File S1

Supporting Materials and Methods

Detailed description of yeast strains and plasmids:

The *CUP1* locus contains several copies of a tandem ~2 kb repeat structure that includes one copy of the *CUP1* gene and one copy of gene *YHR154C*. *YHRC154C* is identical to the 3' end of the *RSC30* gene, the gene located immediately distal to the *CUP1* locus (right side). Therefore, the rightmost repeat unit in the *CUP1* locus overlaps the *RSC30* gene. To obtain a complete deletion of the *CUP1* repeats without disturbing *RSC30* (which plays roles in chromatin remodeling and DNA double strand break repair (SHIM *et al.* 2005)), we modified the *CUP1* locus in two steps. We first deleted all *CUP1* repeats (including the 3' end of *RSC30*) through integration of a PCR fragment containing the *Kl*URA3-*Kan*MX4 CORE cassette (primers JAO99/JAO100; template pCORE (STORICI *et al.* 2001)), followed by transformation with a PCR product that removed the CORE cassette and restored the 3' end of *RSC30* (JAO101/JAO102; WT genomic DNA template). The resulting allele (*cup1Δ RSC30*) had no copies of *CUP1* and one intact copy of *RSC30*, and corresponded to a deletion that included SGD Chr8 coordinates 212,262 to 215,159.

To delete the *SFA1* gene, we built a plasmid (pJA35) containing the *hisG-URA3-hisG* cassette (ALANI *et al.* 1987) flanked by homology regions from both sides of *SFA1*. The homology regions were amplified by PCR from genomic DNA corresponding to SGD Chr4 coordinates 158,742 to 159,349 (JAO326/JAO327; introducing 5' *Eco*RI and 3' *Bam*HI sites) and 160,824 to 160,994 (JAO328/JAO329; introducing 5' *Bam*HI and 3' *Sal*I sites). These two products were digested with *Bam*HI and ligated. A 825 bp ligation product was purified, digested with *Eco*RI and *Sal*I, and ligated to an *Eco*RI/*Sal*I digested pUC19 cloning vector. The resulting plasmid (pJA34) was digested with *Bam*HI, the ends were treated with calf intestinal alkaline phosphatase (CIP), and ligated to a *Bam*HI/*Bgl*II fragment from pNKY51 containing the *hisG-URA3-hisG* cassette (ALANI *et al.* 1987). The resulting pJA35 plasmid was linearized with *Eco*RI and *Sal*I prior to integration into the yeast genome at the *SFA1* locus (*sfa1Δ::hisG-URA3-hisG*), and subsequently a spontaneous *URA3* pop-out event was selected on 5-FOA. The resulting *sfa1Δ::hisG* allele corresponded to a deletion of SGD Chr4 coordinates 159,350 to160,823. The *msh2Δ::hisG* deletion was generated using a similar strategy, through integration of a fragment from linearized plasmid pEAI98, a generous gift from Eric Alani.

The *YMRCTy1-5* element was identified previously as a Ty1 element present in CG379, but absent in the S288c reference genome (ARGUESO *et al.* 2008). We amplified this element by PCR from genomic DNA with primers JAO291/JAO294 and sequenced the product. *YMRCTy1-5* is inserted at the equivalent SGD Chr13 coordinate 748,219. We deleted this element by integration of a PCR product containing the *Kl*URA3-*Kan*MX4 CORE cassette (*ymrcty1-5Δ::CORE*; primers JAO510/JAO511; template pCORE (STORICI *et al.* 2001)), but in this case we did not carry out the secondary transformation to remove CORE.

The *SFA1-CUP1-Hph*MX4 CNV reporter cassette was constructed as follows. Plasmid pAG32 containing the *Hph*MX4 Hygromycin B resistance marker (GOLDSTEIN and MCCUSKER 1999) was linearized with *Bgl*II and ligated to a *Bgl*II digested PCR product containing the *CUP1* gene (JAO104/JAO105; genomic DNA template; 5' and 3' *Bgl*II sites). This plasmid (pJA22) has a 776 bp insertion of *CUP1* including its upstream and downstream regulatory regions (SGD Chr8 coordinates 212,363 to 213,139) in the same transcriptional orientation as the *Hph* marker. pJA22 was used as template to amplify a PCR product (primers JAO113/JAO114) that was integrated into the yeast genome downstream of the *SFA1* gene between SGD Chr4 coordinates 160,823 and 160,842 (*SFA1::CUP1-Hph*MX4). Finally, genomic DNA containing this insertion was used as template to amplify a PCR product (primers JAO143/JAO269) that was integrated on Chr5, between the *DDI1* and *UBP5* genes, at SGD coordinate 457,705 (*DDI1::SFA1-CUP1-Hph*MX4). The cassette includes

