nature communications

Article

https://doi.org/10.1038/s41467-026-74300-9

Probing the limits of genetic recoding using multi-omics-guided evolution

Received: 1 December 2025

Accepted: 3 June 2026

Published online: 22 June 2026

Check for updates

Akos Nyerges[1,11] ✉, Anush Chiappino-Pepe[1,10], Bogdan Budnik[2,10], Maximilien Baas-Thomas[1], Elissa Rhuby[1], Regan Flynn[1], Shirui Yan[1,3], Nili Ostrov[1], Min Liu[4], Meizhou Wang[4], Qingmei Zheng[4], Fangxiang Hu[4], Kangming Chen[4], Alexandra Rudolph[1], Dawn Chen[1], Jenny Ahn[1], Owen Spencer[1], Venkat Ayalavarapu[1], Angela Tarver[5], Miranda Harmon-Smith[5], Matthew Hamilton[5], Ian Blaby[5], Yasuo Yoshikuni[5], Behnoush Hajian[6], Adeline Jin[4], Balint Kintses[7], Monika Szamel[7], Viktoria Seregi[7], Yue Shen[3,8,9], Zilong Li[4] & George M. Church[1,2,11] ✉

Engineering the genetic code—by reassigning multiple of the 64 natural codons—enables making organisms resistant to all viruses, preventing genetic information exchange, and allowing the biosynthesis of genetically encoded unnatural polymers. However, synonymous codon replacement—recoding—is frequently lethal, and how recoding impacts fitness remains poorly explored. Here, we explore these effects using genome synthesis, directed evolution, and genome-transcriptome-translatome-proteome co-profiling on multiple synthetic *Escherichia coli* genomes. We construct six partially recoded *E. coli* strains bearing up to 45.8% of a synthetic genome with a deleterious 57-codon genetic code. As our analyses revealed widespread defects—including unassigned codons in Syn61 and Syn57—we apply multi-omics to revise our genome design and mitigate defects. Using multi-omics, we show that recoding induces transcriptional and translational changes leading to fitness defects under hundreds of conditions. Finally, we develop a multi-omics-guided evolution strategy that rapidly restores fitness, enabling genome synthesis with radical changes.

The genetic code uses 64 codons to encode the 20 canonical amino acids during protein synthesis. However, 18 of these amino acids and the translational start and stop signals are encoded by multiple synonymous codons[1,2] in the standard genetic code. This degeneracy of the genetic code allows coding sequences—through codon selection—to define protein production simultaneously while maintaining cellular homeostasis by regulating gene expression, translation, and protein folding[3,4]. Synonymous codon selection also influences cellular homeostasis by altering messenger (m)RNA structure, translation efficiency, speed, and accuracy[5]. Furthermore, RNA and DNA motifs provide additional noncoding functions, such as regulating DNA and mRNA localization, genome replication, DNA repair, and chromosome segregation[3,6].

[1]Department of Genetics, Harvard Medical School, Boston, MA, USA. [2]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA. [3]BGI Research, Shenzhen, China. [4]GenScript USA Inc., Piscataway, NJ, USA. [5]DOE Joint Genome Institute (JGI), Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [6]Center for the Development of Therapeutics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [7]Institute of Biochemistry, HUN-REN Biological Research Centre, Szeged, Hungary. [8]BGI Research, Changzhou, China. [9]Guangdong Provincial Key Laboratory of Genome Read and Write, BGI Research, Shenzhen, China. [10]These authors contributed equally: Anush Chiappino-Pepe, Bogdan Budnik. [11]These authors jointly supervised this work: Akos Nyerges, George M. Church ✉e-mail: akos_nyerges@hms.harvard.edu; gchurch@genetics.med.harvard.edu

The genetic code's nearly universal nature and degeneracy spurred the exploration of synthetic genomes with artificial genetic codes to enhance biological functions, paving the way for making organisms safely resistant to natural viruses, preventing genetic information flow into and out of genetically modified organisms (GMOs), and enabling the biosynthesis of genetically encoded non-canonical polymers[7–14].

However, despite the wide-ranging benefits of genetic code engineering, studies highlighted its collateral impact on organismal fitness. The creation of Genomically Recoded Organisms (GROs)[7], whose genomes have been systematically redesigned to confer an alternate genetic code, in all cases induced fitness reduction. The creation of the first GRO through the genome-wide removal of 321 TAG stop codons and release factor 1 (RF1, encoded by *prfA*) from *Escherichia coli* K-12 MG1655 resulted in a 60% increase in cell doubling time compared to the parental strain[7,15]. Follow-up projects showed that not only the rarest TAG stop codon but sense codons are also malleable—by attempting to recode 62 codons and eliminate 13 codons in 42 highly expressed prokaryotic genes[16] and, in a separate study, by removing rare AGA and AGG arginine codons from growth-essential genes of *E. coli*[17]—paving the way for sense codon recoding on whole genomes. These studies revealed the occasionally lethal effect of synonymous recoding, indicating that codon preference poses constraints on viable genetic codes[7,15–18].

These early studies suggested that compressing the genetic code in a living organism beyond rare stop codons might be feasible. Driven by this possibility, we initiated the design and synthesis of a complete *E. coli* genome, which we named Ec_Syn57, designed to rely on only 57 genetic codons[19]. Our early tests using select genomic regions spanning 1.1% of the genome confirmed the viability of 63% of the tested recoded genes despite removing seven codons (Fig. 1a). However, these tests also revealed frequent deleterious effects, and our efforts to replace multiple recoded sections in a single strain remained unsuccessful.

In parallel, a strain of *E. coli*, Syn61Δ3, was created by the Chin lab with a synthetic recoded genome from which all annotated instances of two serine codons, TCG and TCA, and the TAG stop codon were eliminated to confer a 61-codon genetic code[18,20]. Successful genome design required testing eight distinct recoding schemes prior to

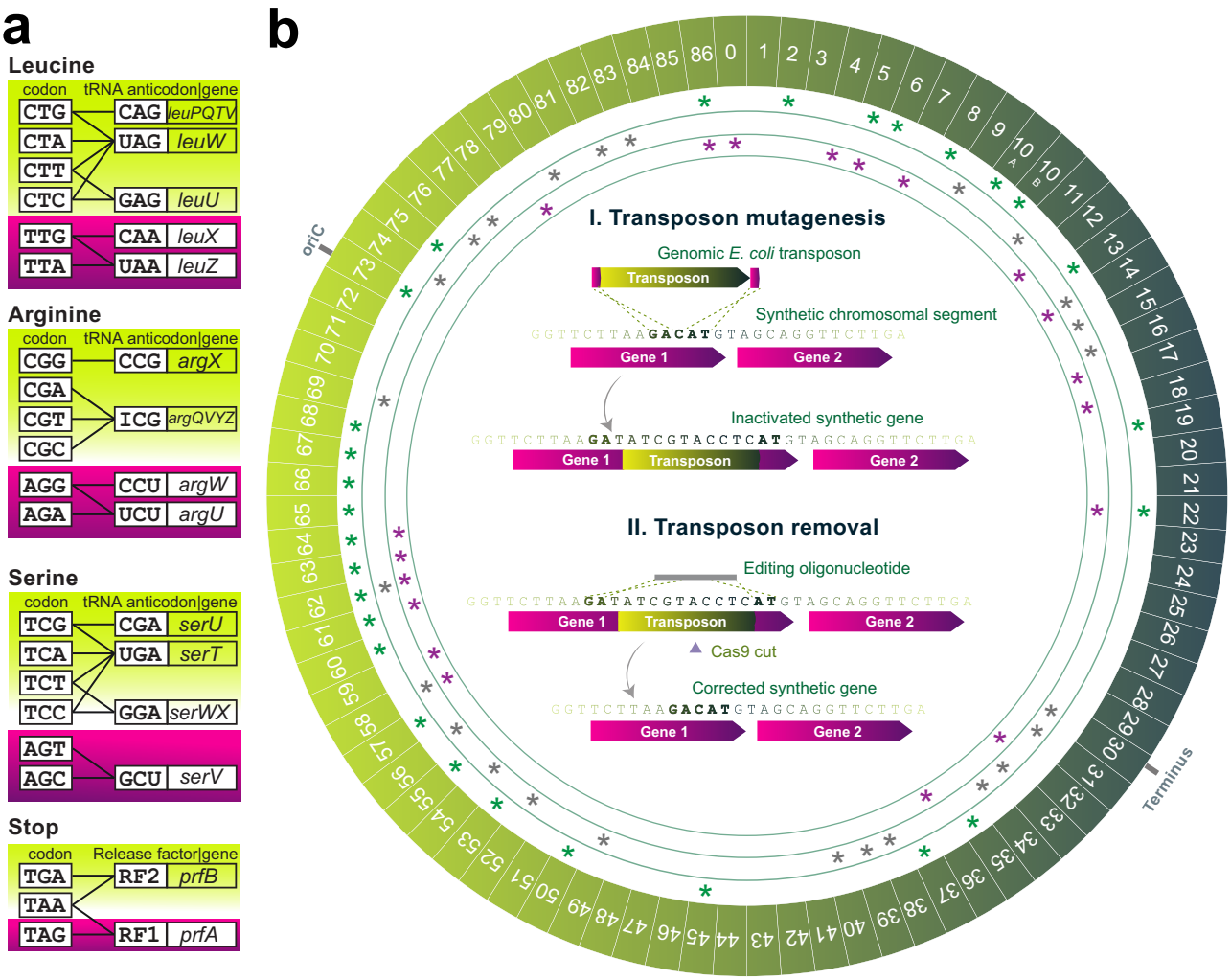

Fig. 1 | Design and synthesis of a synthetic *Escherichia coli* genome with radical genome recoding. a The design of Ec_Syn57's compressed genetic code[19]. Magenta marks codons and their corresponding tRNA genes and release factor I (RF1, encoded by *prfA*) selected for elimination during genome design. Green marks codons selected to remain in the compressed genetic code. **b** The computational genome design of Ec_Syn57 synonymously recoded all previously known[19] instances of the seven target codons highlighted in Fig. 1a and streamlined the synthetic chromosome (Supplementary Fig 1). The refactored, recoded genome was divided into 87 - 50 kbp segments, and the entire genome was synthesized. Stars (*) in green mark newly synthesized chromosomal segments in this study; stars in gray mark segments into which bacterial mobile genetic elements transposed during segment DNA synthesis (see Supplementary Fig 2 and Supplementary Table 1 for their exact location); stars in magenta mark synthetic chromosomal segments that did not support cell growth. Mobile genetic elements were excised using CRISPR/Cas9-assisted MAGE[48], leading to the scarless restoration of the desired synthetic sequence (Fig. 1bI, II).

genome synthesis using replicon excision for enhanced genome engineering through programmed recombination with 2 CRISPR/Cas9 cuts (in short, REXER2), from which only three appeared experimentally viable[18]. Despite a theoretically high-fitness recoding scheme, codon swaps induced lethality at four loci, which were subsequently identified by REXER2[18]. Once combined into a fully synthetic genome, the final strain displayed a 73–153% increase in doubling time compared to its parental strain, depending on growth conditions[20]. Decreased fitness subsequently prevented the immediate deletion of the corresponding serine transfer (t)RNA genes (*serU* and *serT*) and RF1[9,20], necessitating rounds of additional troubleshooting before the first sense-codon-recoded strain could be completed. Recent construction of Syn57[21]—a strain with a distinct recoding scheme and genome-construction workflow compared to our Ec_Syn57 design—followed the same path, resulting in a 270-minute doubling time, 7.71 times longer than the parental strain's[21]. In these studies, the causes of nonviable genetic codes and general fitness defects beyond the lethality of selected edits[18,20] remained unexplored.

