosaccharomyces rouxii       D
Kluyveromyces lactis           D
Pichia angusta                 D
```

Nat Genet 2003

AMN1: negative regulator of mitotic exit network



Wang et al. *Cell* **112**:697-709 2003



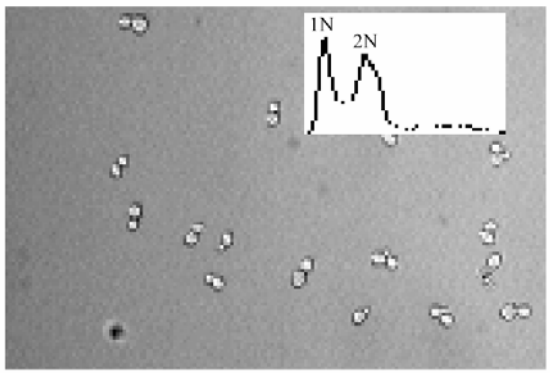
Colman-Lerner et al. *Cell* **107**:739-750 2001

Molecular proof of effect of *AMN1*

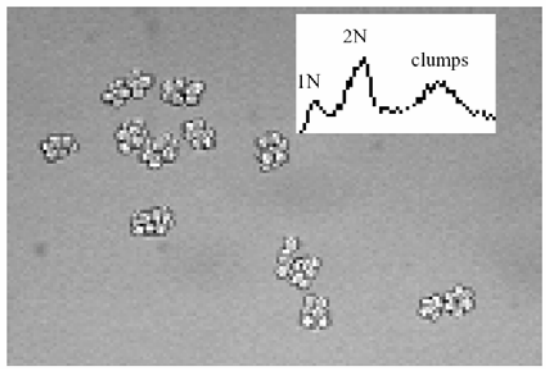
Effect on gene expression

gene	ORF	BY:RM[a]	segregants[b]	RM*amn1*:RM[c]
SCW11	YGL028C	18.0	12.2	16.2
DSE1	YER124C	21.4	18.0	32.9
DSE2	YHR143W	48.2	26.0	40.6

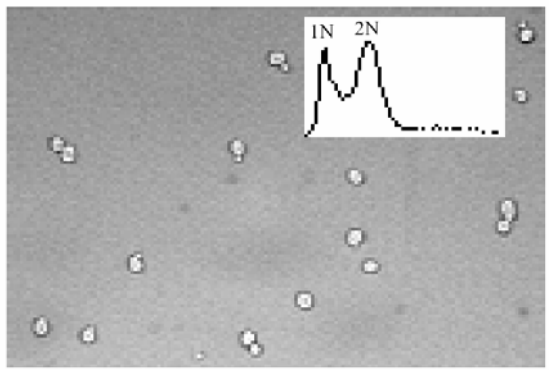
Effect on clumpy growth



BY



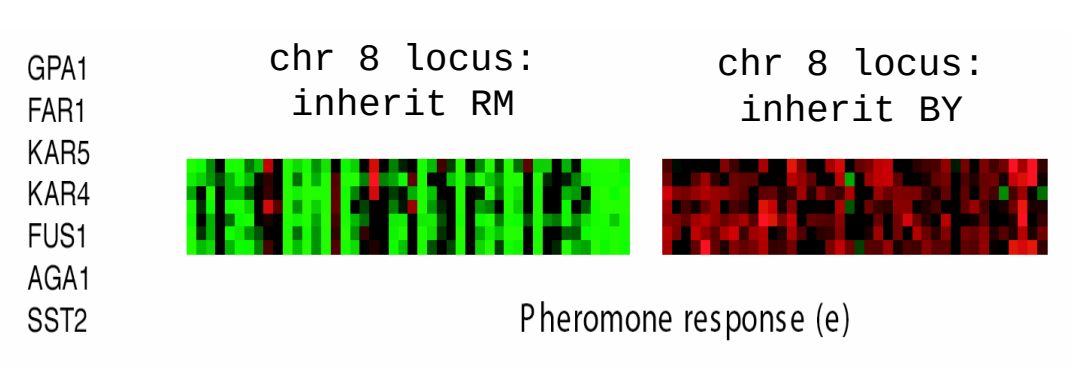
RM

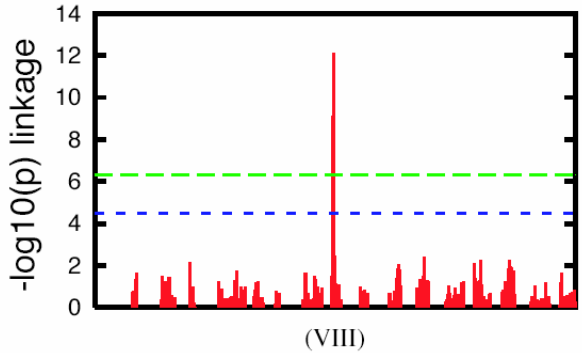


RM Δ*AMN1*

Nat Genet 2003

Positional cloning of *GPA1*





expression of cluster shows linkage

GPA1 maps to region of linkage and encodes a G protein alpha subunit coupled to pheromone receptors

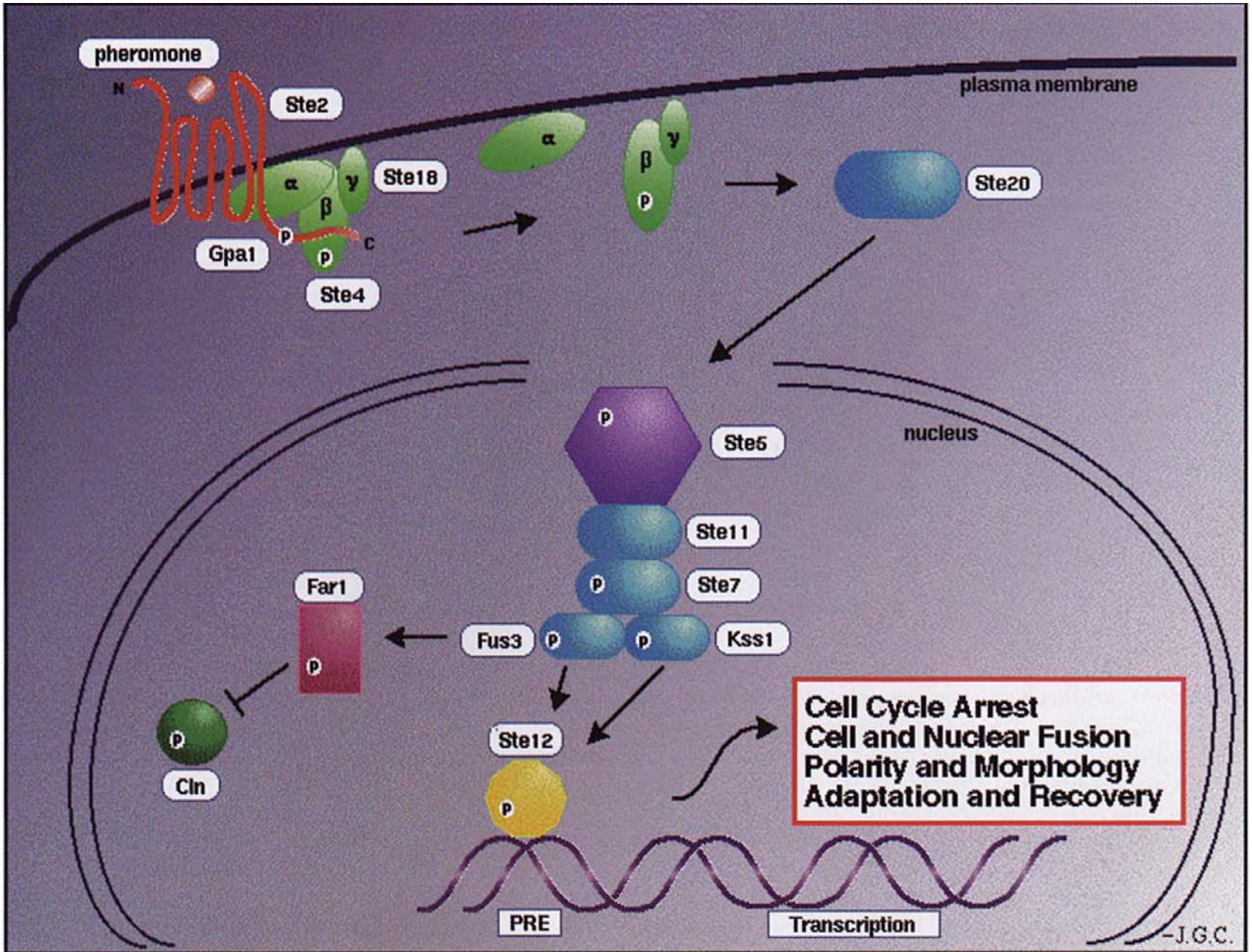
BY allele of GPA1 carries missense mutation at conserved site

```
451  KFVLSAVTDL IIQQNLKKIG II*
451  KFVLSAVTDL IIQQNLKKSG II*
```

S.Paradoxus	S
S.Mikatae	S
S.Bayanus	S
