

Deep functional analysis of *synII*, a 770 kb synthetic yeast chromosome

Authors:

Yue Shen^{1,2,10,11*}, Yun Wang^{1,11*}, Tai Chen^{1,11*}, Feng Gao¹, Jianhui Gong¹, Dariusz Abramczyk², Roy Walker², Hongcui Zhao¹, Shihong Chen¹, Wei Liu², Yisha Luo², Carolin A. Müller³, Adrien Paul-Dubois-Taine², Bonnie Alver², Giovanni Stracquadanio^{4,5}, Leslie A. Mitchell⁶, Zhouqing Luo⁷, Yanqun Fan¹, Baojin Zhou¹, Bo Wen¹, Fengji Tan¹, Yujia Wang¹, Jin Zi¹, Zexiong Xie⁸, Bingzhi Li⁸, Kun Yang⁴, Sarah M. Richardson^{4,5}, Hui Jiang¹, Christopher E. French², Conrad A. Nieduszynski³, Romain Koszul⁹, Adele L. Marston², Yingjin Yuan⁸, Jian Wang¹, Joel S. Bader^{5,6}, Junbiao Dai⁷, Jef D. Boeke⁶, Xun Xu^{1,11}, Yizhi Cai^{2†}, Huanming Yang^{1,10†}

*These authors contributed equally to this work.

Affiliations:

1. BGI-Shenzhen, Shenzhen 518083, China
2. School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, UK
3. Sir William Dunn School of Pathology, University of Oxford, Oxford, OX1 3RE, UK
4. High-Throughput Biology Center, School of Medicine, Johns Hopkins University, Baltimore, Maryland, 21205 USA
5. Department of Biomedical Engineering, School of Engineering, Johns Hopkins University, Baltimore, Maryland 21218 USA
6. Institute for Systems Genetics, NYU Langone Medical Center, ACLSW Room 503, 430 East 29th Street, New York, NY 10016
7. Key Laboratory of Industrial Biocatalysis (Ministry of Education), Center for Synthetic and Systems Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China
8. Key Laboratory of Systems Bioengineering (Ministry of Education), SynBio Research Platform, Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), School of Chemical Engineering and Technology, Tianjin University, Tianjin, 300072, PR China
9. Department of Genomes and Genetics, Institut Pasteur / CNRS UMR3525, 25-28, rue du Docteur Roux, 75724 Paris Cedex 15, France
10. James D. Watson Institute of Genome Sciences, Hangzhou 310058, China.
11. BGI-Qingdao, Qingdao 266555, China

†Corresponding authors:

E-mail: yizhi.cai@ed.ac.uk (Y.C.); yanghm@genomics.cn (H.Y)

Abstract

Herein we report the successful design, construction and characterization of a 770 kb synthetic yeast chromosome *II* (*synII*). Our study incorporates characterization at multiple levels, including phenomics, transcriptomics, proteomics, chromosome segregation and replication analysis to provide a thorough and comprehensive analysis of a synthetic chromosome. Our “Trans-Omics” analyses reveal a modest but potentially significant pervasive up-regulation of translational machinery observed in *synII* is mainly caused by the deletion of 13 tRNAs. By both complementation assays and SCRaMbLE, we targeted and debugged the origin of a growth defect at 37°C in glycerol medium, which is related to misregulation of the HOG response. Despite the subtle differences, the *synII* strain shows highly consistent biological processes comparable to the native strain.

One Sentence Summary:

A near megabase synthetic chromosome, *synII*, was characterized using a “Trans-Omics” approach, which revealed a modest up-regulation of translational machinery; this was reversed by restoring the normal complement of tRNA genes to the strain.

Introduction

Yeasts are among the most important eukaryotic microorganisms, both scientifically and economically, with at least 1500 species identified to date. In the past few decades, a number of different yeast species have been sequenced, and comprehensive genomic comparison studies have revealed answers to several complex evolutionary questions such as whole-genome duplication and horizontal gene transfer during their history (1-5). However, despite the increased knowledge and understanding of the natural genome, our ability to engineer the entire genetic material is still very limited. Rapidly developing DNA synthesis techniques have made a “build to understand” approach feasible, which allows synthetic biologists to re-engineer and construct genes, pathways and even entire genomes (6-12).

Under the collaborative framework of the Sc2.0 project (www.syntheticyeast.org), we report the completion of yeast synthetic chromosome *II*, termed *synII*. The sequence of native *Saccharomyces cerevisiae* chromosome *II* was determined over two decades ago. This native chromosome is 807,888 bp in length and includes 410 open reading frames, 13 tRNAs, and 30 introns(13). *synII* was designed based on the native chromosome *II* following previously reported rules (12, 14, 15), resulting in a “designer” chromosome 770,035 bp in length, 43,149 bp shorter than the native sequence. *synII* was initially synthesized as minichunks (~3 kb), assembled into chunks (~10 kb) and integrated into the yeast genome to replace the native chromosome. Extensive Trans-Omics tests were conducted, including phenomics, transcriptomics, proteomics, chromosome segregation and replication analyses, which indicated that *synII* yeast, despite significant sequence alterations, is virtually indistinguishable from wild type. However, our analyses reveal that up-regulation of many components of the translational machinery (for example ribosomal proteins at both the RNA and protein levels) is a typical feature of synthetic chromosome replacement strains, and the main cause for this subtle defect was shown to be the reduction tRNA gene copy number; it was reversed by restoring the tRNA gene copy number.

Results

Design and synthesis of synII

Complete chromosome re-design of chromosome *II* was performed based on the Sc2.0 project standards (12) using the BioStudio design suite (14) to yield the synthetic yeast chromosome *II* (termed *synII*). Relative to the wild-type reference sequence, *synII* has 33 deletions (53,605 bp in total), 269 insertions (10,456 bp in total) and 14,949 single nucleotide substitutions. While *synII* represents a 5.3% reduction in sheer size compared to the native chromosome *II*, but hundreds of designer features were encoded throughout this synthetic chromosome (table S1, and supplementary Materials and Methods). The extensive changes made to the chromosome naturally raise the question whether there are any subtle phenotypic effects that may have been missed by previous studies.

Compared with previous synthetic strategies (12, 15), we devised an alternative modular assembly method for the construction of *synII*. This new approach enables parallel integration to speed assembly, while at the same time also providing an effective route to debug potential phenotypic defects (see below). We divided *synII* into 25 megachunks (~30 kb each) using BioStudio, and each megachunk was then segmented by an in-house software “Segman” into minichunks (~3 kb, fig. S1A), which are compatible with the Gibson assembly method (16) (fig. S1B) for construction into chunks (Detailed information on megachunks, chunks and minichunk can be found at www.syntheticyeast.org). Instead of performing step-by-step megachunk integration (SwAP-In) (15) from one end (left or right arm), we performed integration in parallel from both ends with two separate parental strains *YS000-L* and *YS000-R* (fig. S1C) bearing *synII-L* and *synII-R* respectively, which overlap by 30 kb, effectively reducing the overall integration time by almost 50%. These two strains were crossed to produce a heterozygous diploid. Despite hundreds of modifications being introduced in *synII*, regions sharing substantial sequence similarity to the native chromosome remain. We anticipate that these might trigger various unwanted meiotic recombination events, which would bring tremendous screening efforts when combining of semi-synthetic chromosomes based on homology-directly recombination. Therefore, to increase the frequency of the mitotic recombination of the two semi-synthetic chromosomes (*synII-L* and *synII-R*), we adopted an I-*SceI* mediated strategy (17) to break *synII-L* and *synII-R* at a designed site, promoting mitotic recombination between them to generate the fully synthetic *synII* (fig. S1D). Induction of the lethal I-*SceI*-mediated double-strand DNA breaks at the junctions of synthetic and wild-type sequence introduced a strong selective pressure for the two semi-synthetic segments to recombine, via a homology-directed repair mechanism. Since our recombination strategy is designed to generate a markerless *synII* and leave the *URA3* marker on wild-type chromosome, we observed a 10-fold increase in the *URA3*-deficient colony numbers upon the I-*SceI* induction, as judged by marker loss screening, suggesting our strategy can greatly bias the crossover events towards the designated regions. The resulting synthetic chromosome *II* strain (version yeast_ *chr02_9.01*) was identified by the PCRTag method (fig. S2, fig. S3).

Sequencing revealed structural variations in synII

Compared with the designed sequence, 61 variations of 4 types were observed by deep sequencing: 50 single nucleotide variations (SNVs), 5 missing loxPsym sites, 4 deletions and 2 structural variations (SVs) (table S2). 28 of the 50 SNVs were found to correspond exactly to the genotype of the native chromosome sequence at these positions, suggesting that these represent residual homologous recombination “patchworks” as seen in *synIII* and other synthetic chromosomes (12, 15, 18-21). When such patchwork regions are very short (i.e. lying entirely between two sets of PCRTags), they can be missed by the PCRTag analysis. For the remaining 22 SNVs, 3 were found to pre-exist in synthetic minichunk DNA indicating that these mutations were introduced during synthesis. The remaining SNVs map to overlapping regions between minichunks or megachunks, suggesting that these SNVs were likely introduced during minichunk assembly or megachunk integration. Since none of these SNVs are in coding regions or noticeably altered phenotype, they were not corrected.

