18 October 2021

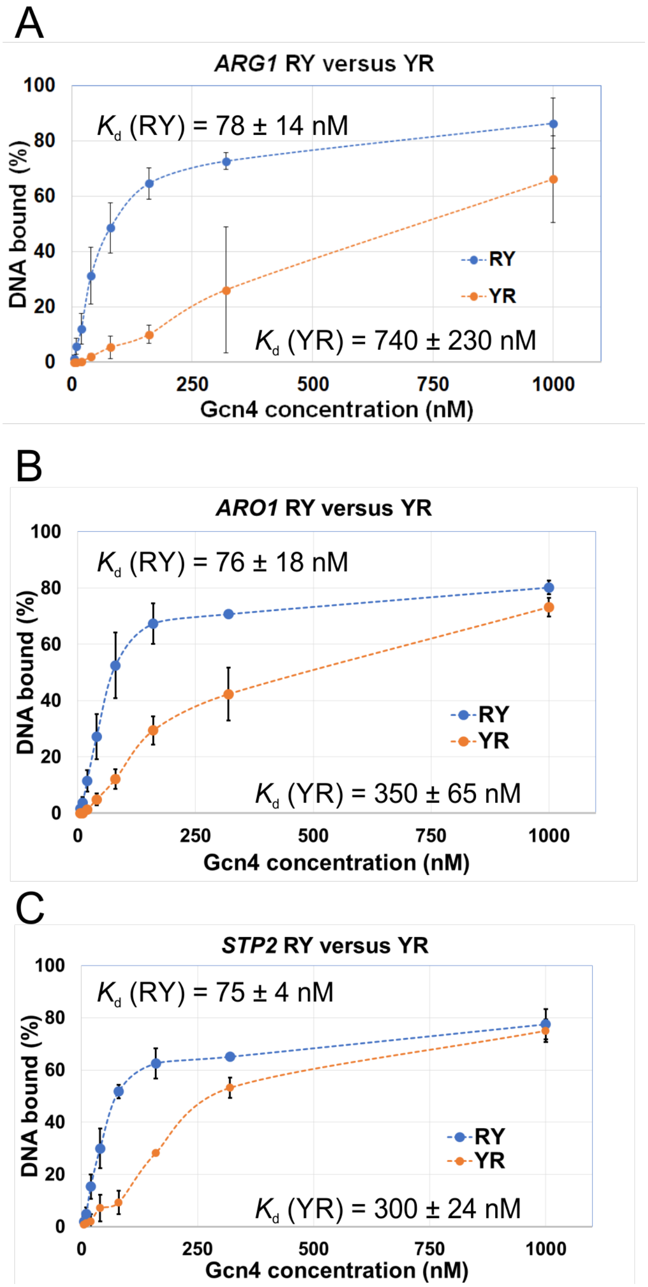
SUPPLEMENTAL DATA

A systematic, genome-wide account of binding sites for the model transcription factor Gcn4

Christopher T. Coey and David J. Clark[1]

Division of Developmental Biology, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA

[1] Corresponding author: clarkda@mail.nih.gov. Tel. 301 496-6966. Building 6A Room 2A02, 6 Center Drive, Bethesda MD 20892.



Supplemental Fig. S1. Quantification of EMSA data in Fig. 1C and Fig. 5. Determination of K_d values (the Gcn4 concentration at which 50% of the probe is bound) for the natural site (RY) and a mutated site (YR); see Supplemental Table S5 for probe sequences. The means and standard deviations for two biological replicate experiments are shown. (*A*) *ARG1* promoter site. (*B*) *ARO1* promoter site. (*C*) *STP2* ORF site.

Supplemental Fig. S2. Comparison of G-SELEX and ChIP-seq data for Gcn4.
See following pages.

IGV tracks showing three G-SELEX Round 3 biological replicate experiments and ChIP-seq data for induced and uninduced wild type cells and induced *gcn4Δ* cells from Rawal *et al.* (2018). The locations of AP-1 motifs (TGACTCA) are indicated at the bottom, together with the motif type. All data are normalised to the genomic average (set at 1).

A. *ARG1* promoter site.
B. *ARO1* promoter site.
C. *MCH1* ORF site.
D. *STP2* ORF site.
E. *ARG2* ORF site.
F. *ARG3* promoter site.
G. *ARG4* promoter site.
H. *ARG5,6* promoter site.
I. *ARG7* promoter site.
J. *ARG8* promoter site.
K. *ATG41/ICY2* ORF site.
L. *CPA1* promoter site.
M. *CPA2* promoter site.
N. *HIS4* promoter site.
O. *ORT1* promoter site.
P. *POS5* ORF site.
Q. *ROT2* ORF site.
R. *SPO21* ORF site.
S. *VPS41* ORF site.
T. *YGL117W* ORF site.
U. *HIS3* promoter.
V. *COG1* and *SDT1*.

Figure S2

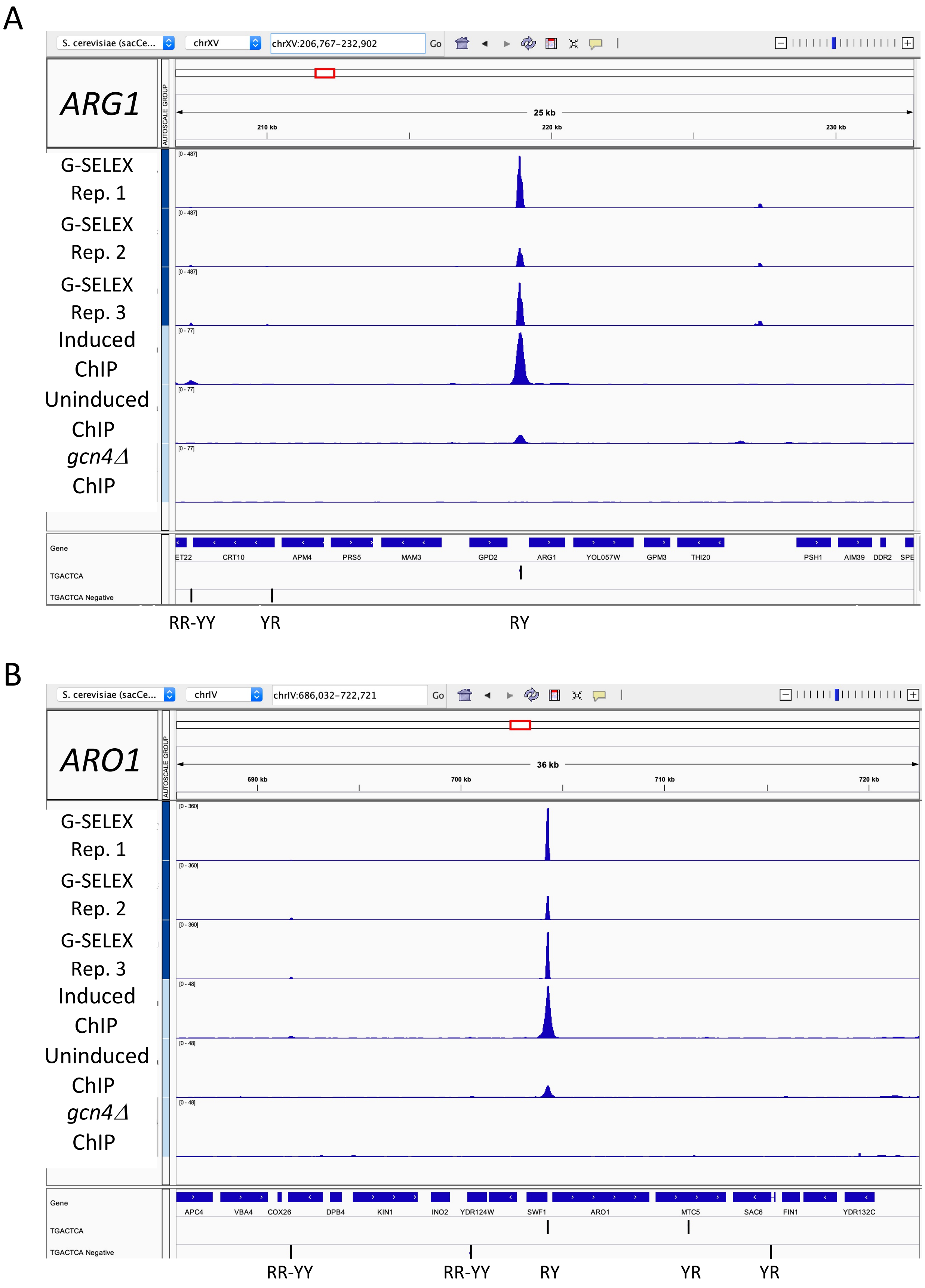

A
S. cerevisiae (sacCe...
chrXV
chrXV:206,767-232,902
Go
ARG1
AUTOSCALE GROUP
25 kb
210 kb
220 kb
230 kb
G-SELEX Rep. 1
G-SELEX Rep. 2
G-SELEX Rep. 3
Induced ChIP
Uninduced ChIP
gcn4Δ ChIP
[0 - 487]
[0 - 77]
Gene
TGACTCA
TGACTCA Negative
ET22
CRT10
APM4
PRS5
MAM3
GPD2
ARG1
YOL057W
GPM3
THI20
PSH1
AIM39
DDR2
SPE
RR-YY
YR
RY
B
S. cerevisiae (sacCe...
chrIV
chrIV:686,032-722,721
Go
ARO1
AUTOSCALE GROUP
36 kb
690 kb
700 kb
710 kb
720 kb
G-SELEX Rep. 1
G-SELEX Rep. 2
G-SELEX Rep. 3
Induced ChIP
Uninduced ChIP
gcn4Δ ChIP
[0 - 360]
[0 - 48]
Gene
TGACTCA
TGACTCA Negative
APC4
VBA4
COX26
DPB4
KIN1
INO2
YDR124W
SWF1
ARO1
MTC5
SAC6
FIN1
YDR132C
RR-YY
RR-YY
RY
YR
YR

Figure S2, cont'd.