S.bayanus var uvarum	S
S.Klyuveri	S
K. thermotolerans	S
Klyuveromyces lactis	S
Dedaromyces hansenii	S
Candida tropicalis	S

Nat Genet 2003

Role of *GPA1* in pheromone response



Effect of *GPA1* S469I on pheromone response genes

gene	ORF	BY:RM[a]	segregants[b]	replacement[c]
GPA1	YHR005C	1.1	1.4	1.4
FAR1	YJL157C	1.1	1.7	1.8
KAR5	YMR065W	1.4	1.4	1.4
KAR4	YCL055W	1.5	1.6	1.6
FUS1	YCL027W	2.5	3.4	3.5
AGA1	YNR044W	6.2	3.7	3.8
SST2	YLR452C	2.4	2.2	1.9

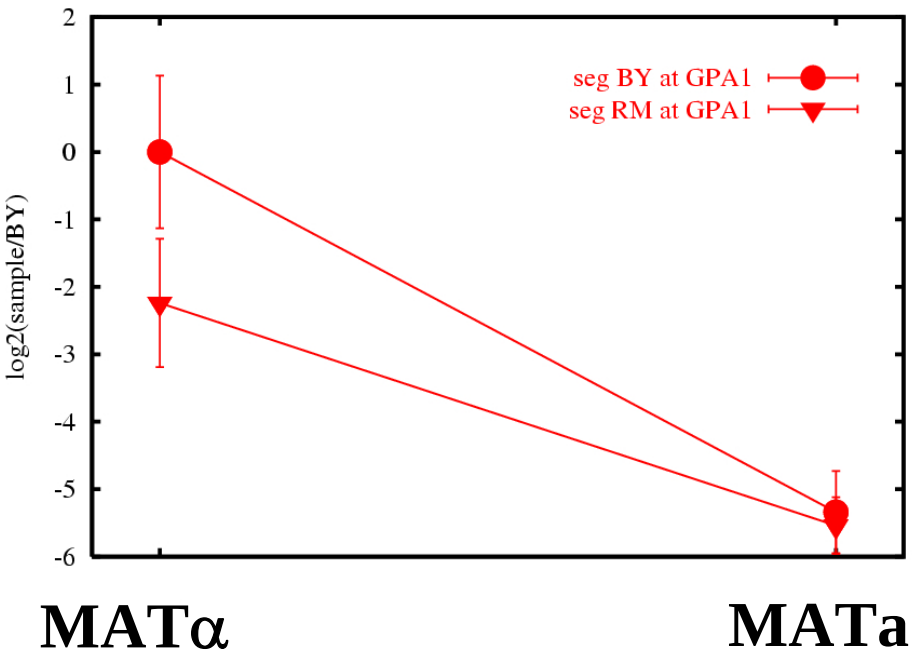
Nat Genet 2003

Two-dimensional clustering of linkages

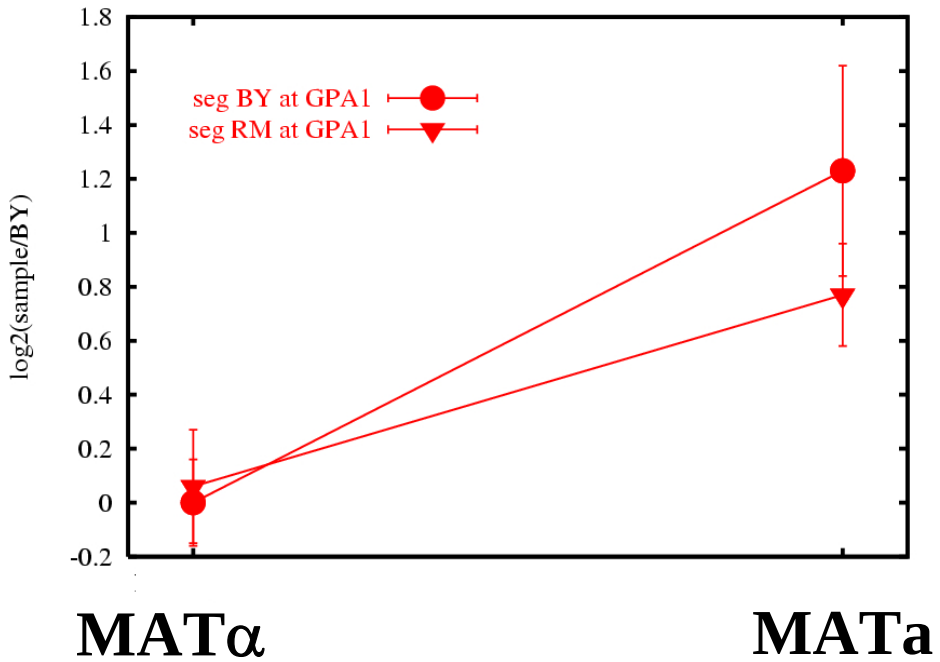
PLoS Biol 2005; Nature 2005

Epistasis between *MAT* and *GPA1 in* segregants

SAG1 α-specific



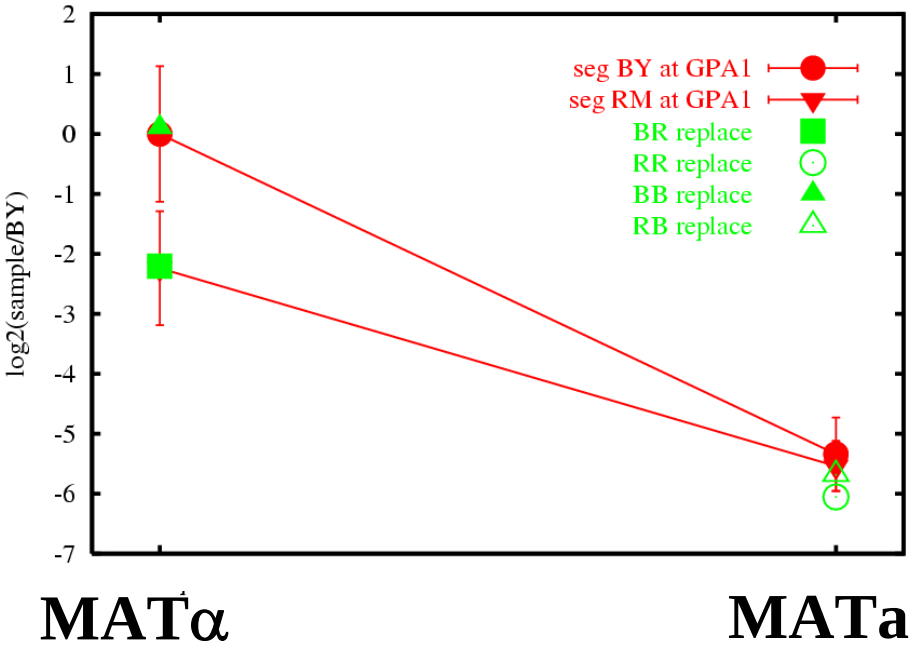
GYP8 a-specific



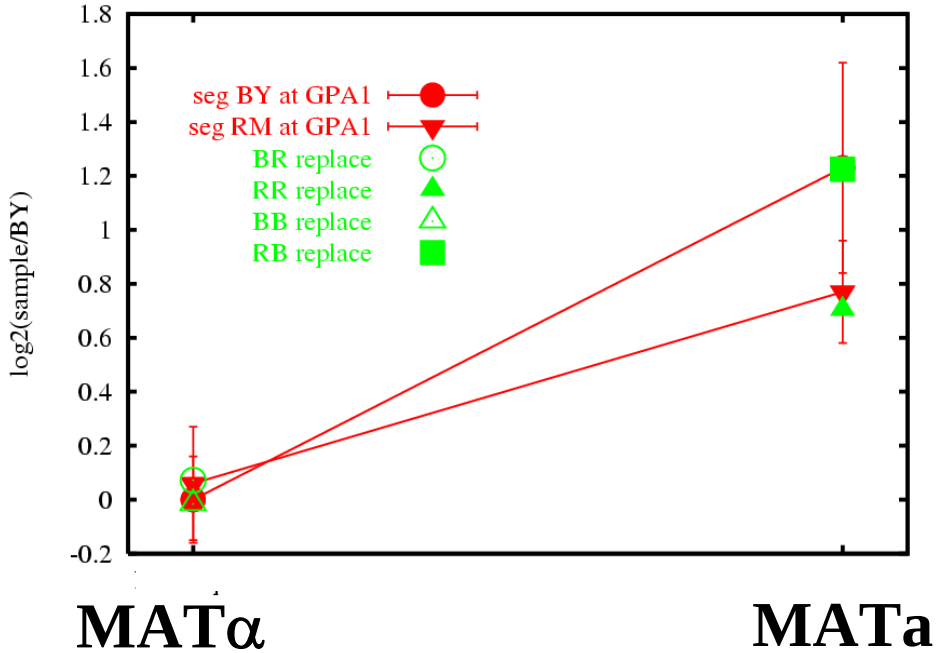
Nature 2005

Epistasis between *MAT* and *GPA1* in isogenic strains

SAG1 α-specific



GYP8 a-specific



Nature 2005

Are trans-regulatory loci transcription factors?

"This project is very creative but I'm afraid it does not make sense … why not simply use a set of knockouts for every yeast transcription factor?"