Prior studies, including the Sc2.0 and *Mycoplasma* genome synthesis projects[22–38] (reviewed in ref. 39.), highlighted two fundamental challenges in front of our ability to construct synthetic genomes with custom, radically refactored genetic codes: i.) the rapid and parallelizable discovery and debugging of fitness-decreasing design issues on synthetic genomes, and ii.) our limited knowledge on the genome-scale effects of synonymous codon changes.

In this paper, we address these challenges to enable radical genome and genetic code refactoring. To this end, we developed a multiomics-guided technology that identifies fitness-decreasing errors in synthetic genomes and enables their debugging via genome editing and directed evolution. We demonstrate that—by integrating multiomics analyses along all steps of the central dogma (i.e., genome, primary- and total-transcriptome, translatome, and proteome profiling) to discover fitness-decreasing errors and by correcting these using multiplexed genome editing—the synthesis of six partially recoded *E. coli* strains is possible, across which we install 162,521 bp changes and remove 62,029 instances of seven codons to confer an initially highly deleterious 57-codon genetic code. These changes include the genome-wide removal of two leucine codons for which no complete recoding was previously achieved using REXER[18], and two arginine codons that frequently confer deleterious effects when changed synonymously[17].

By analyzing the transcriptional and translational effects of a 64-, 61-, and 57-codon genetic code and fitness in hundreds of environments, we show that synonymous codon changes impact mRNA and protein production, influence translation efficiency, induce promoters, and reduce fitness. Beyond discoveries about genomes and the genetic code, this work provides a technology for identifying and resolving defects in synthetic chromosomal regions, which may facilitate the construction and debugging of synthetic microbial genomes with radical changes.

We expect that the combination of synthetic genomics, multiomics, and multiplexed directed evolution will pave the way toward industrially useful organisms with custom functionalities, including high-fitness, virus-resistant cells capable of producing biopolymers with an enhanced chemical repertoire and safe GMOs by eliminating horizontal gene transfer via dependence on genetic codes and amino acids not found in nature[8,9,40,41].

Results

Design and DNA synthesis for a radically recoded genome

We previously designed a genome of *E. coli* MDS42 in which all annotated instances of seven codons were replaced with synonymous alternatives across all protein-coding genes[19]. Based on our previous design principles (Fig. 1a, Supplementary Fig 1)[19], we disentangled overlapping genes and eliminated direct repeats and homopolymer sites, resulting in a total of 162,521 base pair changes compared to the 4 Mbp parental genome. Codons were chosen for replacement based on their low frequency in protein-coding genes to minimize the changes required (i.e., the TAG stop, and the AGA and AGG arginine codons) and their orthogonality within the tRNA pool (i.e., the TTG and TTA leucine, and AGT and AGC serine codons) to ensure that their cognate tRNAs and Release Factor I could be later eliminated. Following recoding, we divided the synthetic genome into operon borders into 88 segments (Fig. 1b). Segments were assembled into bacterial artificial chromosomes (BACs) using chemical DNA synthesis and homologous recombination in *Saccharomyces cerevisiae*[19]. Our genome synthesis workflow reduced the synthesis error rate 19-fold compared to early DNA synthesis for this project and achieved an average error rate of 0.8 compared to 15.4 errors per segment before[19]. In total, the final set of BACs contained 952 instances of DNA synthesis-induced mutations (including 726 SNPs), from which only 19 originated from our updated genome synthesis workflow, while the remaining errors were induced by early DNA synthesis for this project[19]. In line with previous reports showing that endogenous bacterial mobile genetic elements (MGEs)—insertion sequences (ISes) and transposons—frequently invade synthetic DNA constructs[42–44], we detected extensive MGE transposition into synthesized chromosomal regions following segment delivery into *E. coli*. These transposition events were mainly driven by the Tn1000 transposon of *E. coli* DH10B[45,46], a widely used cloning host designed to propagate large BACs (Fig. 1b, Supplementary Table 1). Transpositions occasionally disrupted growth-essential genes in synthetic constructs, preventing genome construction (Supplementary Fig 2). We therefore eliminated transposon mutagenesis using CRISPR/Cas9-assisted MAGE to delete all 24 instances of transposed mobile genetic elements, and utilized mobile-genetic-element-free cloning hosts to maintain recoded segments[47,48]. Taken together, we completed DNA synthesis as 88 separate bacterial artificial chromosomes for an *E. coli* genome in which 62,007 instances of seven codons were recoded to synonymous alternatives.

Genome construction with frequent design issues

With the synthesized set of bacterial artificial chromosomes containing our synthetic genome, we initiated genome assembly and debugging. The development of a combined genome construction and debugging method was necessary based on our previous estimate of approximately one lethal and multiple fitness-decreasing issues in each synthetic segment[19] and the demonstrated lethality of TTG and TTA leucine codon recoding[18]. Several existing bacterial genome construction methods (Supplementary Table 2)[19,20,27,49] do not employ sequence-based counterselection against the parental sequence, rendering them unsuitable for the continuous construction of synthetic genomic regions that contain numerous, closely spaced, fitness-decreasing errors. Without sequence-based counterselection, these methods frequently result in unedited genomes with fitness-impairing changes, preventing the construction of synthetic genomes with less-tolerant recoding schemes[18,27,50] (Fig. 2a, b). We note that the location of these unedited regions can subsequently inform follow-up debugging through analysis of the resulting recoding landscape and identify less-favored recoding schemes, as illustrated by REXER[18,21,27] throughout the construction of Syn61's genome[20]. REXER, serving as both a construction and debugging tool, can quickly identify lethal recoding schemes[18] and construct high-fitness synthetic genomes[20] with distant recoding-induced defects.

We addressed this challenge by developing Synthetic genomes by multiomics-based iterative construction and troubleshooting, or synomics, in short, a genome construction and debugging method that relies on CRISPR/Cas9-counterselection, recombineering (i.e., recombination-mediated genetic engineering), multi-omics, and ALE. In SynOMICS (Fig. 2c), we perform CRISPR/Cas9-counterselection to replace parental chromosomal regions with their synthetic