a segment of *SFA1* corresponding to SGD Chr4 coordinates 159,349 to 160,823, and includes 255 bp of the *SFA1* native promoter. We also generated a version of the CNV reporter marked with G418 resistance by integrating a *Kan*MX4 PCR product (JAO457/JAO458; pFA6-KanMX4 template (WACH *et al.* 1994)) at the site where *Hph*MX4 was originally found, resulting in the *DDI1::SFA1-CUP1-Kan*MX4 marker swap reporter.

In addition to the genetic modifications above constructed for the CNV assay, we also produced constructs that were used to examine allelic mitotic recombination on the right arm of Chr13. We integrated the *KlURA3-Kan*MX4 CORE cassette on the CG379 strain background near the right end of Chr13, at a position distal to the *ADH6* gene, between SGD coordinates 910,882 and 910,902 (*ADH6::CORE*; JAO502/JAO503; template pCORE (STORICI *et al.* 2001)). The resulting strain (JAY405) was mated to the diverged strain YJM799 (isogenic to YJM789 (WEI *et al.* 2007)) to generate the hybrid diploid JAY408.

We also created a similar hybrid diploid that had an integration of a modified version of the original CORE cassette at the same Chr13 position. We created plasmid pJA40 that contains a copy of the *S. cerevisiae URA3* gene (*ScURA3*) in addition to the *KlURA3* marker already present in the original pCORE. The presence of two different, but functionally redundant copies of *URA3* improved the sensitivity of the cells to 5-FOA, and eliminated the occurrence of resistant clones that originate through base pair mutation in *KlURA3*. Strains carrying a CORE2 insertion (*KlURA3-ScURA3-KanMX4*) can only become resistant to 5-FOA following simultaneous inactivation of both copies of *URA3*, possible through a genome rearrangement or LOH event, or extremely rarely through point mutations in both *URA3* genes. pJA40 was constructed by linearizing pCORE with *Bgl*II and *Xma*I. The wild type *ScURA3* gene was amplified from genomic DNA of Ura$^+$ strain JAY291 (ARGUESO *et al.* 2009) using primers JAO336 and JAO650 which had a *Bgl*II site added to the 5' end. This PCR product was then digested with *Bgl*II and *Xma*I and ligated to the pCORE backbone. The resulting pJA40 includes a 1.1 kb fragment (SGD Chr5 coordinates 115,949 to 117,047) that includes the Sc*URA3* gene and its native regulatory regions; and because *ScURA3* was inserted between *KlURA3* and *Kan*, primers originally designed to amplify the CORE cassette also work for CORE2 PCR.

Prior to integration of CORE2 into the genome, we created a derivative of CG379 in which the existing *ura3-52* mutation was replaced with a full deletion of the locus. We first repaired *ura3-52* by transforming a PCR product containing the wild type *URA3* gene from JAY291 (primers JAO487/JAO490) and selecting for the Ura$^+$ phenotype. This Ura$^+$ intermediate strain was transformed with an overlapping PCR product (JAO863/JAO866) that joined two fragments corresponding to sequences upstream (JAO863/JAO864) and downstream (JAO865/JAO866) of *URA3*, followed by selection for 5-FOA resistance. The resulting strain contained a deletion of the *URA3* locus (*ura3Δ0* allele), corresponding to the deletion of SGD Chr5 coordinates 116,069 and 117,036. This strain was used to integrate the CORE2 cassette near the right end of Chr13 (*ADH6::CORE2*; JAO502/JAO503; template pJA40), resulting in strain JAY794 that was mated to YJM799 to generated the hybrid diploid JAY800.