As previously seen in other larger synthetic chromosomes (*synV*, *synX*) (19, 20), we observed two complex structural variants (SVs): the first of these was a ~15 kb tandem-duplication in megachunk *L* with chunk *L2-L4* duplicated and a loxPsym site located between the duplications (fig. S5A & S5B); another SV, identified in megachunk *T*, was fully characterized by PGM sequencing (Life Technologies), and is a >30 kb complex DNA sequence including multiple copies of chunks *T4*, *T5* and a partial chunk backbone plasmid *pSBGAK* (fig. S5A & S5C). We hypothesize that the 34 bp loxPsym sequence can serve as a homologous region during homologous recombination-mediated integration, albeit at a very low frequency, which led to the formation of the first SV. The mechanism of formation of the second SV remains unknown.

Correction of structural variants in synII

We designed a straightforward strategy to repair the two structural variations, again by applying the I-SceI system (17). The 18bp I-SceI recognition sequence was designed to carry a selective marker (*URA3*) and overlap with both ends of the tandem repeat junction as a donor fragment (fig. 1A). Upon induction of I-SceI digestion, the *synII* chromosome was broken and thereafter was repaired through homology-directed recombination between the repeat sequences, and as a consequence it effectively looped out the duplicated region.

Using this strategy, we sequentially repaired the two SVs in synthetic chromosome *II*, yielding strains yeast_chr02_9.02 and yeast_chr02_9.03 respectively (fig. 1B). PCR analysis (fig. S5D) and deep sequencing (fig. 1C) validated the successful repair of the duplicated regions in *synII*. Duplications were also observed in other synthetic chromosomes (*synV*, *synX*, *synXII*). Thus the I-SceI mediated strategy could serve as an efficient strategy especially suitable for repair of large duplications in synthetic chromosome construction.

In addition to the SVs, pulsed field gel analysis (PFGE) showed an abnormal chromosome karyotype of native *chrXIII* and *chrXVI* (fig. S6). This was further confirmed by DNA sequencing and Hi-C analysis (22). As similar instances also been found during the construction of *synIII* (15), we surmise that this represents a spontaneous low-frequency event occurred at some point during assembly. PFGE shows that the *chrXIII* and *chrXVI* karyotype of the *synII* strain with SVs (yeast_chr02_9.01) is correct. Therefore, we repaired this

chromosomal crossover by first switching the mating type of the repaired *synII* (yeast_chr02_9.03) and then back crossing to *synII strain* (yeast_chr02_9.01) to avoid crossover between synthetic chromosome *II* and native chromosome *II*. After tetrad dissection, the correct karyotype was verified by PFGE (fig. S6).

Comprehensive characterization of synII (yeast_chr02_9.03) phenotype

Growth curves, phenotype and morphology tests under various cultivation and stress conditions were performed to check the fitness of the repaired *synII strain* compared to the wild-type counterparts (*BY4741* and *BY4742*) (fig. 2A & 2B). Results show that *synII* strains are largely on par with wild-type strains (fig. S7, S8, S9). We did observe a significant fitness defect appear after megachunk *E* integration. Further investigation revealed that the defect was introduced by the insertion of the selective marker *URA3* (for integration selection) adjacent to the *NCL1* gene, which encodes a tRNA:m5C-methyltransferase. Loss of function of *NCL1* gene is previously reported to be involved in slow growth and temperature sensitivity (23, 24), which explains the observed phenotypic defect. This auxotrophic marker interference phenotype was reversed in the next step of SwAP-IN, in which the *URA3* marker was overwritten by an intact *NCL1* gene. Similar instances of such reversible fitness defects were also found after integration of megachunk B, U, and W (fig. S8).

In addition, a defect at high temperature (37°C) in medium with glycerol as the carbon source was observed: colony size is much smaller than wild-type. We confirmed that the slow-growth phenotype is recessive by crossing *synII* with wild type (*BY4742*) (fig. S10A). To ascertain the origin of this defect, we examined the phenotype of each intermediate strain under this particular condition and found that the defect appeared after the integration of megachunk *X* (fig. S10B). Using complementation assays, we identified a modification resulting from PCRTag recoding in *YBR265W* (fig. S10C). As an essential gene, *YBR265W* encodes 3-ketosphinganine reductase (Tsc10p) that catalyzes the second step in the pathway for sphingolipid synthesis in *S. cerevisiae* (25). Sphingolipids are reportedly involved in the regulation of the yeast high-osmolarity glycerol (HOG) response pathway (26). We reasoned that the defect might be caused by modifications in *YBR265W*. On replacement with the wild-type gene, the defect was greatly mitigated (fig. S10D). This highlights the effectiveness of the genome debugging mechanism provided by the modular chromosome construction strategy in the Sc2.0 project. Given that some defects might involve more than one bug, making the aforementioned method inadequate. In this case, SCRaMbLE may be a powerful alternative approach for targeting the defect. We have demonstrated the feasibility of this approach by quickly evolving *synII* strain to identify and overcome a design problem. By turning on the SCRaMbLE system in *synII* cells and directly plating the SCRaMbLED population on the defect-causing condition, we identified a few large colonies that resemble wild type (fig. S11A). From two large colonies we isolated for sequencing, no rearrangement was observed within the *YBR265W* region. However, all genomic rearrangements on *synII* involve genes coding for regulatory proteins that have either genetic or physical interactions with HOG regulatory proteins (fig. S11B). Interestingly, we observed that a phenotypic defect can be recovered through a number of alternative routes in the complex cellular interactome network.

As more and more individual synthetic chromosomes are finishing and will be merged into a single haploid cell, we successfully merged *synII* and *synIII* using method established previously (18). The resulting *synII/III* cell grows just like native cells (fig. S12), indicating that *synII* can be successfully incorporated into another synthetic background without introducing growth defects.

Segregation and replication of synII

Once *synII* was constructed and phenotypically profiled, a logical next question is to ask how this synthetic chromosome replicates and segregates. Therefore, to further investigate whether the modifications to *synII* might introduce subtler chromosome damage not detected in large-scale phenotypic assays, we examined and compared DNA replication and segregation processes during the cell cycle in *synII* and WT (*BY4741*) strains by tagging *synII* with an array of *tet* operators and using the previously described TetR-GFP method (27). Here, *synII* and WT cells were released from a *G1* block synchronously into the cell cycle (fig. S13). The progression of cells from metaphase to anaphase was identified by spindle morphology. We scored separation of sister chromatids at *CEN2*, by counting the percentage of cells in which two GFP dots were visible in both *synII* and *BY4741* strains and found them to be comparable (fig. 2C). Calculation of the overall ratio of metaphase to anaphase spindles and flow cytometry analysis (FACS) of DNA content provided further evidence for normal cell cycle progression in the *synII* strain (fig. 2C, fig. S14).

Early activating replication origins are frequently adjacent to tRNAs and transposable elements containing Long Terminal Repeats (LTRs) (28). In *synII*, two early origins are no longer linked with tRNAs and LTRs, as they were removed by design. Therefore, *synII* offers an opportunity to test whether this correlation represents a functional relationship. We determined the replication dynamics for *synII* and demonstrated that the synthetic chromosome showed identical replication dynamics to the wild-type chromosome *II* (fig. 2D). This finding is consistent with the linkage between early activating origins, tRNAs and LTRs not being a functional requirement for early origin activation.

Our results show that *synII* exerts no gross negative effect on key cell cycle transitions; including chromosome replication (*S* phase) and segregation of sister chromatids (anaphase). In additional, Hi-C analysis of *synII* also revealed no substantial changes between synthetic and native chromatin, suggesting that the designed sequence has little of no effect on the average global folding of the chromosome (22).

synII Trans-Omics analysis reveals genome plasticity

We used a systems biology approach (“Trans-Omics” analysis) to probe the impact of *synII* on the genomics, transcriptomics, proteomics and metabolomics of *S. cerevisiae*. The insertion of 267 loxPsym sites in *synII* could affect genomic stability; therefore we evaluated genome integrity and the loss frequency of chromosome segments in the absence of Cre expression. PCRTag analysis showed that no deletions were observed in 27 independent isolates derived from over 130 mitotic generations from 9 independent lineages (fig. S15A & S15B). The overall loss rate was estimated at lower than 5.9×10^{-6} . In addition, the deep sequencing

analysis of 9 derived single strains showed that no mutation or genome rearrangement was observed, indicating that genome stability was maintained faithfully even after 100 generations of nonselective growth (fig. S15C). Transcriptome profiling identified only 18 out of 6,561 genes to have differential mRNA expression in comparison to the wild-type, with 7 up-regulated and 11 down-regulated ($FDR < 0.01$ and $p\text{-value} < 7.62E-06$, fig. 3A, table S3). In the proteomics analysis, mass spectrometry was performed on *synII* and *BY4741* strains. Mass spectrometry provided abundance data for 3,965 out of 6,682 protein-coding genes, and only 6 proteins showed substantial differential abundance in *synII* compared with wild-type (fig. 3B, table S4). Finally, metabolic LC-MS analysis was performed for metabolomics profiling. In total 4,941 and 6,417 mass spectra peaks with $CV < 30\%$ were identified in positive and negative mode respectively, mapping to 1,032 unique metabolites. Fewer than 0.78% of the metabolites were differentially represented (fig. 3C & 3D). Potential differentially regulated metabolites included sphingolipids, glycerophospholipids, steroids and steroid derivatives (table S5, table S6). Interestingly these are all molecules associated with membranes.