https://doi.org/10.1038/s41467-026-74300-9

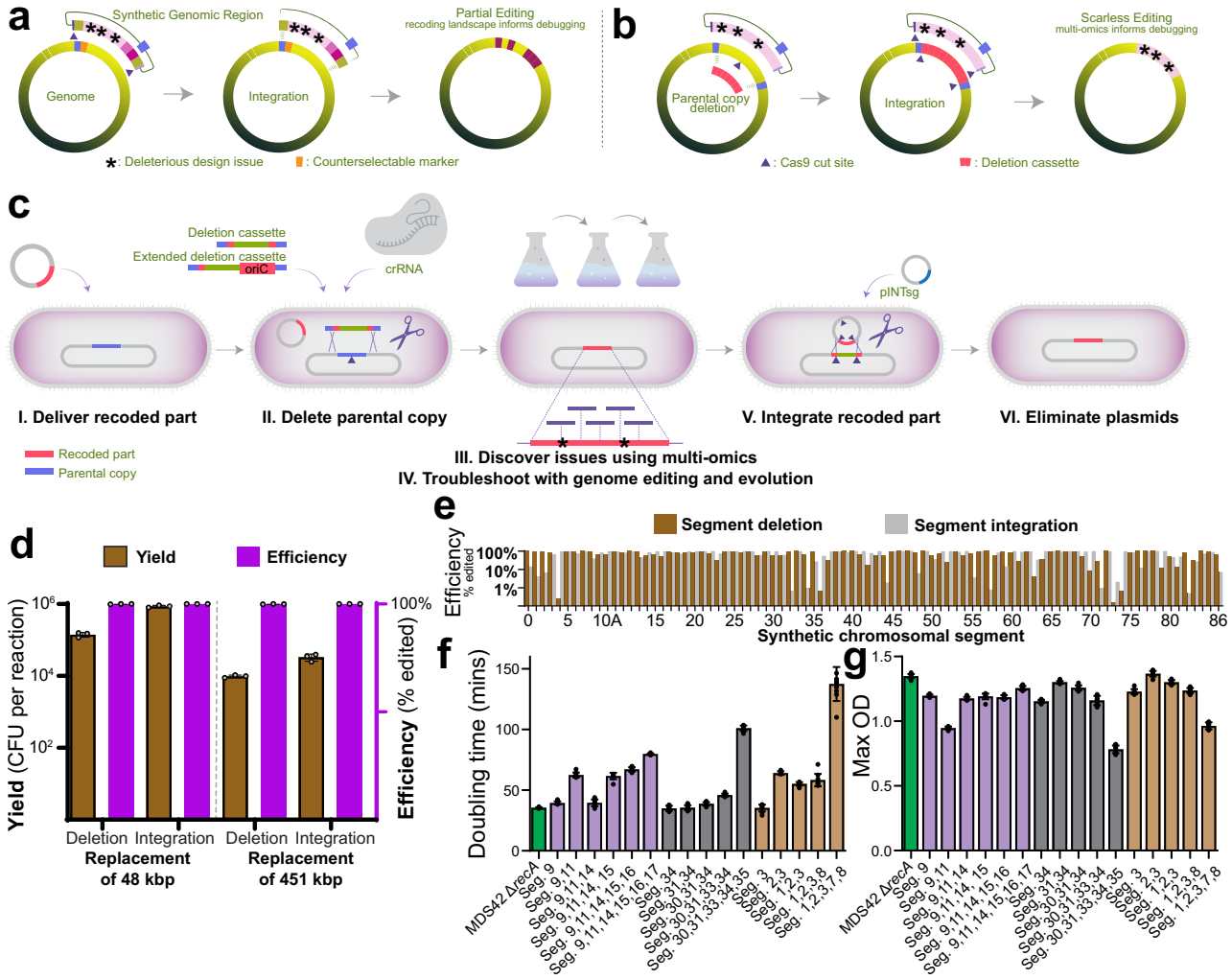

Fig. 2 | Genome construction with a challenging recoding scheme. In the presence of multiple closely spaced deleterious design issues (**a**) without sequence-based counterselection against the parental genome (i.e., CRISPR/Cas9-cut within the parental copy, marked as ▲), genome construction methods result in chimeric genomes and the reversion of deleterious codon changes (marked as *)[18,27]. **b** CRISPR/Cas9-assisted counterselection against the parental locus allows the generation of deleterious recoding schemes. **c** The workflow of SynOMICS-based genome construction. SynOMICS utilizes a recombination-deficient parental strain, leading to increased CRISPR/Cas9 selection stringency and preventing unwanted recombination between the parental and recoded copies. Following recoded segment delivery (**I.**), the parental segment copy is deleted using CRISPR/Cas9-assisted Lambda-Red recombineering (**II.**). Following the deletion of the parental copy, fitness-decreasing design errors are discovered using multi-omics analyses (**III.**) and debugged using multiplexed genome editing and laboratory evolution (**IV.**). Finally, the recoded segment is integrated by delivering a 6-plex sgRNA plasmid (**V.**) that site-specifically integrates the synthetic part (**VI.**) and eliminates the genome-targeting crRNA plasmid of Step II., preparing cells for the next SynOMICS cycle. **d** Efficiency and yield of SynOMICS's deletion and integration step at a 48 kbp and 451 kbp locus. Data are mean ± SD based on $n = 3$ independent biological replicates. **e** Synthetic segment integration using SynOMICS. Figure shows segment deletion and integration efficiency along the genome of Ec_Syn57. Segment integration experiments were performed once. **f**, **g** Growth of partially recoded strains, carrying increasing synthetic sections. Doubling time (**f**) and maximum optical density (OD at 600 nm) (**g**) were measured in 2×YT broth at 37 °C. Data are mean ± SD based on $n = 10$ independent biological replicates. Color groups (i.e., purple, gray, brown) represent parallelly constructed strain series. Source data are provided as a Source Data file.

counterpart in recombination-deficient (i.e., Δ*recA*) cells, followed by omics-guided directed evolution and ALE to debug design issues. We hypothesized that the deletion of the parental copy of a synthetic genomic segment, without induced crossover between the parental and synthetic genomes, would allow us to construct viable recoded variants even when recoding-induced flaws are densely distributed (Fig. 2b). The deletion of the parental copy, in turn, poses a fitness-based selection pressure on the synthetic region, allowing the multi-omics-based discovery of issues and the directed evolution of beneficial mutants.

In SynOMICS, we achieve these goals using a sequence of optimized recombineering events. Following the delivery of a synthetic-segment-containing BAC (Fig. 2c), CRISPR/Cas9-assisted recombineering enables the complete deletion of the corresponding parental locus. The use of a recombination-deficient host strain[51] allowed us to achieve nearly 100% efficiency when deleting up to 451 kbp chromosomal DNA (Fig. 2d), 3.3-fold longer than alternative methods (Supplementary Table 2). During the deletion step, we sporadically—in 9 out of 88 segment deletion experiments (Supplementary Table 3)—observed the reversion of a few recoded codons within the short terminal homology of the deletion cassette due to recombination errors, further confirming the importance of sequence-based selection against the parental copy.

We next explored whether SynOMICS could construct a genomic region bearing a recoding scheme previously identified as highly deleterious using REXER[18]. The *mraZ-ftsZ* locus in Segment 2 of the 57-

codon genome served as an ideal target, as i.) it harbors the entire, previously tested[18] *mraZ-ftsZ* locus, ii.) attempts the elimination of the same 157 instances of TTG and TTA leucine codons that led to catastrophic failure in earlier REXER-based recoding efforts[18], and iii.) contains a total number of 585 codon changes, including the synonymous recoding of AGT, AGC serine, the challenging rare AGG, AGA arginine[17], and TAG stop codons besides leucine codon recoding, 273% more codon changes than what was previously[18] attempted (Supplementary Table 4). SynOMICS-based deletion of the parental copy of Segment 2 yielded 1047 colonies. Edited clones were easily identified based on their antibiotic resistance profile and colony PCR, followed by genotyping using Multiplex Allele-Specific Colony (MASC) PCR[15,19]. 92% of all clones surviving antibiotic selection contained only the recoded allele. The *mraZ-ftsZ* recoded strain displayed a 7.7% increase in doubling time and a 14% reduction in final OD600 in rich bacterial medium compared to the parental strain's 35.4 ± 0.38 min doubling time and 1.35 ± 0.02 maximum optical density under identical conditions (Supplementary Data 2)). On agar plates, the *mraZ-ftsZ* recoded strain produced tiny, barely visible colonies after 24 h of aerobic growth under the standard growth condition used during our SynOMICS experiments, while the parental MDS42 displayed regular-sized colonies (Supplementary Fig. 3). This experiment demonstrated SynOMICS's ability to achieve an exceptionally challenging recoding scheme.

Following deletion of the parental locus, scarless integration of the recoded segment is initiated by delivering a CRISPR/Cas9 guide (g) RNA-expressing plasmid. This plasmid expresses six single guide RNAs (sgRNAs), simultaneously cutting the two ends of the SynOMICS deletion cassette, two ends of the recoded segment, and the backbone of the BAC vector (Fig. 2c), thereby forcing cells to integrate the recoded region into their genome. Once edited clones are identified based on their antibiotic resistance profile, a short growth step prepares the cells for the next integration cycle. After extensive plasmid optimization (Methods, Supplementary Fig 4), we achieved near-100% integration efficiency during SynOMICS's integration step (Fig. 2d, e). We note that while large DNA integration is a key step, the rate-limiting challenge in SynOMICS arises during the deletion of parental copy, as this is the step where recoding-induced fitness defects are first exposed (Fig. 2c).

Using SynOMICS, we initiated iterative single-segment replacements but observed drastic fitness impairment. 19 out of 88 chromosomal segments of Ec_Syn57 did not support cell growth without their parental copy (Fig. 1b), and recoding one to six neighboring ~50-kbp segments (Fig. 2f, g) resulted in severe fitness reduction, confirming the fitness-decreasing effect of Ec_Syn57's recoding scheme. As these fitness-decreasing errors prevented the replacement of neighboring genomic regions, we next investigated the sources of fitness issues and attempted their correction.

Identification of lethal design issues

As the next step of SynOMICS, we identified potential fitness-decreasing issues using simultaneous genome, transcriptome, translatome, and proteome profiling. We hypothesized that synonymous codon swaps that overlap with i.) promoter motifs within coding sequences would decrease promoter activity, or impact translation initiation via changes in ii.) mRNA folding or iii.) ribosomal binding site strength.

Based on the role of the eliminated, rare arginine codons in translation initiation[17], we first explored the effects of synonymous recoding on translation initiation. Although our genome design aimed to minimize changes in mRNA folding around the ribosomal binding site (RBS) and preserved 15–20 base pairs of the 5′ untranslated region (UTR) for overlapping genes (Supplementary Fig. 1), the in vivo effects of these changes remained unexplored. Therefore, we explored translation initiation rate (TIR) changes by predicting TIR for every protein-coding gene on the 57-codon genome using the RBS Calculator[52,53] and compared them to the TIR of the parental variant. Our analysis identified 816 genes with reduced and 507 genes with increased translation initiation rates, representing 36.3% of all protein-coding genes (Supplementary Fig. 5).

Following the prediction of translational effects, we sought to explore the impact of synonymous recoding on bacterial promoters. Using deep primary transcript sequencing, we identified all active promoters of MDS42 in lysogeny broth, a commonly used rich bacterial medium. We performed Cappable-seq[54,55], a transcriptomic method that captures the 5′ end of transcripts and determines transcription start sites (TSS) at single-base resolution. Deep short-read Cappable-seq identified 39,140 TSS, from which 17,504 resided in protein-coding genes (i.e., intragenic TSS). We also identified the principal promoters for all operons on MDS42's genome using long-read Pacific Biosciences SMRT-Cappable seq[55] (Supplementary Fig. 6). This latter method extends the Cappable-seq workflow by performing long-read sequencing on entire primary transcripts, allowing the identification of all genes driven by a single promoter. Using Cappable-seq, we identified all promoters that overlap with codon changes on Ec_Syn57's genome. Next, we individually synthesized 12 impacted promoters in their wild-type and 57-codon recoded form, cloned them into a single-copy BAC vector, and analyzed the mRNA output of these promoter variants using transcriptomics. Transcriptomics revealed that synonymous recoding significantly decreased the activity of 7 out of the 12 assayed promoter pairs (Fig. 3a). Notably, promoters that drive growth-essential genes and conferred decreased mRNA output in our assay (i.e., *accA*p, *dnaT*p, *msbA*p, *pgsA*p, *ribF*p, *rpsT*p) were all located in recoded chromosomal regions that had previously failed to replace their corresponding parental copy (Fig. 1b), confirming these damaged promoters' impact on fitness.

Based on this observation, we validated the discovered promoter mutations' fitness effects using Segment 21's *fabH-ycfH* locus. The deletion of Segment 21's parental copy induced a drastic fitness drop, and our analysis predicted accidental promoter-recoding in the growth-essential *fabH-ycfH* region as the cause of fitness decrease (Supplementary Fig 6). We restored gene expression by performing MAGE[48,56] and inserted a library of constitutive promoters in front of each impacted gene. Following editing, we identified fast-growing variants. In line with our hypothesis, fitness increase required the insertion of three promoters that replaced all predicted damaged promoters within the *fabH-ycfH* locus.

Using transcriptome profiling and computational analyses, we identified hundreds of potential design issues and fitness-decreasing synonymous codon changes and validated a subset of these using multiplexed genome editing.

