Manipulating the 3D organization of the largest synthetic yeast chromosome

Weimin Zhang[1], Luciana Lazar-Stefanita[1†], Hitoyoshi Yamashita[3†], Michael J. Shen[1†], Leslie A. Mitchell[1], Hikaru Kurasawa[3], Max A. B. Haase[1], Xiaoji Sun[1], Qingwen Jiang[4], Stephanie L. Lauer[1], Laura H. McCulloch[1], Yu Zhao[1], David Ichikawa[1], Nicole Easo[1], S. Jiaming Lin[1], Viola Fanfani[5], Brendan R. Camellato[1], Yinan Zhu[1], Jitong Cai[6], Zhuwei Xu[1], Maya Sacasa[1], Ryan Accardo[7#], Ju Young Ahn[8#], Surekha Annadanam[8#], Leighanne A. Brammer Basta[7#], Nicholas R. Bello[6#], Lousanna Cai[6#], Stephanie Cerritos[7#], MacIntosh Cornwell[6#], Anthony D'Amato[7#], Maria Hacker[7#], Kenneth Hersey[7#], Emma Kennedy[7#], Ardeshir Kianercy[9#], Dohee Kim[6#], Hong Seo Lim[6#], Griffin McCutcheon[10#], Kimiko McGirr[8#], Nora Meaney[7#], Lauren Meyer[8#], Ally Moyer[6#], Maisa Nimer[8#], Carla Sabbatini[7#], Lisa Scheifele[7#], Lucas S. Shores[6#], Cassandra Silvestrone[7#], Arden Snee[7#], Antonio Spina[6#], Anthony Staiti[7#], Matt Stuver[7#], Elli Tian[6#], Danielle Whearty[7#], Calvin Zhao[6#], Tony Zheng[6#], Vivian Zhou[6#], Karen Zeller[7], Joel S. Bader[6], Giovanni Stracquadanio[5], Samuel Deutsch[11], Junbiao Dai[4*], Yasunori Aizawa[3*], Jef D. Boeke[1,2*]

[1] Institute for Systems Genetics and Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York, NY, USA

[2] Department of Biomedical Engineering, NYU Tandon School of Engineering, Brooklyn, NY, USA

[3] School of Life Science and Technology, Tokyo Institute of Technology, Yokohama, Japan

[4] CAS Key Laboratory of Quantitative Engineering Biology, Guangdong Provincial Key Laboratory of Synthetic Genomics and Shenzhen Key Laboratory of Synthetic Genomics, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

[5] School of Biological Science, The University of Edinburgh, Edinburgh, UK

[6] Department of Biomedical Engineering, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, USA

[7] Department of Biology, Loyola University Maryland, 4501 N Charles St, Baltimore, MD, USA

[8] Department of Biology, Krieger School of Arts and Sciences, Johns Hopkins University, Baltimore, MD, USA

[9] Brady Urological Institute, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[10] Department of Chemical and Biomolecular Engineering, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, USA

[11] DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

* Correspondence to: junbiao.dai@siat.ac.cn, yaizawa@bio.titech.ac.jp, jef.boeke@nyulangone.org

[†]Equal contributions
[#]Build a Genome class

Current address:

JYA: Texas A&M University, College Station, TX, USA

RA: New York-Presbyterian Hospital, New York, NY, USA

SA: Department of Internal Medicine Resident, University of Michigan, Ann Arbor, MI, USA

LABB: Chemistry Department, US Naval Academy, Annapolis, MD, USA

NRB: Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA, USA

LC: Facebook, Greater Seattle Area, WA, USA

MC: Institute for Systems Genetics, NYU Langone Health, New York, NY, USA

SC: Maryland Department of Health, Baltimore, MD, USA

AD: Xaverian Brothers High School, Westwood, MA, USA

SD: Nutcracker Therapeutics, San Francisco, CA, USA

VF: Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

MH: Branford Dental Care, Connecticut, Branford, CT, USA

KH: Capelli Sport, Mülheim an der Ruhr, Germany

DI: Pandemic Response Lab, Long Island City, NY, USA

QJ: Molecular Pharmacology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA

AK: Arkansas College of Osteopathic Medicine, Fort Smith, AK, USA

DK: Otorhinology, Head and neck surgery resident, Seoul national university hospital, Seoul, Republic of Korea

EK: Eli Lilly and Company, San Diego, CA, USA

HK: Kanagawa Institute of Industrial Science and Technology, Ebina, Kanagawa, Japan.

HSL: Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

AM: Tri-Institutional MD–PhD Program, New York, NY, USA

GM: Boston University, Boston, MA, USA

KM: Applied BioMath, LLC, Concord, MA, USA

LM: Vilcek Institute, NYU School of Medicine, New York, NY, USA

LAM: Neochromosome, Inc., Long Island City, NY, USA

NM: Memorial Sloan Kettering Cancer Center, New York, NY, USA

MN: Department of Surgery, UT southwestern, Dallas, TX, USA

Anthony S: Northern Neck Orchestra Foundation, Kilmarnock, VA, USA

Antonio S: Boston Consulting Group, Boston, MA, USA

Arden S: Appalachian State University, Boone, NC, USA

Carla S: NYIT College of Osteopathic Medicine, New York, NY, USA

Cassandra S: New York-Presbyterian Hospital, New York, NY, USA

LSS: University of Chicago, Chicago, IL, USA

Matt S: Abbott Laboratories, Chicago, IL, USA

ET: University of Nevada, Las Vegas School of Medicine, Las Vegas, NV, USA

DW: MedStar Union Memorial Hospital, Baltimore, MD, USA

ZX: NICHD, NIH, Bethesda, MD, USA

HY: Chitose Laboratory Corp., Kawasaki, Kanagawa, Japan.

CZ: NYU Grossman School of Medicine, New York, NY, USA

KZ: Department of Biology, Beverly K. Fine School of the Sciences, Stevenson University, Baltimore, MD, USA

TZ: Oregon Health & Science University, Portland, OR, USA

VZ: Salesforce, Greater Seattle Area, WA, USA

Summary

Whether synthetic genomes can power life has attracted broad interest in the synthetic biology field, especially when the synthetic genomes are extensively modified with thousands of designer features. Here we report *de novo* synthesis of the largest eukaryotic chromosome thus far, *synIV*, a 1,454,621-bp *Saccharomyces cerevisiae* chromosome resulting from extensive genome streamlining and modification. During the construction of *synIV*, we developed a megachunk assembly method, combined with a hierarchical integration strategy. This strategy significantly increased the accuracy and flexibility of synthetic chromosome construction and facilitated chromosome debugging. In addition to the drastic sequence changes made to *synIV* by rewriting it, we further manipulated the three-dimensional structure of *synIV* in the yeast nucleus to explore spatial gene regulation within the nuclear space. Surprisingly, we found few gene expression changes, suggesting that positioning inside the yeast nucleoplasm plays a minor role in gene regulation. Therefore, our manipulation of the spatial structure of the largest synthetic yeast chromosome shed light on higher-order architectural design of the synthetic genomes.

Keywords: *Saccharomyces cerevisiae*, *synIV,* megachunk assembly, chromosome debugging, three-dimensional structure, inside out chromosome, nuclear organization, transcriptomics.