-Anonymous NIH reviewer

Analysis of *trans*-acting loci by molecular function

GO Molecular Function	Number of Genes[a]	Observed[b]	Expected[c]	Chi Squared[d]
Binding				
cytoskeleton protein binding	59	20	21.1	0.06
ATP binding	160	58	56.3	0.05
guanyl nucleotide binding	67	30	24.4	1.29
RNA binding	334	127	120.1	0.39
Chaperone	68	23	24.5	0.09
Enzyme				
Transferase				
transferase, 1 carbon	63	22	23.1	0.05
transferase, phosphorus	298	105	107.6	0.06
transferase, glycosyl	87	33	30.2	0.27
Kinase	216	74	77.9	0.20
Nuclease	90	28	31.3	0.34
Ligase	112	38	39.7	0.07
Isomerase	43	13	14.4	0.14
Oxidoreductase	227	88	80.5	0.70
Hydrolase				
hydrolase, acting on ester bonds	206	66	72.4	0.56
hydrolase, acting on acid anhydrides	217	86	76.0	1.30
hydrolase, acting on glycosyl bonds	59	18	14.1	1.05
peptidase	103	39	37.1	0.10
Enzyme Regulator	101	35	36.5	0.06
Signal Transducer				
receptor	31	12	11.0	0.09
receptor signaling protein	78	30	27.9	0.15
Structural Molecule	335	116	116.9	0.01
Transporter				
protein transporter	39	8	14.1	2.61
amino acid transporter	26	11	9.2	0.36
carbohydrate transporter	31	6	10.1	1.65
ion transporter	99	35	34.9	0.00
carrier	164	59	56.5	0.11
Translational regulator	59	20	20.9	0.04
Triplet codon adaptor	300	101	109.1	0.60
Transcriptional regulator				
transcription cofactor	35	16	10.6	2.78