Genome annotation errors leading to unassigned codons

Given the deleterious effect of non-recoded codons on organisms with compressed genetic codes[57], we next explored the presence of non-recoded codons on existing recoded chromosomes. We hypothesized that unassigned non-recoded and, therefore, forbidden[19,57] codons that lack translational elements (cognate release factors or tRNAs)[57] can originate from annotation errors on bacterial chromosomes and—following the elimination of tRNAs or release factors recognizing these codons—will impact growth by arresting protein synthesis[57]. Therefore, we revised the gene annotations of our 57-codon genome using the genome of *E. coli* K-12 MG1655, the parental strain of MDS42[44]. Genome reannotation revealed the presence of 109 previously unannotated protein-coding genes and an additional 100 genes with altered coordinates, containing 644 forbidden codons. Performing the same analysis on Syn61's and the Robertson et al. Syn57's genome[21] indicated the presence of 105 previously unannotated protein-coding genes and altered gene-coordinates, containing 186 forbidden codons, including forbidden codons in *mukE*, *ykfM*, *yjbS* and *safA* Syn61's essential

https://doi.org/10.1038/s41467-026-74300-9

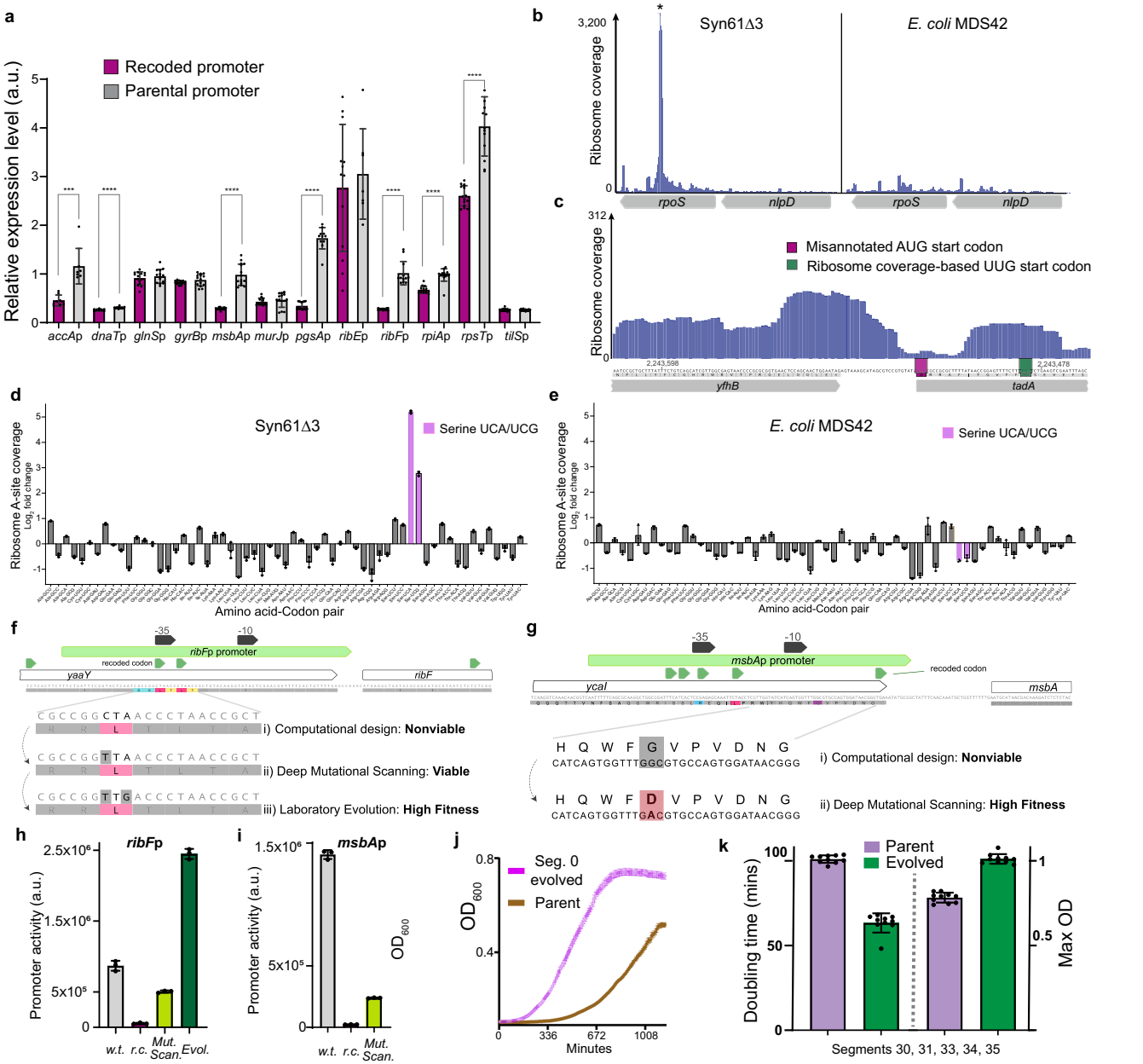

Fig. 3 | Detection and debugging of synonymous recoding-induced defects. **a** Negative effect of synonymous recoding on promoter activity. Figure shows the transcript output of promoters that overlap with synonymous codon swaps in Ec_Syn57. Data are mean-based on $n = 10$ biological replicates, error bars denote SD; *** indicates a $P = 0.0002$, **** indicates $P \leq 0.0001$ based on unpaired two-sided Student's t-test. **b** Ribosomes stall at forbidden codons present in Syn61Δ3. Figure shows ribosome footprint coverage along *rpoS* containing a TCA codon (marked with *) in Syn61Δ3 lacking $tRNA^{Ser(UGA)}$ and $tRNA^{Ser(CGA)}$, and in *E. coli* MDS42 bearing the canonical genetic code. **c** Ribosome footprint coverage at the previously annotated (in magenta) and revised (in green) start codon of *tadA*. **d**, **e** Ribosome A-(aminoacyl)-site coverage of sense codons in Syn61Δ3 (**d**) and MDS42 (**e**). Ribosomes in the compressed genetic code of Syn61Δ3 accumulate at TCA and TCG codons—in magenta—present due to genome annotation errors and hypermutagenesis-induced mutations. Data indicates the mean, Ribo-seq data is based on $n = 3$ independent biological replicates, error bars indicate SD. **f** Troubleshooting recoding-induced lethality within Segment 0 using multiplexed genome editing and laboratory evolution. The nonviable genome design (i) was troubleshot using (ii) targeted mutagenesis and (iii) ALE to identify optimal variants within the growth-essential promoter of the *ribF-ispH* operon. **g** Troubleshooting synonymous recoding-induced lethality within Segment 18 using targeted mutagenesis. The nonviable computational design was debugged using DIvERGE mutagenesis and growth-based selection. **h** *ribF* promoter activity changes during promoter debugging, based on sfGFP expression. Data represents the mean, error bars denote SD based on $n = 3$ independent biological replicates. **i** *msbA* promoter activity changes in Segment 18 during promoter debugging, based on sfGFP expression assay. Data represents the mean, error bars denote SD based on $n = 3$ independent biological replicates. **j** Growth kinetics of the ALE-evolved Segment 0 variant, based on $n = 10$ biological replicates. Data represents the mean, error bars denote SD. **k** Growth and maximal OD of the parental and ALE-evolved strain containing Segments 30,31,33,34,35 of Ec_Syn57, based on $n = 10$ independent biological replicates. Data represents the mean, error bars denote SD. Source data are provided as a Source Data file.

genes[58], and 89 non-recoded genes containing 539 non-recoded codons in Syn57[21], respectively (Supplementary Note 1).

We next explored the consequences of unassigned codons by performing genome-wide ribosome occupancy analysis using ribosome profiling on MDS42 and Syn61Δ3. As Syn61Δ3 lacks $tRNA^{Ser(UGA)}$ and $tRNA^{Ser(CGA)}$ needed to translate TCA and TCG codons, stalled ribosomes are expected to accumulate at these codons. In line with this hypothesis, we detected ribosome A-(aminoacyl)-site

footprint peaks centered at forbidden codons in translated genes (Fig. 3b, Supplementary Fig 7), and the analysis of ribosome occupancy across the genetic code table highlighted the arrest of ribosomes at the TCA and TCG serine codons (Fig. 3d, e). Based on ribosome profiling data, we estimate forbidden codons sequester 2.1% of all translating ribosomes in Syn61Δ3.

Genome reannotation followed by ribosome profiling revealed a severe design error in Ec_Syn57's genome. Due to the erroneous annotation of the growth-essential *tadA*'s start codon and, consequently, the accidental recoding of this gene's TTG start (Fig. 3c), our analysis predicted no expression for *tadA*. In line with this hypothesis, eliminating *tadA*'s parental copy resulted in drastic fitness loss, i.e., a doubling time of 429 ± 63.2 min, preventing segment integration. By correcting the start codon with ssDNA-mediated recombineering, we restored translation initiation[59] and the integrability of this segment (Fig. 2e), resulting in a 9.57-fold decrease in doubling time (i.e., 44.88 ± 1.9 min, based on $n = 10$ replicates). Finally, due to the extent of genome annotation-induced recoding errors and recent reports indicating the presence of multiple cryptic translated open reading frames (ORFs) on the genome of *E. coli* MG1655[60–65], we performed an untargeted proteomics-based screen to identify translated sequences on our genome, revealing dozens of previously unannotated sequences containing unassigned codons (Supplementary Note 2, Supplementary Figs. 8, 9).

Multi-omics experiments jointly revealed that existing recoded organisms harbor and tolerate hundreds of forbidden codons (Supplementary Fig. 10), and these codons might be partially responsible for the observed fitness defects. Based on the detected adverse effect of unassigned codons on bacterial translation, we selected all protein-coding genes that are actively transcribed in MDS42 and recoded these 73 genes containing 311 forbidden codons to match Ec_Syn57's genetic code (Supplementary Data 1).

Editing-based debugging of lethal design errors

Based on the list of uncovered design errors, we initiated the debugging of all lethal errors that directly prevented genome construction. These errors were computationally predicted disruptions to promoters, ribosome-binding sites, and operon structures in synthetic segments that failed to complement their parental copy during SynOMICS's deletion step. Using the list of recoding-impacted promoters and RBSes and the operon structure of *E. coli* MDS42, we debugged predicted lethal issues using three strategies: In Segments 4 and 7, we utilized CRISPR/Cas9-assisted MAGE[48] to simultaneously insert a promoter and an optimized RBS in front of impacted operons. As expected, the predicted 12-fold increase in gene expression restored viability for both segments. In the case of Segments 5, 22, 61, 64, and 86, in which multiple distant changes were necessary, we resynthesized the entire recoded segment with a modified design that restored gene expression at all indicated loci (see Supplementary Methods 1 for the detailed list of edits). To compensate for recoding-induced expression defects, we targeted an approximately 10-fold increase in gene expression compared to the recoded variant by inserting an RBS Calculator[52,53] optimized RBS and, in the case of damaged promoters, a synthetic constitutive promoter sequence[66,67]. We intentionally overcompensated predicted gene expression changes to increase the likelihood of successful designs and relied on multi-omics-guided directed evolution for subsequent fine-tuning to rebalance expression levels. Segment resynthesis restored the viability of all but one segment (i.e., Segment 64, which we later debugged using a combination of all three methods).