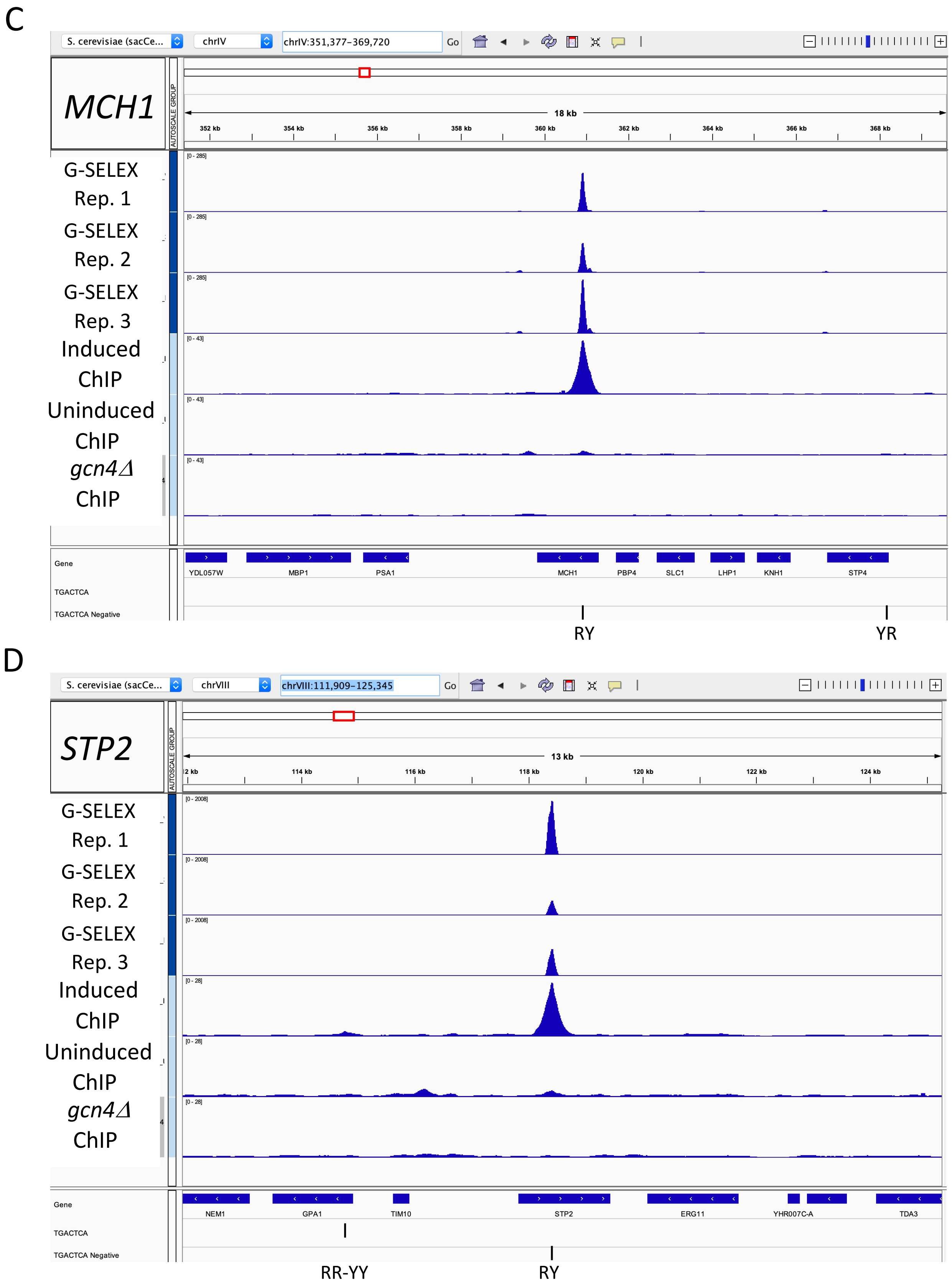
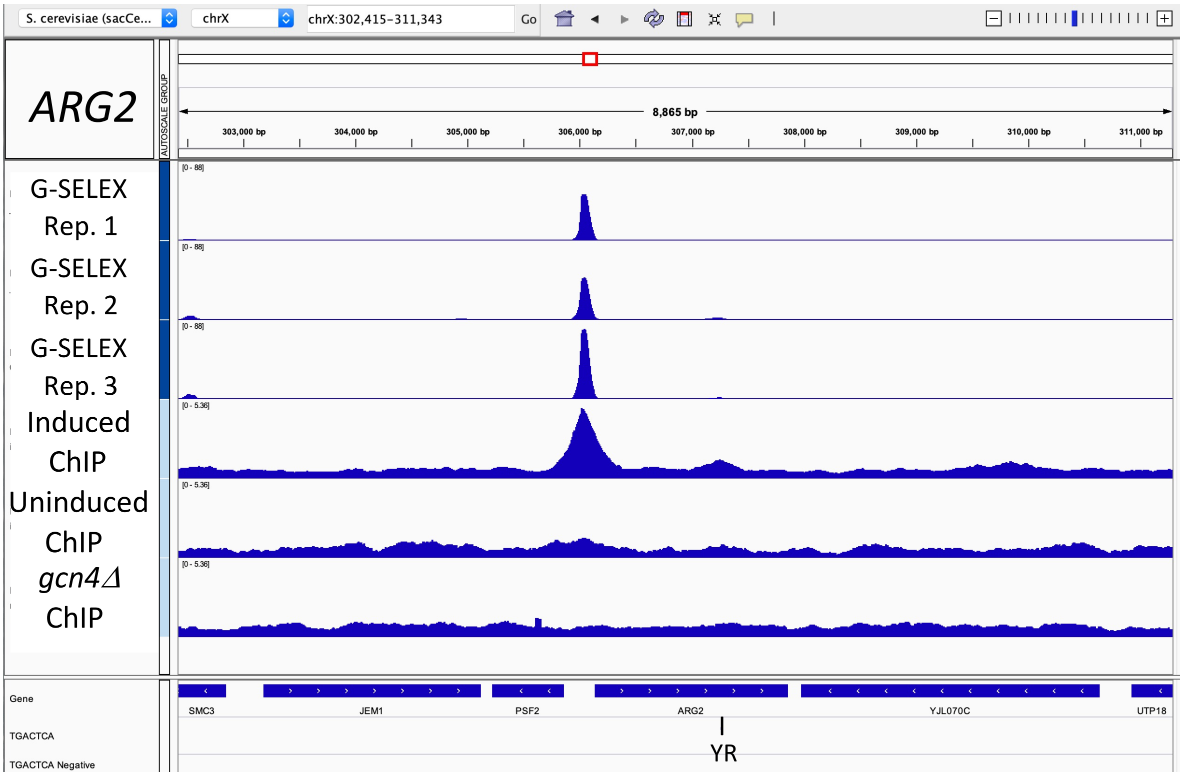


Figure S2, cont'd.