Introduction

Reprogramming an entire native genome facilitates the understanding of many fundamental questions regarding genome content essentiality, genome regulation and genome evolution (Gibson et al., 2008, 2010) while also increasing its versatility by incorporating new design features (Fredens et al., 2019; Ostrov et al., 2016; Richardson et al., 2017). Although our ability to synthesize genomes has increased dramatically in the past two decades (Zhang et al., 2020), it is still challenging to accurately and efficiently build megabase-sized DNA segments in a bottom-up fashion, hindering studies aiming to probe the functional basis of genomes from a synthetic biology perspective. *Saccharomyces cerevisia*e was the first eukaryotic organism to have its genome sequenced, which profoundly facilitated genetic studies of yeast. Further, the Synthetic Yeast Genome Project (Sc2.0) represents the first and largest eukaryotic synthetic genome to be built. The project aims to drastically alter the yeast genome and assesses whether the native yeast genome can be reprogrammed by removing retrotransposons and other repetitive sequences in order to increase genome stability, whether splicing systems can be eliminated in an intron-less yeast genome and whether synthetic yeast can gain new properties when the entire gene content is “shuffled” randomly. Synthesis of the individual synthetic chromosomes have answered many fundamental questions (Annaluru et al., 2014; Dymond et al., 2011; Mitchell et al., 2017; Shen et al., 2017; Wu et al., 2017; Xie et al., 2017; Zhang et al., 2017), yet much more remains open as Sc2.0 is driven to completion.

In the process of building Sc2.0, the focus thus far has been on sequences of the chromosomes, and less so on how synthetic chromosomes are spatially organized in a living cell. Numerous studies have shown that genome misfolding and dysregulation are strongly associated with human disease (Fudenberg et al., 2011; Hnisz et al., 2016; Lupiáñez et al., 2015), suggesting fundamentally important roles of large scale chromatin 3D organization. It is crucial to have a thorough understanding of how synthetic chromosomes are organized in the nucleus so that we have a better control of chromosome-wide gene regulation. In yeast, chromosomes are organized in a relatively conserved structure called the Rabl orientation (Rabl, 1885)–the spindle pole body (SPB) resides on the opposite side of the nucleus relative to the nucleolus and tethers the 16 centromeres throughout the cell cycle. The 32 telomeres are clustered to form three to eight foci anchored on the inner nuclear envelope (Palladino et al., 1993; Taddei and Gasser, 2012). Such nuclear organization forms different sub-nuclear compartments that influence genome-wide gene expression. For example, the peripheral domain of the nuclear pore complex corresponds to an “active expression” compartment, whereas telomere clusters on the inner nuclear membrane correspond to a repressive compartment. The nucleolus is paradoxical due to its nuclear apposition, where silencing is seen of RNAP2 reporter genes, but ribosomal DNA itself is actively transcribed by RNAP1 and RNAP3. However, to our knowledge, current findings are limited to compartments close to the nuclear envelope. It is still unclear whether the yeast nucleoplasm contains multiple, elaborately structured sub-nuclear compartments.

Here, we describe the design, construction and characterization of a 1,454,641 bp yeast chromosome, *synIV*, the largest synthetic eukaryotic chromosome reported. Despite the incorporation of thousands of designer features, *synIV* is still able to provide near wild-type

fitness to its host strain under various growth conditions. *SynIV* shows remarkably similar 3D structure to wild-type IV, and thanks to its lack of repetitive DNA, smoother intra-chromosomal contact maps are produced by chromosome conformation capture assay. Lastly, we redefined the 3D structure of *synIV* in the nucleus by "intramolecular centromere relocation." The resulting *synIV* is flipped in arm orientation relative to the other 15 wild-type chromosomes, giving rise to the reorientation and repositioning of all 796 genes carried on *synIV* in the nucleus, and forcing centromere proximal sequences to telomere proximal and vice versa – producing an "inside out" chromosome. By manipulating the 3D organization of *synIV*, we observed surprisingly minor gene expression changes relative to the original *synIV*, indicating that *synIV* organization is highly plastic and suggesting that there are no functional subnuclear compartments that influence gene expression in the interior space of the yeast nucleoplasm. Our work pushes the upper size boundary of synthetic eukaryotic chromosome construction and provides the first designer 3D structure of a synthetic chromosome in yeast.

Results

Hierarchical assembly of *synIV*

Chromosome IV of *Saccharomyces cerevisiae* encodes the highest number of genes among the 16 chromosomes. Following the design pipeline of the Sc2.0 project (Richardson et al., 2017), the synthetic chromosome IV (*synIV*) becomes the largest eukaryotic chromosome to be built thus far. A total of 479 loxPsym sites were inserted downstream of non-essential genes and at the positions of deleted tRNA genes, 23 introns and 28 tRNA genes were removed, 183 TAG stop codons were swapped to TAA and 989 pairs of PCRTags were generated to serve as synthetic "watermarks" (Figure 1A) (Annaluru et al., 2014; Dymond et al., 2011; Mitchell et al., 2017; Shen et al., 2017; Wu et al., 2017; Xie et al., 2017; Zhang et al., 2017). To minimize the possibility of genome instability, subtelomeric repeats and retrotransposons were removed, and seven open reading frames (ORFs) containing tandemly repeated regions (Verstrepen et al., 2005) were synonymously recoded (repeat smashed ORFs, Table S1).

The 1,454,641 bp *in silico* designed *synIV* was segmented into 36 megachunks of approximately 60 kb in length (named megachunks A through JJ), which were further subdivided into 646 minichunks of about 3 kb each. Minichunks were assembled from oligonucleotides by DNA synthesis vendors. Adjacent minichunks share 100 bp of identity, allowing assembly of three to four minichunks into 8-10 kb DNA chunks (Figure S1A). Each chunk was flanked by two rare cutting restriction enzyme (RE) sites previously designed into ORFs without changing the encoded protein sequences (Table S2). Yeast homologous recombination or Gibson assembly were used to assemble the chunks (Figure 1B) (Annaluru et al., 2014; Gibson et al., 2009). Initially, five or six chunks were ligated *in vitro* to form a megachunk, and used directly for SwAP-In (Switching Auxotrophies Progressively for Integration), a methodology used to "overwrite" each wild-type chromosome segment with its synthetic counterpart (Richardson et al., 2017). We noted that the SwAP-In success rate varied considerably among different megachunk integrations due to inconsistent ligation efficiency/yield of full length megachunk products, and furthermore, the restriction fragment

end ligation accuracy was never validated prior to integration, leading to a higher than average mutation rate within the RE site junctions.

In order to better standardize the synthetic DNA materials used for SwAP-In, we attempted to assemble megachunks as cloned building blocks for this chromosome (Figure S1B, Table S3). The megachunk assembly approach produced a significantly higher integration efficiency when compared to the previous method (Figure 1C) and, since all megachunk clones were sequenced prior to use, this improved methodology eliminated the problem of errors arising from RE end misligation. To accelerate construction of *synIV*, we performed SwAP-In using 11 entry strains in parallel, among which any two adjacent entry strains have the opposite mating type and different auxotrophic markers. Ten of the 11 semi-synthetic strains have only four megachunks, while the megachunk J-O strain contains six megachunks, and each pair of adjacent strains shares a single overlapping megachunk. This design strategy facilitates Meiotic Recombination-mediated Assembly (MRA) (Figure 1D) (Zhang et al., 2017). After completing the integration of 11 semi-synthetic strains, we performed fitness characterization under various growth conditions (Figures S2A-B). We found that strains with G-J megachunks had a severe growth defect in rich medium (YPD) at high temperature; those with J-O megachunks grew poorly under all conditions we tested; while the strain with U-X megachunks showed low fitness in liquid cultures (Figure S2B). The unhealthy strains were subjected to debugging (see section below “Debugging of *synIV*”) until fitness met a near wild-type criterion. Karyotyping analysis of all semi-*synIV* chromosomes revealed no large-scale genomic rearrangement was created during the construction process (Figure S2C). These data suggest that megachunk cloning combined with hierarchical assembly strategy significantly increased the accuracy and flexibility of *synIV* construction.