To troubleshoot all remaining design issues that induced fitness defects but did not immediately prevent the deletion of the parental copy, we leveraged the fitness-based selection pressure imposed by deleting the parental chromosomal region in SynOMICS and performed targeted mutagenesis to increase fitness. We utilized DIvERGE, an ssDNA-recombineering-based targeted mutagenesis method that allows an up to a million-fold increase in mutation rate along the entire lengths of multiple defined loci—ranging from promoters up to entire operons—without inducing off-target mutations, and capable of targeting hundreds of sites simultaneously[68,69]. We synthesized 324 DIvERGE oligonucleotides targeting promoter issues and an additional 442 MAGE oligonucleotides to correct 447 DNA synthesis errors. Next, using a pooled mixture of all DIvERGE and MAGE oligos targeting all potential promoter issues and DNA synthesis errors in each partially recoded strain, we performed rounds of diversification followed by growth-based selection. Whole-genome sequencing of the resulting clones indicated the reversion of multiple DNA synthesis errors, up to ten in a single strain, and newly emerged promoter mutations within DIvERGE oligo targets (Fig. 3f,g). In Segment 0, where promoter transcriptomics identified the recoded promoter of the *ribF-ispH* operon in the upstream *yaaY* as the cause of nonviability (Fig. 3a), mutagenesis identified a single point mutation (Fig. 3f). In *yaaY*, two neighboring recoded leucine codons overlap with the −35 region of the growth-essential *ribF-ispH* promoter and DIvERGE mutagenesis resulted in the reversion of one of the recoded leucine codons, restoring viability. Similarly, diversifying the damaged promoter of the growth-essential *msbA-ycaQ* operon in Segment 18 (Fig. 3g) and the *pgk-fbaA* operon in Segment 56 identified beneficial mutations (Supplementary Fig 11).

Finally, we validated the positive effect of the DIvERGE-identified mutations on downstream gene expression. We individually synthesized and cloned parental and evolved *ribF* and *msbA* promoter variants, followed by a superfolder GFP (sfGFP) marker gene, and measured sfGFP production. Recoded variants conferred a marked decrease in sfGFP expression compared to their parental promoter, while the DIvERGE-evolved promoter variants partially restored the lost promoter activity (Fig. 3h, i).

These experiments indicated that while targeted DIvERGE editing can rapidly discover beneficial variants, additional fine-tuning will likely be needed to fully mitigate fitness defects.

Laboratory evolution allows genome recoding in 11 sections

Following genome editing, we initiated Adaptive Laboratory Evolution (ALE) to further increase the fitness of Ec_Syn57's synthetic chromosomal segments. ALE is a powerful method that enables the discovery of compensatory mutations in a nontargeted manner, and we expected it to identify additional compensatory mutations that remained hidden in our previous screens. To perform ALE without preferentially evolving compensatory mutations outside recoded regions and to minimize off target mutations leading to forbidden codons (as shown in the case of Syn61Δ3; Supplementary Note 1, Fig. 3b, d, e), we relied on natural mutagenesis and growth-based selection using a large population size (5×10^9–10^{10} cells per transfer and a population size of approximately 10^{12} cells per ALE step) without inducing hypermutagenesis. Prior studies[70,71] showed that relying on the natural, low mutation rate of *E. coli*—approximately 10^{-3} mutations per genome per generation[72,73]—and using a population size of ~10^{12} cells allows the exploration all possible single-step mutations across the entire genome at every ALE step[74], which in turn suppresses the evolution of off-target mutations by allowing the evolution of exceedingly rare, high-fitness on-target mutations[71].

Strain-fitness rapidly improved during ALE experiments (Fig. 3j, k), allowing the integration of additional recoded segments (Fig. 2e, Supplementary Method 1). Within only 140–450 generations, performed in 16–50 days, we identified improved variants. For example, ALE identified a beneficial mutation within the previously detected deleterious *ribF-ispH* promoter. This ALE-evolved sense mutation in *yaaY* increased the *ribF-ispH* promoter activity five-fold and drastically improved fitness (Fig. 3f, h, j). As expected based on the low mutation rate applied during evolution[70], whole-genome sequencing of the ALE-

evolved variants indicated the accumulation of only a few mutations within the recoded region (Supplementary Data 5, Supplementary Note 3).

During genome recoding, we iteratively applied multiplexed editing and ALE to maintain viability while integrating additional recoded segments. In total, we performed 89 SynOMICS integration rounds, 113 genome editing cycles, and 680 cumulative days of ALE to construct partially recoded strains (Fig. 4b). We estimate that these experiments cumulatively explored approximately 6.1 trillion variants (Supplementary Note 4). Iterative SynOMICS and debugging cycles in eleven *E. coli* strains overcame lethal fitness issues, restored the integrability of all synthetic chromosomal segments, and enabled us to replace the entire genome across the eleven partially recoded strains with their synthetic 57-codon counterparts.

Debugging enables partial genome assembly but pleiotropic defects remain

Following assembly and testing of the 57-codon genome across 11 partially recoded strains, we used SynOMICS to merge recoded chromosomal regions. We merged chromosomal regions by reversing the integration step in SynOMICS, thereby liberating the entire recoded region from the genome. The liberated linear chromosomal region (150–451 kbp in size, containing 4–12 segments) was then transposed into a BAC vector within the same cells using recombineering (Fig. 4a). Following chromosome fission, we delivered the fission BACs into a recipient cell containing a partially recoded genome. Finally, SynOMICS without any additional modifications allowed us to combine recoded regions (Fig. 2d).

SynOMICS allowed us to assemble recoded chromosomal regions sequentially from distinct donor strains (Fig. 4c). Despite transferring recoded chromosomal segments to a clean genomic background and thus eliminating a few compensatory mutations from previous debugging steps, recoded regions remained viable. Strains with increasingly recoded genomes displayed lower overall fitness, quantified as the maximal attainable optical density in rich bacterial growth medium (i.e., OD at 600 nm, measured in 2×YT broth; Fig. 4c). Strains harboring more than 2.2% of the synthetic genome lost their ability to grow in minimal M9 broth (Fig. 4d).

Iterative genome replacement and debugging allowed us to combine 35 segments from five donor strains, representing 39% of the entire genome. During debugging (Fig. 4c), we observed only 16 newly emerged mutations, of which six resided within the synthetic genome. Out of the 16 mutations, three resulted in a codon targeted for synonymous recoding outside the recoded region, potentially indicating the benefits of the reacquisition of forbidden codons. As a reference, the hypermutagenesis-based troubleshooting of Syn61Δ3 in an earlier study[9] led to the accumulation of 482 off-target mutations[9] and 31 forbidden codons (Supplementary Note 1, Fig. 2b).

As doubling time and the maximal attainable optical density only partially capture the multidimensional nature of fitness landscapes and potential underlying metabolic deficiencies[75,76], we also characterized strain-fitness under 480 additional environments using the Biolog Phenotype MicroArray system[77]. The investigated environments included various carbon, nitrogen, and phosphorus sources, osmotic and chemical stressors. Strains with increasingly recoded genomes displayed more severe deficiencies, which we hypothesize might be behind the observed poor growth and the inability of the performed initial debugging cycles to fully restore fitness (Fig. 4c, d, e, Table 1, Supplementary Note 5).

In sum, with 210 genome editing cycles and 1120 days of laboratory evolution that explored approximately ten trillion variants (Supplementary Note 4), we successfully constructed seven partially recoded strains representing the entire Ec_Syn57 genome and consolidated five synthetic chromosomal segments into a single strain. However, despite extensive debugging, we observed severe fitness reduction.

Multi-omics-based discovery of recoding effects

Fitness degradation during genome assembly indicated cumulative defects that remained undetected in prior analyses. Therefore, we next explored the transcriptome- and translatome-level consequences of synonymous recoding by performing simultaneous genome-transcriptome-translatome profiling across a wide array of recoded strains, including Syn61Δ3. RNA-seq-based comparative transcriptome profiling revealed large-scale changes. Strains harboring the replacement of 33%, 6.9%, 11.7% of their genome with the 57-codon version contained 1122, 663, and 1247 differentially expressed genes (i.e., showing at least two-fold relative transcript level difference and $P < 0.05$, based on $n = 3$), respectively. The comparison of Syn61Δ3's transcriptome to MDS42, in the same manner, indicated 1208 differentially expressed genes (Fig. 5a, Supplementary Fig 12). Comparative translatome profiling using Ribo-seq confirmed the extent of transcriptomics-indicated effects. Ribosome profiling indicated 1118, 627, and 716 differentially translated genes (i.e., showing at least two-fold relative ribosome occupancy difference and $P < 0.05$, based on $n = 3$) on the same set of genomes, respectively. The comparison of Syn61Δ3's translatome to MDS42 indicated 1129 differentially translated genes (Supplementary Fig 13, 14). The availability of RNA- and Ribo-seq data for the same strains allowed us to measure changes in translation efficiency for every parental and recoded gene and compare the resulting dataset to our previous prediction. Unexpectedly, in vivo translation efficiency differences between the recoded and parental genome showed no correlation with in silico predictions ($R^2 = 4.86 \times 10^{-3}$, Supplementary Fig 15).