E



F

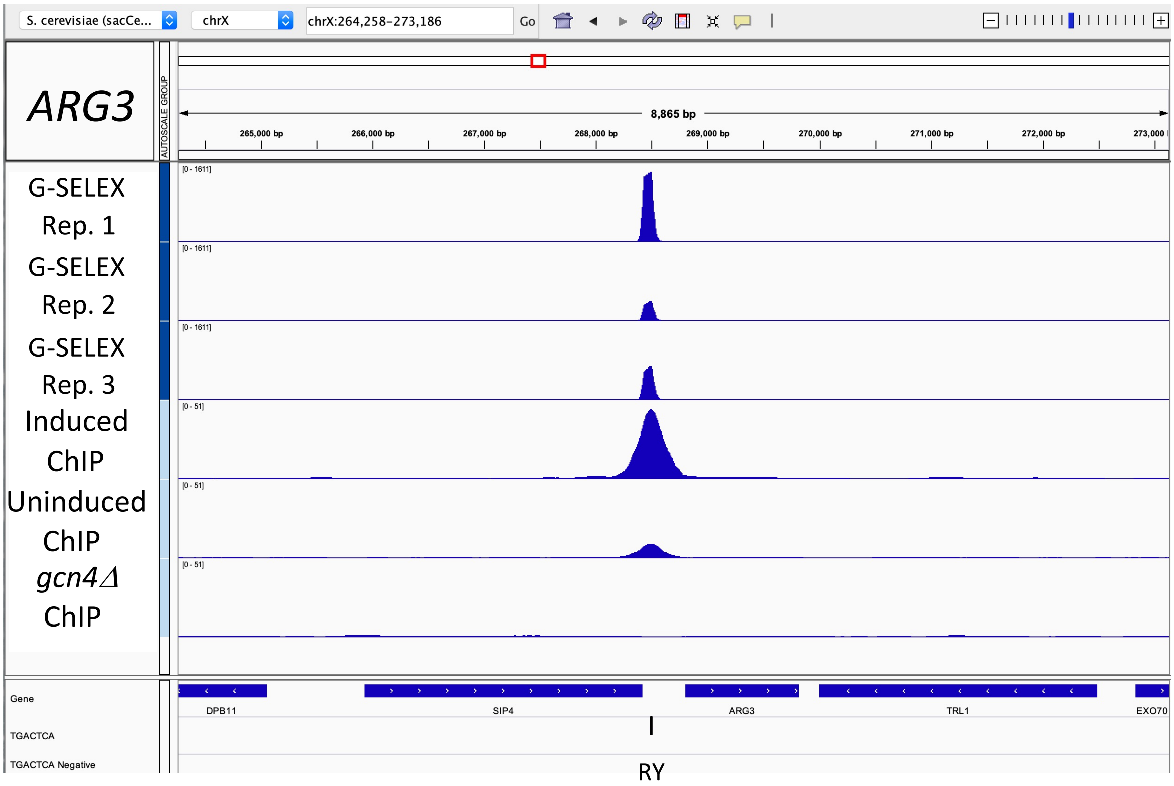
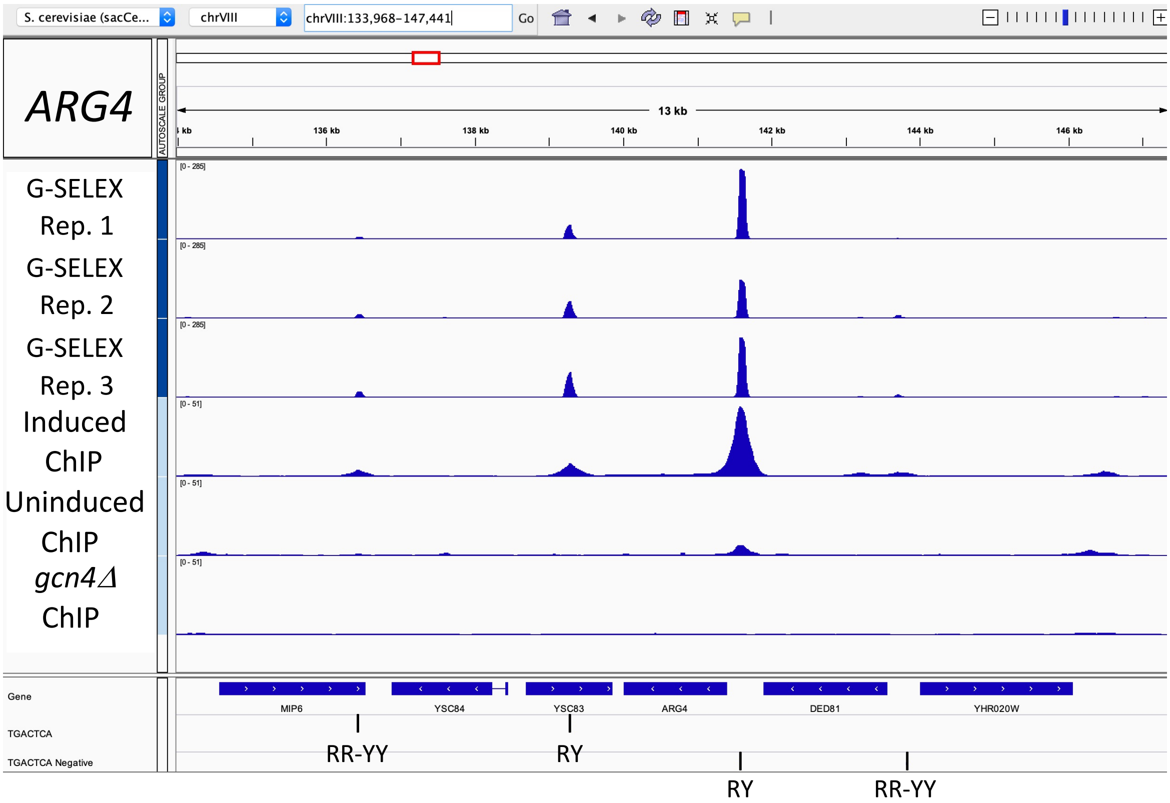


Figure S2, cont'd.

G



H

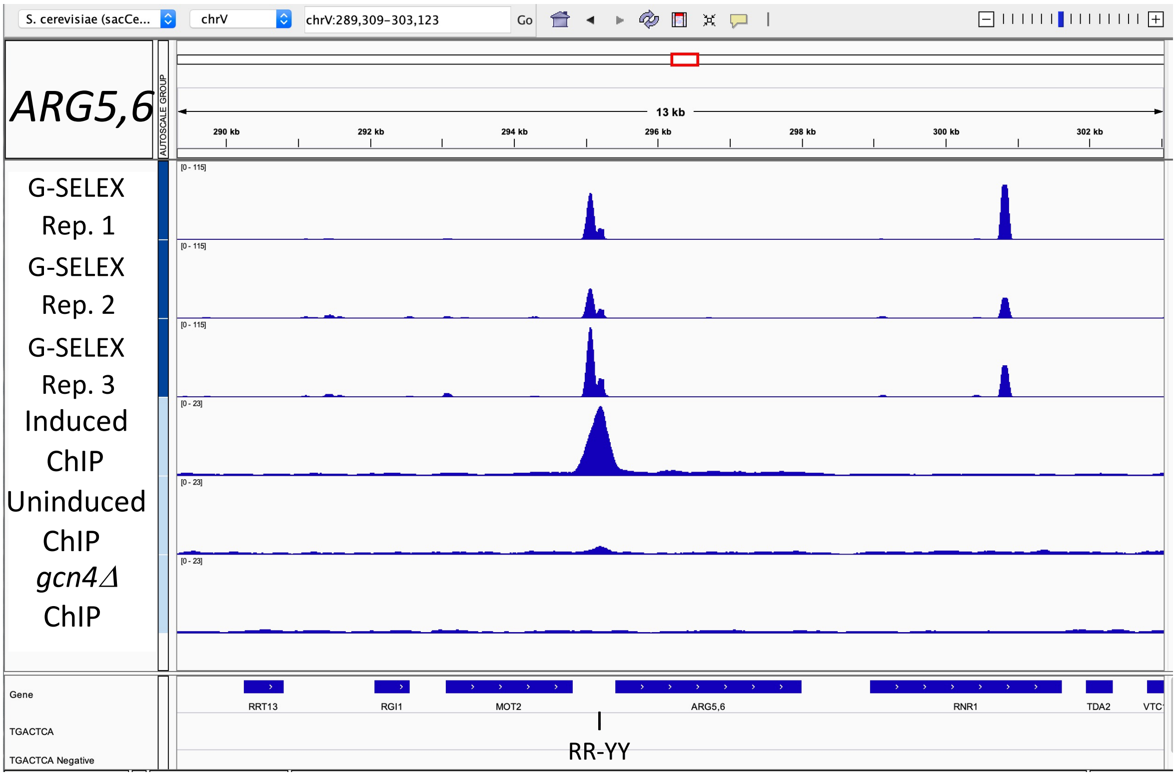
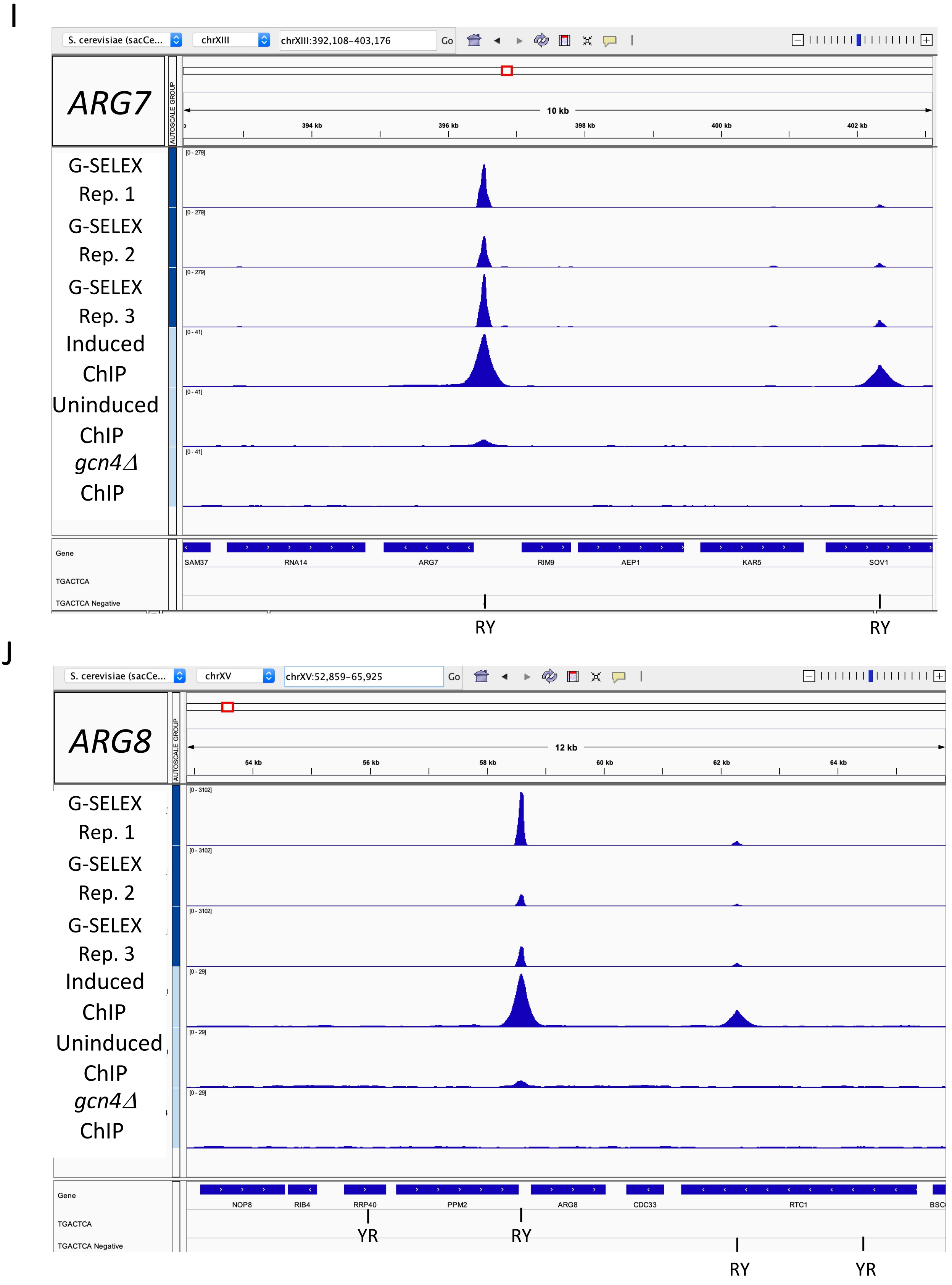


Figure S2, cont'd.