Moreover, we also observed cellular perturbations that potentially arise from the Sc2.0 overall design principles. Compared to the WT strain, a subtle but potentially biologically significant up-regulation of genes with the GO terms “ribosome” and “cytoplasmic translation” was observed in the *synII* strain on both the transcriptome and proteome level (fig. 3E, fig. S16A, B, C, D & E). Similar up-regulation was also observed in the *synV* and *synX* transcriptome (fig. S17). As previously reported, deletion of multi-copy tRNA genes can lead to increased expression of translation machinery (29), thus we reasoned that the up-regulation of translational function might be caused by the 13 tRNA genes deleted in *synII*, which are all from multi-copy tRNA gene families. RNAseq analysis of *synII* with the tRNA array (yeast_ chr02_3_9.03, strain *YCy1193*) verified our hypothesis: the introduction of tRNA array to *synII* greatly diminished upregulation of translational functions (fig. 3F).

In summary, despite these very specific and subtle differences, the Trans-Omics analysis provides explicit evidence that biological processes within the *synII* strain are highly consistent with the wild-type strain. Therefore, the yeast genome displays a great degree of plasticity, and can readily cope with the large degree of editing encoded into *synII*.

Conclusions

In previous reports, the entire *Mycoplasma genitalium* genome was synthesized from oligonucleotides, assembled in budding yeast into a complete genome (9, 30). The genome transplantation method heavily depends on accurate design of a viable genome, and the low success rate of the transplantation step (and a lack of knowledge about how generically it can be done) makes it challenging to restore fitness when defects arise. In contrast, the Sc2.0 consortium uses a modular construction approach to progressively swap the wild-type chromosome with its designer counterpart as illustrated for the first time (15), and many finished chromosomes (18-21). The modular approach provides an important mechanism to systematically dissect and repair phenotypic defects, exemplified by the elimination of a growth defect on a specific carbon source that we here found to be caused by a 25 bp PCRTag sequence in a 770 kb chromosome. Similar debugging success has been reported in other synthetic chromosomes (18, 19, 21). Another advantage of modular assembly is that it enables multiple teams to collaborate and construct synthetic chromosome in parallel. The modular approach also permits parallel construction of individual chromosomes. In this study, we have constructed synthetic chromosome II from both ends. Other innovative approaches with the shared goal of parallelized construction are possible; *synXII* was constructed in 6 parallel production strains (a method termed Meiotic Recombination-mediated Assembly), which can further promote the construction efficiency (21). Future efforts can be planned through the attempt to join the complementary advantage of our approach to this method to develop more efficient construction method.

Our comprehensive phenotypic assays of *synII* strains revealed the extensive plasticity of the yeast genome. An important Sc2.0 design goal was to be conservative by avoiding changes likely to affect fitness-related phenotypes. Results with *synII* and other single synthetic chromosome strains suggest that the design has been successful in avoiding major fitness defects. However, interesting subtle patterns of differential gene expression were revealed by looking for enriched GO terms in the transcriptome and proteome analyses. It may be premature to suggest that the Sc2.0 design is too conservative, since strains hosting multiple synthetic chromosomes may have more profound phenotypic differences from the wild type due to the interactions among designer features in different chromosomes. A Trans-Omics approach will be a powerful way to capture subtle changes in the interactome network at various levels. Our design objective for the Sc2.0 genome is not to create a strain that encodes a particular phenotype; rather our goal is to create a robust, high fitness, engineerable chassis for unbiased exploration of the viable genotype to phenotype space using SCRaMbLE induction under various conditions. This goal is in contrast to more standard synthetic biology goals to engineer a single, specific phenotype or metabolic capability. Prior pilot studies provide strong evidence that SCRaMbLE works as designed as an effective approach to generate an unbiased exploration of the enormous design space for synthetic chromosomes (11, 31).

In conclusion, we have demonstrated that *synII* segregates, replicates and functions (at the Trans-Omics level) in a very similar way to its wild-type counterpart which has naturally evolved over millions of years. We conclude that *synII* fits the design objective and is ready to

be integrated into the final Sc2.0 genome with the joint effort of the entire Sc2.0 community in the near future.

Materials and Methods

SynII Design

We designed *synII* based on the version of S288C-derived strains available as of [30/11/2011] on SGD (http://downloads.yeastgenome.org/sequence/S288C_reference/genome_releases/, Version ID: S288C_reference_genome_R64-1-1_20110203). Genome editing suite *BioStudio* (14) was used to conduct the in silico design of *synII*, with the final version being yeast_ *chr02_3.25*. Compared to native chromosome II, the resulting 770,035 bp *synII* sequence is altered on average once in every 467 bp, and has a total of ~10.26% sequence alteration. 22 of 30 introns from protein-coding genes were removed from *synII*, with the remaining 8 retained because they are either known to lead to a fitness defect when deleted or they are ribosomal proteins genes and deletion of such introns may lead to fitness defects (32). More design details can be found in Table S1 and on the synthetic yeast project website (<http://www.syntheticyeast.org>).

synII Segmentation

Once *synII* design was completed, segmentation of *synII* into megachunks (~30 kb) and chunks (~10 kb) was performed using *BioStudio* (14). By applying an in-house developed program called SegMan, each chunk was segmented into minichunks (~3 kb). Minichunks were designed to be assembled into chunks by Gibson assembly (16) (see “Minichunk Assembly” section), and therefore 40 bp overlaps between each adjacent minichunks were added by SegMan. Terminal restriction endonuclease sites were also added on both ends of each minichunk to allow excision from the plasmid with the requirement that only 5' sticky or blunt ends could be generated by digestion. The 40 bp overlap regions were chosen based on following criteria: minimal free energy > -3 kcal/mol, and melting temperature (T_m) of 68 ± 4 °C. Higher free energy of overlap sequence leads to reduced probability of self-folding of single-stranded DNA and results in higher efficiency for overlap-based in vitro assembly methods (33), such as Gibson assembly (16) and USER assembly (34). Here the RNAfold program of the ViennaRNA Package 2.0 (35) was used to calculate the minimal free energy of the DNA sequence. A simplified formula (36) was applied to estimate the melting temperature: $T_m = 945 * \Delta H / (\Delta S + R * \log(0.0001)) - 273.15$, in which ΔH is enthalpy (kJ/mol), ΔS is entropy, R is the molar gas constant (1.9872 cal/mol-K). In addition, to reduce synthesis cost, *LEU2* and *URA3* markers were designed separately as independent cassettes and used repeatedly in minichunk assembly.

synII Synthesis and Assembly

Minichunk Assembly

Synthesis of ~3 kb minichunks was outsourced to Invitrogen, Genscript and BGI Tech. *pSBGAA* or *pSBGAK* (sequence information can be found at www.syntheticyeast.org) was chosen as the accepting vector of chunks. Minichunks were excised from the plasmids using the terminal restriction sites. *Bam*HI was used to linearize the chunk-accepting vectors *pSBGAA* (*Amp^R*) and *pSBGAK* (*Kan^R*). The chunks were assembled using Gibson assembly strategy (16) with a modified recipe: 1 μ L of Taq ligase (New England Biolabs) was added in the final volume of 20 μ L for each reaction. The molar ratio of accepting vector to minichunk DNA was 1:5. After thorough mixing, a one-hour incubation at 50°C was performed. Then 10 μ L of the reaction mixture was transformed to 50 μ L *E.coli* DH5 α competent cells (TAKARA). For assembly verification, single colonies were selected for overnight culture at 37°C. After miniprep, restriction digestion of the terminal restriction sites was performed to verify the assembly result.

Chunk and Junction Preparation

Once verified by restriction digestion, the component chunks of each megachunk were digested using the terminal restriction sites and gel purified. For each megachunk, 1 μ L of each chunk DNA, 0.5 μ L of T4 DNA ligase (New England Biolabs) and ddH₂O were added to a final volume of 10 μ L and mixed well for an overnight ligation at 16°C. Then the reaction product was diluted 1:10 and used as template to amplify ~1 kb junction fragments with the corresponding primers (primer sequence information can be found at www.syntheticyeast.org) using Phusion DNA polymerase (New England Biolabs). After gel verification, junction fragments were purified using a PCR cleanup kit (Axygen).

Replacement of WT Yeast Chromosome II with Synthetic Chunks

BY4741 (*MATa his3 Δ 1 leu2 Δ 0 LYS2 met15 Δ 0 ura3 Δ 0*) containing the *KanMX* marker (strain ID: YS000-L) and BY4742 (*MATa*) containing the *URA3* marker (strain ID: YS000-R) were used as the initial strains for synthetic megachunk replacement from the left arm and right arm, respectively. The *URA3* and *LEU2* selectable markers were used iteratively for replacing the native sequences of chromosome II with synthetic chunks. Together with the ~1 kb junction fragments (consisting of ~500 bp overlaps with two adjacent chunks), the chunks constituting each megachunk were co-transformed using the LiOAc transformation protocol (15) with 300 ng of each chunk DNA and 200 ng of each junction DNA added. The transformation products were re-suspended in 100 μ L 5 mM CaCl₂ and plated on an appropriate selectable media (SC-Ura or SC-Leu) with serial dilutions where appropriate. After 18 successive rounds (left arm initiated) and 8 successive rounds (right arm initiated) of replacements, two semi-synthetic chromosomes were successfully constructed, with one containing synthetic megachunks *A-R* (*synIIA-R*, strain ID: YS018) and the other containing synthetic megachunks *R-Y* (*synIIR-Y*, strain ID: YS026). The selectable marker *URA3* in *synIIR-Y* was removed through yeast transformation with the corresponding markerless synthetic fragment and screened with 5-FOA (37).