transcription factor	53	19	18.5	0.01
RNA polymerase II transcription factor	116	40	42.1	0.11

Nat Genet 2003

Comparison of whole genomes at nucleotide resolution using Affymetrix™ yeast tiling microarrays
OR
Genome resequencing without a genome center

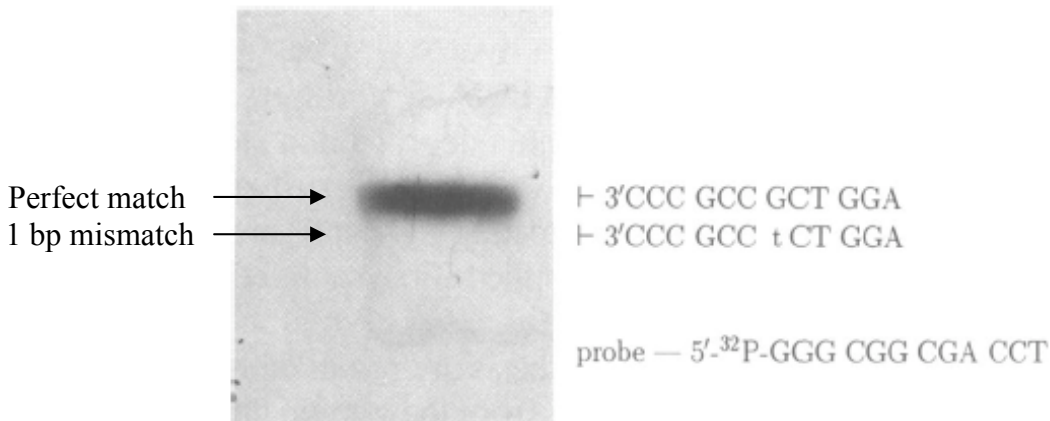
```
5′-CTGAATATGCATTGAAATAAGATCC
       ATATGCATTGAAATAAGATCCAAAC
           GCATTGAAATAAGATCCAAACAGCT
               TGAAATAAGATCCAAACAGCTAAGA
                   ATAAGATCCAAACAGCTAAGAACAG
                       GATCCAAACAGCTAAGAACAGGAAA

3′-GACTTATACGTAACTTTATTCTATGTTTGTCGATTCTTGTCCTTT
```

probes

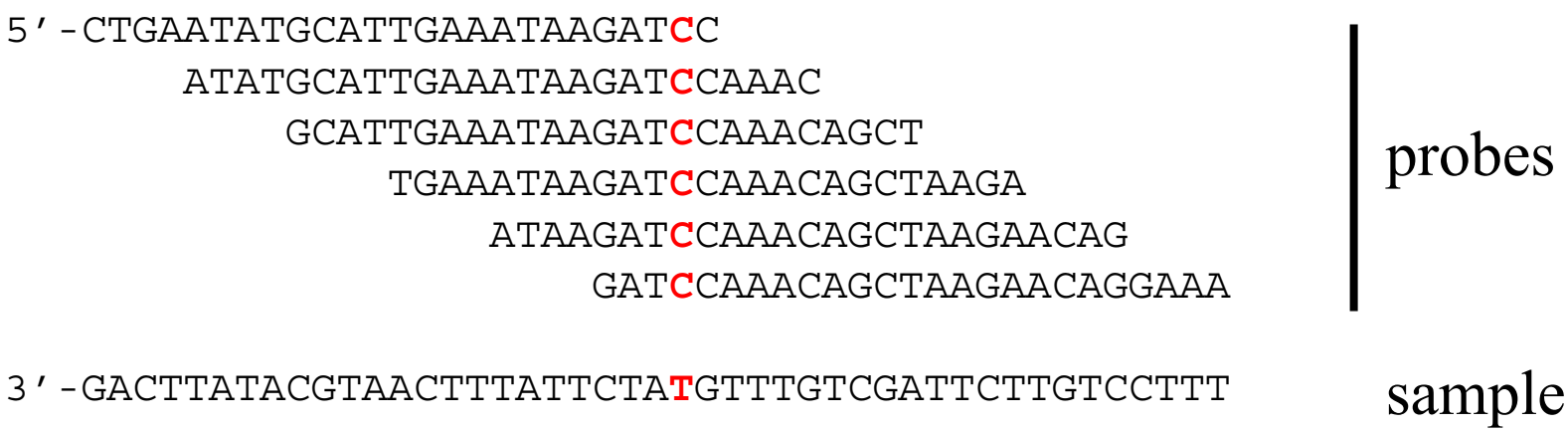
sample

Hybridization efficiency is acutely sensitive to mismatches when oligonucleotides are short



Maskos and Southern, NAR (1992)

Affymetrix Yeast tiling arrays provide complete and redundant coverage of the yeast genome



Comparison of two genomes to model decrease in hybridization due to SNPs

Nonpolymorphic strain

S288C

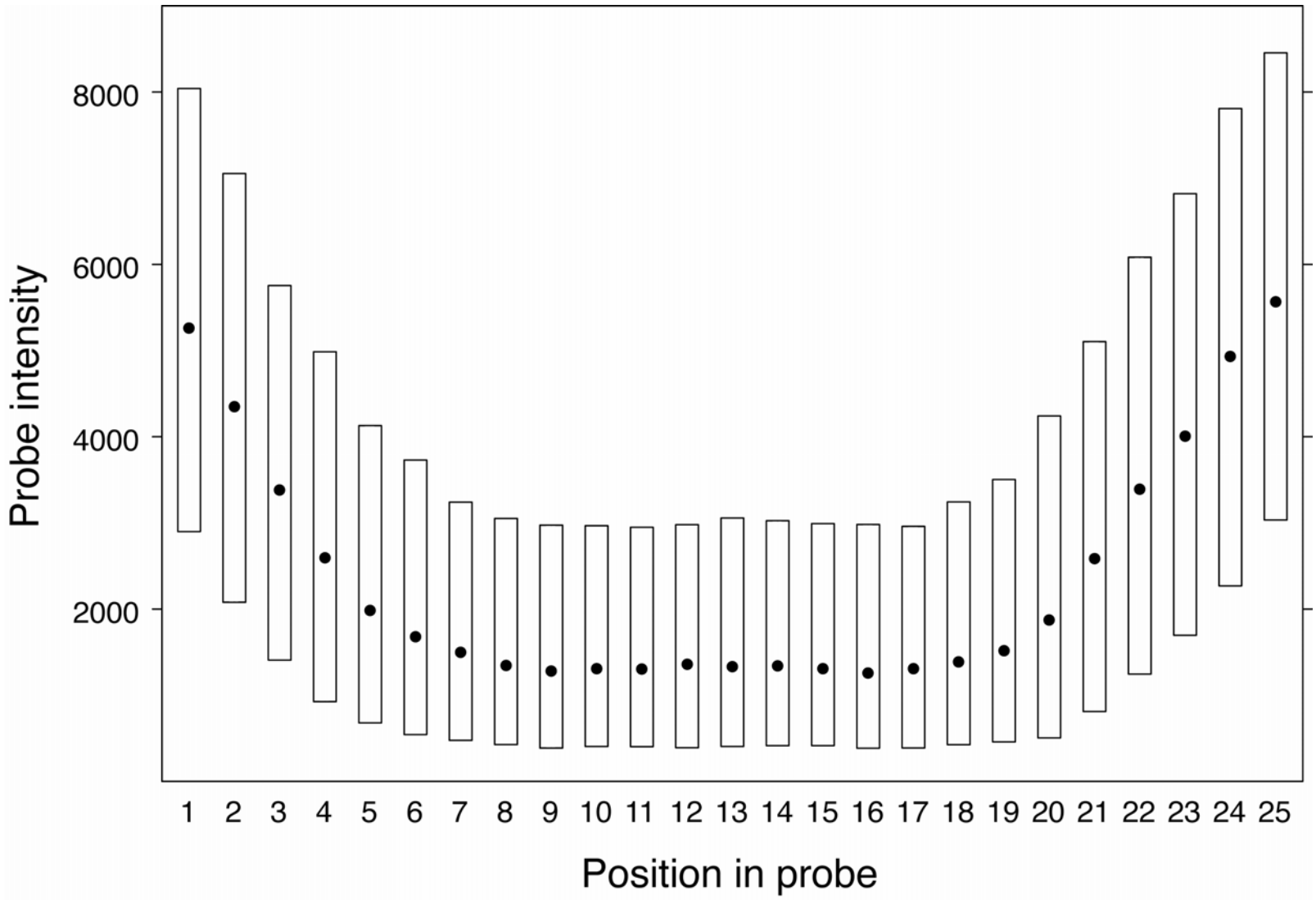
- Reference sequence represented on array
- Hybridization intensities reflect maximal binding of complementary DNA

Polymorphic strain

RM11-1A

- High quality sequence of wild strain
- 24,848 isolated SNPs overlapped by 123,016 probes
- Hybridization intensities reflect effect of mismatch on maximal binding