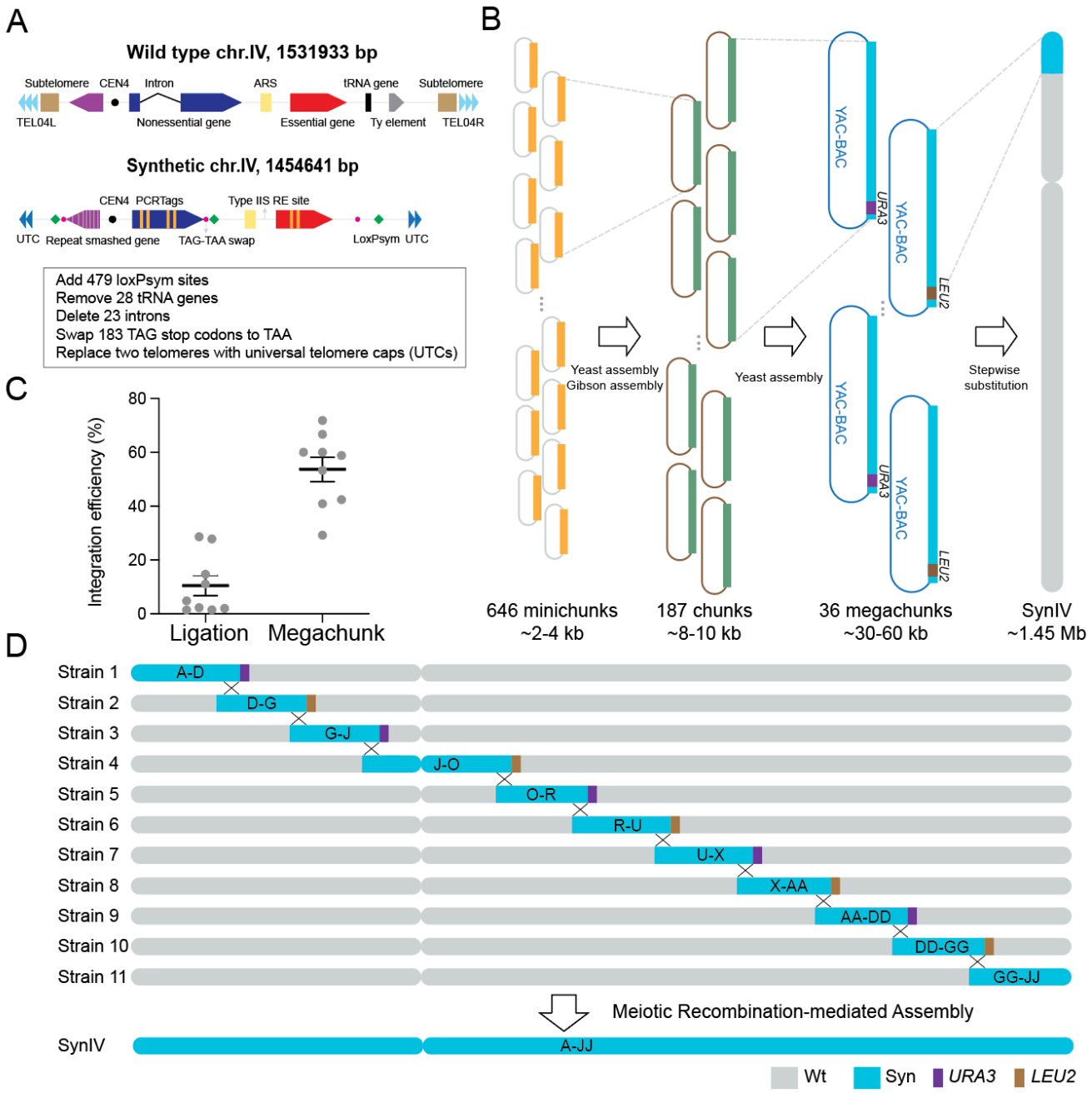

Fig. 1. Design and assembly of *synIV*. (A) Schematic illustration of the design features of *synIV*, including subtelomeric deletion, REPEATSMASHER recoding, TAG to TAA stop codon swaps, PCRTag recoding, intron deletion, tRNA gene deletion and loxPsym site insertion. *CEN*, centromere; RE, restriction enzyme; UTC, universal telomere cap; ARS, autonomously replicating sequence; *TEL*, telomere. Genes are categorized by different colors for nonessential gene (dark blue), repeat smashed gene (purple), and essential gene (red). (B) *SynIV* hierarchical assembly workflow. Yellow bars are ~2-4 kb minichunk DNAs, green bars are ~8-10 kb chunk DNAs, blue bars are ~30-60 kb megachunk DNAs. The DNA assembly methods are indicated below the rightward arrow. YAC-BAC, yeast artificial chromosome-bacterial artificial chromosome shuttle vector. (C) Integration efficiency comparison between conventional ligation method and the new megachunk assembly method (n=9 for each method). Bars are mean±SEM. (D) Full length *synIV* was split into 11 intermediate strains for parallelization. Semi-*synIV* segments from different intermediate strains were consolidated via Meiotic Recombination-mediated Assembly.

SynIV debugging

Characterization of semi-*synIV* strains revealed that strains containing megachunks G-J, J-O and U-X were defective in growth under at least one tested condition (Figures S2A-B). As the megachunks were incorporated in consecutive order, we characterized all strains generated at each SwAP-In step, searching for the defect(s) caused by megachunk insertion in each unhealthy strain. Eventually we narrowed down the bugs to megachunks J, N, O and V and classified them according to three types of alterations: repeat-smashed ORFs, PCRTag recoding, and unexpected mutations (Figure 2A).

Two *synIV* bugs resulted from use of a piece of code, REPEATSMASHER (Richardson et al., 2017), designed to recode a set of a few dozen protein coding genes containing tandem repeat sequences that create challenges for assembly (Table S1). This early version of REPEATSMASHER recoded the entire ORF rather than only the repetitive segment, resulting in a pervasively recoded ORF. For example, the “repeat-smashed” essential gene, *NOP1*, was found to cause a severe growth defect, especially when grown at 37°C on YPD (Figure 2B). This growth defect was fully reversed when repeat-smashed *NOP1* was replaced with wild-type *NOP1* (Figure 2B). We examined *NOP1* mRNA levels in repeat-smashed and wild-type strains and found that mRNA levels were not significantly different at both 30°C and 37°C YPD conditions (Figure 2C). However, the protein level of repeat-smashed Nop1p was substantially decreased at 37°C, and slightly decreased at 30°C (Figure 2D). It is worth noting that the two *NOP1* genes encode identical amino acid sequences even though they only share 74% identity at the DNA sequence level (Figure S3A). Importantly, repeat-smashed *NOP1* contains more non-optimal (based on codon usage frequency) codons for *S. cerevisiae* (Figure S3B), which may interfere with mRNA stability, translation initiation/elongation, and/or protein folding (Chu et al., 2014; Hanson and Coller, 2018; Presnyak et al., 2015). Thus, we speculate that the fitness defect is probably caused by altered translation of pervasively recoded *NOP1*. Similarly, the incorporation of *PCF11*, a repeat-smashed essential gene in megachunk V, caused growth defects under multiple conditions, which were rescued by reverting the repeat-smashed *PCF11* to the wild type (Figure S3C).