Integration of parallel constructed semi-synthetic chromosomes

An I-SceI (38)-mediated method was developed to combine the two semi-synthetic chromosomes, *synIIA-R* and *synIIR-Y*. For *synIIA-R*, a fragment (I-SceI-*URA3*) containing the I-SceI recognition sequence and *URA3* was designed to have a 40 bp overlap upstream and downstream of the *LEU2* marker on *synIIA-R*. A previously described method (39) was used to integrate the I-SceI-*URA3* cassette into *synIIA-R* to replace *LEU2* marker (strain ID: YS027). Following the same strategy, the I-SceI-*URA3* cassette was inserted into *synIIR-Y* with the *URA3* marker residing in wild-type sequence at the upstream end of megachunk R. In addition, a *LEU2* marker was inserted into *synIIR-Y* at the upstream of the centromeric region of *synIIR-Y* within gene YBL005W (strain ID: YS028). Mating of the *synIIA-R* strain (*MATa*, strain ID: YS027) and *synIIR-Y* (*MATa*, strain ID: YS028) was performed by overnight co-culturing in 3mL YPD medium at 30°C. The I-SceI expression vector *pRS413-pGAL-I-SceI* was transformed into the diploid cells followed by plating on SC-His plate. Single colonies were selected for overnight culture in 3 mL SC-His/glucose medium. The overnight culture was then added to 20 mL SC-His/raffinose (2% raffinose and 0.1% glucose) to reach an initial OD₆₀₀ of 0.1 and incubated until OD₆₀₀ reached 0.4 (~4-hour). Cells were harvested by centrifugation at 6,200 rcf for 5 min and re-suspended in 20 mL SC-His/galactose (2% galactose). After a 2-hour I-SceI induction on galactose media, 20 μ L cells were inoculated into 3 mL YPD medium, followed by overnight culture at 30°C. Overnight culture (1 mL) was harvested by centrifugation at 13,800 rcf for 1 min. The pellets were re-suspended in 200 μ L ddH₂O, patched on a SPOR plate and incubated at room temperature for

1 day. Then 5-7 days incubation at 30°C was performed until a significant number of tetrads was observed.

Cells were re-suspended in 25 µL Zymolyase-20T (25 mg/mL in 1 M sorbitol), then incubated at 37°C, with shaking at 800 rpm for 60 min. Then 500 µL ddH₂O was added and mixed thoroughly, followed by plating on 5-FOA plates and incubating at 30°C for 3 days. Then replica plating on SC-Leu and YPD plates was performed, followed by 24 hours incubation at 30°C. FOA^R Leu⁻ colonies were selected from YPD plates, inoculated into 3 mL YPD medium, and cultured overnight. To verify the integration of semi-synthetic chromosomes, one pair of synthetic PCRTag and wild-type PCR tag were chosen from each megachunk (in total 25 pairs) to perform PCRTag analysis. Then sequencing was performed for further verification.

Yeast Genomic DNA Preparation for PCRTag analysis

Yeast cells were collected from 0.5 mL of overnight culture by centrifugation at 13,800 rcf for 1 min. Pellets were resuspended in 100 µL breaking buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.1 M NaCl, 2% Triton x-100, 1% SDS). Glass beads (0.1 g, 0.5 mm of diameter) and 200 µL PCI (Phenol: Chloroform: Isoamyl alcohol= 25:24:1) were added and the mixture was vortexed at maximum speed for 3 min. Then 100 µL ddH₂O was added and the mixture was centrifuged at 13,800 rcf for 5 min. Supernatant (150 µL) was transferred to a microcentrifuge tube.

PCRTags Analysis Reaction setting

PCRTag amplification was performed using rTaq Polymerase (TAKARA). Forward and reverse PCRTag primer pairs (400 nM each, detailed information of PCRTag primers can be found at www.syntheticyeast.org), 1 µL BY4741/*synII* genomic DNA and 2.25 µL ddH₂O were added to a final reaction volume of 12.5 µL. The PCR thermal-cycler program setting was as follows: 94°C/5 min, 30 cycles of (94°C/30 sec, 55°C/30 sec, 72°C/30 sec), and a final extension of 72°C/5 min. Detection of PCRTags was carried out by gel electrophoresis. PCR product (3 µL) was loaded onto a 2% agarose gel, and electrophoresis was performed at 180V for 20 min.

Yeast Genomic DNA Preparation for DNA Sequencing

Yeast cells were grown in 5 mL YPD medium using a 14 mL round bottom tube for 2 days till saturation. Pellets were collected by centrifugation at 13,800 rcf for 1 min. Breaking buffer (400 µL) was added to re-suspend the pellet. Glass beads (0.2 g, 0.5 mm of diameter) and 400 µL PCI (Phenol:Chloroform:Isoamyl alcohol=25:24:1) were added, the re-suspension was vortexed at maximum speed for 3 min, and then centrifuged at 13,800 rcf for 10 min. 400 µL of the aqueous layer was transferred to a new 1.5 mL tube. The genomic DNA was precipitated by adding 400 µL of isopropyl alcohol and kept at room temperature for 5 min. Then genomic DNA was pelleted by centrifugation at 13,800 rcf for 10 min. The pellet was washed with 500 µL 70% ethanol, followed by 5 min drying at 37°C. The genomic DNA was re-suspended in 50 µL TE buffer (10 mM Tris-HCl pH8.0, 1mM EDTA) with RNase (25 µg/mL) and incubated at 37°C for 30 min.

Nucleotide Sequence Analysis of *synII* with Hiseq2500 sequencing platform

Library preparation and whole genome sequencing

Paired-end whole genome sequencing was performed for the *synII* (yeast_ *chr02_9.01*, strain ID: YS029) on the HiSeq2500 platform. A 500-bp library was prepared according to standard Illumina DNA preparation protocols.

Sequencing quality control and mapping to genome sequence

Before mapping of reads, quality control of sequencing reads was performed. Reads with adapters or shorter than 90 bp were removed. Reads with more than one base having a Phred-score below 10 or with more than one unknown base were removed, leaving 673 Mbp cleaned paired-end reads, with 55.6-fold sequencing depth of the genome. Cleaned reads of each sample were mapped to yeast reference sequences (the original sequence of chromosome II being replaced by synthetic chromosome II sequence) using BWA 0.5.6 (40) with standard settings. For each alignment result, local realignment was performed with GATK 2.7-2 RealignerTargetCreator and IndelRealigner tools (41) to clean up mapping artifacts caused during reads mapping on the edges of indels. The resulting files in BAM format were then prepared for initial SNV/indel calling.

Identification of SNVs and indels

Both Samtools (42) and GATK 2.7 (41) pipelines were used to identify the SNVs and indels using default parameters. The variants were filtered by the criteria of QUAL < 50 or PV4 < 0.1,0.1,0.1,0.1 or MQ <10 or DP < 10 or DP4 < 0,0,3,3 for Samtools results and QUAL < 50 or FS >3 or BaseQRankSum >3 or MQRankSum>3 or ReadPosRankSum>3 or MQ <10 or DP < 10 for GATK results. The variants identified by either tool were merged with CombineVariants implemented in GATK. The merged variants in the synthetic chromosome were checked manually using Tablet (43) to exclude false-positive results caused by sequencing or mapping errors. Annotation was performed for observed variants, with the following types: synonymous type, non-synonymous type, frameshifts, and variant outside coding region (Table S2).

Identification of loxPsym sites

In *synIII*, *loxPsym* sites can be absent from expected locations (15). To check whether all *loxPsym* sites are present in *synII*, sequencing reads containing *loxPsym* sites were extracted from the whole read library for identifying the presence of all expected *loxPsym* sites. For each *loxPsym* site, the read mapped span the *loxPsym* site of upstream and downstream flanking sequence >10 nucleotides and the sequenced *loxPsym* site with bases of >=24 matching or mismatching *loxPsym* sequence was recognized as a *loxPsym* read and bases of >=15 deleted was counted as a “*loxPsym_lost*” read (Table S2). The *loxPsym* site with p-value of < 0.001, which estimated by Poisson mode, and supporting read number >=5 was identified as a loss of *loxPsym* site.

Structural variant detection with the PGM sequencing platform

Library preparation and whole genome sequencing

A 400-bp DNA library of a *synII* (yeast_ chr02_9.01, strain ID: YS029) was prepared for single-end whole genome sequencing according to the Life Tech standard preparation protocol using the Ion Xpress™ Barcode Adapter 1-96 Kit (Cat.no.4474517) and sequenced on the Ion PGM™ platform. Quality control of sequencing reads was performed. Reads shorter than 30bp or duplicated were removed. Reads with more than 1% of bases having a Phred-based quality score <10 or with unknown bases were trimmed first to meet the filtering criteria and removed if the first trimming step failed, leaving 488 Mbp cleaned reads, with 32.6-fold sequencing depth of the genome and 248 bp average length. Cleaned reads of each sample were mapped to synthetic yeast genome sequences using bowtie2-2.0.0 (44) with standard settings.