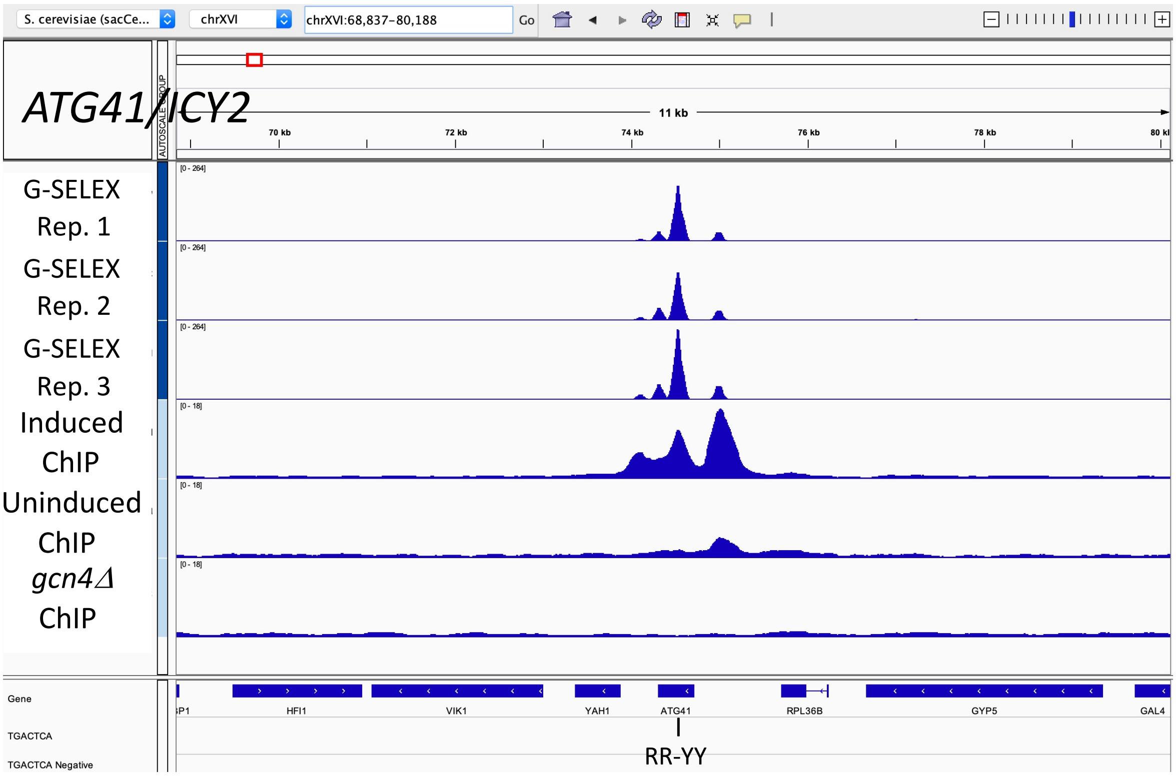


S. cerevisiae (sacCe... chrXIII chrXIII:392,108-403,176

S. cerevisiae (sacCe... chrXV chrXV:52,859-65,925

Figure S2, cont'd.

K



L

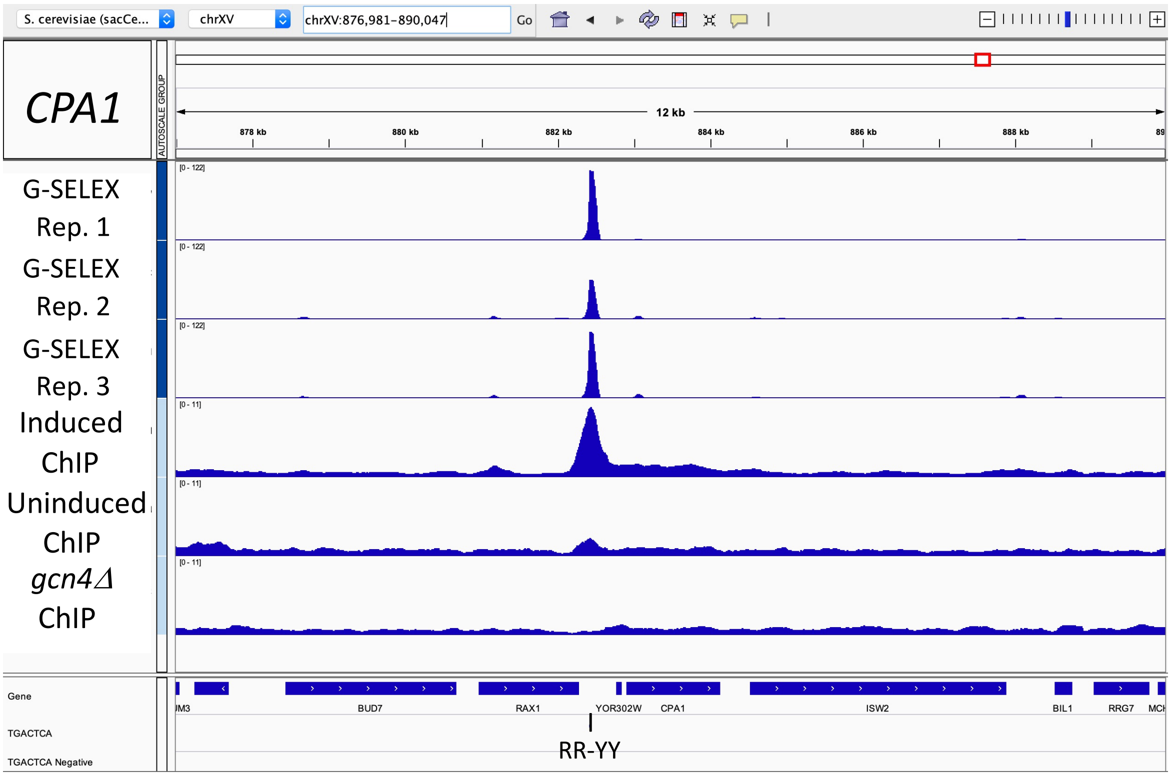


Figure S2, cont'd.

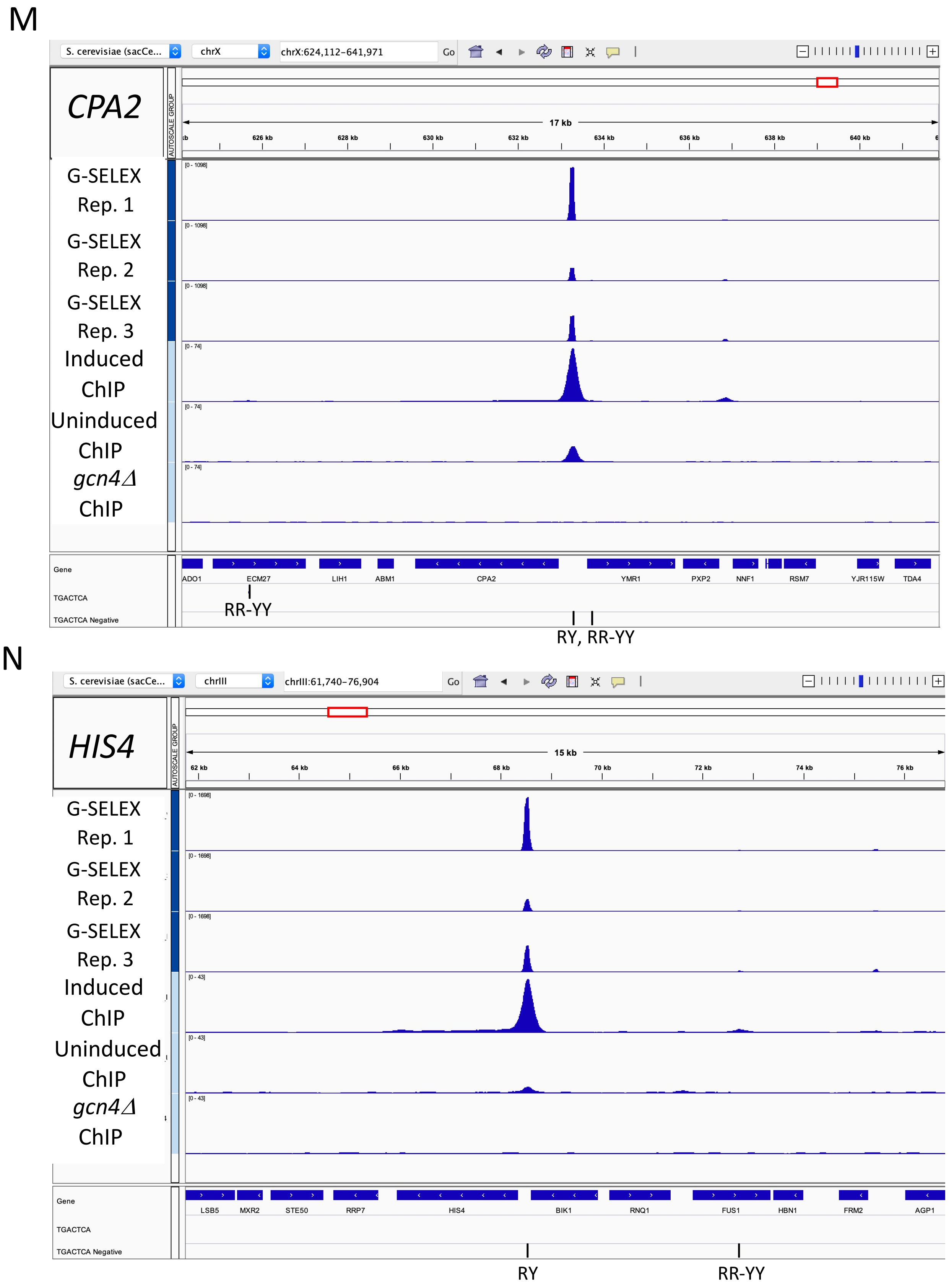
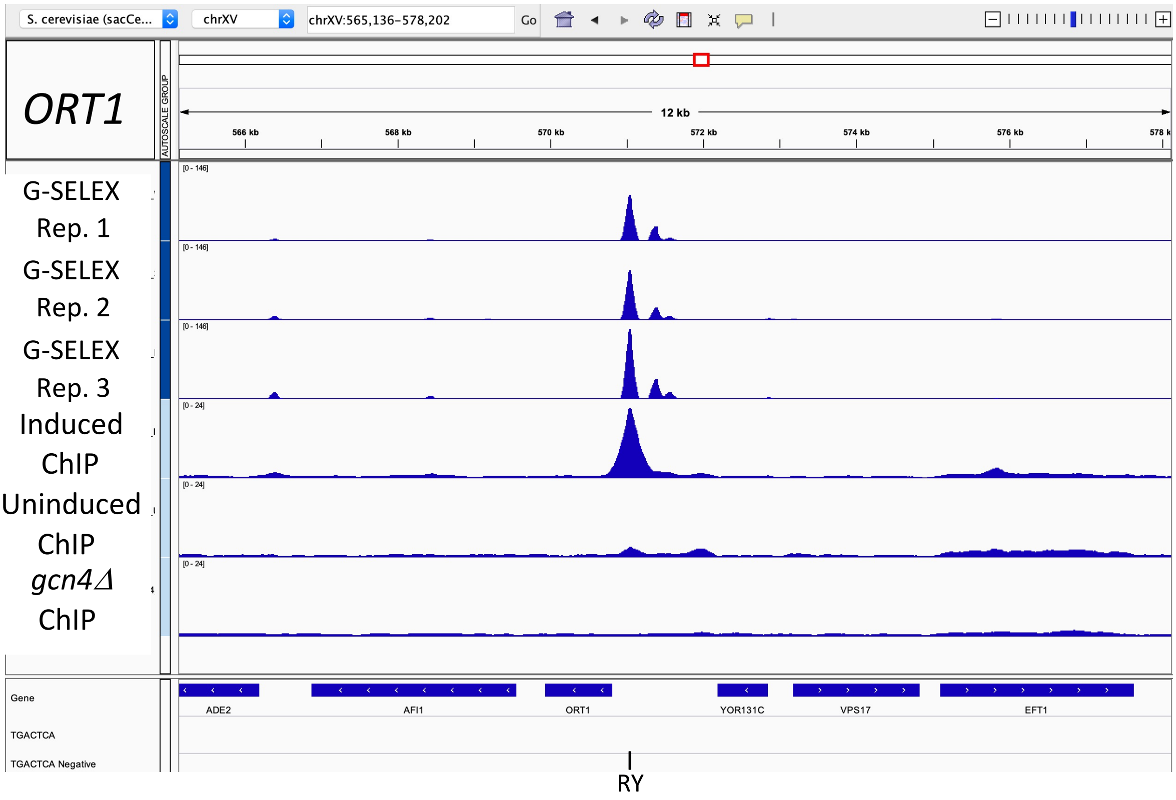