A PCRTag bug was discovered when megachunk O was integrated. The growth defect phenotype was especially notable when the strain was grown at low temperature (22°C) (Figure S4A). To narrow in on the bug, we integrated each chunk of megachunk O individually to a wild-type strain. We found that introducing chunk O4 alone to was sufficient to produce a similar fitness defect to that seen following whole megachunk O integration. Further experiments mapped the bug to the sequence of the *RRP8* gene (Figure S4B). The only unique feature of synthetic *RRP8* is a pair of PCRTags, and we eventually pinpointed the causative bug to the forward synthetic PCRTag, a 28-bp sequence in *RRP8* (Figure 2E). The incorporation of this 28-bp sequence did not compromise mRNA expression of *RRP8* (Figure 2F). mRNA secondary structure prediction indicates the PCRTag recoding produces a strong stem loop structure that may interfere with the translation of Rrp8p (Figure S4C), similar to the situation described for the *PRE4* gene on *synVI* (Mitchell et al., 2017).

Some bugs were caused by unexpected (i.e. non-designed) mutations in the synthetic DNA (Figure 2A). In the case of the *MAK21* gene, we found a single unplanned base-pair substitution, lying 186 bp upstream of *MAK21* in its promoter region (Figure 2G). This mutation resulted in a >50% reduction of *MAK21* mRNA expression (Figure 2H). The restoration of the point mutation rescued the growth defects (Figure 2I). We subsequently searched for *S. cerevisiae* transcription factor (TF) binding motif consensus sequences in *MAK21*'s promoter using the Yeastract+ tool (Monteiro et al., 2020). However, we failed to identify any known TFs that bind this -186 bp motif, indicating that another functional element may be disrupted by this point mutation.

After bug-fixing, the 11 healthy semi-*synIV* strains were consolidated via multiple rounds of MRA (Figure 3A). The MRA combined synthetic chromosome segments efficiently until we approached the last step. We found an *E. coli* IS1 transposon had inserted into the *RTT103* gene (Figure S6A), which did not cause any growth defect in strain yWZ462 (containing megachunk A-M and X-JJ) (Figure S6B), but led to severe growth defects on SC medium in strains containing almost all the megachunks of *synIV* (Figure S6C, top3 rows). We fixed this "combinatorial" bug by removing the IS1 sequence from the *RTT103* gene, producing a healthy *synIV* strain (Figure S6D).

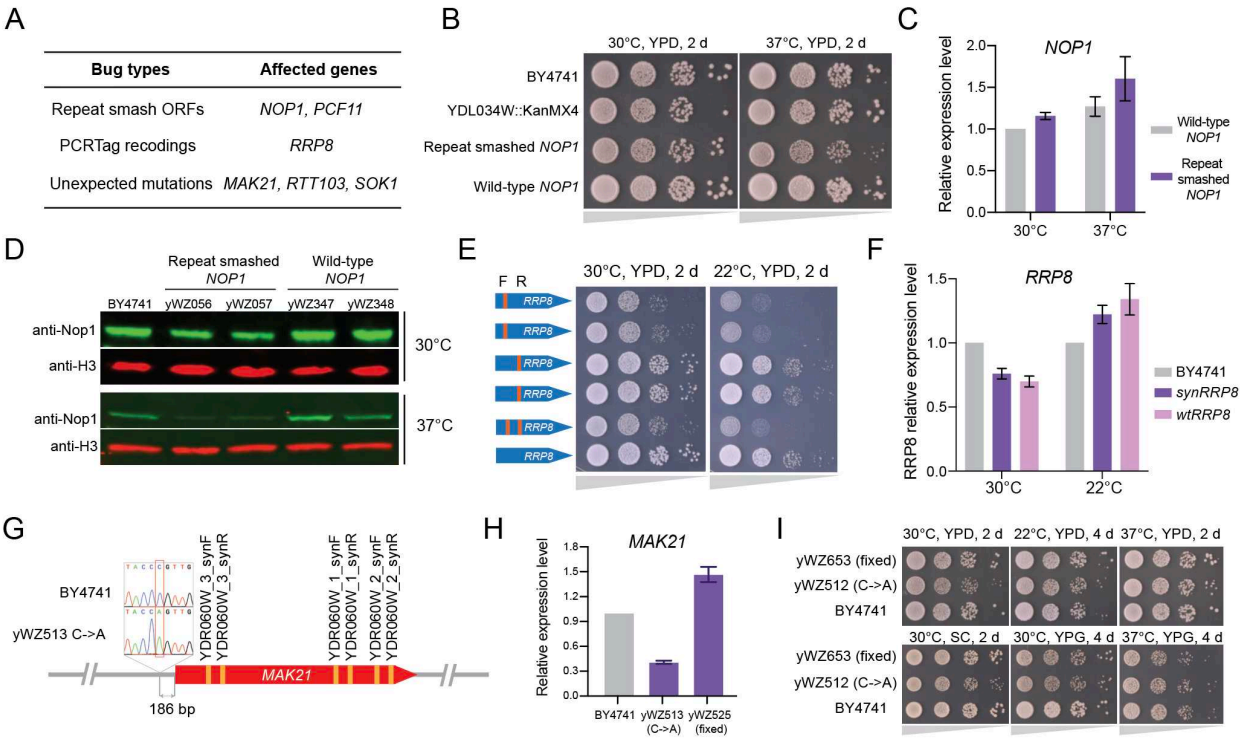

Fig. 2. *SynIV* debugging. (A) Summary of bugs and genes affected in *synIV*. (B) Serial dilution spot assay of strains with different versions of *NOP1*. BY4741 and *YDL034W* knock-out (entry strain of megachunk J) strains serve as controls. (C) RT-qPCR analysis of wild-type and repeat smashed *NOP1* mRNA expression at 30°C and 37°C. Bars represent mean ± SD of three technical replicates. (D) Immunoblotting analysis of Nop1p; two independent clones were used for each strain. (E) Serial dilution assay shows fitness of strains with *RRP8* with only forward synthetic PCRTag, only reverse synthetic PCRTag or both forward and reverse PCRTags. F, forward synthetic PCRTag; R, reverse synthetic PCRTag. (F) Relative expression of synthetic

and wild-type *RRP8* at 30°C and 22°C on YPD. Bars represent mean ± SD of three technical replicates. (G) Illustration of synthetic *MAK21* in *synIV*, a cytosine to adenine mutation is highlighted in the promoter region of *MAK21.* Synthetic PCRTags are indicated as orange. (H) Relative expression of *MAK21* with or without intergenic C>A mutation. Bars represent the mean ± SD of three technical replicates. (I) Serial dilution assay shows growth fitness of strain with or without intergenic C>A mutation.