Since translation efficiency measurements assess the correlation between the number of mRNA reads and ribosome footprints on the given gene, we hypothesized that promoter location and activity changes could influence translation efficiency results. According to this hypothesis, newly generated intragenic promoters would generate artifacts in transcriptomic data by increasing transcript levels without upregulating translation, while antisense transcripts downregulate translation through dsRNA formation, increased mRNA degradation, and the inaccessibility of the ribosome binding site[78,79]. We tested our hypothesis by repeating Cappable-seq on a partially recoded 57-codon strain carrying 36.7% of the entire synthetic genome. We analyzed the location and expression intensity of transcriptional start sites (TSS) and compared the parental and recoded genomes (Fig. 5b, c). In line with our expectation that synonymous recoding only impacts intragenic promoters, from intergenic TSS we detected no difference in total mRNA output between the parental and recoded variant (that is, a total mRNA output of 303,477 transcripts per million (TPM) in the case of the recoded strain, compared to 292,890 TPM in the case of the parental MDS42). In contrast, intragenic TSS were markedly affected. The total sense transcript output from intragenic TSS—which are producing transcripts with the matching direction of the gene they reside in—increased 2.23-fold in the recoded strain, while the transcript output from antisense intragenic TSS increased 4.01-fold compared to the antisense transcription of the parental MDS42. Some of these recoding-induced antisense promoters' mRNA output was comparable to sense transcription for the same gene. Our analysis revealed the same effect on Syn61Δ3's genome (Supplementary Figs. 16, 17, Supplementary Note 6). We also note that during laboratory evolution, one of the sixteen mutated genes in the partial Ec_Syn57 genome was *rnc* (encoding ribonuclease III) (Fig. 4c), the principal endonuclease responsible for dsRNA degradation, indicating that antisense transcription likely contributes to fitness defects. We hypothesize that antisense transcripts in recoded genomes increase mRNA degradation and prevent

https://doi.org/10.1038/s41467-026-74300-9

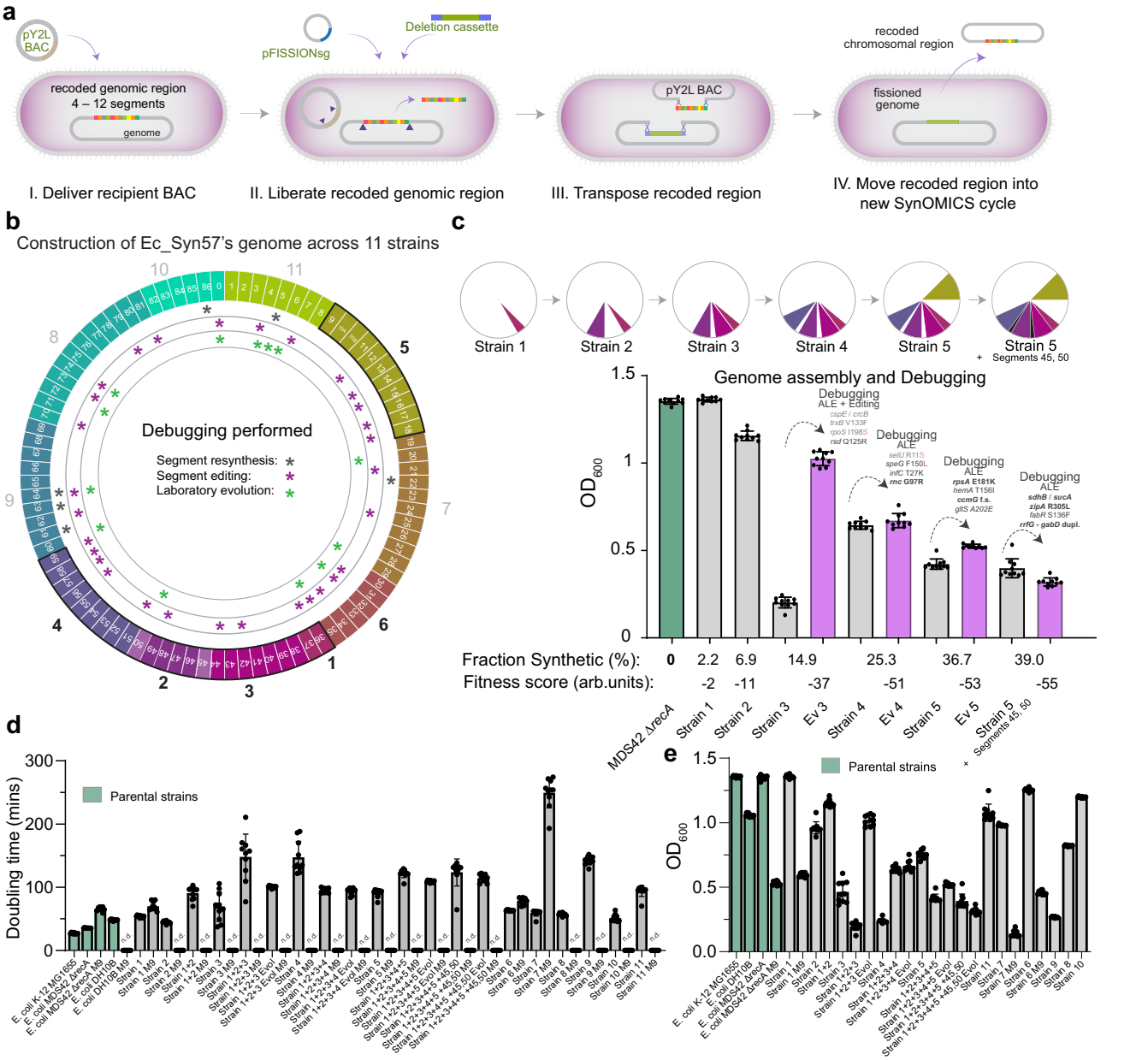

Fig. 4 | Assembly of recoded chromosomal regions indicates cumulative defects. a The reversal of SynOMICS's steps and direction allows genome fission, fusion, and assembly. Chromosome fission is achieved by delivering recipient BAC (pY2L BAC) into cells bearing a partially recoded genome (**I.**), and (**II.**) following the expression of Lambda-Red and Cas9 proteins, chromosome fission is initiated by delivering a four-plex sgRNA expression plasmid (pFISSIONsg) and a SynOMICS deletion cassette. (**III.**) CRISPR/Cas9-cuts liberate the recoded genomic region and linearize the recipient BAC allowing the Lambda-Red-mediated transposition of the recoded region into the recipient BAC while simultaneously sealing the genomic cut. Finally, (**IV.**) the fissioned chromosomal region is delivered into a new recipient cell where it is integrated using the standard SynOMICS workflow (Fig. 2c). **b** The synthetic genome of Ec_Syn57 across 11 partially recoded strains. Numbers indicate the steps in which genomic sections are merged to assemble the final, fully synthetic genome. Segments 45 and 50 (marked as ▲) were added following the merging of sections 1 to 5. To date, sections 1–6 have been combined into a single strain. **c** SynOMICS-based sequential genome assembly and debugging. Pie chart shows genome assembly steps; colored sections mark recoded chromosomal regions transferred. Mutations below the debugging method indicate mutations acquired during debugging. Mutations in gray indicate mutations replaced during genome construction; mutations in red indicate mutations to forbidden codons; while mutations in bold indicate mutations acquired within the recoded region. Bar graph shows mean maximal optical density at 600 nm (OD_{600}) in 2×YT, error bars denote SD based on $n = 10$ independent biological replicates. Magenta bars represent debugged strains while gray indicates strains before debugging. Green represents parental strain. Fitness scores indicate mean fitness defect compared to *E. coli* MDS42 across 480 environments, based on $n = 2$ independent biological replicates. **d**, **e** Growth (**d**) and maximal OD (**e**) of the parental and partially recoded strains and their evolved derivatives in 2×YT and M9 broth. N.d. indicates no growth detected after 14 days based on three independent replicates. Data represents the mean, error bars denote SD based on $n = 10$ independent biological replicates. Source data are provided as a Source Data file.

ribosome docking to mRNAs via dsRNA formation within the RBS region, therby reducing translation and increasing the fraction of a fast-decaying mRNA population inside cells[80]. To provide additional evidence for this hypothesis, we also attempted to delete *rnc* in our most-recoded Strain 5, but observed no viable deletants (i.e., a deletion frequency below 1.1×10^{-10}). Using the same method, the deletion of *rnc* in MDS42 yielded thousands of colonies with 100% deletion efficiency, in line with prior studies[81].

Table 1 | Fitness of strains under diverse environmental conditions

Strain	Fitness score (arb. units) based on 480 environments
Syn61Δ3	−31.87
Strain 1	−2.07
Strain 2	−11.2
Strain 3 Evolved	−37.06
Strain 4 Evolved	−50.61
Strain 5 Evolved	−52.86
Strain 5 + Segments 45 and 50 Evolved	−55.49
Strain 6	−8.03
Strain 7	−20.62
Strain 10	−10.26
Strain 11	−29.7

Fitness scores indicate mean fitness defect compared to *E. coli* MDS42 across 480 environments, measured using the Biolog system, based on $n = 2$ independent biological replicates. Biolog source data is provided in Supplementary Data 2.

Following the sequencing-based discovery of antisense promoters on Ec_Syn57's genome, we individually validated and confirmed the promoter activity of five antisense promoter-regions (Fig. 5d). To explore which specific codon change generates an antisense promoter, we performed codon-scanning at the recoded Seg93 codon of *hypF*. Promoter activity measurements revealed that while the parental wild-type codon conferred one of the lowest antisense transcription levels, the Ser93 AGC → TCT codon change of our 57-codon code induced, on average, a 20.2-fold increase in antisense transcription, the highest across all six serine codons nature uses to encode serine (Fig. 5e).

Based on these measurements and prior studies[78,79,82], we hypothesized that recoding-induced antisense transcription contributes to synonymous-recoding-induced fitness defects, jointly with perturbations in natural intragenic promoters (Fig. 3a) and recoding-induced changes in translation initiation (Supplementary Fig 5). Therefore, we next attempted to multiplex the genome-scale correction of these issues.

Multi-omics-guided evolution fully restores fitness

Based on multi-omics, we revised our genome construction and debugging strategy. Promoter-, translation initiation-, and antisense transcript-mediated effects are all expected to reduce protein production. Therefore, we hypothesized that restoring protein production based on ribosome profiling data would mitigate fitness defects. To debug translation changes, we selected all differentially translated genes of Strain 6, containing Segments 30–35 of Ec_Syn57. We then designed 114 genome editing oligonucleotide libraries to restore translation (Supplementary Fig. 18). Four rounds of ten genome editing cycles, followed by growth-based selection using our terascale protocol, quickly identified beneficial variants. Doubling time measurements for independently evolved variants indicated doubling times matching the parental strain's doubling time (Fig. 5f). Population whole-genome sequencing over the course of editing indicated positive selection for edited variants at almost all target sites, with select edits fixing quickly (Fig. 5g, Supplementary Fig. 19). Selected fast-growing variants contained up to 24 RBS edits and newly inserted promoters upregulating 20 genes. Repeated ribosome profiling indicated that inserted edits frequently reduced gene expression for previously overexpressed genes while previously downregulated genes' expression increased, indicating that multiplexed editing optimized and rebalanced gene expression (Supplementary Fig. 20). For example, at the previously overexpressed *pheS-pheT-ihfA* operon where the debugging of a lethal error necessitated the insertion of a strong constitutive promoter and RBS at prior steps of genome construction resulting in 12.4-fold overexpression for *pheS* compared to the parental expression level, multi-omics-guided debugging restored expression to near parental level (i.e., a 13.9-fold decrease compared to the upregulated state).