Structural variant reconstruction

To identify whether structural variations exist in *synII*, reads that did not map to the reference genome were

split to pairwise ends (split reads, at least 30 bp) by scanning over all intermediate positions at least 30bp from the ends of the read. Then these pairwise ends were then aligned to the reference using Bowtie2 (44) by single-end mapping with parameter $-k$ 100. The pairwise ends that matched parental sequence were analyzed for breakpoints to provide direct evidence for structural variants. One read was assigned as the most probable mapping type based on five priorities: no recombination events (top priority), intra-chromosome recombination, inter-chromosome recombination within wild-type chromosomes and external chromosome recombination between synthetic and wild-type chromosomes, and single end mapping. For each identified breakpoint, a 5 bp error range was allowed. These reads at a breakpoint site with at least 3 supporting reads were locally assembled using PolyPhred (45), and then the assembly results were annotated by Blastn against the reference containing the selection marker and acceptor vector and wild-type chromosome *II* sequence to identify the potential extrinsic sequence. Combining the split-read mapping, sequencing depth and local assembly contigs, the structural variations were reconstructed (Fig. S5).

synII Structural Variant Correction

An I-SceI mediated transformation strategy was designed to correct the two structural variations observed in Megachunks *L* and *T* of *synII* (yeast_ *chr02_9.01*, strain ID: *YS029*). For the duplication in Megachunk *L*, a donor fragment containing the *URA3* sequence and an I-SceI site (TAGGGATAACAGGGTAAT) was designed to carry 40 bp overlapping regions at the end of the original chunk *L4* and the beginning of duplicated chunk *L2* (Fig. 1B). The donor fragment was produced by PCR using primers including the I-SceI site and overlap sequence. I-SceI expression vector *pRS413-pGAL-I-SceI* and the donor fragment were co-transformed into the *synII* strain (strain ID: *YS029*). After I-SceI induction on galactose media, cells were plated directly on 5-FOA plates directly and incubated at 30°C for 3 days. Then FOA^R colonies were selected and cultured overnight in YPD medium at 30°C. Confirmation of repair was performed by PCR using a pair of primers designed to amplify across the *L4-L2* breakpoint (primer sequence information can be found at www.syntheticyeast.org).

For the complex variation in Megachunk *T*, a donor sequence was designed to contain a *URA3* cassette and the I-SceI site, with the left end overlapping with the right end of chunk *T4* by 500 bp, and the right end overlapping with the left end of chunk *T5* by 500 bp (Fig. 1B). Both 500 bp overlapping sequences and the I-SceI-*URA3* cassette were produced by PCR with 40 bp overlaps with each other and assembled together through the Gibson assembly method (16) to form the donor fragment. An additional 100 bp of chunk *T4* was added between the I-SceI-*URA3* cassette and 500 bp overlapping sequence of *T5* to facilitate the homologous recombination. The donor fragment and I-SceI expression vector *pRS413-pGal-I-SceI* were then co-transformed into the *synII* (yeast_ *chr02_9.02*, strain ID: *YS030*) in which the structural variation in Megachunk *L* had been repaired. I-SceI induction was performed on galactose-containing media and colonies were screened on 5-FOA-containing plates. PCR verification was then performed with 8 pairs of verification primers designed across each breakpoint observed in this structural variation (Fig. S5D). The *synII* (yeast_ *chr02_9.03*, strain ID: *YS031*) was sequence-verified again using the PGM sequencing platform, with library preparation and sequencing method as described in “structural variant detection with the PGM sequencing platform”.

Mating Type Switch of Yeast

pJD138 (pGAL-HO) was transferred into the *SynII* strain (yeast_ *chr02_9.03*, strain ID: *YS031*) using the LiOAc transformation protocol (15), followed by overnight culture in 5ml SC-Ura/galactose (2% galactose)

liquid at 30°C. The mating type was then determined by crossing with *MATa* and *MATα* tester strains and confirmed by microscopy.

Mating and Sporulation

A *synII* (yeast_ *chr02_9.01*, strain ID: *YS029*) and a second *synII* (yeast_ *chr02_9.03*) strain that had its mating type switched to alpha (strain ID: *YS032*) were recovered on YPD plates. *synII* (yeast_ *chr02_9.01*) colonies and *synII* (yeast_ *chr02_9.03*) colonies were picked and patched in the same section (mating section) on a new YPD plate with an inoculating loop, followed by overnight culture at 30°C. Then diploid colonies were patched on SPOR plates [10 g potassium acetate (Sigma), 1.25 g yeast extract (Oxoid), 1 g glucose (Sangon Biotech), 20 g agar (Sangon Biotech), ddH₂O up to 1 L] and left at room temperature for one day, followed by 8 days incubation at 30°C. Then cells were taken from SPOR plates and re-suspended in 25 μL zymolyase solution (MP Biomedicals, 1 mg/ml Zymolyase in 1M sorbitol), followed by 25 min incubation at 37°C. Cells were checked under the microscope to determine zymolyase efficiency. After tetrads were appropriately digested, 500 μL ddH₂O was added and the suspension mixed vigorously by vortexing to break up the tetrads. Then a series of 10-fold dilutions were made before plating on YPD plates. The plates were incubated at 30°C for 2-3 days. Single colonies were picked and streaked on fresh plates, followed by overnight incubation. Mating type of random spores was confirmed by crossing with *MATa* and *MATα* tester strains and confirmed by microscopy. Spore-derived colonies were selected for PCR verification. The nine pairs of primers used for identifying breakpoints in *SynII* (yeast_ *chr02_9.01*) were used here to verify the correct *synII* sequence (primer sequence information can be found at www.syntheticyeast.org). Pulsed-field gel electrophoresis (PFGE) was performed for karyotype analysis. Finally, candidates were further verified using the PGM sequencing platform, with library preparation and sequencing method as described in “structural variant detection with the PGM sequencing platform” above.

Pulsed-field Gel Electrophoresis

Samples were prepared for pulsed-field gel electrophoresis (46). Identification of chromosomes was inferred from the karyotype of WT (*BY4741*, *BY4742*) on the same gel. Samples were analyzed on a 1.0% agarose gel in 1 × TAE (pH 8.0) for 22 hrs at 14°C on a CHEF apparatus. The voltage was set to 5 V/cm, at an angle of 120°. Switch time was set to 60 – 90 sec ramped over 22 hrs.

Genome Stability Analysis

The *synII* (yeast_ *chr02_9.03*, strain ID: *YS031*) was streaked on YPD plate for 2 days and incubated at 30°C. Nine single colonies were selected for successive subculture in YPD medium for ~130 generations, followed by plating on YPD plates overnight. Three single colonies of each initial isolate were selected. Genomic DNA preparation was performed for all 27 isolated single colonies. 48 pairs of PCRTags were chosen to perform PCRTag assay for all 27 isolated single colonies. Sequencing on the PGM platform was conducted for 9 of the 27 isolated single colonies (one from each initial isolate), with library preparation and sequencing method as described in “structural variant detection with PGM sequencing platform” above. Sequence data were analyzed according to the method described in “Nucleotide Sequence Analysis of *synII* with Hiseq2500 sequencing platform” above.

Cre SCRaMbLE of synII

BY4741 and *synII* (yeast_ *chr02_9.03*, strain ID: *YS033*) with *pSCW11-CreEBD-His3* plasmid (47) were cultured in SC-His liquid media overnight at 30°C, with duplicates for each strain. Cells were inoculated into

SC–His liquid medium with 1 μ M EST and cultured for 24 hrs at 30°C. Then cells were washed, serial dilution and spotting were performed on YPG plates, followed by 4 days incubation at 37°C.

Yeast Total RNA Isolation for RNA Sequencing

The *synII* (yeast_chr02_9.03, strain ID: YS031) and BY4741 strain both with 3 biological replicates were cultured overnight in 3 mL YPD medium at 30°C. The cultures were added to 10 mL fresh YPD medium and incubated until the OD₆₀₀ reached ~0.8. The cells were harvested by centrifugation at 900 rcf for 5 min. Total RNA was isolated using the RiboPure-yeast kit (Ambion) according to the manufacturer's instructions.

RNA-seq Analysis of synII

The 330 bp cDNA libraries were prepared according to the manufacturer's instructions (Illumina Inc.) and were paired-end sequenced using the Illumina Hiseq2500. Raw read data were filtered using the following criteria: no N bases, no adaptor sequences, minimum read length 100, bases of low quality (<10) was no more than 1% in a read. An average of 7.0 M and 8.6 M clean reads were obtained from BY4741 and *synII* (yeast_chr02_9.03, strain ID: YS031) samples respectively, with 95.7% and 95.2% mapped to the corresponding genome reference. Cleaned reads were mapped to genomes by tophat v2.0.10 (48), with the parameter $-r$ ($--mate-inner-dist$) of 130. After reads counting, RPKM was calculated for each gene. Differential gene expression was analyzed by DEseq v1.20.0 (49), with the no replicates scenario (parameters were: $--method$ blind, $--sharingMode$ fit-only, $--fitype$, local). For each gene, we obtained a raw p-value and adjusted p-value using the Benjamini–Hochberg procedure. Genes were assessed for statistical significance by rejecting the null hypothesis if the adjusted p-value (FDR) was <0.01 and if the raw p-value fell below the threshold of the 5% Family Wise Error Rate (FWER) after Bonferroni correction (threshold=7.62E-06). False positive results were inferred by the following rules and removed: dubious genes, transposable genes, genes with low coverage (<60%) for native and synthetic strains.

Proteome and Metabolome Analysis of synII

BY4741 and *synII* (yeast_chr02_9.03, strain ID: YS031) strains with 3 biological replicates were cultured overnight in 3 mL YPD medium at 30°C and re-inoculated in 10 mL fresh YPD medium for further incubation until the OD₆₀₀ reached ~0.8. The cells were collected for both proteome and metabolome analysis. Proteins from the yeast cell were extracted with Urea, reduced, alkylated, digested with trypsin and iTRAQ labeled- BY4741 (113, 115, 117, 121) and *synII* (114, 116, 118 and 119) (AB SCIEX, Framingham, MA, USA). After labeling, the peptides were fractionation with SCX method and analysis by an Orbitrap Q Exactive mass spectrometer (Thermo Fisher Scientific, San Jose, CA) coupled with an online HPLC. Mascot and IQuant (50) software were used for protein identification and quantification.