Molecular karyotyping analysis: PFGE and array-CGH:

Yeast cultures were grown in 7 ml liquid YPD at 30^o for 48 hr. The cells were then immobilized in LMP-agarose plugs to prepare full-length chromosomal DNA for PFGE. Ten plugs were prepared for each clone. Genomic DNA for array-CGH was prepared by extraction from the same agarose plug batches prepared for PFGE. This was ensure that the PFGE and array-CGH analyses were done using the same DNA preparation, minimizing the effects of differential loss of unstable chromosomal rearrangements between independent DNA preparation cultures. To prepare genomic DNA for array-CGH, we briefly dried four PFGE agarose plugs (~70 μl each) with paper wipes to remove excess 1x TE storage buffer and transferred them to a 15 ml polypropylene conical tube. The DNA was purified using a modified Fermentas GeneJET gel extraction kit protocol. 280 μl of binding buffer was added to melt the agarose plugs.

Tubes were vortexed to homogenize the suspension and then transferred to a Diagenode Bioruptor UCD-200 multisample sonicator to shear the DNA. The tubes were incubated for 30 minutes at 4°, at high power setting, and cycles of 30 sec ON and 30 sec OFF. After sonication, the normal Fermentas-recommended gel extraction protocol was followed, concluding with collection of the sheared DNA in 20 μl of elution buffer. The size of the sheared fragments (typically 1-2 kb) was examined in a 1% agarose gel, and the DNA concentration (typically 150-250 ng/μl) was measured in an Invitrogen Qubit fluorometer. DNA from the parental strain was prepared using the same protocol. One μg of sheared CFR clone DNA was labeled with dUTP-Cy5 (GE-Healthcare) using the Invitrogen BioPrime array-CGH labeling system, and parental DNA was labeled with dUTP-Cy3. The labeled DNAs were co-hybridized to one of two types of Agilent microarrays designs. Agilent catalog design G4810A-14810 slides have four array sectors (4x44k), each with 41,775 60-nucleotide single stranded probes distributed across the genome with a median spacing of 257 bp. We also used lower density custom-designed Agilent slides that had eight array sectors per slide (8x15k), each containing essentially every third probe from the 4x44k catalog design, totaling 14,965 probes and a median spacing of 774 bp. This custom 8x15k design (AMID 028943) has sufficient genomic coverage to locate CNV breakpoints at a resolution comparable to that of the higher density 4x44k catalog design. The hybridized microarray slides were scanned and the hybridization signal was quantified with Genepix 6.0 software. Subsequently, the hybridization data was analyzed with Biodiscovery Nexus Copy Number software to identify the CNVs present in each CFR.

Supporting cited literature:

ALANI, E., L. CAO and N. KLECKNER, 1987 A method for gene disruption that allows repeated use of *URA3* selection in the construction of multiply disrupted yeast strains. Genetics **116:** 541-545.

ARGUESO, J. L., M. F. CARAZZOLLE, P. A. MIECZKOWSKI, F. M. DUARTE, O. V. C. NETTO *et al.*, 2009 Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. Genome Research **19:** 2258-2270.

ARGUESO, J. L., J. WESTMORELAND, P. A. MIECZKOWSKI, M. GAWEL, T. D. PETES *et al.*, 2008 Double-strand breaks associated with repetitive DNA can reshape the genome. Proceedings of the National Academy of Sciences of the United States of America **105:** 11845-11850.

GOLDSTEIN, A. L., and J. H. MCCUSKER, 1999 Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. Yeast **15:** 1541-1553.

SHIM, E. Y., J. L. MA, J. H. OUM, Y. YANEZ and S. E. LEE, 2005 The yeast chromatin remodeler RSC complex facilitates end joining repair of DNA double-strand breaks. Molecular and Cellular Biology **25:** 3934-3944.

STORICI, F., L. K. LEWIS and M. A. RESNICK, 2001 In vivo site-directed mutagenesis using oligonucleotides. Nature Biotechnology **19:** 773-776.

WACH, A., A. BRACHAT, R. POHLMANN and P. PHILIPPSEN, 1994 New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. Yeast **10:** 1793-1808.

WEI, W., J. H. MCCUSKER, R. W. HYMAN, T. JONES, Y. NING *et al.*, 2007 Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. Proceedings of the National Academy of Sciences of the United States of America **104:** 12825-12830.