We next scaled up our method to debug entire genomes. We initiated ribosome profiling-based debugging of our most-recoded strain (Strain 5), which contains 39% of the synthetic genome and 1118 differentially translated genes. Nineteen genome editing cycles resulted in variants doubling every 64.8 ± 3.09 min and reaching a maximum optical density of 1.01 ± 0.02, a drastic increase compared to the parental strain's 110.6 ± 5.83 min doubling time and 0.31 ± 0.02 maximum optical density. The improved fitness of the evolved variants allowed us to transfer Segments 30-35 of Ec_Syn57 into the debugged Strain 5, generating a partially recoded strain that contains 45.8% of Ec_Syn57's genome (Fig. 5h, Supplementary Method 1).

Using multi-omics analyses across multiple recoded bacterial strains, we discovered that synonymous codon replacements perturb noncoding sequence motifs that regulate gene expression and induce drastic transcriptome and translatome perturbations. To mitigate these effects, we developed a multi-omics-guided directed evolution method that simultaneously troubleshoots hundreds of target genes genome-wide in weeks to restore fitness. We estimate that this method accelerates debugging more than 30-fold compared to ALE alone (Supplementary Note 7). In the remaining five steps of genome construction (Fig. 5h), we are utilizing multi-omics-guided evolution to optimize partially recoded strains as we assemble the fully synthetic genome of Ec_Syn57.

Discussion

We have demonstrated that multi-omics-guided genome synthesis and directed evolution enable genome recoding with a highly deleterious codon set, thereby allowing the detailed exploration of collateral effects of synonymous codon changes. To reveal synonymous recoding-induced effects on synthetic genomes and guide future recoding efforts—using multi-omics—we explored the transcriptional, translational, proteomic, and phenotypic consequences of genome recoding using an array of *E. coli* strains bearing synthetic genomes with modified genetic codes. Using these tools, we discovered previously unseen effects that pose constraints on functional genetic codes and developed genome-scale methods to enable the construction of genomes with a previously unattainable genetic code[16–19]. Our multi-omics-guided debugging strategy relies on three synergistic layers: first, genome reannotation, untargeted proteomics, and ribosome profiling identify gene targets requiring recoding; second, transcriptional start site profiling using Cappable-seq identifies promoters and operon structures that must be preserved during sequence design or restored during debugging; third, ribosome-profiling-guided accelerated directed evolution of differentially translated genes, in combination with laboratory evolution, enables the creation of high-fitness variants.

Our results provide five important insights for future genome recoding projects.

First, we revealed and experimentally confirmed that both our 57-codon genome design and previously constructed recoded organisms harbor non-recoded codons due to frequent genome annotation errors and cryptic translated sequences—a widespread issue not previously recognized in recoded genomes. These codons lead to unassigned codons in the final, recoded organism and induce widespread ribosome stalling[57] (Fig. 3b, d) as they lack translation components for their decoding, sequestering more than 2% of the entire ribosome pool in a recently developed synthetic recoded *E. coli* strain[9], Syn61Δ3. As genome annotation errors and cryptic translated sequences yield unassigned codons, we suggest the proteomics- and ribosome-

https://doi.org/10.1038/s41467-026-74300-9

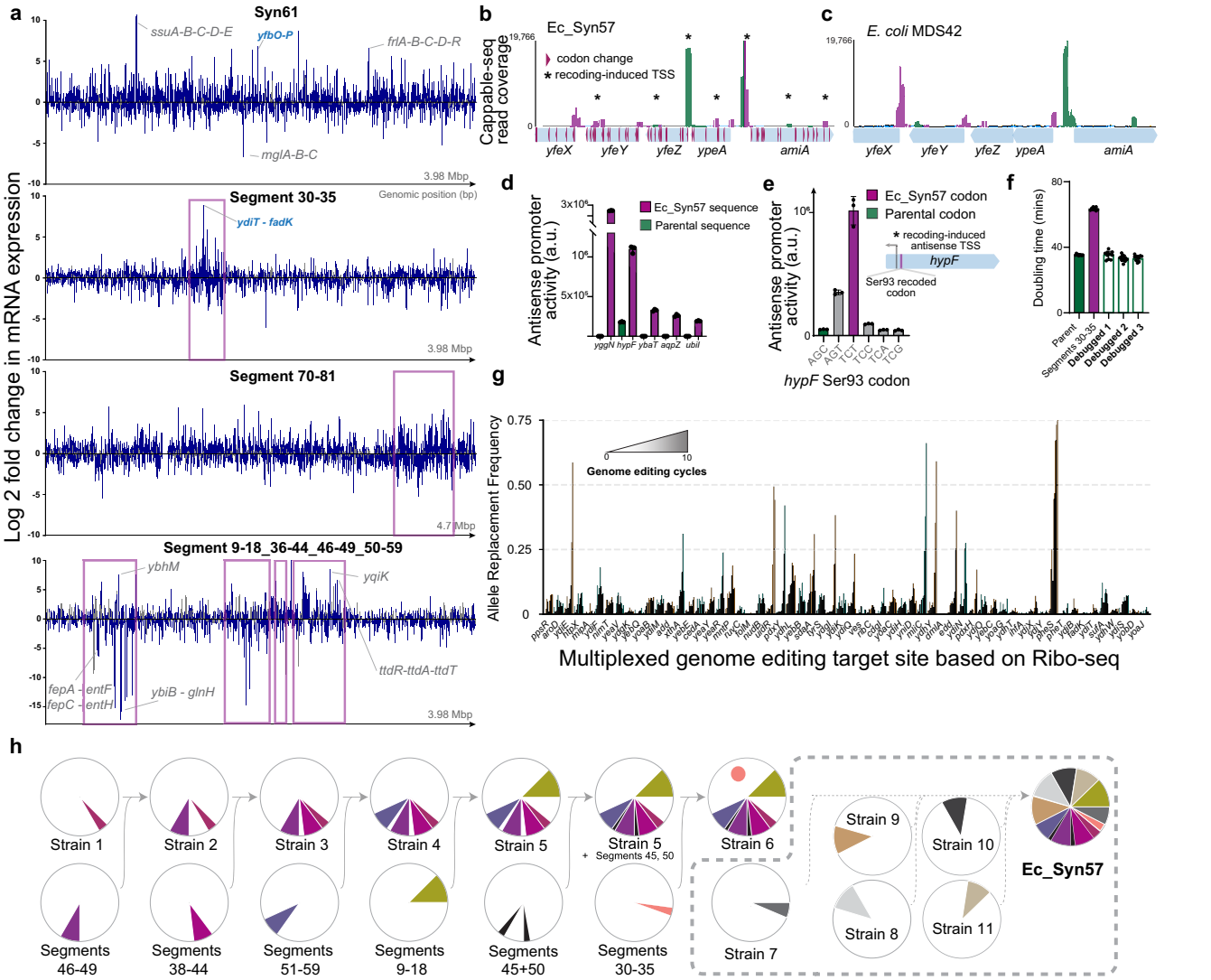

Fig. 5 | Multi-omics reveals widespread gene expression changes due to synonymous recoding and enables rapid fitness debugging. a The recoded genome of Syn61Δ3 and recoded chromosomal regions of Ec_Syn57 display marked differences in mRNA expression. The figure shows differential mRNA expression for all protein-coding genes along the recoded genome of Syn61Δ3 and partially recoded genomes of Ec_Syn57, compared to their parental variant. Recoded, synthetic regions on partially recoded genomes are marked in magenta. Example operons displaying the most significant changes in mRNA expression are highlighted. Bars indicate mean differential expression, calculated using EdgeR from $n = 3$ independent biological replicates. Bars in blue indicate a $P < 0.05$, while bars in gray indicate $P \geq 0.05$ based on EdgeR. **b**, **c** Synonymous recoding induces the widespread appearance of cryptic promoters. Figure shows the position and direction of primary transcripts based on primary transcriptome sequencing on Ec_Syn57 (**b**) and MDS42 (**c**), aligned to their corresponding genome. Recoding-induced transcriptional start sites are marked with a star (*). Magenta marks reverse transcript orientation, while green marks forward transcript orientation. Magenta triangles indicate recoded codons. **d** Antisense promoter activity of intragenic Ec_Syn57 sequences and their parental variant. Data represents mean sfGFP expression levels based on $n = 3$ independent biological replicates, error bars indicate SD. **e** Antisense promoter activity conferred by all six serine codons of the standard genetic code at Ser93 of *hypF* on Ec_Syn57's genome. Data represents average sfGFP expression levels based on $n = 3$ independent biological replicates, error bars indicate SD. **f** Growth of the parental *E. coli* MDS42, the partially recoded, and the final, debugged strains carrying Segments 30-35 of Ec_Syn57 in 2×YT broth at 37 °C. Data represents the mean, error bars denote SD based on $n = 10$ independent biological replicates. **g** Allelic replacement frequencies during MAGE-based multiplexed editing-and-selection cycles using genome-editing oligos targeting all downregulated genes of Strain 6 based on Ribo-seq. Figure is based on a representative sample. Source data are provided as a Source Data file. **h** Genome construction of Ec_Syn57. To date, Strains 1–6 have been constructed, and the integration of the extrachromosomal Segments 30-35 is ongoing. Dashed line indicates future construction steps.

profiling-based discovery of translated sequences[60,63] before genome and genetic code refactoring. As a step towards a generalizable solution, we have developed a ribosome-profiling- and untargeted proteomics-based workflow that detects translated sequences and tracks genome-wide ribosome stalling in engineered genetic codes.

Second, we have shown that synonymous codon replacements induce widespread promoter perturbations in recoded organisms, leading to frequent changes in gene expression and the generation of additional transcripts. As the elimination of seven codons from the standard genetic code resulted in the appearance of ~1 additional promoter in each recoded gene, providing an up to 715% increase in global antisense transcription in a 57-codon genetic code and a 301% increase in a 61-codon code, our results confirm a prior hypothesis[83–85] about how codon preference naturally evolves to reduce transcriptional noise. Further evidence for this hypothesis is provided by the emergence of mutations that generate codons targeted for elimination during genome recoding (Fig. 4c), the preferred reversion of synonymous codon changes of our 57-codon code (Supplementary Fig 17),

and the deleterious effect of select leucine (Fig. 3f-i) and serine codon changes (Fig. 5e). As certain sense codons are highly enriched in bacterial intragenic promoter motifs[86], including the rare AGA and AGG arginine and the TTA leucine codons, their elimination in our 57-codon genetic code might contribute to the observed defects. Importantly, parallel bacterial genome recoding efforts[13,20,50] all targeted codons that contribute less frequently to internal promoter motifs, which might explain their mild fitness effect. In addition, the known role of rare arginine codons in bacterial translation initiation[17] likely further exacerbates gene expression changes in the 57-codon code presented in this work. While intragenic promoter disruption by synonymous codon changes was previously noted for individual loci, the genome-wide creation of additional promoters as a general consequence of synonymous genome recoding—on both the previously reported Syn61 and our Ec_Syn57's genome—is a key finding of this study. We expect that our combination of transcriptional start site and ribosome profiling will provide a tool for future genome synthesis projects to circumvent these effects by preserving important intragenic promoters and regulatory motifs, and through rational codon selection within 5′ UTRs to preserve translation initiation.