For the global metabolomics, metabolites were extracted with buffer (50% methanol and 50% water), separated with a BEH C18 column. For the lipidomics, metabolites were extracted with buffer (75% dichloromethane and 25% methanol), separated with a CSH C18 column. The metabolites were detected by a XEVO-G2XS QTOF mass spectrometer (Waters, Manchester, UK) and the raw spectrums were processed by Progenesis QI 2.0 software (Nonlinear Dynamics, Newcastle, UK) for peak picking, alignment, normalization and identification. Further statistical analysis was performed on the resulting normalized peak intensities using in-house developed software metaX. Pathway analysis was conducted using the MetaboAnalyst pathway tool (51).

Yeast KEGG pathway and GO enrichment

To identify differentially active biological processes between *synII* (yeast_ *chr02_9.03*, strain ID: *YS031*) and *BY4741* strains, gene enrichment and co-expression enrichment analyses were performed using yeast KEGG pathways and yeast Gene Ontology (GO) annotations integrating transcriptomics, proteomics and metabolomics data respectively. Genes and metabolites with differential expression of \log_2 (fold-change) >0 and down-regulation of \log_2 (fold-change) <0 were considered as up-regulated and down-regulated for enrichment analysis. The significance of each KEGG pathway and GO term in genes, metabolites and co-expression was individually identified using the hyper-geometric test and Chi-squared test with false discovery rate (FDR) correction(52) and the threshold P-value <0.001 . The KEGG category and interaction with genes of metabolites were classified based on the Yeast Metabolome Database (53) (<http://www.ymdb.ca/system/downloads/current/ymdb.json.zip>).

Growth Curve Assay

The growth curve analysis of *BY4741*, *BY4742*, and *synII* (yeast_ *chr02_9.03*, strain ID: *YS031*) was carried out using a Bioscreen C system (Oy Growth Curves Ab Ltd). Overnight cultures with 3 replicates were sub-cultured into 300 μ L of the different media to final OD₆₀₀ of ~ 0.1 in the 96 well micro-plates. Medium was also added as blank.

Serial dilution assay on various types of media

BY4741, *BY4742*, all intermediate strains (strain ID: *YS001-YS026*), and *synII* (yeast_ *chr02_9.03*, strain ID: *YS031*) were inoculated in YPD medium at 30°C overnight. Then 1:10 serial dilutions were performed using these samples onto different medium plates, which including: YPD, YPD with MMS (testing for defective DNA repair), YPD with benomyl (microtubule inhibitor), YPD with Camptothecin (topoisomerase inhibitor), YPD with Hydroxyurea (testing for defective DNA replication), YPD with various final concentrations of Sorbitol (osmotic stress 0.5 M, 1 M, 1.5 M, 2 M, to cause osmotic stress), synthetic complete (SC) medium, SC with 6-Azaauracil (testing for defective transcription elongation), YPEG (respiratory defects) containing 2% Glycercol and 2% Ethanol (testing for respiratory defects), and YPD adjusted to pH 4.0 and pH 9.0 with HCl and NaOH (testing for vacuole formation defects. Two specific drugs, Hydrogen peroxide (testing for oxidative stress) and Cycloheximide (testing for defective protein synthesis), were added to overnight cultures to treat the cells for two hours. Yeast cells were collected by centrifugation and resuspended in water before serial dilution and plating. For different medium types, plates were incubated at 25°C/30°C/37°C for 2/3/4 days (Fig. S8).

Cell Morphology

Cells were grown to log phase in YPD at 30°C. DIC Images were collected using a Nikon microscope Ti-E (100X) with an Andor Zyla 5.5 camera (Fig. S9).

CEN2-GFP Strain Construction and Culturing

To create the *CEN2-GFP* strain, a 950 bp region to the right of *CEN2* was cloned into *Amp327* (27) and integrated into the *synII* strain (yeast_ *chr02_9.03*, strain ID: *YS031*) and *BY4741* carrying *tetR-GFP*, generating a GFP label 15 kb to the right of *CEN2* (Strain ID: *YCy922*, *YCy1018*). Cell cycle synchronization of *CEN2-GFP* strains was performed by adding alpha factor in YEP medium. Alpha factor was removed by filtration and washing with YEP medium. Three hour mitotic time courses were performed at room temperature unless otherwise stated.

Microscopy

Methods of fixing cells for GFP-labeled chromosome visualization and indirect immunofluorescence described previously were used in this study (27). Tubulin was visualized using an anti-rat antibody at a dilution of 1:50 in PBS/BSA and an anti-rat FITC antibody at a dilution of 1:16.67 in PBS/BSA. Microscopy was performed on a Zeiss Axioplan 2 microscope and images were captured using a Hamamatsu camera operated through Axiovision software. To score GFP dots and spindles in mitosis, cells (Strain ID: *YCy922*, *YCy1018*) were co-stained with DAPI to visualize nuclear morphology (Fig. S13). For each sample obtained at each 15 min time point, 200 cells were counted in the field to score GFP dots and spindles.

FACS Analysis

Methods of flow cytometry described previously were used in this study (27). Cells were fixed in 70% ethanol at 4°C overnight. Cells (Strain ID: *YCy922*, *YCy1018*) were then treated with RNase (20 mg/ml, Sigma), and digested with proteinase K (20 mg/ml, Amresco). Samples were briefly sonicated before analysis. FACS analysis was performed according to the manufacturer's instructions (BD FACS Calibur). Data were analyzed using FlowJo software. (Fig. S14).

synII Replication Timing Profiles Analysis

Replication profiles were obtained using methods described previously (54). Asynchronous cells were fixed with 70% ethanol, treated with RNaseA and Proteinase K and stained with 10 × Sytox. The MoFlo XDP was used to enrich for replicating and non-replicating cells. Sorted cells were treated with Zymolase, RNaseA and Proteinase K. DNA was purified using phenol-chloroform extraction and ethanol precipitation. Single end Illumina HiSeq2500 sequencing yielded a minimum of 29.6 million uniquely mapped reads per sample. The ratio between uniquely mapped reads from the non-replicating and the replicating sample were calculated for every 1000 bp window. The ratio was normalized to a baseline of one to control for differences in number of reads between the samples.

References

1. O. Ozier-Kalogeropoulos, A. Malpertuy, J. Boyer, F. Tekaia, B. Dujon, Random exploration of the *Kluyveromyces lactis* genome and comparison with that of *Saccharomyces cerevisiae*. *Nucleic Acids Res* **26**, 5511 (Dec 1, 1998).
2. R. B. Langkjaer, M. L. Nielsen, P. R. Daugaard, W. Liu, J. Piskur, Yeast chromosomes have been significantly reshaped during their evolutionary history. *J Mol Biol* **304**, 271 (Dec 1, 2000).
3. V. Wood *et al.*, The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**, 871 (Feb 21, 2002).
4. S. Brachat *et al.*, Reinvestigation of the *Saccharomyces cerevisiae* genome annotation by comparison to the genome of a related fungus: *Ashbya gossypii*. *Genome Biol* **4**, R45 (2003).
5. M. Kellis, B. W. Birren, E. S. Lander, Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617 (Apr 8, 2004).
6. H. G. Menzella *et al.*, Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat Biotechnol* **23**, 1171 (Sep, 2005).
7. D. K. Ro *et al.*, Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940 (Apr 13, 2006).
8. G. Posfai *et al.*, Emergent properties of reduced-genome *Escherichia coli*. *Science* **312**, 1044 (May 19, 2006).
9. D. G. Gibson *et al.*, Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215 (Feb 29, 2008).
10. C. Lartigue *et al.*, Genome transplantation in bacteria: changing one species to another. *Science* **317**, 632 (Aug 3, 2007).
11. J. Dymond, J. Boeke, The *Saccharomyces cerevisiae* SCRaMbLE system and genome minimization. *Bioeng Bugs* **3**, 168 (May-Jun, 2012).
12. J. S. Dymond *et al.*, Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471 (Sep 22, 2011).
13. H. Feldmann *et al.*, Complete DNA sequence of yeast chromosome II. *The EMBO journal* **13**, 5795 (Dec 15, 1994).
14. S. M. Richardson *et al.*, Design of a synthetic yeast genome. *Science, this issue* ***,***-**, (2017).
15. N. Annaluru *et al.*, Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55 (Apr 4, 2014).
16. D. G. Gibson *et al.*, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343 (May, 2009).
17. J. Parenteau *et al.*, Deletion of many yeast introns reveals a minority of genes that require splicing for function. *Molecular biology of the cell* **19**, 1932 (May, 2008).
18. L. A. Mitchell *et al.*, Synthesis, debugging and consolidation of synthetic chromosomes in yeast: synVI and beyond. *Science, this issue* ***,***-**, (2017).