Figure S2, cont'd.

O



P

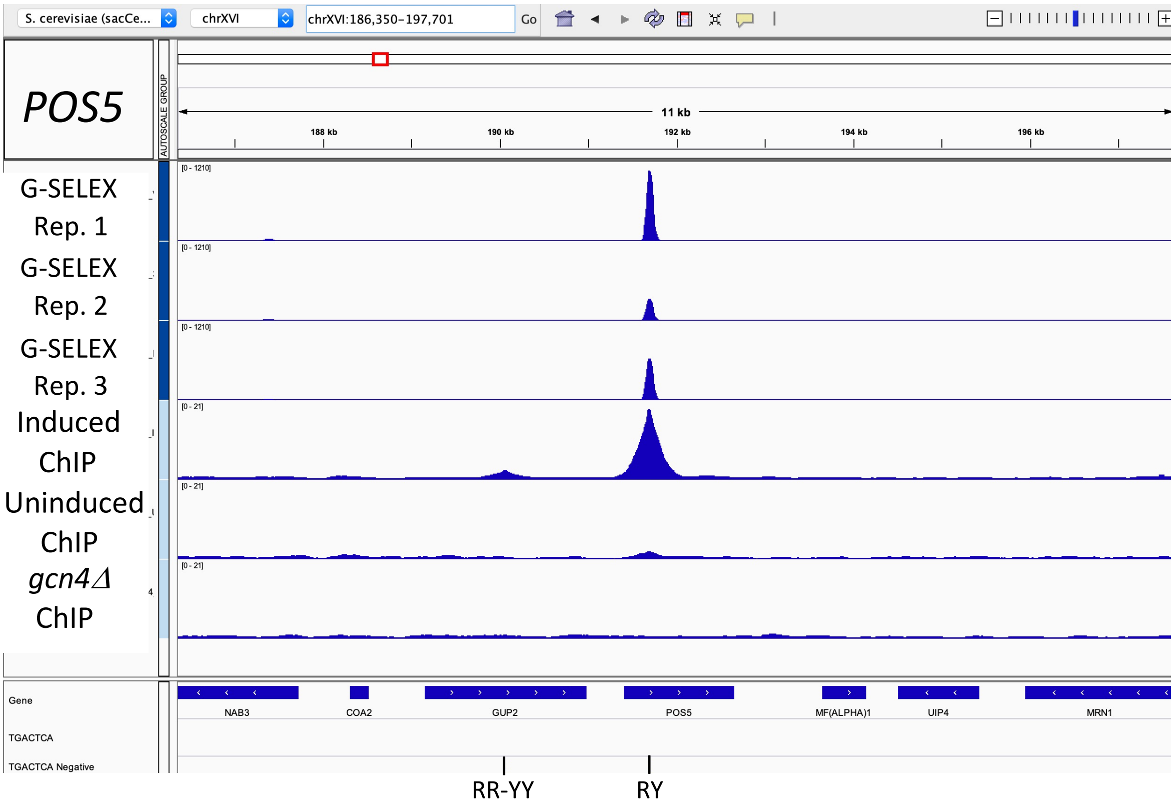
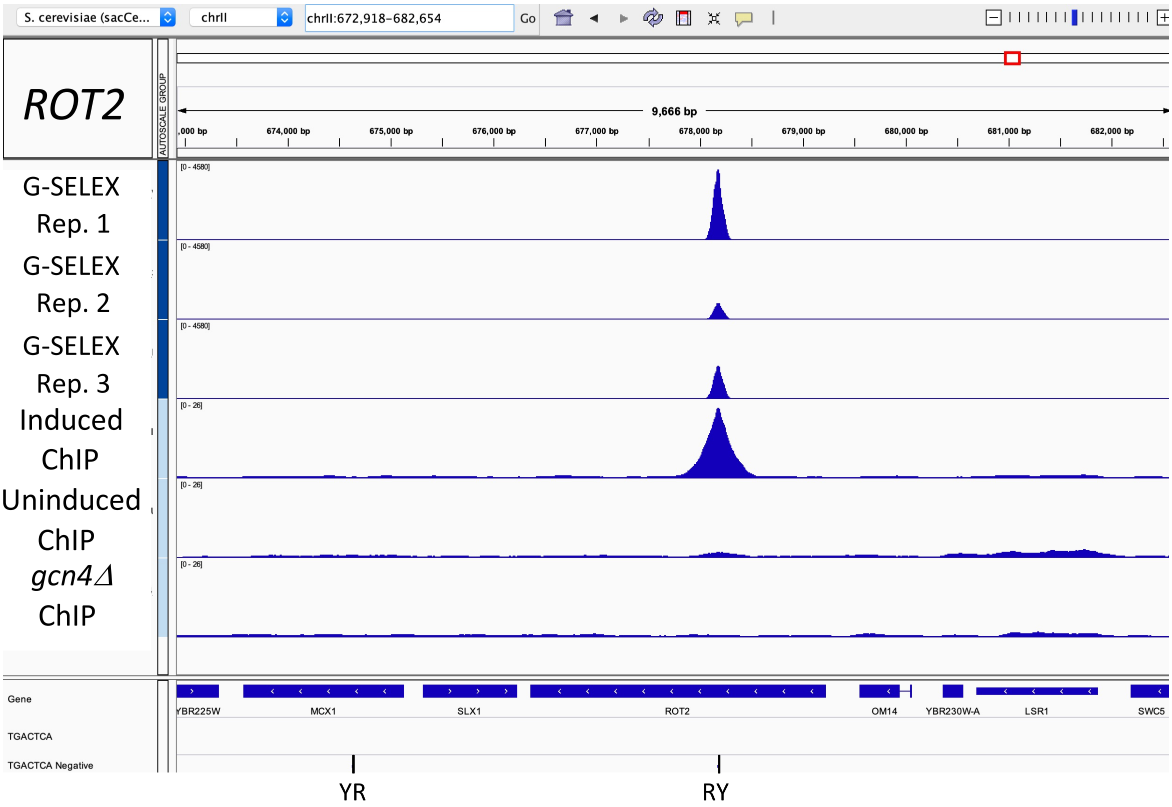


Figure S2, cont'd.