Sequence and structure of *synIV*

Meiotic recombination mediated assembly created a variety of intermediate *synIVs* that contained interspersed wild-type patches. Eventually, those intermediate *synIVs* were intercrossed to produce a full length *synIV* draft strain, yeast_chr04_9_01. A PCRTagging assay showed that the wild-type chromosome IV was completely replaced by *synIV* (Figure 3B). Whole genome sequencing of *synIV* revealed two structural variants: *ENA5-ENA2-ENA1* duplication and *HXT7-HXT6* duplication (Figure S7). The *ENA* genes are more than 90% identical to each other and are arranged in tandem in the genome, providing a favorable configuration for duplication in yeast. We successfully removed the duplicated *ENA* genes by first inserting a *URA3* gene near the duplicated region, and then replacing the duplicated region as well as the *URA3* gene with the correct DNA chunk. The *HXT7-HXT6* duplication is a well-known genome rearrangement event previously seen during evolution experiments under glucose limitation (Brown et al., 1998), suggesting that the *synIV* strain may have experienced low glucose stress during construction. Besides these structural variants, 12 TAG stop codons were incorporated erroneously due to bookkeeping errors. We employed SpCas9-NG, a SpCas9 variant that recognizes the NG protospacer adjacent motif (Nishimasu et al., 2018) to precisely swap TAG stop codons to TAA. The SpCas9-NG achieved a high editing efficiency for 10 out of 12 TAG stop codons when swapping to TAA (Figure S8). We then swapped the two remaining TAGs by a *URA3* in-and-out strategy, similar to the one described for the *ENA* genes. A detailed table summarizes the correction steps made to each version of *synIV* (Table S4). Whole genome sequencing revealed that the final *synIV* strain has several variants from the original design of *synIV*, all of which are listed in Table S5.

The fitness of yWZ703 (the final synIV strain) is near wild type under various conditions (Figure 3C), indicating that the *in silico* designed *synIV* is capable of powering a yeast cell to near normal growth rates. It is well documented that the yeast genome is well-organized within the three-dimensional nuclear space (Duan et al., 2010). Consequently, we wondered whether the drastic changes we made by re-writing chromosome IV could affect the 3D structure of *synIV*. We performed Hi-C on both wild-type and *synIV* strains, and as expected due to the lack of Ty element repeats, we found that the intra-chromosomal contact map of *synIV* was almost completely free of unmappable regions, which appear as white stripes on the wild-type *chrIV* map (Figure 3D, S9). Notably, in *synIV* we found increased interactions between *Ty*-flanking sequences reflecting the fact that the intervening *Ty* elements were removed during synthesis. Furthermore, the 3D projections of the Hi-C maps support that both *synIV* and wild-type IV chromosomes are organized similarly within the intranuclear space (Figure 3E, supplementary movies), suggesting that Sc2.0 modifications have only minor effects on 3D structure of synthetic DNA molecules, but increase the contact map mappability of Hi-C, consistent with previous findings (Mercy et al., 2017).

Transcript profiling revealed a small number of genes that were differentially expressed in the *synIV* strain (Figure 3F). These encode uncharacterized open reading frames (ORFs) according to the *Saccharomyces* gene database (SGD). Briefly, for upregulated genes in *synIV*, *YDL007C-A*'s promoter is adjacent to a serine tRNA gene, and *YDR278C* overlaps with a glutamate tRNA gene in wild-type IV. It is well known that tRNAs silence nearby polymerase II transcription (Bolton and Boeke, 2003; Hull et al., 1994), so we speculate that the upregulation of *YDL007C-A* and *YDR278C* results from removal of the flanking tRNA genes. For down-regulated genes in *synIV*, *YDR541C* is relocated adjacent to the new universal telomere cap (UTC) after deleting the subtelomeric region in *synIV*, so the down-regulation of *YDR541C* is likely due to telomere silencing.

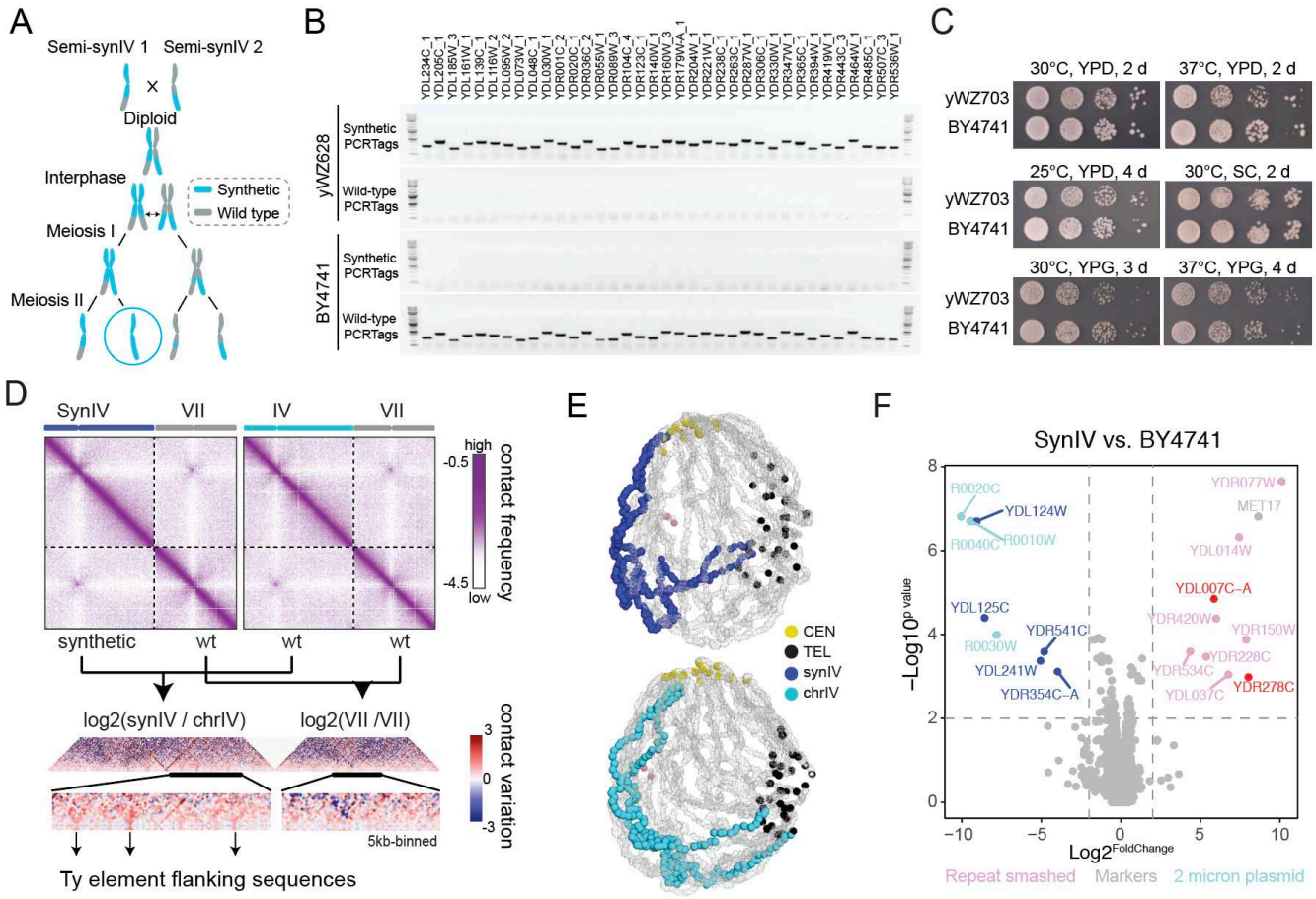

Fig. 3. *SynIV* characterization. (A) Schematic illustration of Meiotic Recombination-mediated Assembly. (B) PCRTagging analysis of *synIV* (yWZ628) and wild-type (BY4741) strains. One PCRTag per megachunk was used for representation. Amplicons were loaded in 1.5% agarose gel for electrophoresis. (C) *SynIV* and BY4741 fitness comparison by ten-fold serial dilution (spot) assays under six growth conditions. YPD, yeast extract, peptone, dextrose. YPG, yeast extract, glycerol. SC, synthetic complete. (D) Hi-C contact maps of yeast with synthetic or wild-type chromosome IV. Chromosome VII is shown as an internal control. Chromosomes are annotated atop the maps in blue, cyan, and gray, respectively. Left and right maps (5 kb-binned) were generated by aligning Hi-C reads on a reference sequence containing the designed *synIV* chromosome. Violet to white color scale reflects high to low contact frequency (log10). Contact variations in *synIV* vs. wild-type IV are shown in the bottom panels where blue to red color scale reflects contact enrichment in *synIV* (log2-ratio maps 5 kb bin). (E) 3D average representations of the Hi-C contact maps in