19. Y. Wu *et al.*, Bug-mapping and fitness testing of chemically synthesized chromosome X in Yeast. *Science, this issue* ***,***-**, (2017).
20. Z. Xie *et al.*, "Perfect" designer chromosome V and behavior of a ring derivative. *Science, this issue* ***,***-**, (2017).
21. W. Zhang *et al.*, Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science, this issue* ***,***-**, (2017).
22. G. Mercy *et al.*, 3D organization of synthetic and scrambled chromosomes. *Science, this issue* ***,***-**, (2017).
23. A. Alexandrov *et al.*, Rapid tRNA decay can result from lack of nonessential modifications. *Mol Cell* **21**, 87 (Jan 6, 2006).
24. J. M. Dewe, J. M. Whipple, I. Chernyakov, L. N. Jaramillo, E. M. Phizicky, The yeast rapid tRNA decay pathway competes with elongation factor 1A for substrate tRNAs and acts on tRNAs lacking one or more of several modifications. *RNA* **18**, 1886 (Oct, 2012).
25. S. K. Kim, Y. H. Noh, J. R. Koo, H. S. Yun, Effect of Expression of Genes in the Sphingolipid Synthesis Pathway on the Biosynthesis of Ceramide in *Saccharomyces cerevisiae*. *J Microbiol Biotechn* **20**, 356 (Feb, 2010).
26. M. Tanigawa, A. Kihara, M. Terashima, T. Takahara, T. Maeda, Sphingolipids Regulate the Yeast High-Osmolarity Glycerol Response Pathway. *Mol Cell Biol* **32**, 2861 (Jul, 2012).
27. J. Fernius, A. L. Marston, Establishment of cohesion at the pericentromere by the Ctf19 kinetochore subcomplex and the replication fork-associated factor, Csm3. *PLoS Genet* **5**, e1000629 (Sep, 2009).
28. S. C. Di Rienzi, D. Collingwood, M. K. Raghuraman, B. J. Brewer, Fragile genomic sites are associated with origins of replication. *Genome Biol Evol* **1**, 350 (2009).
29. Z. Bloom-Ackermann *et al.*, A Comprehensive tRNA Deletion Library Unravels the Genetic Architecture of the tRNA Pool. *Plos Genetics* **10**, (Jan, 2014).
30. D. G. Gibson *et al.*, One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc Natl Acad Sci U S A* **105**, 20404 (Dec 23, 2008).
31. Y. Shen *et al.*, SCRaMble generates designed combinatorial stochastic diversity in synthetic chromosomes. *Genome Res* **26**, 36 (Jan, 2016).
32. J. Parenteau *et al.*, Preservation of Gene Duplication Increases the Regulatory Spectrum of Ribosomal Protein Genes and Enhances Growth under Stress. *Cell reports* **13**, 2516 (Dec 22, 2015).
33. M. Muller, K. Daoulas, Calculating the free energy of self-assembled structures by thermodynamic integration. *The Journal of chemical physics* **128**, 024903 (Jan 14, 2008).
34. J. Bitinaite *et al.*, USER (TM) friendly DNA engineering and cloning method by uracil excision. *Nucleic Acids Res* **35**, 1992 (Mar, 2007).
35. R. Lorenz *et al.*, ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**, 26 (2011).
36. W. Rychlik, W. J. Spencer, R. E. Rhoads, Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic Acids Res* **18**, 6409 (Nov 11, 1990).

37. J. D. Boeke, F. LaCroute, G. R. Fink, A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol Gen Genet* **197**, 345 (1984).
38. L. Colleaux, L. D'Auriol, F. Galibert, B. Dujon, Recognition and cleavage site of the intron-encoded omega transposase. *Proc Natl Acad Sci U S A* **85**, 6022 (Aug, 1988).
39. U. Guldener, S. Heck, T. Fielder, J. Beinhauer, J. H. Hegemann, A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res* **24**, 2519 (Jul 1, 1996).
40. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754 (Jul 15, 2009).
41. A. McKenna *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297 (Sep, 2010).
42. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078 (Aug 15, 2009).
43. I. Milne *et al.*, Using Tablet for visual exploration of second-generation sequencing data. *Brief Bioinform* **14**, 193 (Mar, 2013).
44. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
45. D. A. Nickerson, V. O. Tobe, S. L. Taylor, PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* **25**, 2745 (Jul 15, 1997).
46. J. Herschleb, G. Ananiev, D. C. Schwartz, Pulsed-field gel electrophoresis. *Nat Protoc* **2**, 677 (2007).
47. D. L. Lindstrom, D. E. Gottschling, The Mother Enrichment Program: A Genetic System for Facile Replicative Life Span Analysis in *Saccharomyces cerevisiae*. *Genetics* **183**, 413 (Oct, 2009).
48. C. Trapnell, L. Pachter, S. L. Salzberg, TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105 (May 1, 2009).
49. S. Anders, W. Huber, Differential expression analysis for sequence count data. *Genome Biol* **11**, R106 (2010).
50. B. Wen *et al.*, IQuant: an automated pipeline for quantitative proteomics based upon isobaric tags. *Proteomics* **14**, 2280 (Oct, 2014).
51. J. Xia, D. S. Wishart, Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. *Nat Protoc* **6**, 743 (Jun, 2011).
52. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* **57**, 289 (1995).
53. T. Jewison *et al.*, YMDB: the Yeast Metabolome Database. *Nucleic Acids Res* **40**, D815 (Jan, 2012).
54. C. A. Muller *et al.*, The dynamics of genome replication using deep sequencing. *Nucleic Acids Res* **42**, e3 (Jan, 2014).

Acknowledgements:

The project was funded by China National High Technology Research and Development Program ["863" Program: 2012AA02A708] to H.Y. and National Natural Science Foundation of China [21621004, 21390203] to Y.Y.; a Chancellor's Fellowship from the University of Edinburgh, a start up fund from SULSA and BBSRC grants [BB/M005690/1; BB/M025640/1 and BB/M00029X/1] to Y.C.; the Wellcome Trust to A.M. [090903; 092076]; US NSF grants [MCB-1026068, MCB-1158201] to J.D.B.; US NSF grant [MCB-1445545] to J.S.B. and US Department of Energy grant [DE-FG02097ER25308] to S.M.R. Work at Tsinghua University was supported by National Science Foundation of China grant 31471254 and Chinese Ministry of Science and Technology grant 2012CB725201 to J.D. **Funding from ERASynBio and Agence Nationale pour la Recherche (IESY ANR-14-SYNB-0001-03) to R.K.** We thank Weimin Zhang for many comments on the manuscript. J.D.B. and J.S.B. are founders and directors of Neochromosome Inc. J.D.B. serves as a scientific advisor to Recombinetics Inc. and Sample6, Inc. These arrangements are reviewed and managed by the committees on Conflict of interest at NYU Langone Medical Center (J.D.B.) and Johns Hopkins University (J.S.B.). Y.C., H.Y. and J.D.B. conceived the concept and designed the experiments. L.M., G.S., S.M.R., J.S.B., and J.D.B. designed the *synII* chromosome. Y.S., Y.W., T.C., J.G., Y.L. and Y.C. designed experiments. T.C., F.G., H.Z., Y.S., S.C., D.A., Z.L., R.W., W.L., and F.T. performed experiments. Y.W. and J.G. analyzed genomic and RNA-seq data. Y.F., B.Z., and J.G. analyzed mass spectrometry and LC-MS data. Y.S., and D.A. analyzed *synII* cell cycle data. C.M. and C.N. analyzed *synII* replication timing profiles. G.M., H.M. and R.K. analyzed *synII* 3C/3D structure. The manuscript was written by Y.S., Y.W., T.C., F.G., J.D. and Y.C., with input from C.M., C.N., R.K., C.F., A.M., J.S.B, J.D.B. and H.Y., with comments provided by all other authors. *synII* genomic sequencing data and RNA-seq data have been deposited in the Sequence Read Archive (SRA) sequence database (<http://www.ncbi.nlm.nih.gov/sra/>), with accession code SRP062892. Proteomics data have been submitted to the ProteomeXchange (<http://www.proteomexchange.org/>), with accession code PXD003435. The metabolomics and lipidomics data have been deposited to the MetaboLights metabolomics repository (accession number: MTBLS296, accessed via link: <http://www.ebi.ac.uk/metabolights/MTBLS296>). The designed and sequenced *synII* information is uploaded with accession codes CP013607 and CP013608 on GenBank. Additional information (*synII* design diagram, PCRTag sequences, Variants in physical strain (yeast_chr02_9_01), PCR primers, Summary of megachunks, chunks and minichunks) related to *synII* can be accessed on the Sc2.0 website (www.syntheticyeast.org).

Supplementary Materials

Figures. S1 to S17

Tables S1 to S7

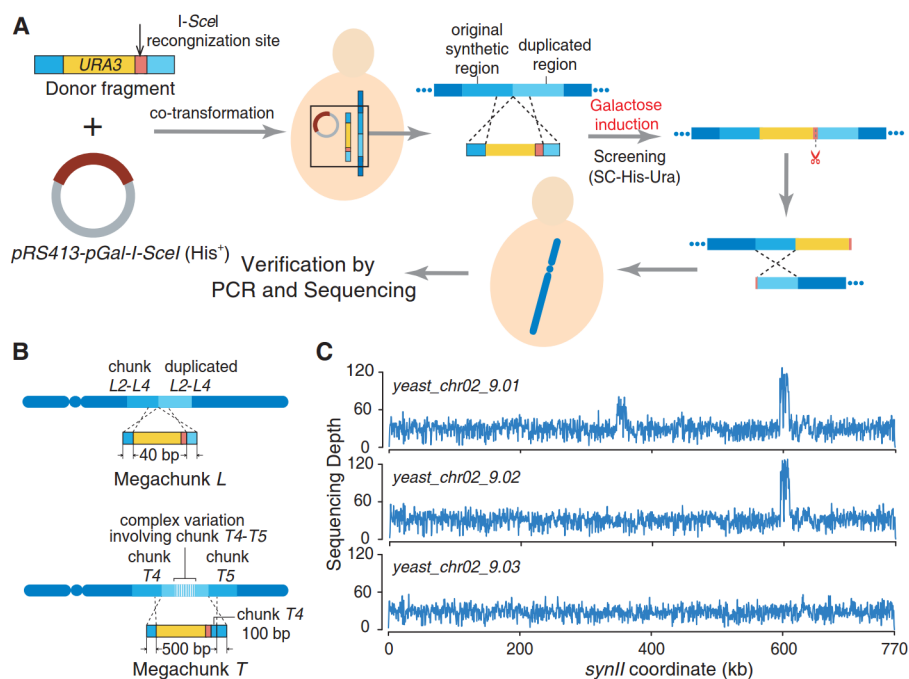


Fig. 1. Structure variation repair through chromosome breakage.