Hybridization decrease in presence of SNP is related to position within probe

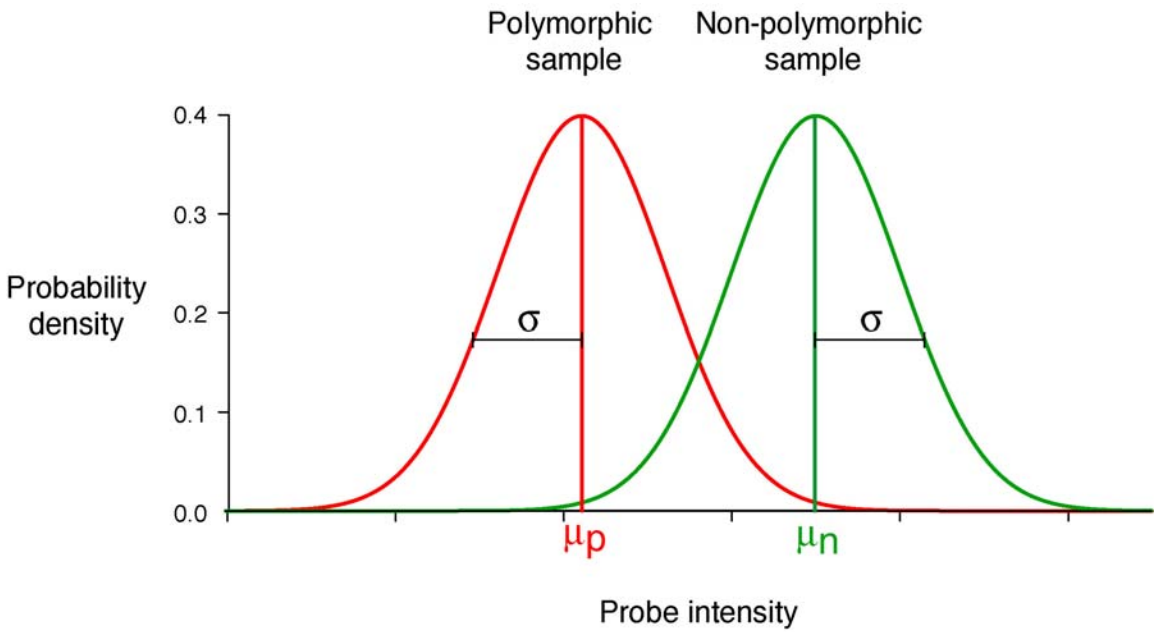


Model decrease in hybridization due to SNP at Probe(*i*)

- Position in SNP (*j*)
- PM-MM for reference sequence
- GC content
- Local sequence context (*t*)

$$D_{ijt} = \alpha_{ij} + \beta_{ij}\left(GC_i\right) + \gamma_{ij}\left(PM_i - MM_i\right) + \delta_{ij}PM_i + I_{ijt}$$

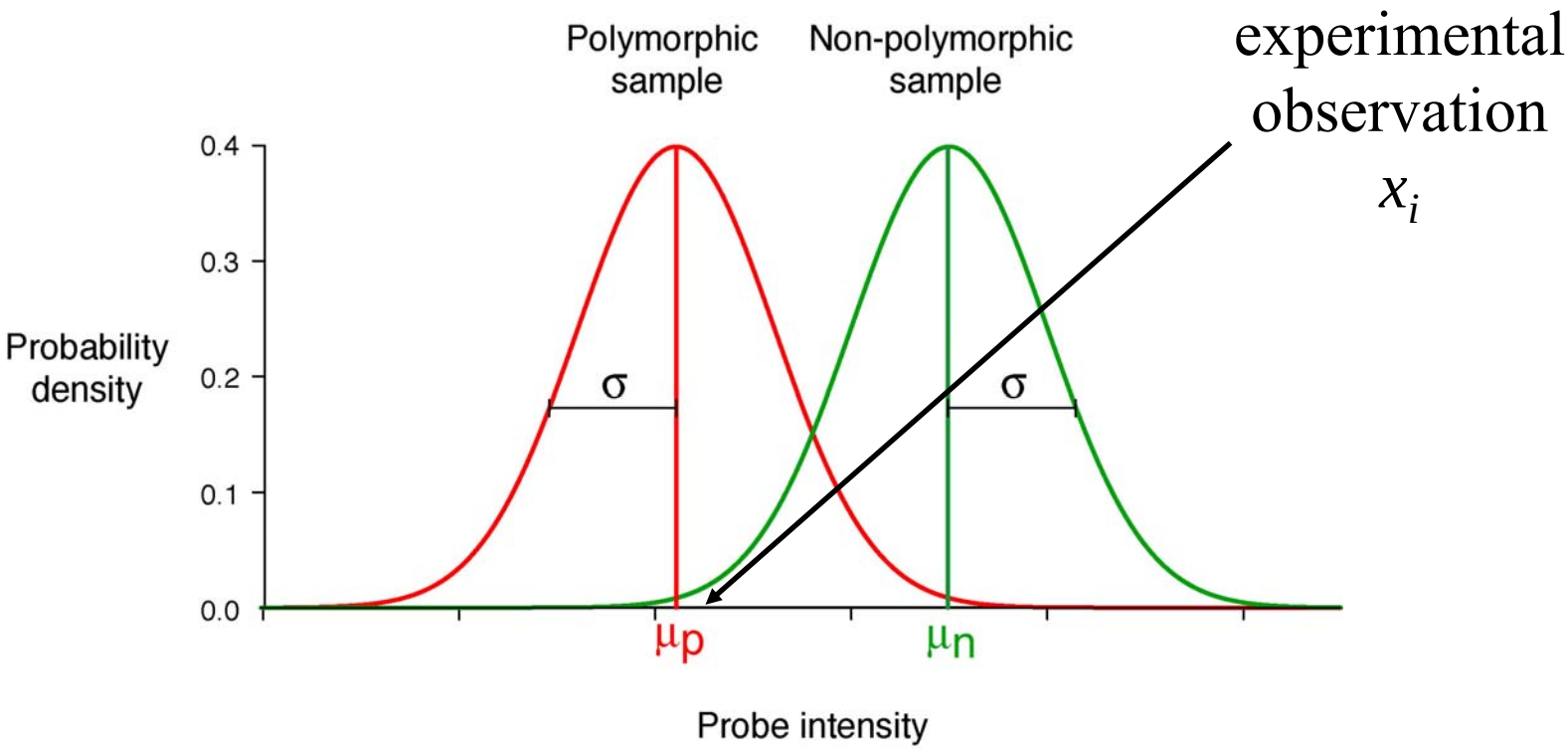
Model decrease in hybridization due to SNP



Model inputs

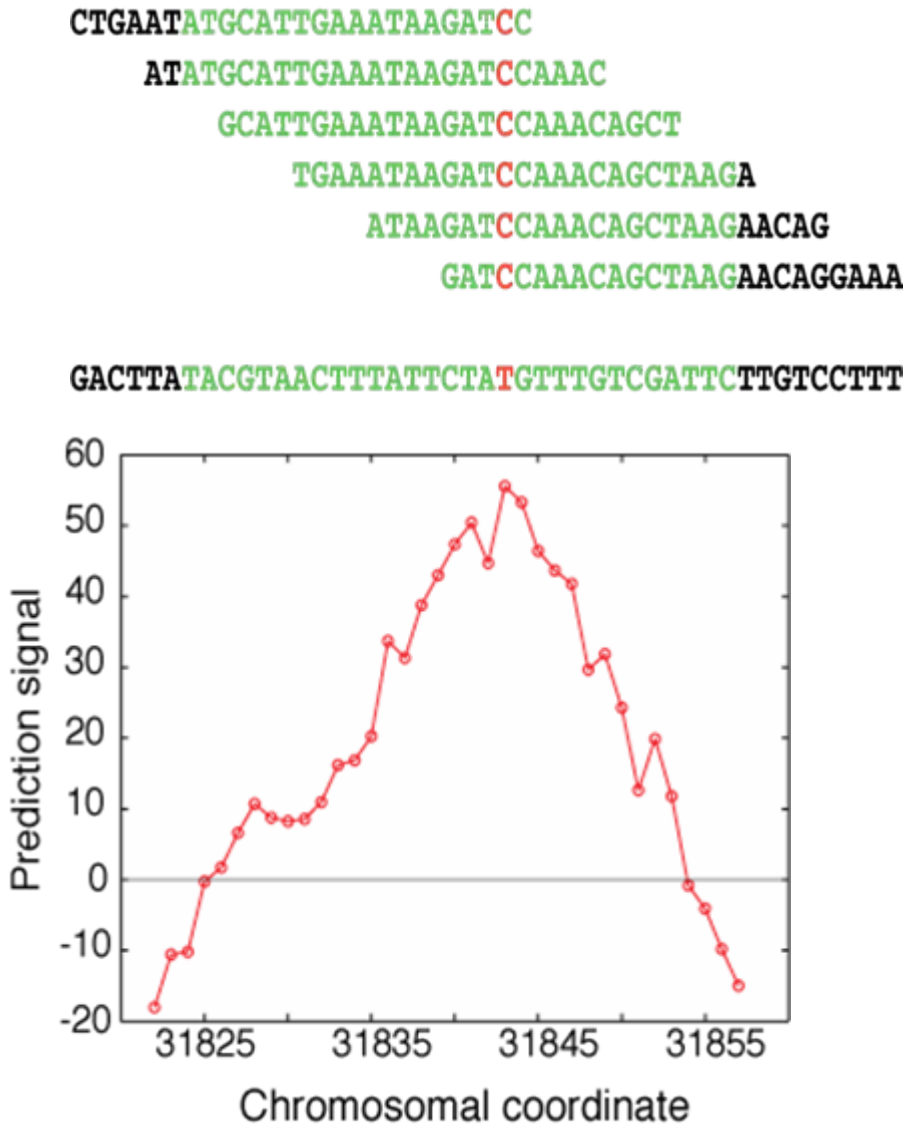
- position of SNP in probe
- PM minus MM intensity for reference sequence
- probe GC content
- local sequence context

Use model to call polymorphic sites

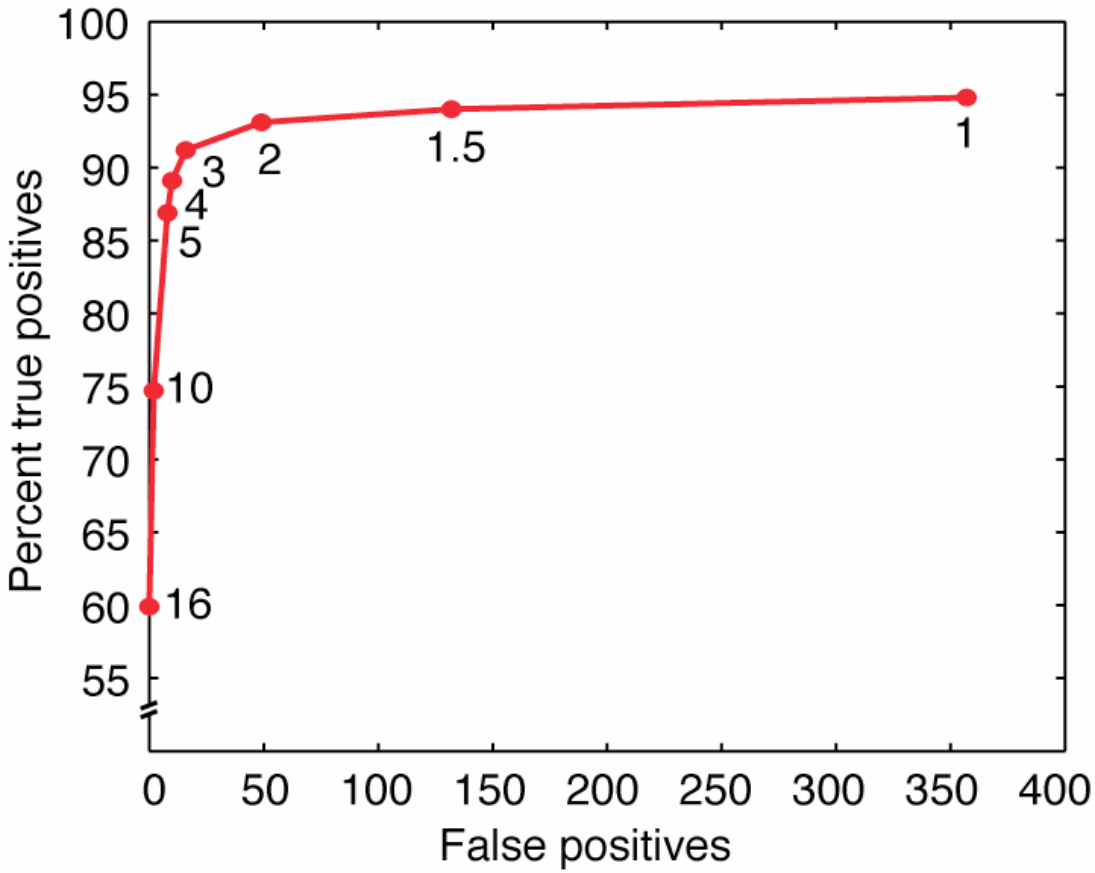


$$\text{Prediction signal} = \log \frac{\text{Likelihood site is polymorphic}}{\text{Likelihood site is nonpolymorphic}}$$

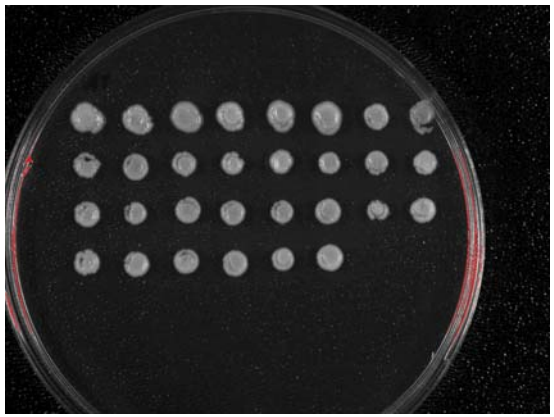
A positive SNP prediction is associated with a region of elevated signal



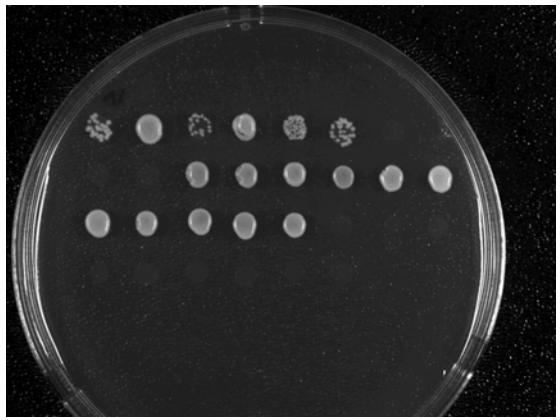
Analysis of sequenced strain YJM789 identifies >90% of 30,690 known SNPs



Collection of independent spontaneous mutants

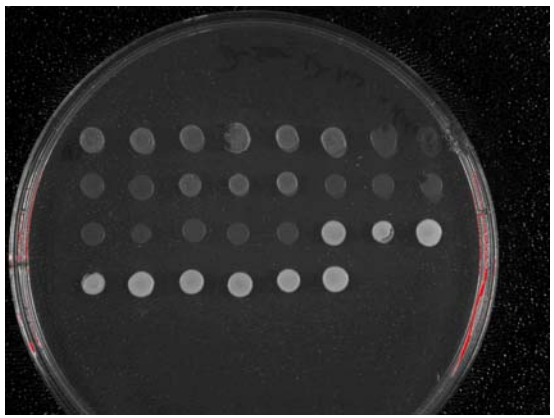


YPD



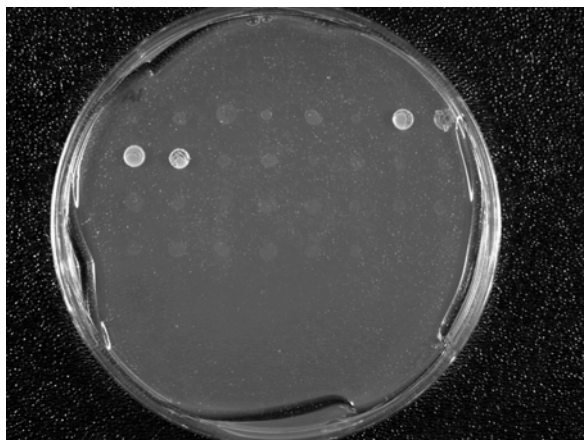
→ *CAN1*

60mg/L canavanine



→ *GAP1*

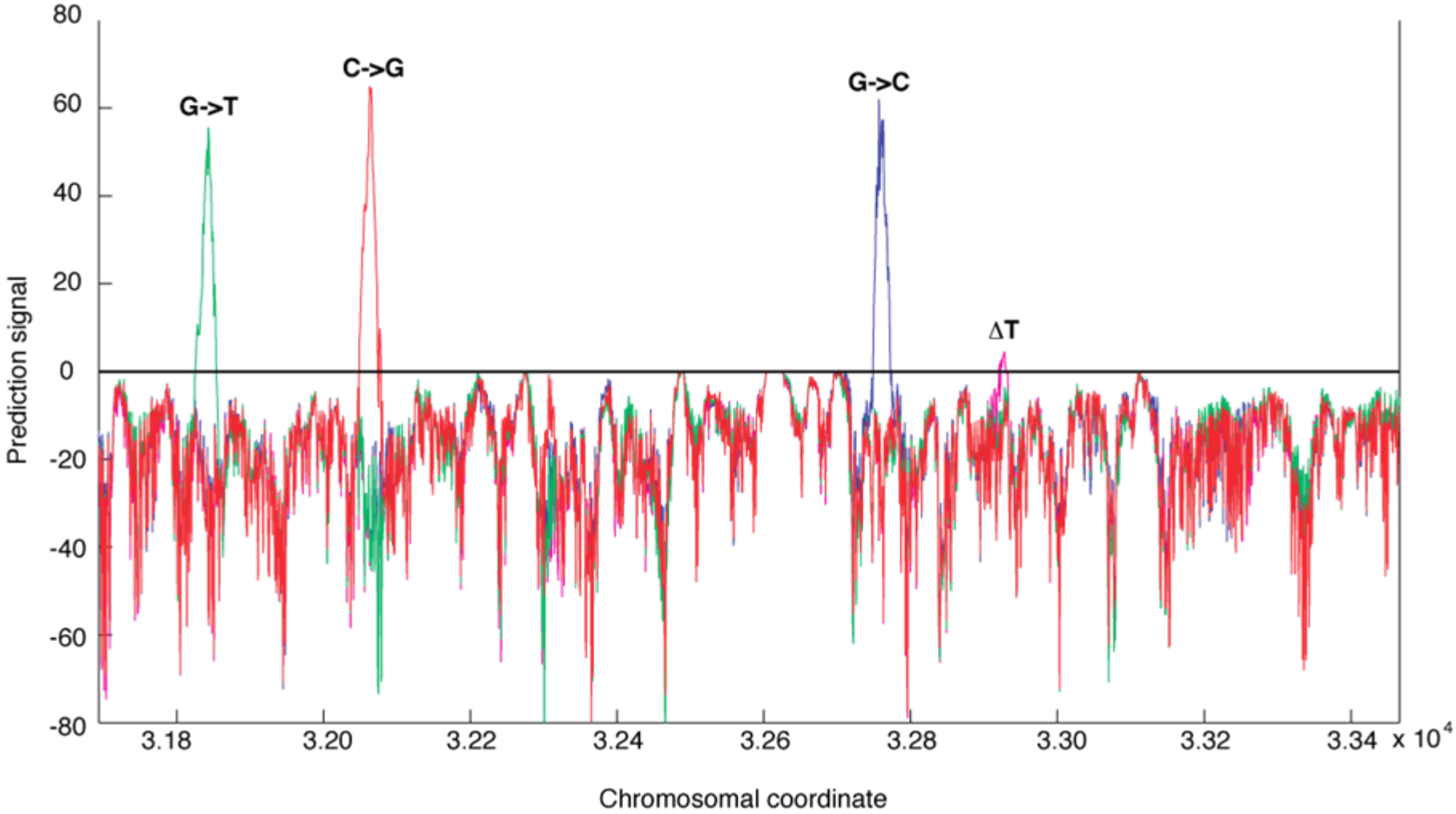