(D), color code highlights *synIV* and wild-type IV, in addition to centromeres, telomeres and rDNA. (F) Volcano plot of *synIV* vs. BY4741. Genes are colored red for upregulated and blue down regulated in *synIV*, when compared to BY4741. Auxotrophic gene *MET17* was present in *synIV* but not in BY4741, whereas *YDL124W* and *YDL125C* were deleted in this yeast_chr04_9_04 *synIV* strain, but were restored in subsequent strain versions. Subtelomeric *YDL241W* was removed by design *synIV*. Gray colored genes indicate genetic markers, pink genes that were repeat smashed, and light blue were 2-micron plasmid genes. Fold change cutoff is 4, p-value cutoff is 0.01.

"Inside out" *synIV* revealed limited gene expression changes

Chromosome organization influences gene regulation. Next, we asked whether three-dimensional structure of *synIV* can be artificially and dramatically restructured in a living yeast cell. As a proof of concept, we wanted to flip the linear orientation of chromosome arms, we chose to relocate the most structurally impactful sequence, *CEN4*, to a context as far-removed from its native location as possible, namely the telomere-adjacent sequence, thus moving the telomere-adjacent sequence towards the centromere. The goal was to create an "inside out" *synIV* chromosome, assuming that the newly relocated *CEN4* remains connected to the spindle pole body. This move places the former centromere proximal sequences adjacent to the (former) subtelomeres, and the former telomere-proximal sequences adjacent to the centromere. We refer to this as an "inside out" (IO) chromosome.

To construct the inside out *synIV* strain, we first circularized the original "rightside out" (RO) *synIV* by providing a *LEU2* marker with flanking homology arms. The left homology arm maps to the last chunk of *synIV* (chunk JJ5) whereas the right homology arm maps to the first chunk of *synIV* (chunk A1). DNA double-strand breaks were introduced by CRISPR-Cas9 near both UTCs, resulting in their excision and the simultaneous bridging of the sequence by the *LEU2* donor DNA (Figure 4A, Figure S10A). Next, we performed intramolecular centromere relocation by simultaneously swapping *LEU2* to *CEN4* and replacing *CEN4* with the telomerator, a synthetic cassette designed to convert circular DNA to linear DNA *in vivo* (Mitchell and Boeke, 2014) (Figures S10B-C). We expected that the relocated *CEN4* sequence would retain contact with the SPB, forcing the original telomere-proximal sequence close to the SPB and flipping the orientation of both *synIV* chromosome arms. Lastly, we transiently expressed I-SceI endonuclease in the IO circle *synIV* strain to linearize the IO circle at the telomerator (Figure 4A, Figure S10D). Hi-C maps were generated to validate these extraordinary structural reorganizations of *synIV*. Telomere fusion resulted in the formation of a circular RO chromosome, whose formerly peri-telomeric ends gave rise to strong intra-chromosomal contacts, while their interactions with the remaining active telomeres were lost (Figure 4B, first panel). After flipping the entire circular *synIV* inside out, we observed that the ectopic relocation of *CEN4* within the former peri-telomeric region was sufficient to convert this latter region into a neo-pericentromere (Figure 4B, second panel). Finally, the re-linearization of the inside out circle *synIV* at the former peri-centromere gave rise to an inside out linear chromosome carrying both ectopic *CEN* and *TEL* sequences that appear structurally functional (Figure 4B, third panel). Pulsed field gel electrophoresis was performed for the two linear *synIV* strains (RO and IO) and two circular *synIV* strains (RO circle and IO circle). Both RO and IO *synIV*s are slightly smaller than wild-type IV in size, as we deleted 77.3 kb in *synIV* compared

to the wild type IV. Both RO circle and IO circle *synIV*s are trapped in the agarose plug, as circular DNAs are unable to migrate out of the plugs through PFGE (Lartigue et al., 2007) (Figure 4C).

Lastly, we asked whether the chromosome-wide gene repositioning of IO *synIV* caused gene expression changes. We found that only *YDR540C* and *YDR541C* were upregulated when we compared the IO *synIV* transcription profile to that of RO *synIV* (Figure 4D, left). These two genes were immediately adjacent to RO's right telomere, and were relocated to the centromere cluster near the SPB in IO *synIV*, indicating that relocation of *CEN4* brings these two genes from a telomeric repression compartment to a more active compartment. Although other genes on IO *synIV* were all relocated in space, we did not detect any significant gene expression changes. A similar result was obtained when we compared the two circular *synIV* strains (Figure 4D, right). Our results suggest that relative gene positions in the interior space of the nucleus have only minor effects on gene expression regulation.