Q



R

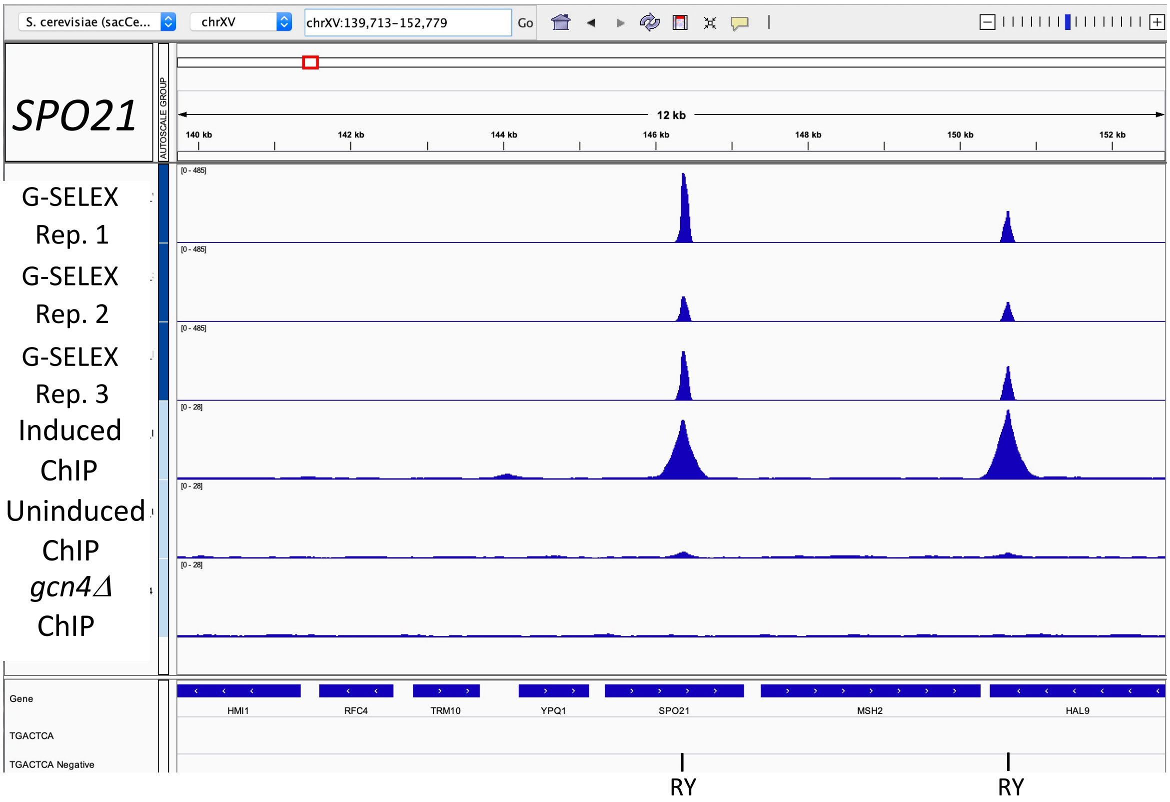
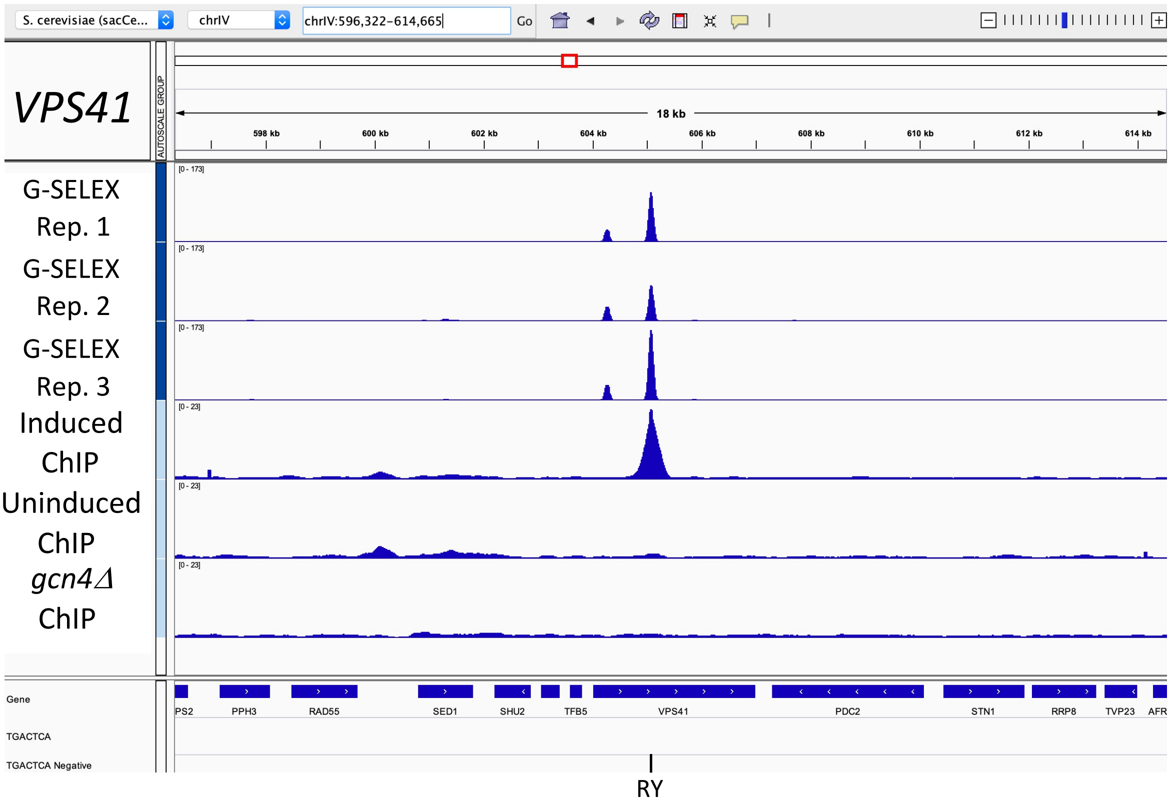


Figure S2, cont'd.