3mM D-His +1.5mM D-Ser



→ *FCY1*

1mM 5-fluorocytosine

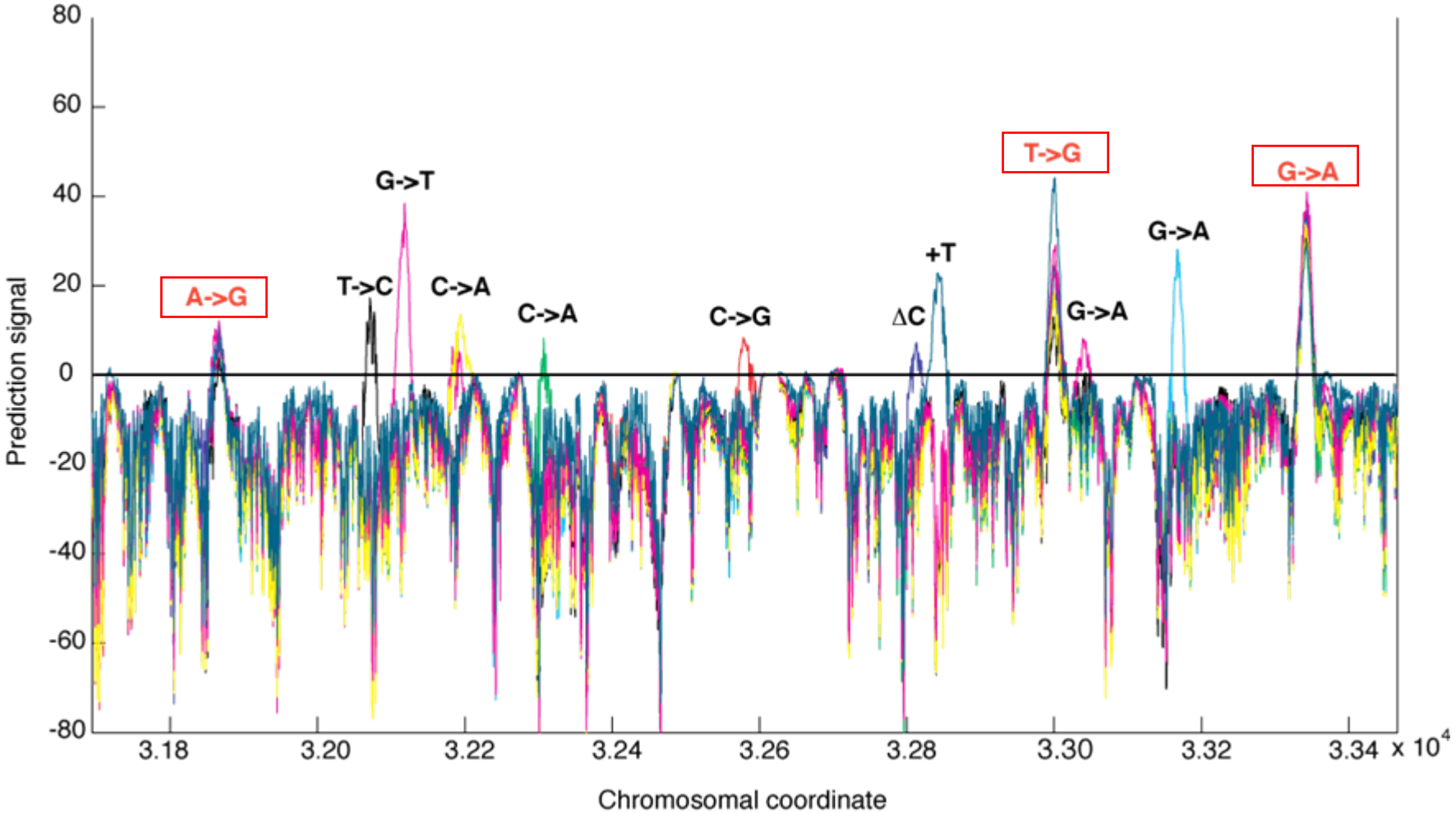
A positive signal prediction at *CAN1* for four CAN^R mutants



Selection of spontaneous mutants associated with small number of additional SNPs genome-wide

Mutant	Prediction at expected locus	Confirmed Mutation	Genome-wide predictions	Rank of known mutation	Sequence confirmed mutations
Can1-1001	chrV:32,758 (62)	chrV:32,758G→C	5	1	3
Can1-1002	chrV:32,929 (4.5)	chrV:32,924ĘT	5	-	2
Can1-1003	chrV:31,844 (55.6)	chrV:31,844G→A	3	1	3
Can1-1004	chrV:32066 (65)	chrV:32,064C→G	4	1	3
Gap1-1002	chrXVI:514,919 (105)	chrXI:514,919C→G	4	1	2
Fcy1-1001	chrXVI:677,256 (178)	chrXVI:677,256C→T	120	5	2

CEN.PK CAN^R mutants have unique mutations and common SNPs in *CAN1*

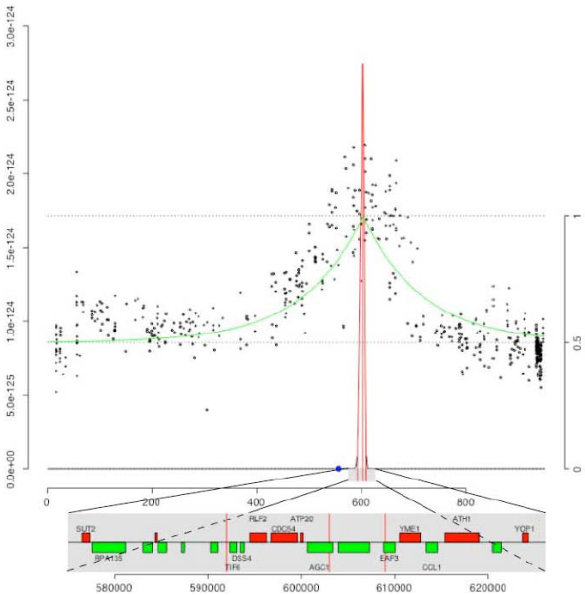


Small number of predictions genome-wide for mutants in non-reference background

Strain	**Prediction at *CAN1***	**Confirmed mutation**	**Predictions genome-wide**	**Rank of known mutation**
Can1-2003	chrV:32,486 (0.66)	chrV:32,487G→T	0	-
Can1-2004	chrV:32,119 (38.5)	chrV:32,119G→T	16	1
Can1-2005	chrV:33,168 (28.1)	chrV:33,169G→A	4	2
Can1-2007	chrV:32,579 (8.4)	chrV:32,580C→G	5	1
Can1-2008	chrV:32,308 (8.2)	chrV:32,304C→A	3	-
Can1-2009	chrV:32,814 (7.3)	chrV:32811ΔC	5	5
Can1-2010	chrV:32,072 (17.3)	chrV:32,077T→C	15	-