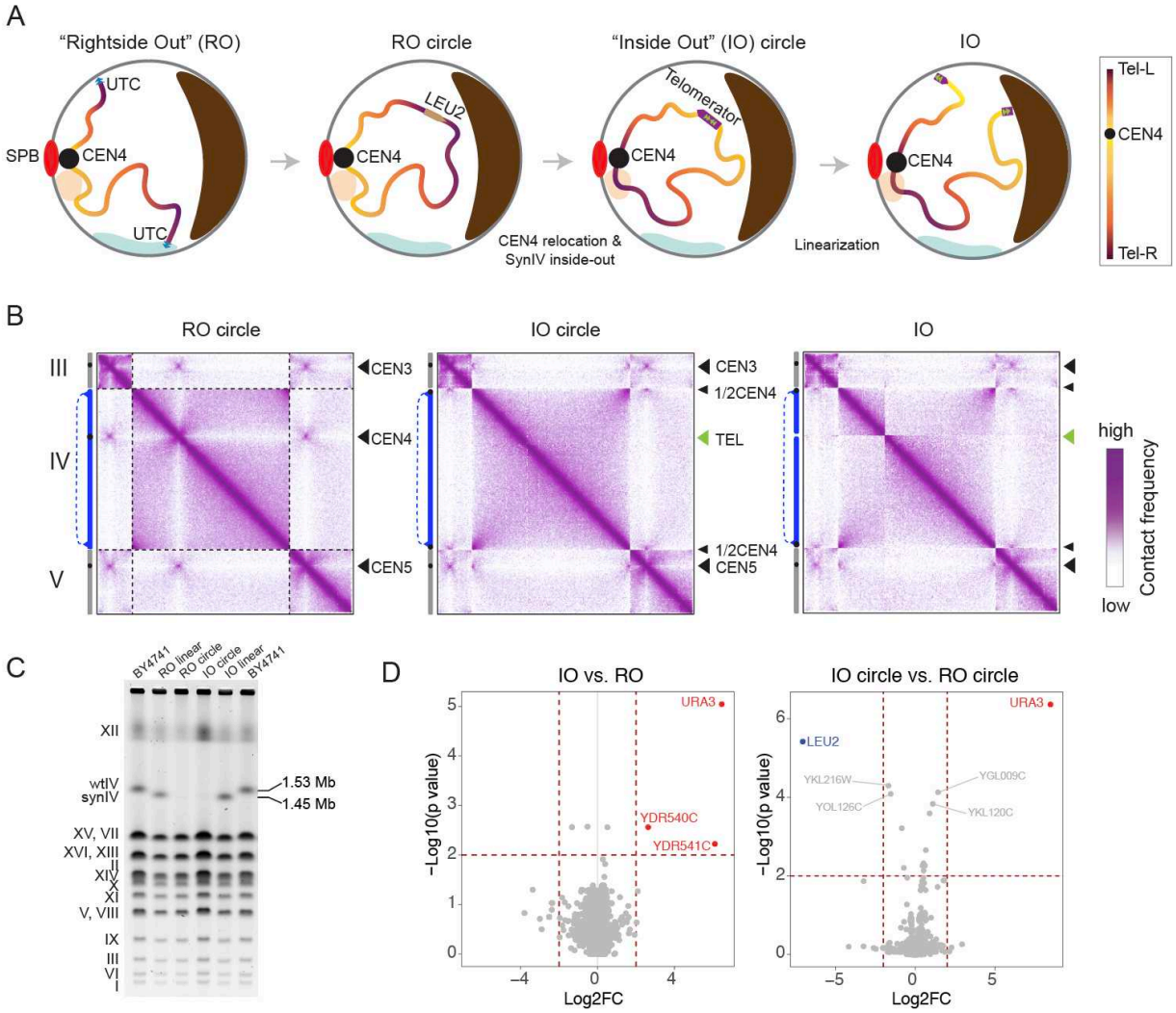

Fig. 4. "Inside out" *synIV* engineering and characterization. (A) Schematic illustration of the "inside out" *synIV* construction. *SynIV* chromosome is shown as a gradient color line, with chromosome regions near the wild-type centromere shown by light yellow and chromosome regions near the wild-type telomeres shown by

dark purple. Two hypothetical sub nuclear compartments are shown as orange and light green regions near the nuclear envelope. The nucleolus is depicted in brown. SPB, spindle pole body. (B) Hi-C contact maps of yeast with *synIV* in different configurations annotated in blue on the left side of each map (5 kb-binned). Black and green arrowheads indicate centromere and telomere positions, respectively. Violet to white color scale reflects high to low contact frequency (log10). (C) Linear *synIV*, circular *synIV* and linear wild-type IV were separated by pulsed field gel electrophoresis. CHEF mapper auto algorithm was used for separating the chromosomes, low molecular weight, 200 kb; high molecular weight, 2500 kb; voltage, 6 V/cm; total electrophoresis time was 29 hours. (D) Transcript profiling of both linear and circular *synIV*s. *URA3* marker was present in IO circle *synIV*, *LEU2* marker was present in RO circle *synIV*. Fold change ($log2^{FoldChange}$) cut off is 2, $-log10^{p\text{-}value}$ cutoff is 2.

Discussion

The assembly of *synIV*, the largest synthetic yeast chromosome, is a key step towards the completion of the first synthetic eukaryotic genome. *SynIV* was built in a hierarchical fashion, and the megachunk cloning strategy described here offered multiple advantages: 1) increased efficiency and reproducibility of megachunk incorporation, 2) sequencing cloned megachunks eliminated ligation errors up front and 3) bugs were corrected in parallel prior to the assembly phase. Although *synIV* was modified massively – with PCRTag recoding, loxPsym site insertion, and repeat element deletion – the 3D organization of *synIV* is remarkably similar to its native counterpart, revealing the sequence changes made had only minor effects on 3D chromosome organization.

Proper positioning of synthetic chromosomes in the yeast nucleus ensures the physiological expression of genes on the chromosome. How to define the organization of a synthetic chromosome is a worthwhile question. We manipulated the 3D organization of a megabase-sized chromosome in yeast by design, effectively turning it inside out, representing a major shift from sequence rewriting to structural redefinition of a synthetic yeast chromosome.

Mammalian chromatin is packaged into topological association domains (TADs) in the interior space of the mammalian nucleus and plays important roles in regulating gene spatiotemporal expression (Dixon et al., 2012). In contrast, yeast sub-nuclear compartments are seemingly restricted to the main nucleoplasm and the part adjacent to the nuclear envelope. Whether yeast has mammalian TAD-like territories that reside in the interior space of the nucleus is unclear. Through the construction and characterization of the “inside out” *synIV* strain, we made drastic spatial changes to all the genes on *synIV*, and we found only minor gene expression changes, leading to the conclusion that specific positioning in the interior space of yeast nucleus plays a negligible role in gene expression regulation, and suggesting that most regulation is local rather than regulated by global intranuclear position.

In summary, we deployed a “design-build-test-learn” cycle during the synthesis of the largest yeast synthetic chromosome. *SynIV* is a valuable resource for studying fundamental questions regarding evolution, chromosome stability and chromosome organization. On the other hand, the yeast chromosome rewriting technologies we acquired here can readily be deployed into higher order organisms, such as plant, mouse and human, to tackle intractable questions otherwise challenging to address by existing methods.

References and Notes

Annaluru, N., Muller, H., Mitchell, L.A., Ramalingam, S., Stracquadanio, G., Richardson, S.M., Dymond, J.S., Kuang, Z., Scheifele, L.Z., Cooper, E.M., et al. (2014). Total synthesis of a functional designer eukaryotic chromosome. Science *344*, 55–58. https://doi.org/10.1126/science.1249252.

Bolton, E.C., and Boeke, J.D. (2003). Transcriptional interactions between yeast tRNA genes, flanking genes and Ty elements: a genomic point of view. Genome Res. *13*, 254–263. https://doi.org/10.1101/gr.612203.

Brown, C.J., Todd, K.M., and Rosenzweig, R.F. (1998). Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. Mol. Biol. Evol. *15*, 931–942. https://doi.org/10.1093/oxfordjournals.molbev.a026009.

Chu, D., Kazana, E., Bellanger, N., Singh, T., Tuite, M.F., and von der Haar, T. (2014). Translation elongation can control translation initiation on eukaryotic mRNAs. EMBO J. *33*, 21–34. https://doi.org/10.1002/embj.201385651.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380. https://doi.org/10.1038/nature11082.

Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y.J., Lee, C., Shendure, J., Fields, S., Blau, C.A., and Noble, W.S. (2010). A three-dimensional model of the yeast genome. Nature *465*, 363–367. https://doi.org/10.1038/nature08973.

Dymond, J.S., Richardson, S.M., Coombes, C.E., Babatz, T., Muller, H., Annaluru, N., Blake, W.J., Schwerzmann, J.W., Dai, J., Lindstrom, D.L., et al. (2011). Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. Nature *477*, 471–476. https://doi.org/10.1038/nature10403.

Fredens, J., Wang, K., de la Torre, D., Funke, L.F.H., Robertson, W.E., Christova, Y., Chia, T., Schmied, W.H., Dunkelmann, D.L., Beránek, V., et al. (2019). Total synthesis of Escherichia coli with a recoded genome. Nature *569*, 514–518. https://doi.org/10.1038/s41586-019-1192-5.

Fudenberg, G., Getz, G., Meyerson, M., and Mirny, L.A. (2011). High order chromatin architecture shapes the landscape of chromosomal alterations in cancer. Nat. Biotechnol. *29*, 1109–1113. https://doi.org/10.1038/nbt.2049.

Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A., et al. (2008). Complete Chemical Synthesis, Assembly, and Cloning of a Mycoplasma genitalium Genome. Science *319*, 1215–1220. https://doi.org/10.1126/science.1151721.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345. https://doi.org/10.1038/nmeth.1318.

Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.-Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M., et al. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. Science *329*, 52–56. https://doi.org/10.1126/science.1190719.

Hanson, G., and Coller, J. (2018). Codon optimality, bias and usage in translation and mRNA decay. Nat. Rev. Mol. Cell Biol. *19*, 20–30. https://doi.org/10.1038/nrm.2017.91.

Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.-L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A., et al. (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science *351*, 1454–1458. https://doi.org/10.1126/science.aad9024.

Hull, M.W., Erickson, J., Johnston, M., and Engelke, D.R. (1994). tRNA genes as transcriptional repressor elements. Mol. Cell. Biol. *14*, 1266–1277. https://doi.org/10.1128/mcb.14.2.1266-1277.1994.

Lartigue, C., Glass, J.I., Alperovich, N., Pieper, R., Parmar, P.P., Hutchison, C.A., Smith, H.O., and Venter, J.C. (2007). Genome Transplantation in Bacteria: Changing One Species to Another. Science *317*, 632–638. https://doi.org/10.1126/science.1144622.

Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012–1025. https://doi.org/10.1016/j.cell.2015.04.004.

Mercy, G., Mozziconacci, J., Scolari, V.F., Yang, K., Zhao, G., Thierry, A., Luo, Y., Mitchell, L.A., Shen, M., Shen, Y., et al. (2017). 3D organization of synthetic and scrambled chromosomes. Science *355*. https://doi.org/10.1126/science.aaf4597.

Mitchell, L.A., and Boeke, J.D. (2014). Circular permutation of a synthetic eukaryotic chromosome with the telomerator. Proc. Natl. Acad. Sci. U. S. A. *111*, 17003–17010. https://doi.org/10.1073/pnas.1414399111.

Mitchell, L.A., Wang, A., Stracquadanio, G., Kuang, Z., Wang, X., Yang, K., Richardson, S., Martin, J.A., Zhao, Y., Walker, R., et al. (2017). Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. Science *355*, eaaf4831. https://doi.org/10.1126/science.aaf4831.

Monteiro, P.T., Oliveira, J., Pais, P., Antunes, M., Palma, M., Cavalheiro, M., Galocha, M., Godinho, C.P., Martins, L.C., Bourbon, N., et al. (2020). YEASTRACT+: a portal for cross-species comparative genomics of transcription regulation in yeasts. Nucleic Acids Res. *48*, D642–D649. https://doi.org/10.1093/nar/gkz859.

Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T., Abudayyeh, O.O., Gootenberg, J.S., Mori, H., et al. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science *361*, 1259–1262. https://doi.org/10.1126/science.aas9129.

Ostrov, N., Landon, M., Guell, M., Kuznetsov, G., Teramoto, J., Cervantes, N., Zhou, M., Singh, K., Napolitano, M.G., Moosburner, M., et al. (2016). Design, synthesis, and testing toward a 57-codon genome. Science *353*, 819–822. https://doi.org/10.1126/science.aaf3639.

Palladino, F., Laroche, T., Gilson, E., Axelrod, A., Pillus, L., and Gasser, S.M. (1993). SIR3 and SIR4 proteins are required for the positioning and integrity of yeast telomeres. Cell *75*, 543–555. https://doi.org/10.1016/0092-8674(93)90388-7.

Presnyak, V., Alhusaini, N., Chen, Y.-H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., et al. (2015). Codon Optimality Is a Major Determinant of mRNA Stability. Cell *160*, 1111–1124. https://doi.org/10.1016/j.cell.2015.02.029.

Rabl, C. (1885). Über Zelltheilung, Morphologisches Jahrbuch.

Richardson, S.M., Mitchell, L.A., Stracquadanio, G., Yang, K., Dymond, J.S., DiCarlo, J.E., Lee, D., Huang, C.L.V., Chandrasegaran, S., Cai, Y., et al. (2017). Design of a synthetic yeast genome. Science *355*, 1040–1044. https://doi.org/10.1126/science.aaf4557.

Shen, Y., Wang, Y., Chen, T., Gao, F., Gong, J., Abramczyk, D., Walker, R., Zhao, H., Chen, S., Liu, W., et al. (2017). Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. Science *355*, eaaf4791. https://doi.org/10.1126/science.aaf4791.

Taddei, A., and Gasser, S.M. (2012). Structure and Function in the Budding Yeast Nucleus. Genetics *192*, 107–129. https://doi.org/10.1534/genetics.112.140608.

Verstrepen, K.J., Jansen, A., Lewitter, F., and Fink, G.R. (2005). Intragenic tandem repeats generate functional variability. Nat. Genet. *37*, 986–990. https://doi.org/10.1038/ng1618.

Wu, Y., Li, B.-Z., Zhao, M., Mitchell, L.A., Xie, Z.-X., Lin, Q.-H., Wang, X., Xiao, W.-H., Wang, Y., Zhou, X., et al. (2017). Bug mapping and fitness testing of chemically synthesized chromosome X. Science *355*, eaaf4706. https://doi.org/10.1126/science.aaf4706.

Xie, Z.-X., Li, B.-Z., Mitchell, L.A., Wu, Y., Qi, X., Jin, Z., Jia, B., Wang, X., Zeng, B.-X., Liu, H.-M., et al. (2017). “Perfect” designer chromosome V and behavior of a ring derivative. Science *355*, eaaf4704. https://doi.org/10.1126/science.aaf4704.

Zhang, W., Zhao, G., Luo, Z., Lin, Y., Wang, L., Guo, Y., Wang, A., Jiang, S., Jiang, Q., Gong, J., et al. (2017). Engineering the ribosomal DNA in a megabase synthetic chromosome. Science *355*. https://doi.org/10.1126/science.aaf3981.

Zhang, W., Mitchell, L.A., Bader, J.S., and Boeke, J.D. (2020). Synthetic Genomes. Annu. Rev. Biochem. *89*, 77–101. https://doi.org/10.1146/annurev-biochem-013118-110704.

Acknowledgements and competing interests

We thank the NSF for grants MCB-1026068, MCB-1443299, MCB-1616111 and MCB-1921641 to JDB for supporting this work. This work is also supported by National Natural Science Foundation of China (31725002), Shenzhen Science and Technology Program (KQTD20180413181837372), Guangdong Provincial Key Laboratory of Synthetic Genomics (2019B030301006), Shenzhen Outstanding Talents Training Fund and Bureau of International Cooperation, Chinese Academy of Sciences (172644KYSB20180022) to JD. The synthesis of the left arm starting materials was supported by the DOE through the DOE’s Joint Genome Center Community Science Project. Jef Boeke is a Founder and Director of CDI Labs, Inc., a Founder of and consultant to Neochromosome, Inc, a Founder, SAB member of and consultant to ReOpen Diagnostics, LLC and serves or served on the Scientific Advisory Board of the following: Sangamo, Inc., Modern Meadow, Inc., Rome Therapeutics, Inc., Sample6, Inc., Tessera Therapeutics, Inc. and the Wyss Institute. Yasunori Aizawa is a Founder and CSO of Logomix, Inc. We thank Meghan O’Keefe for collecting all BAG authors’ information.