S



T

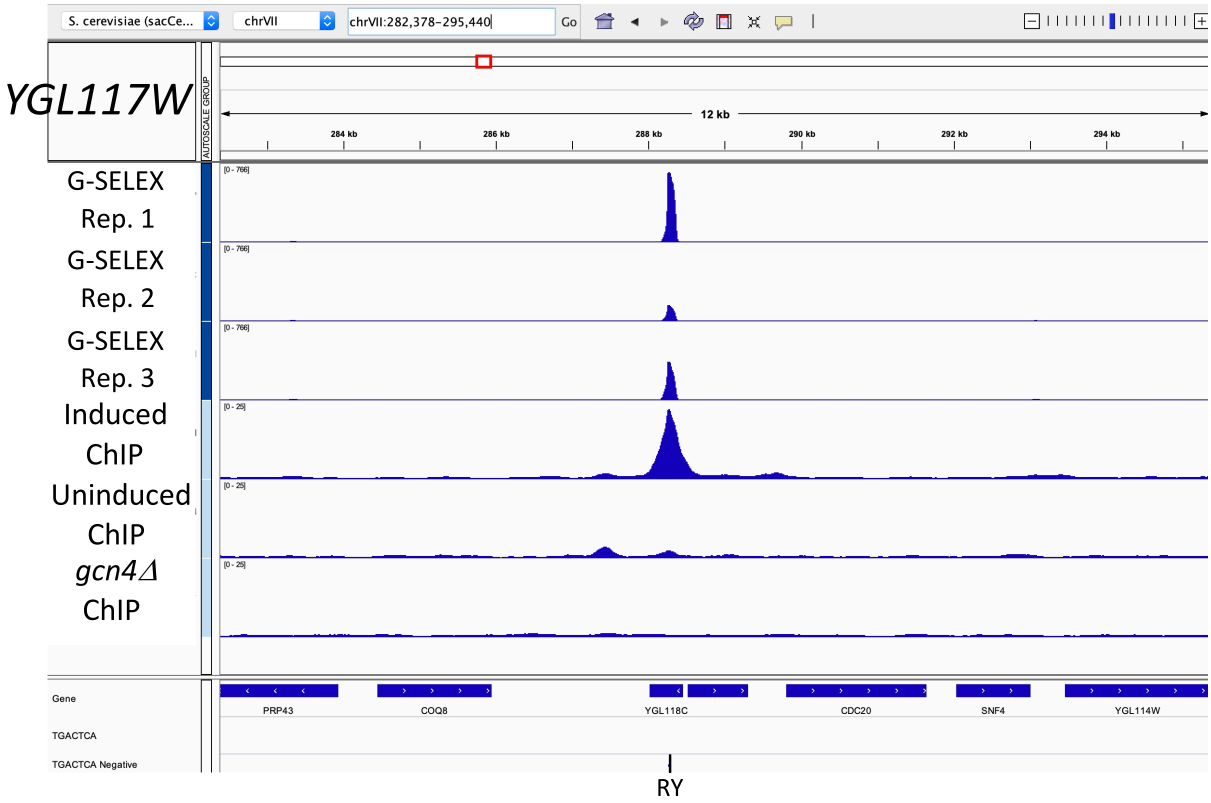
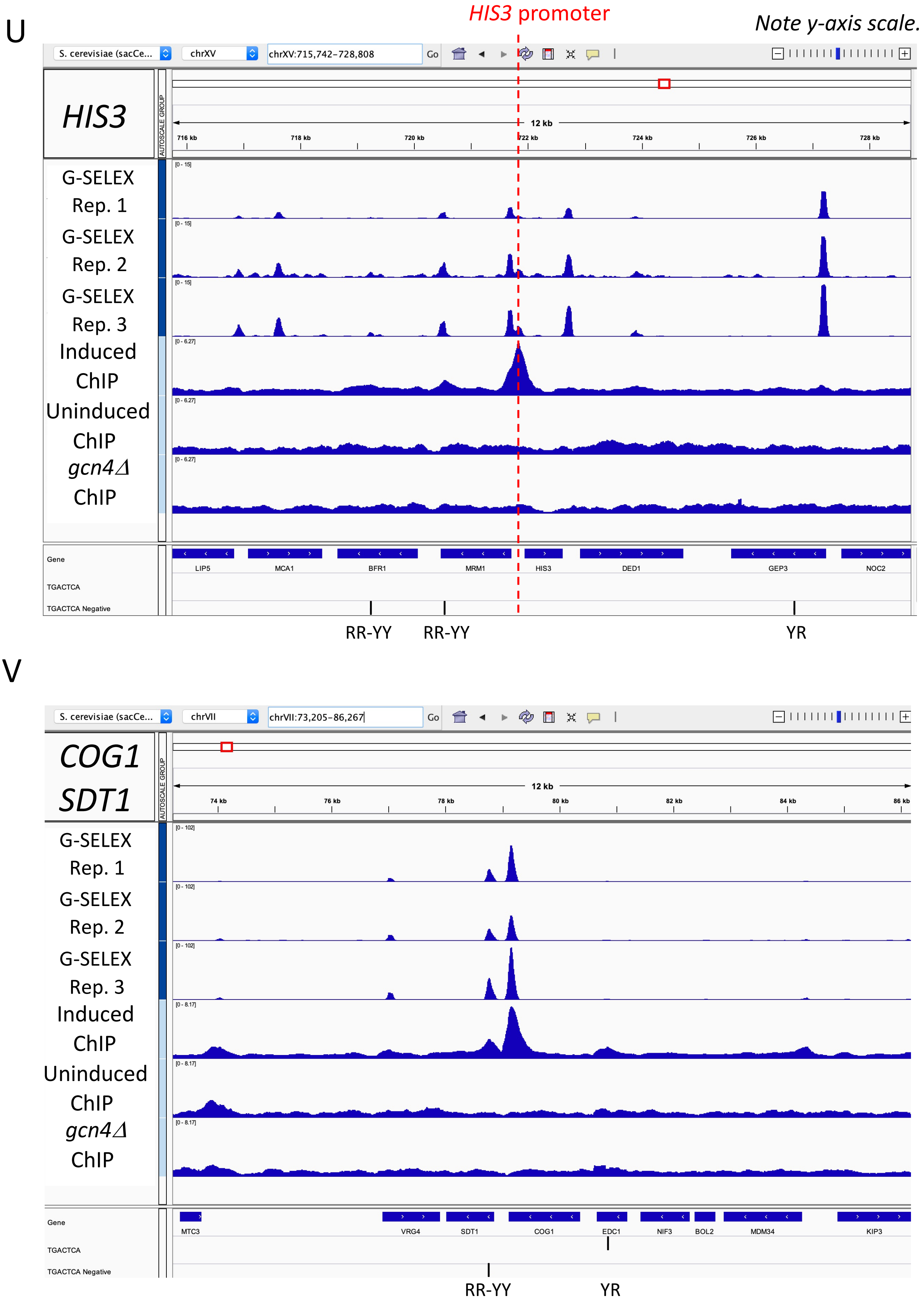
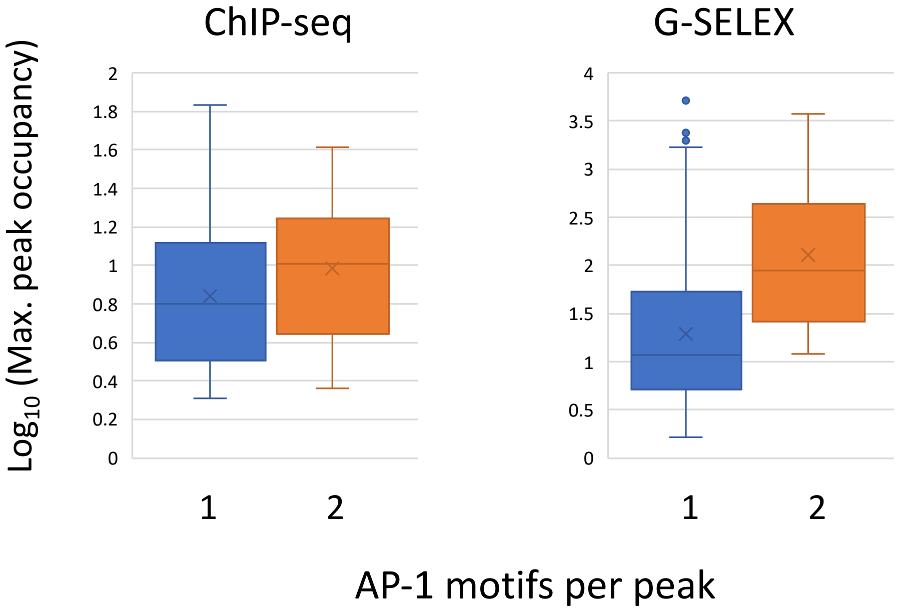


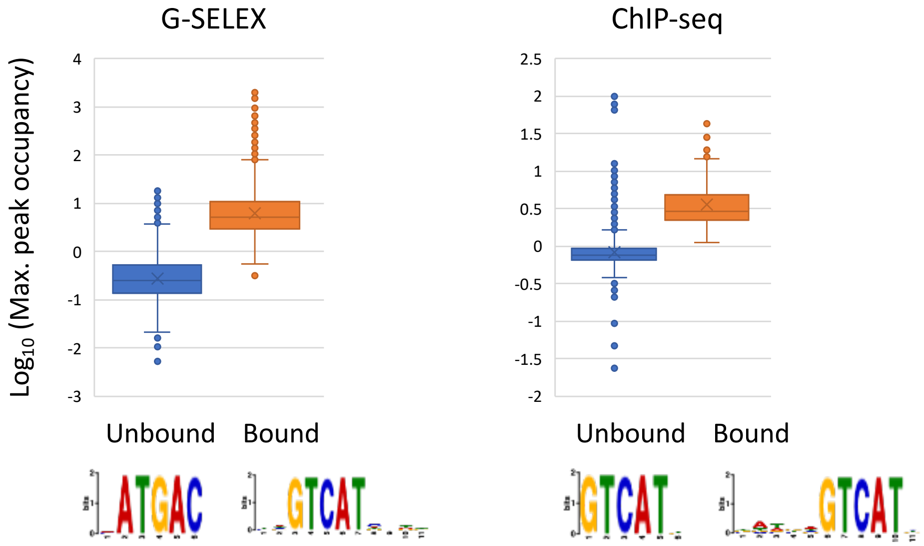
Figure S2, cont'd.



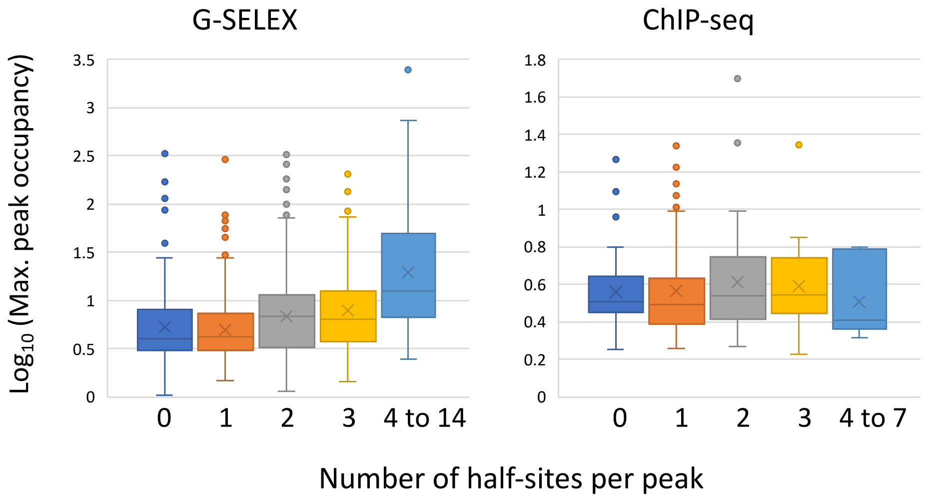


Supplemental Fig. S3. Gcn4 occupancy is higher if a peak contains two motifs. There are 353 ChIP peaks with one motif and 18 ChIP peaks with 2 motifs; 592 G-SELEX peaks have one motif and 12 G-SELEX peaks have 2 motifs.