Can1-2011	chrV:31,195 (13.6)	chrV:32195C→A	2	1
Can1-2013	chrV:33040 (8.2)	chrV:33043G→A	56	10
Can1-2014	chrV:32843 (22.9)	chrV:32842+T	556	3

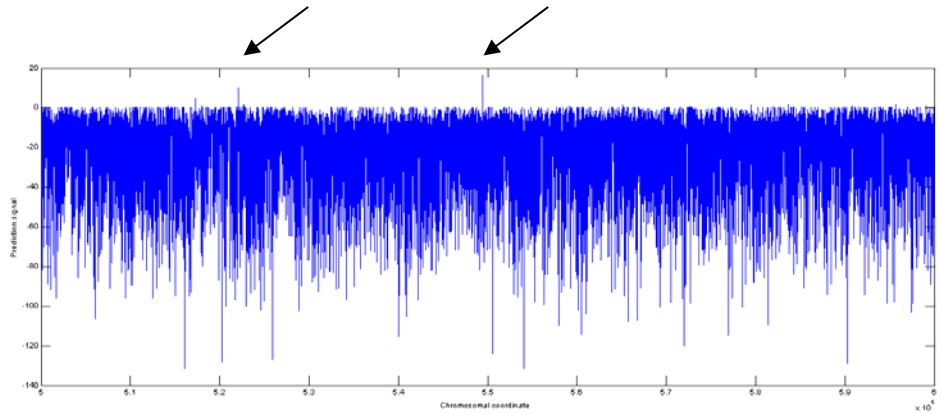
Genome-wide mutation detection aids positional cloning

Respiratory growth defect locus localized near centromere on chromosome XVI using S98 array



Brauer et al., 2006

Two candidate SNPs identified in 100kb critical region

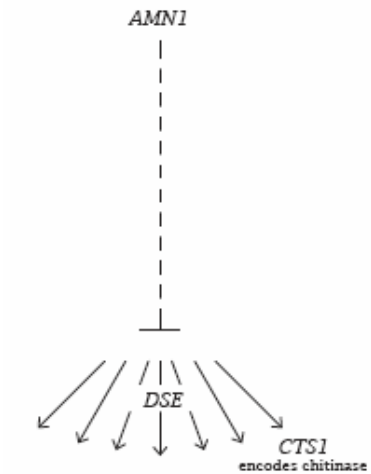


Deleterious mutation in *AEP3* in mutant strain

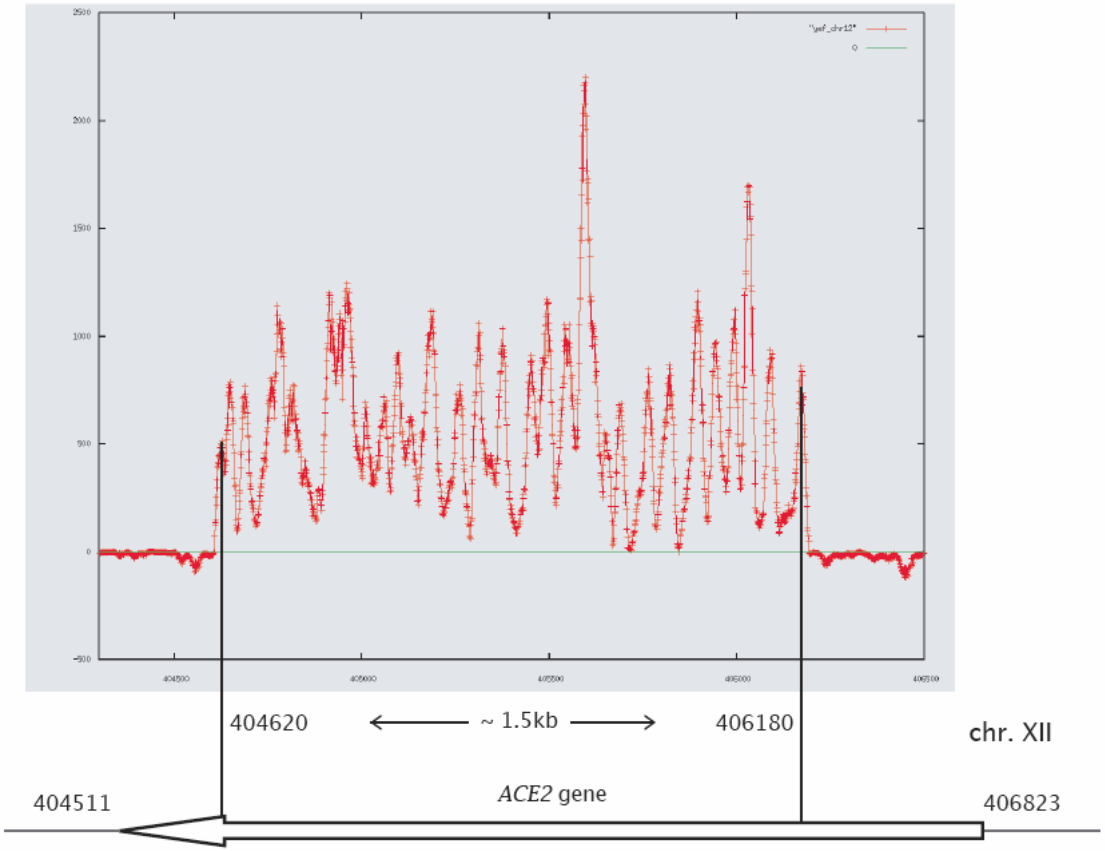
Strain	**Phenotype**	**Sequence in *AEP3***
FY3	Acetate+	ÉTTCAAATGGGTG AACAÉ
CP1AB	Acetate-	ÉTTC **C**AATGGGTGAACAÉ

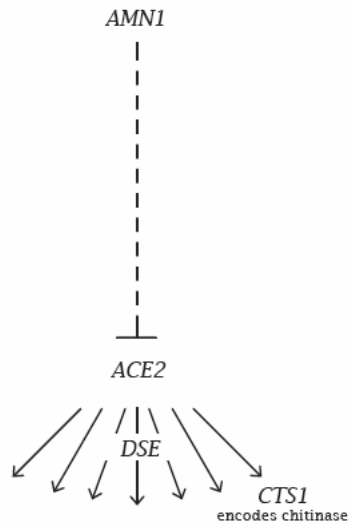
Spontaneous mutation supresses *amn1*

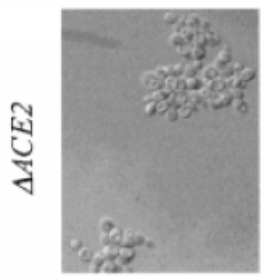
RM11	*AMN1*	DSE ↓	clumpy
BY4716	*amn1*	DSE ↑	not clumpy
RM11 *amn1*Δ		DSE ↑	not clumpy
S288c *amn1*Δ		DSE ↑	not clumpy
BY4716 *amn1*Δ (YEF1695)		DSE ↓	clumpy
BY4716 *amn1*Δ (YEF 1703)		DSE ↑	not clumpy
BY4724 *amn1*Δ (YEF 1706)		DSE ↑	not clumpy



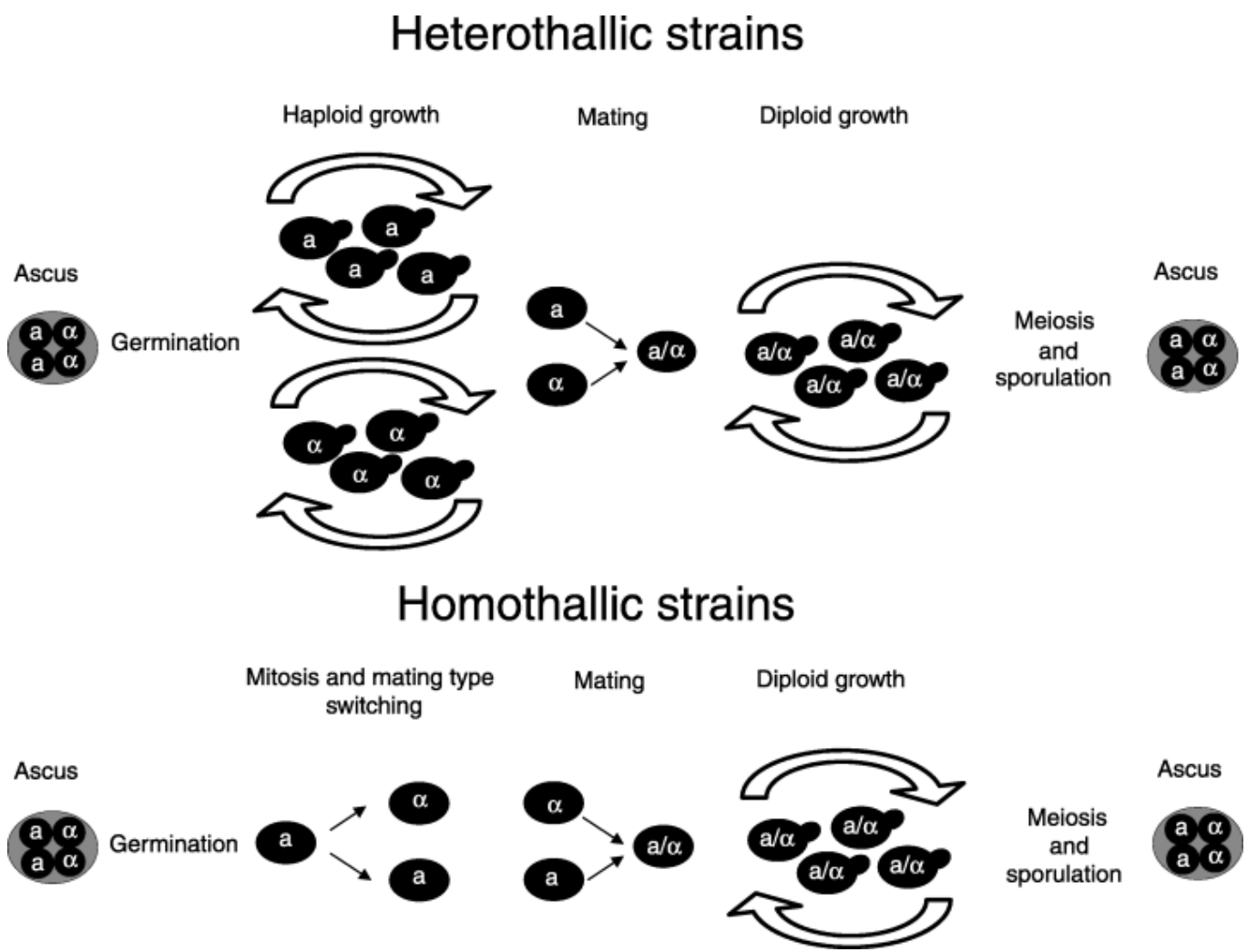
Detection of *ACE2* deletion on a tiling array







Yeast life cycle

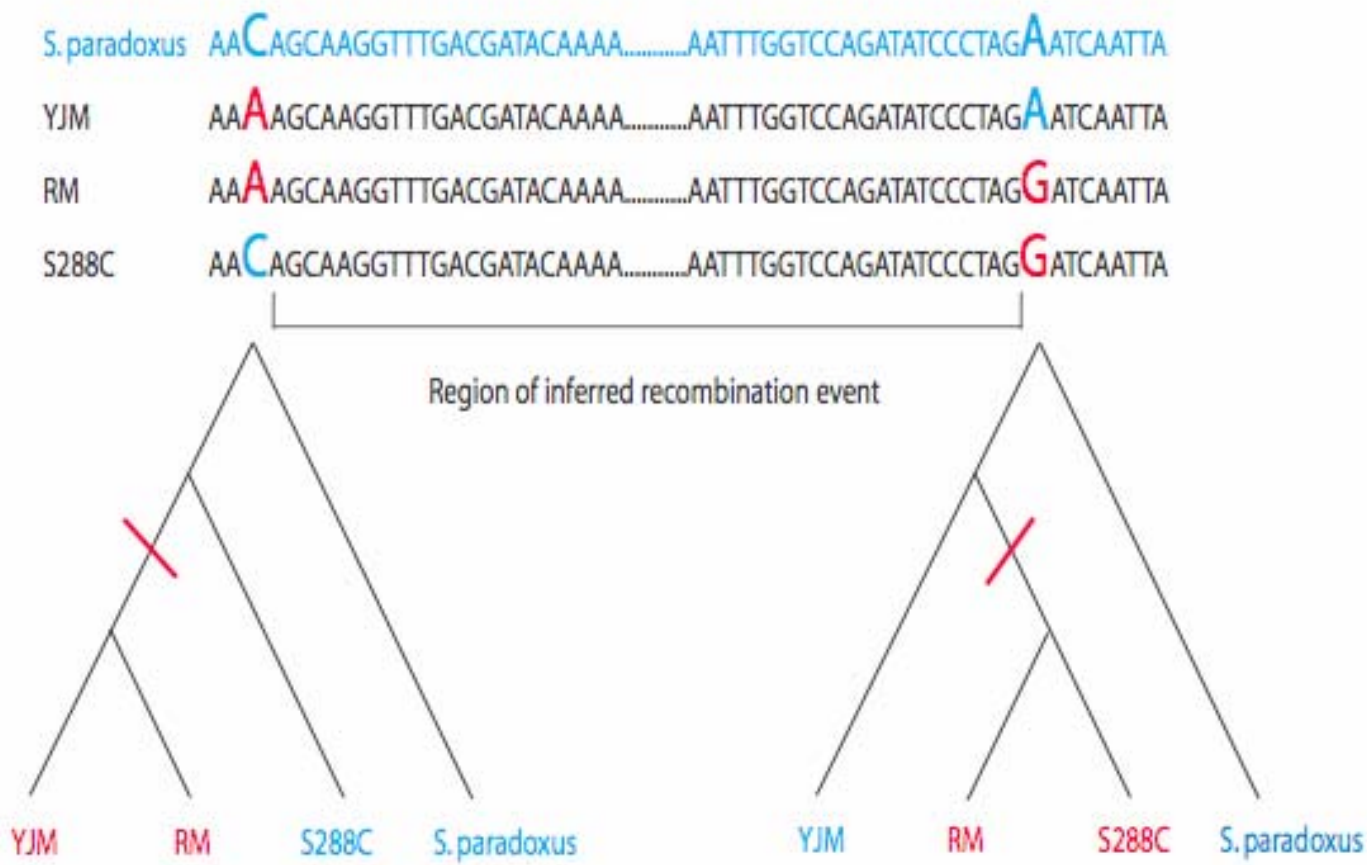


Landry et al. Molecular Ecology 15:575-591.

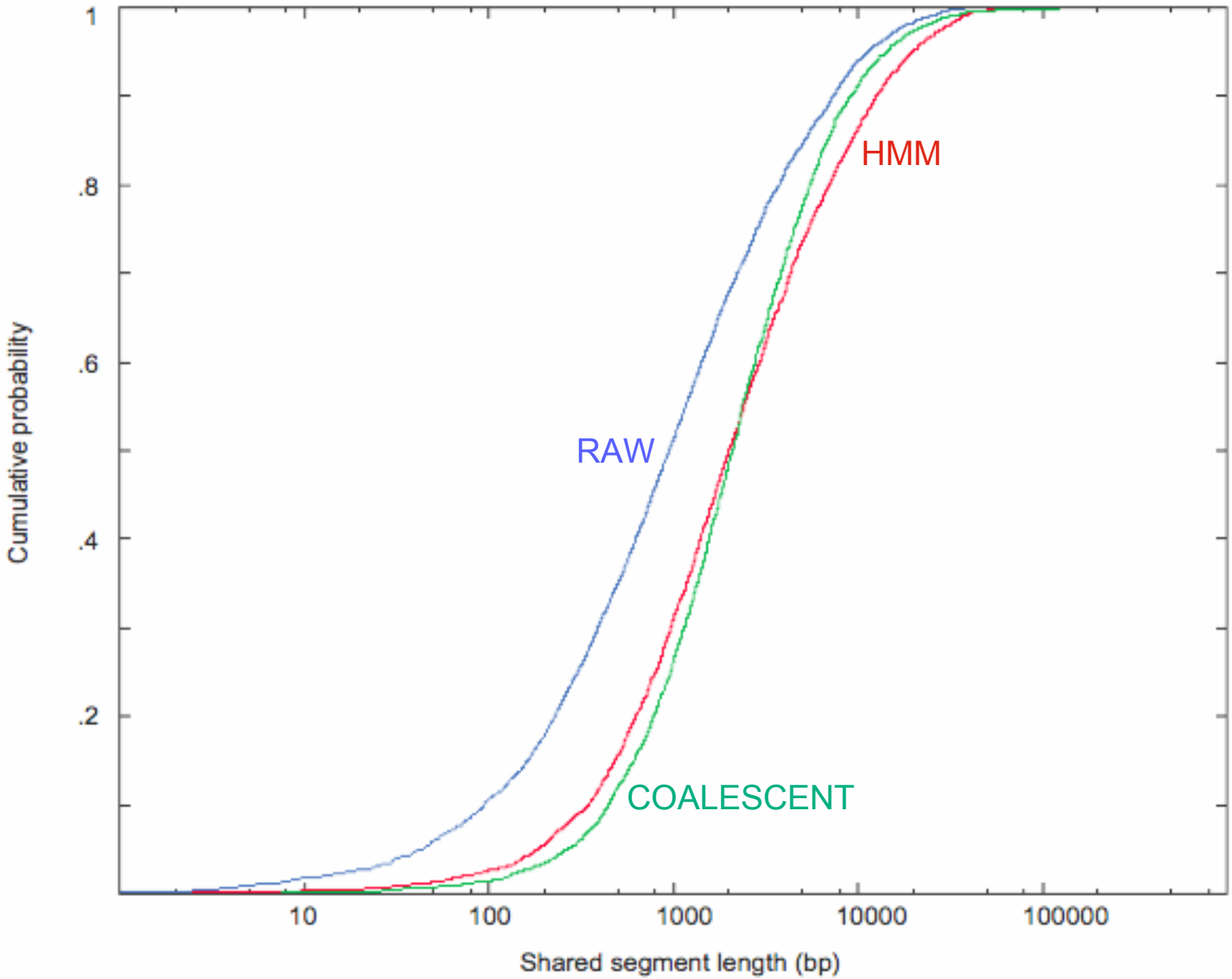
Do diverse yeast strains mate in nature?