(A) I-SceI mediated repair strategy for *synII* structure variations. The donor fragment was designed to carry a *URA3* cassette (yellow) and I-SceI recognition site (red), with both ends overlapping the structural variation sequences observed in *synII* (yeast_*chr02_9.01*). The donor fragment was integrated into *synII* between the two tandem repeats through homologous recombination, then an episomal plasmid *pRS413-pGal-I-SceI* was transformed into the cell. A double strand break at the I-SceI site was induced in galactose medium, and the homologous recombination of two partial chromosomes of *synII* eliminated the duplication.

(B) Structure variations in megachunks *L* and *T* and their corresponding donor sequence design. In megachunk *L*, a copy of chunk *L2-L4* was observed following the original *L2-L4* sequence that generated a tandem-duplication. The donor fragment was inserted directly between two duplications. In megachunk *T*, a complicated variation involving sequence of chunk *T4-T5* was inserted into *synII* between chunk *T4* and *T5* and the donor fragment was inserted to remove the complex variation.

(C) Deep sequencing read depth analysis revealed the successful sequential removal of duplications. The starting synthetic chromosome *synII* (yeast_*chr02_9.01*) has both duplication regions, and after the first round of repair at megachunk *L*, the resulting chromosome *synII* (yeast_*chr02_9.02*) has only one duplication. The finished chromosome *synII* (yeast_*chr02_9.03*) was obtained after the second repair at the megachunk *T* region.

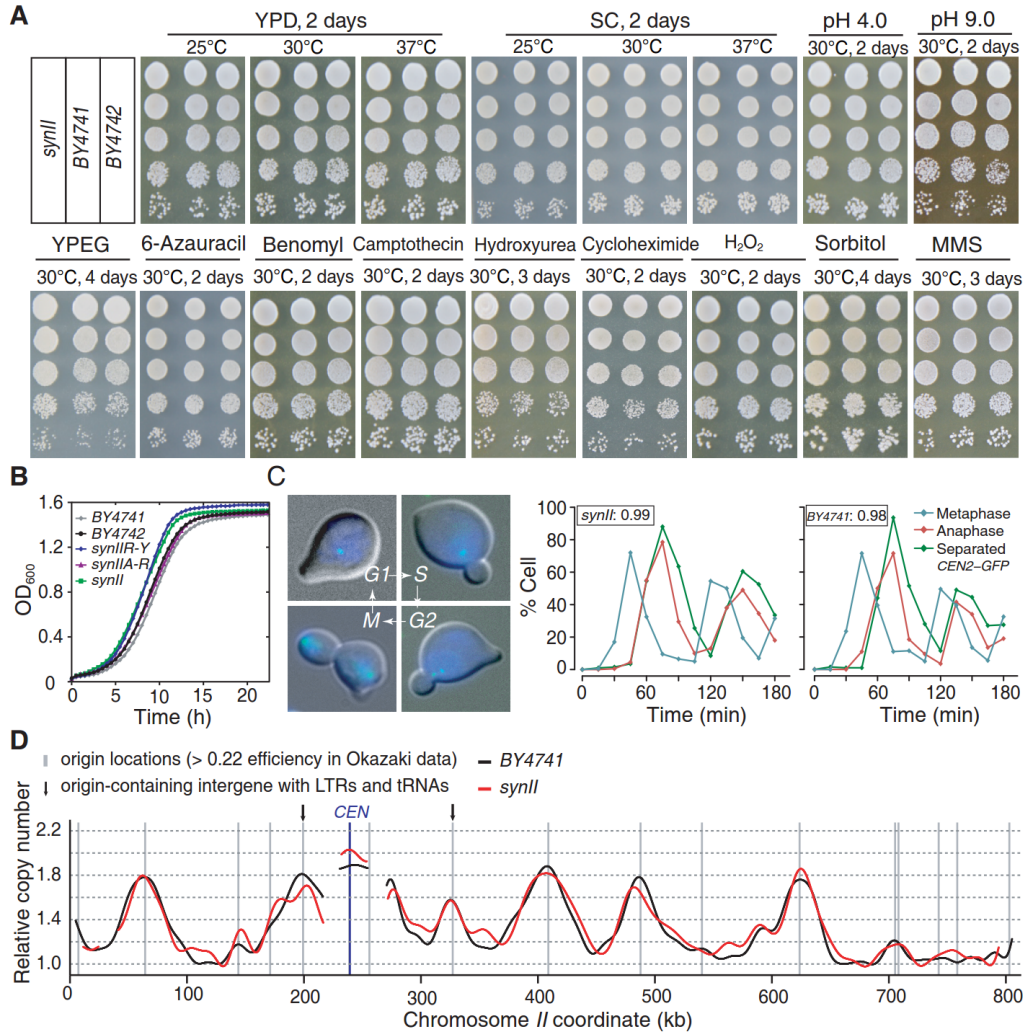


Fig. 2. Phenotypic profiling of *synII*.

(A) Phenotype tests of *synII* on different media. 10-fold serial dilutions of overnight cultures of *synII* and wild-type (*BY4741* and *BY4742*) strains were used for plating. From left to right: YPD at 25°C, 30°C, and 37°C; SC at 25°C, 30°C, and 37°C; low pH YPD (pH 4.0) and high pH YPD (pH 9.0); YPEG; SC+6-Azauracil; YPD+Benomyl; YPD+Camptothecin; YPD+Hydroxyurea; YPD+Cycloheximide (10 µg/ml, 2 hrs pretreatment); YPD+H₂O₂ (1 mM, 2 hrs pretreatment); YPD+Sorbitol; YPD+MMS, (YPD, yeast extract peptone dextrose; YPEG, yeast extract peptone glycerol ethanol; MMS, methyl methane sulfone; SC, synthetic complete).

(B) Growth curves of *synIIA-R*, *synIIR-Y* and *synII* strains compared with those of *BY4741* and *BY4742* strains in YPD at 30°C.

(C) Cell cycle comparison between *synII* and *BY4741*. Images show cell morphology at different stages during the cell cycle after release from *G1* block. DNA staining is shown in blue; CEN2-GFP in *G1*, *S*, *G2* and *M* phase are shown in green. Graphs show the percentage of *synII* cells with separated CEN2-GFP dots, metaphase spindles and anaphase spindles during the cell cycle. For each time point at least 200 cells were counted. The inset numbers of 0.99

indicate the overall ratio of metaphase to anaphase cells throughout the time course for *synII* and *BY4741* strains.

(D) *synII* (red) and *BY4741* (black) replication time expressed as relative copy number by deep sequencing.

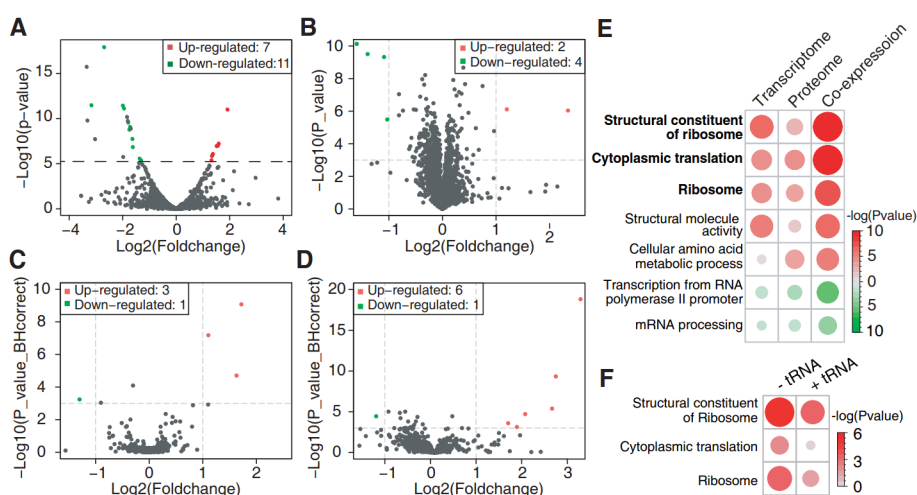


Fig. 3. *synII* strain Trans-Omics profile (*BY4741* as reference) demonstrates that the synthetic chromosome design has minimal impact on cell physiology, despite a modest up-regulation of translational machinery triggered by tRNAs removal.

(A-D) Identified dysregulated genetic features at (A) transcriptome level, (B) proteome level and (C and D) metabolome level (Metabolic and lipid profiling in LC-MS positive mode respectively) of *synII* cells, compared to *BY4741* cells. The detailed methods of these comparisons are described in the materials and methods section of the SOM. The total number of differentially expressed ($p\text{-value} < 0.001$) features in transcriptome, proteome and metabolome are also presented as well. Up-regulated and down-regulated features are labeled in red and green respectively.

(E) Enriched pathways and the co-expression profile revealed by transcriptome and proteome in yeast GO terms. Up-regulated features are labeled in red and down-regulated features are labeled in green.

(F) RNAseq analysis of *synII* with/without tRNA array. By adding back the tRNA array of *synII*, the up-regulation of translational functions is greatly mitigated. For E and F, significance level is indicated by heatmap color intensities and symbol sizes. Up-regulated and down-regulated features are labeled in red and green respectively.