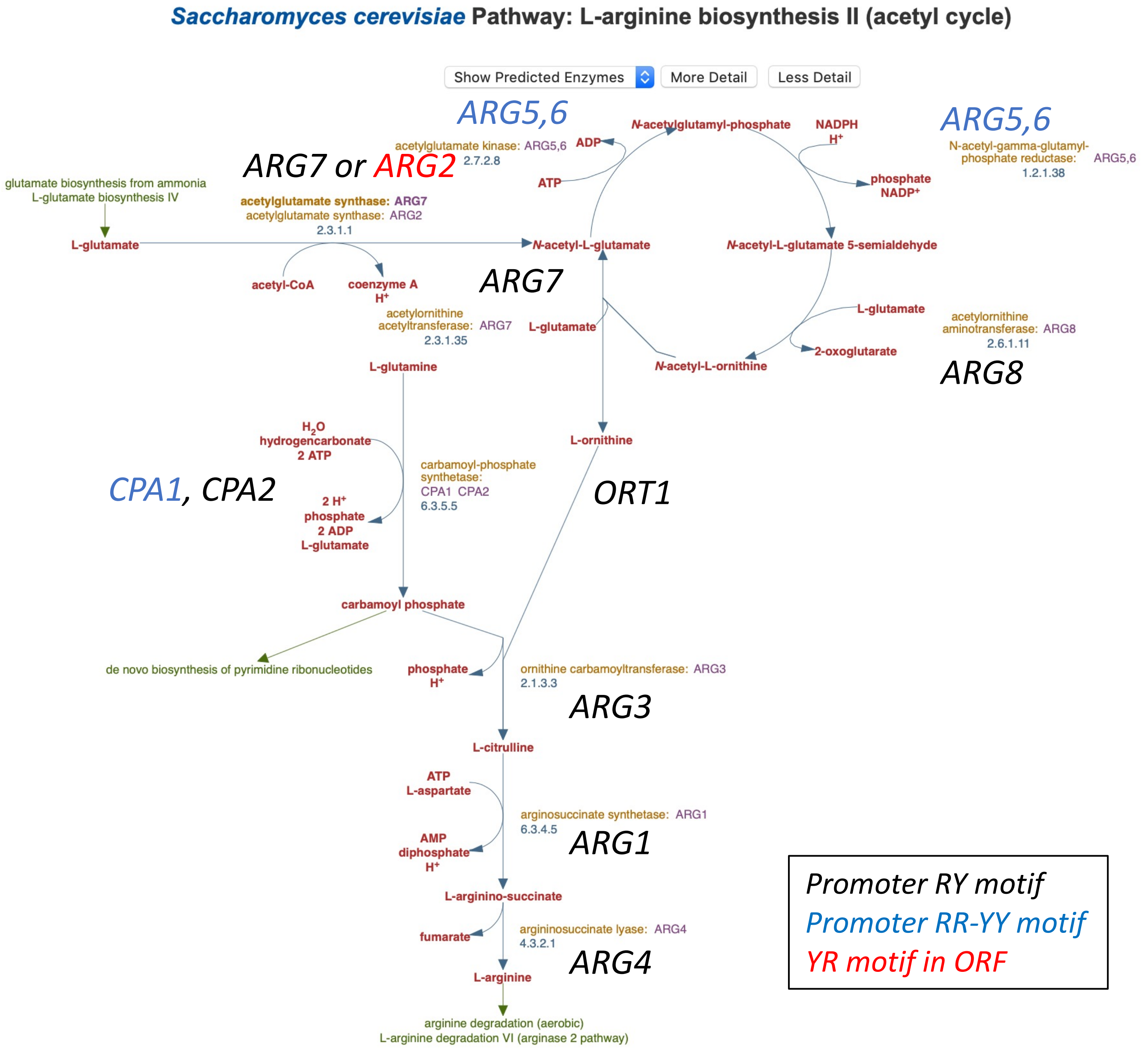
A Bound and unbound half-site occupancies



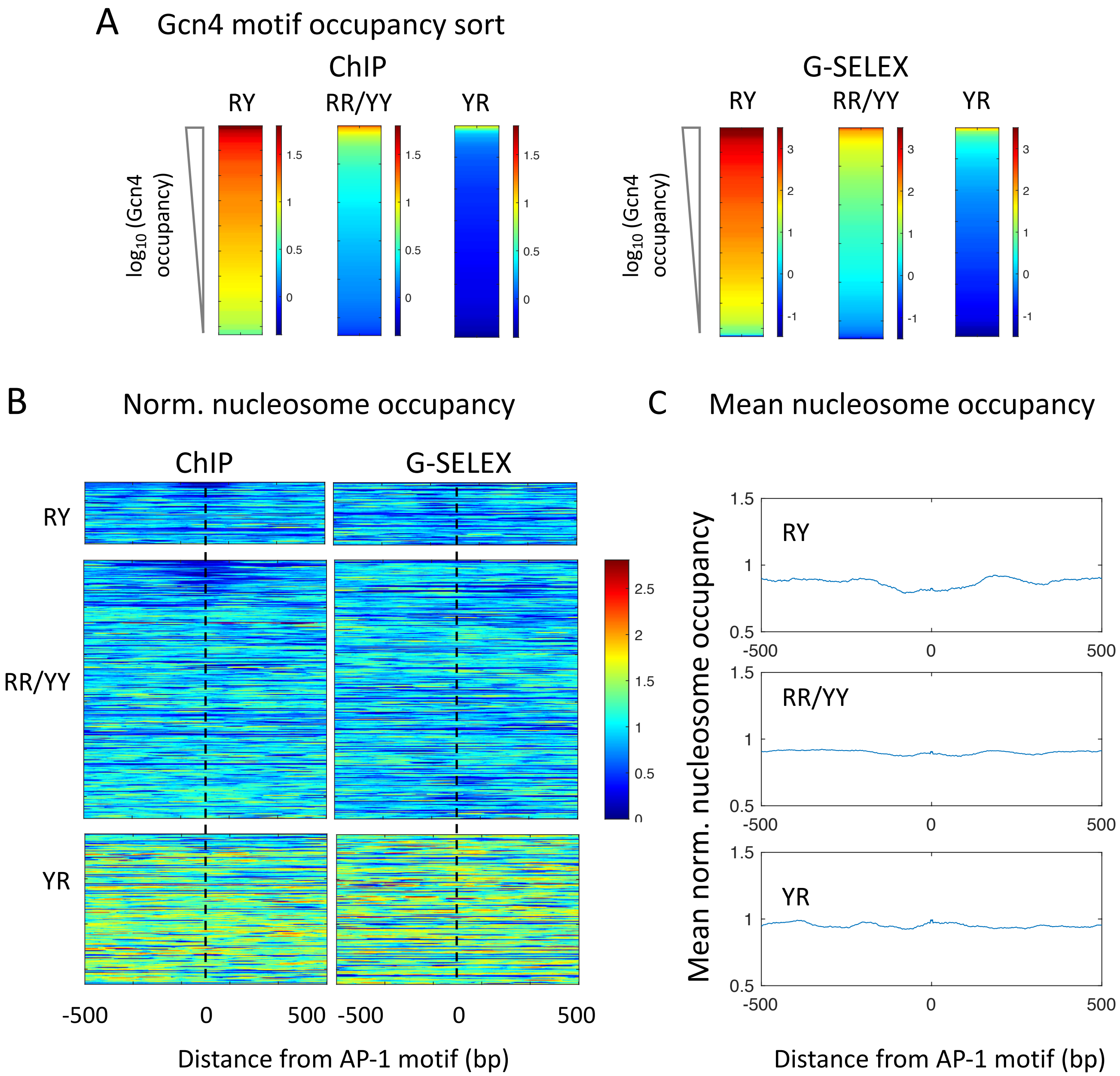
B Half-sites in peaks



Supplemental Fig. S4. Analysis of Gcn4 half-site motifs (ATGAC) detected by ChIP-seq and G-SELEX. (*A*) Occupancies and MEME motifs for bound and unbound half-sites in vitro (G-SELEX) and in vivo (ChIP-seq). (*B*) The occupancy of bound half-site peaks increases with the number of half-sites in the peak in vitro but not in vivo.



Supplemental Fig. S5. Almost all of the major enzymes in the metabolic pathway for arginine biosynthesis in yeast have Gcn4-bound RY motifs in their promoters. Gene names are colour-coded according to their motif type (see box). The figure was obtained from SGD YeastPathways: https://pathway.yeastgenome.org/YEAST/NEW-IMAGE?type=PATHWAY&object=ARGSYNBSUB-PWY



Supplemental Fig. S6. Analysis of nucleosome occupancy in the vicinity of Gcn4 sites. (*A*) Gcn4 motifs were divided by motif type and sorted by their Gcn4 occupancies in vivo using the same colour scale (left), or in vitro (right). (*B*) Heat maps showing normalised nucleosome occupancy (derived from MNase-seq data for induced cells; Cole et al. 2014) sorted as in *A*, either by Gcn4 occupancy in vivo (left) or in vitro (right) as a function of distance from the central nucleotide in the motif. The heat map scale is linear. (*C*) Mean nucleosome occupancy plots for the three sections of the heat maps in *B*. There are 166 RY sites, 546 RR/YY sites and 366 YR sites.

Supplemental Table S5. DNA oligonucleotides used in this study.

PCR primers used to obtain the GCN4 gene as a NdeI-XhoI fragment for insertion into pET21b(+) to construct p789:

1. 5'-CATATGTCCGAATATCAG
2. 5'-CTCGAGGCGTTCGCCAAC

Complementary oligonucleotides used to insert the 6xHis and Flag tags at the C-terminus of Gcn4 using the XhoI site in p789 (XhoI sticky ends):

3. 5'-TCGAGACCGGTCATCATCATCATCATCACGGTACCGACTACAAGGACGATGACGACAAGTAAC
4. 5'-TCGAGTTACTTGTCGTCATCGTCCTTGTAGTCGGTACCGTGATGATGATGATGATGACCGGTC

Oligonucleotides used for EMSA probes:

All site-specific DNA oligomers were taken directly from the *S. cerevisiae* genome with no changes if possible, and all Gcn4 motifs (shown in bold) were placed as centrally as possible within the oligomer. Where necessary, any guanine residues located at the 3' end of a sequence were changed to another base to avoid quenching of the 3' 6-FAM label. All base changes that converted a 5' purine/3' pyrimidine (RY) configuration to a 5' pyrimidine/3' purine (YR) configuration were made with careful attention paid to the nearby sequence, and the generation of new potential binding sites was avoided.

NB. The complementary oligonucleotides are not shown (they have no 6-FAM label and yield blunt-ended double-stranded probes after annealing):

ARG1 promoter
RY (native) sequence: 5'-TTTTTA**GTGACTCAT**GTCGCATTT-6-FAM-3'
YR (swapped) sequence: 5'-TTTTTA**TTGACTCAG**GTCGCATTT-6-FAM-3'

ARO1 promoter
RY sequence: 5'-ATTGTCATGCAT**ATGACTCAT**CCCGAACAT-6-FAM-3'
YR sequence: 5'-ATTGTCATGCAT**CTGACTCAG**CCCGAACAT-6-FAM-3'

MCH1 ORF
RY sequence: 5'-CGGAGGAAATG**GTGACTCAT**CATTCAATCT-6-FAM-3'
YR sequence: 5'-CGGAGGAAATG**TTGACTCAG**CATTCAATCT-6-FAM-3'

STP2 ORF
RY sequence: 5'-AAAGTTTCGGA**ATGACTCAT**CAGCAATCC-6-FAM-3'
YR sequence: 5'-AAAGTTTCAGA**CTGACTCAG**CAGCAATCC-6-FAM-3'