- Mating shuffles polymorphisms via recombination
- Ancestral recombination events can be detected
- Three *S. cerevisiae* strains with sequenced genomes
  - S288C, lab strain, progenitor isolated from rotting fig
  - RM11-1a, haploid segregant from vineyard isolate
  - YJM789, originally isolated from the lung of an AIDS patient
- Sequence of *S. paradoxus* defines ancestral state

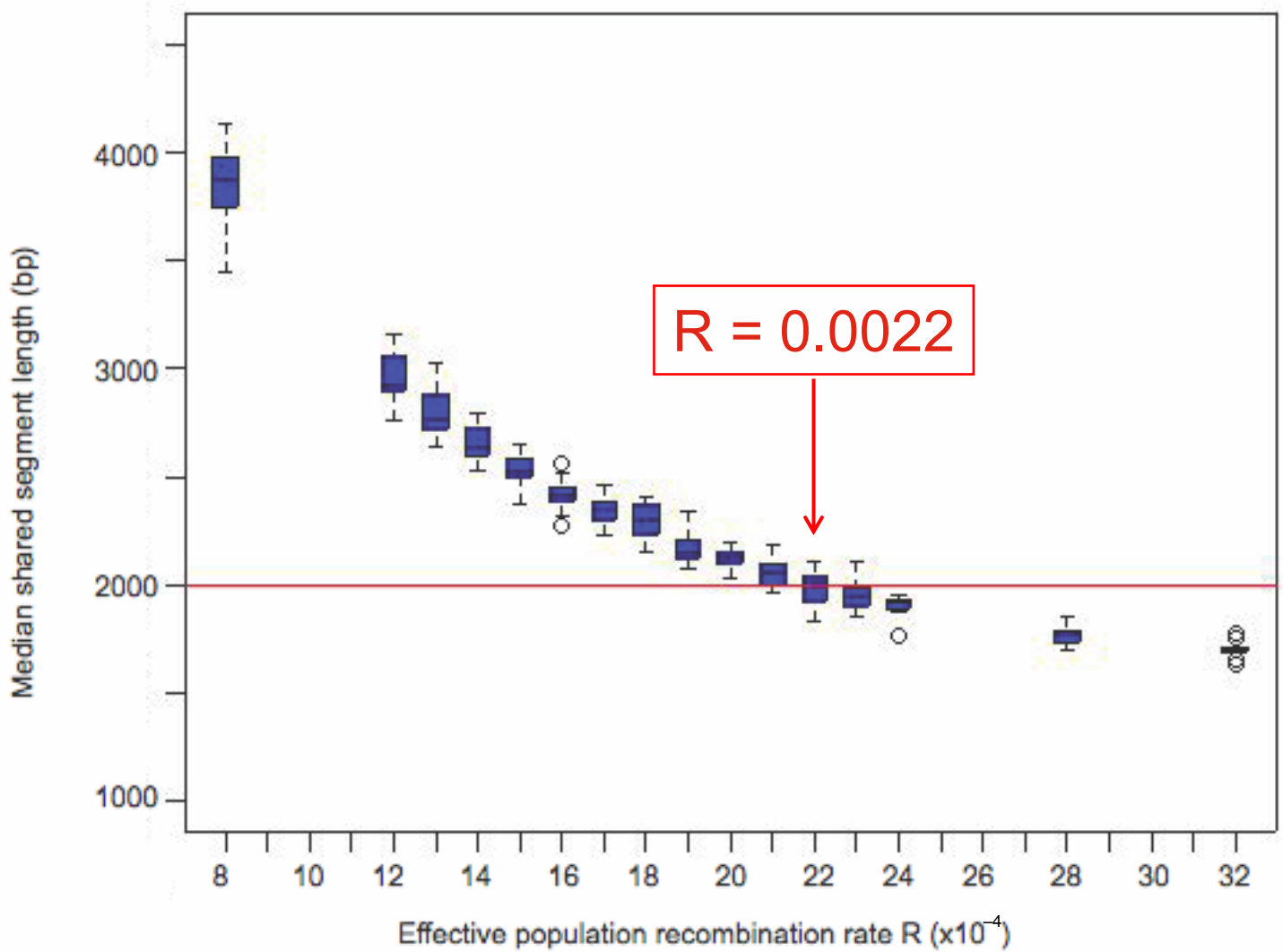
Detection of ancestral recombination



Distribution of shared segment lengths



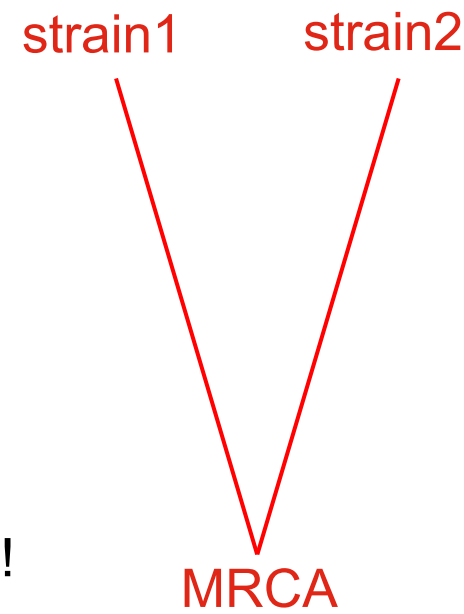
Estimating effective recombination rate



Yeast population history

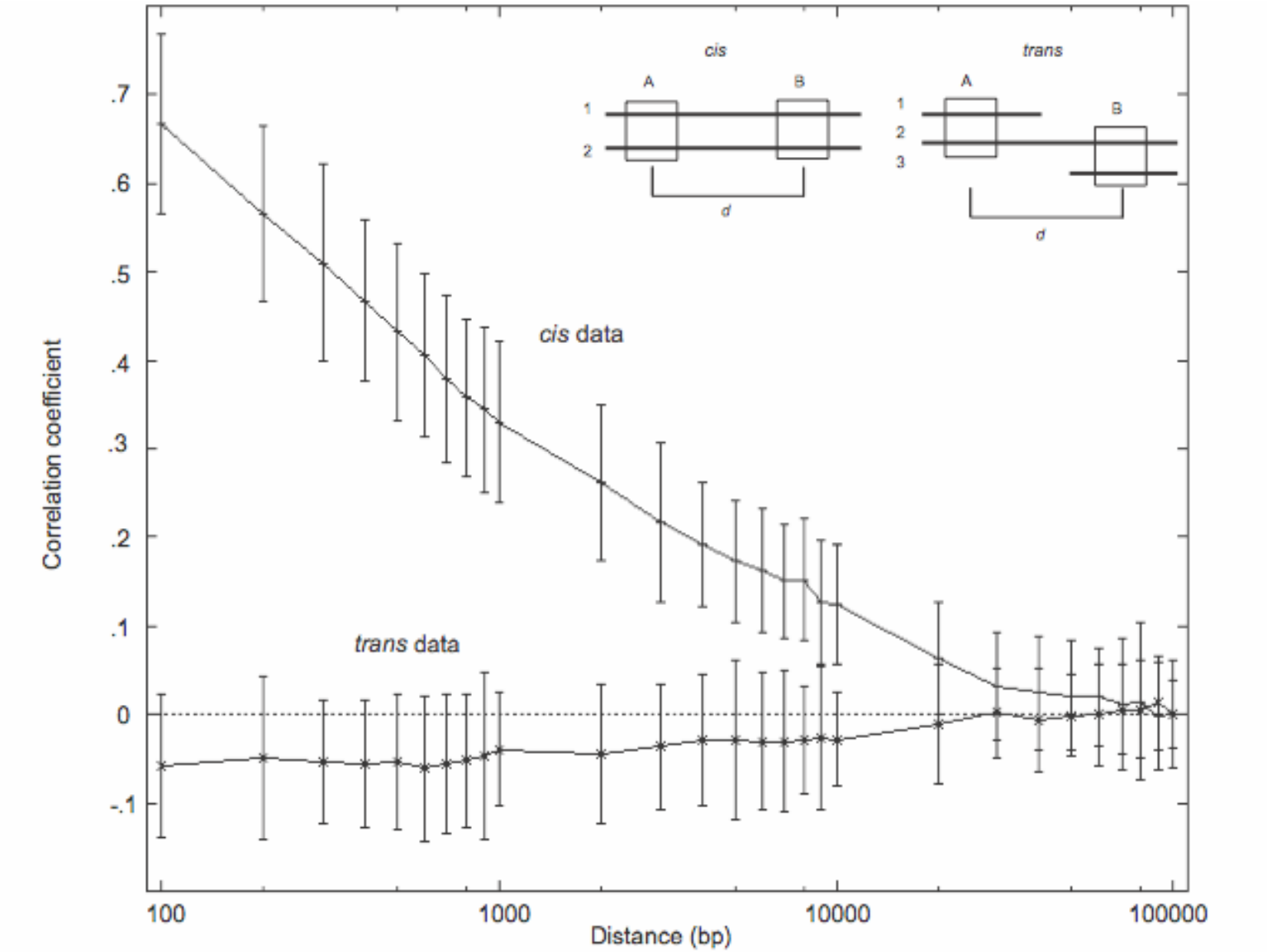
- Population mutation parameter $\theta = 0.0058$
- Population recomb. parameter $R = 0.0022$

- $\theta = 2\, G_T\, \mu$ $\quad \mu = 1.84 \times 10^{-10}$ $\quad G_T = 1.58 \times 10^{7}$
- $R = 2\, G_O\, r$ $\quad r = 3.5 \times 10^{-6}$ $\quad G_O = 314$



- One outcrossing event every 50,000 cell divisions!
- If yeast divide once per day in nature, any pair of strains diverged 43,000 years ago, with <1 outcrossing per century
- Outcrossing may be underestimated due to population structure

Correlation in polymorphism decays over kilobases



Summary

- Large fraction of genome is differentially expressed between strains
- Mapped over a thousand loci that affect expression
- Most differences are due to multiple loci
- 25% of genes have local regulatory variation
- Find *trans*-acting polymorphisms with widespread effects
- Identify genetic interactions
- Use tiling arrays for accurate polymorphism detection in yeast
- Use multiple sequenced strains to infer population history
- Diverse yeast strains mate in nature, albeit infrequently
- Recombination shuffles genomes into segments measured in kilobases

Acknowledgments

Lab Alumni
Rachel Brem
Gael Yvert
James Ronald
Jacqueline Whittle
Stephen Pratt
Eric Foss
Josh Akey
Becky Clinton
Rachel Mackelprang

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

Current Lab Members
Douglas Ruderfer
Joseph Schacherer
Erin Smith

Collaborators
David Gresham
David Botstein
Maitreya Dunham
Donna Storton

John Storey

Lewis-Sigler Institute for Integrative Genomics, Princeton