Third, our analyses indicate that studying recoding-induced effects on megabase-sized genomic sections can reveal fitness-impacting factors that remain hidden in smaller-scale tests[18,19], including the cumulative impact of antisense transcription and issues caused by genomic annotation errors. This complements genome construction and debugging methods, such as REXER[18,20] and SIRCAS[49], which enable both recoding scheme selection and issue identification.

Fourth, we have demonstrated that omics-guided multiplexed genome engineering in combination with carefully designed laboratory evolution experiments offers a powerful strategy for the debugging of synthetic genomic regions with deleterious recoding schemes. Our results revealed that ribosome profiling enables target identification for subsequent multiplexed genome-editing-based debugging of megabase-sized synthetic chromosomal regions, allowing fitness restoration even for highly deleterious changes that require dozens of compensatory mutations. The utility of ribosomec profiling to discover fitness-decreasing issues on synthetic genomes has also been highlighted during the construction of the Sc2.0 genome[30]. Following genome editing, laboratory evolution allowed us to fine-tune selected genotypes. While previous evolution experiments frequently utilized hypermutator variants[9] or insufficient population sizes[31,50] to evolve compensatory mutation, the terascale (i.e., 10^{12} variant-scale) selection strategy[70,71] established in this work enabled us to evolve high-fitness variants with reduced frequency of deleterious off-target mutations, allowing the parallelized debugging of synthetic sequences. This technology may provide a broadly applicable strategy for industrial use of whole-genome synthesis.

Fifth, we have observed that endogenous mobile genetic elements of commonly used cloning strains frequently invade synthetic genomes, preventing genome assembly. Future genome synthesis projects should therefore avoid using cloning hosts containing active mobile genetic elements or prophages, i.e., *E. coli* DH10B or BL21(DE3)[42,44,47,87], and strains with elevated mutation rates, as also demonstrated by the mobile-genetic-element-free MDS42-based assembly of Syn61's genome.

Limitations of this study include the low number of recoding schemes tested compared to the theoretically possible, astronomical number of possible recoding schemes, i.e., x^y for a single codon, in which x is the number of alternative codons (2 to 6 in the standard genetic code) and y is the number of codons targeted for recoding genome-wide, 2.8×10^2 to 6.4×10^4. Furthermore, despite assessing fitness under 482 conditions, recoded cells might face drastically different environmental conditions outside laboratory environments or in their natural habitat, where *E. coli* cells double once approximately every 15 h, 45 times slower than under laboratory conditions[88]. Finally, the current work focuses on *E. coli* genome construction, and the generalizability of our findings to other organisms remains to be established. Promoter architecture, antisense transcription tolerance, and translation initiation signals vary substantially across bacterial species[89,90], and the fitness consequences of synonymous recoding can therefore depend on the target organism. However, we note that the strong preference for the TTG motif in the −35 promoter element is conserved not only in *E. coli* but also in other *Gammaproteobacteria*, *Firmicutes*, and *Alphaproteobacteria*, suggesting that the elimination of the TTG codon from an organism's genetic code would influence intragenic promoters across a broad range of bacterial phyla—a design constraint that future recoding efforts in these organisms will likely need to account for and analyze using multi-omics prior to genome design.

We also note that eliminating certain genomic codons, followed by deleting their corresponding tRNAs and release factors, might not directly translate to fully liberated codon channels. The existence of hundreds of TCA and TCG codons on the recently developed Syn61Δ3 strain lacking $tRNA^{Ser(UGA)}$ and $tRNA^{Ser(CGA)}$, the efficient translation of TCA and TCG codons by the *E. coli serU* $tRNA^{Ser(CGA)}$ and TCT codons by overexpressed viral $tRNA^{Ser(UGA)}$ [8], and recent reports showing how near-cognate suppression can impact the fidelity of protein production in recoded organisms[13,57], indicate that crosstalk between remaining translational components can mask liberated codon channels. If this happens, we envision that the use of engineered tRNAs that avoid RNA modifications known to promote relaxed tRNA selection by the ribosome[13,91,92] and the replacement of cellular ribosomes with hyperacute variants[93] to increase translational fidelity will provide a solution.

We expect these results to enable troubleshooting cells harboring refactored genomes, including engineered genetic codes for therapeutic and industrial applications. Below, we conclude our observations for future genome designs and guidelines for implementing SynOMICS in other organisms:

Genome design

- As genomic annotations errors, incomplete genome databases, and cryptic translated sequences yield unassigned codons, proteomics- and ribosome-profiling-based discovery of translated sequences[60,63,94] is necessary before genome and genetic code refactoring. In addition to omics-based methods, high-resolution Tn-seq can serve as an additional confirmation for coding-sequence boundaries and reveal the fitness consequences of loss-of-function mutations, as demonstrated during the design and minimization of the *Mycoplasma mycoides* JCVI-syn3.0 genome[26,58].
- Prior to genome design, accurately identifying promoters and operon structures is necessary using simultaneous short- and long-read Cappable-seq for prokaryotic genomes and ReCappable-seq for eukaryotic genomes[54,55,95].
- Following promoter identification, experimental and computational transcriptional factor and ribosome binding site analyses can reveal essential regulatory sequences that need to be preserved during genome refactoring and sequence design[86,96–98].

Genome synthesis, assembly, and troubleshooting

- Projects should minimize DNA synthesis errors and avoid using cloning hosts containing active mobile genetic elements, prophages, or hypermutator alleles[44,87].
- Transcriptome and ribosome profiling, in combination with genomic machine learning and metabolic models[76,99,100] and phenotypic microarray screens[76] to discover and prioritize fitness-decreasing errors for multiplexed troubleshooting, can yield high-fitness variants, paving the way for industrial use.

We anticipate that future bacterial genome design projects attempting to achieve codon compression beyond the elimination of a select few codons[20,21,101] should focus on developing and implementing algorithms that reduce changes in promoter activity, regulatory functions, and translation rate and avoid generating unwanted cryptic promoters and translated ORFs. We hypothesize that the development of genome language models capable of forecasting transcriptional and translational effects and integrating these predictions to refactor codon composition would minimize the negative fitness impact of synonymous genome recoding[52,53,100,102–109]. Toward this goal, we provide high coverage, high-quality, openly available multi-omics datasets for a range of partially and fully recoded genomes.

The updated design of Ec_Syn57 contains 62,347 target codons for recoding—excluding 342 codons in silent genes based on our omics analyses—of which 62,029 have been recoded to date, 99.4% of all (Supplementary Fig. 10). A potential challenge in completing Ec_Syn57's genome is the time needed to iteratively debug partially recoded strains in the next five steps of genome construction. However, based on our omics-based workflow, which now allows significantly faster growth restoration than ALE before, we expect no difficulty finishing genome assembly. Once complete, the resulting Ec_Syn57Δ7 strain will provide a broadly applicable chassis for genetic code expansion in fundamental research and biotechnology. We also anticipate that our genome construction workflow and the omics datasets generated in this work will significantly reduce the cost and time required for future genome synthesis efforts. Follow-up works might translate and validate our results in ongoing prokaryotic and eukaryotic genome recoding projects[25,27,32,34,49,110–113].

Methods

Bacterial media and reagents

Adaptive Laboratory Evolution experiments utilized Lysogeny Broth Lennox (LBL) supplemented with Tris/Tris-HCL as a buffer, prepared by dissolving 10 g/l tryptone, 5 g/l yeast extract, 1.5 g/l Tris/Tris-HCL, and 5 g/l sodium chloride in deionized H_2O and sterilized by autoclaving. Bacterial cell cultures and genome editing experiments utilized 2×YT media consisting of 16 g/l casein digest peptone, 10 g/l yeast extract, and 5 g/l sodium chloride. 2×YT agar plates were prepared by supplementing LBL medium or 2×YT with agar at 1.6% w/v before autoclaving. Super Optimal Broth (SOB) was prepared by dissolving 20 g/l tryptone, 5 g/l yeast extract, 0.5 g/l sodium chloride, 2.4 g/l magnesium sulfate, and 0.186 g/l potassium chloride in deionized H_2O, and sterilized by autoclaving. Minimal M9 broth with 2% glucose and 0.5 μg/ml thiamine was purchased from Teknova Inc (M8000).

Bacterial genome sequencing and annotation

Genomic DNA from overnight saturated cultures of isogenic bacterial clones was prepared using the MasterPure™ Complete DNA and RNA Purification Kit (Lucigen) according to the manufacturer's guidelines and sequenced at SeqCenter (Pittsburgh, PA, USA). BACs containing individual recoded chromosomal regions were isolated using the ZR BAC DNA Miniprep Kit (Zymo Research) based on the manufacturer's protocol. Sequencing libraries were prepared using the Illumina DNA Prep kit and IDT 10 bp UDI indices with a target insert size of 280 bp and sequenced on an Illumina NextSeq 2000 or on an Illumina NovaSeq X Plus, producing 150 bp paired-end reads. Demultiplexing, quality control, and adapter trimming were performed with bcl-convert (v4.2.4). Reads were then trimmed to Q28 using BBDuk from BBTools and aligned to their corresponding reference by using Bowtie2 2.3.0[114] in --sensitive-local mode. Single-nucleotide polymorphisms (SNPs) and indels were called using breseq (version 0.36.1)[115]. Only variants with a prevalence higher than 75% were voted as mutations. Following variant calling, mutations were also manually inspected within the aligned sequencing reads using Geneious Prime® 2023.2.1.

We performed de novo sequencing and hybrid genome assembly on all parental strains, strains with completed recoded chromosomal sections, and SynOMICS fissioned and fused, partially recoded strains. Hybrid genome assemblies were performed as described earlier[8]. Specifically, we generated 300–600 Mbp of Oxford Nanopore (ONT) long-reads by PCR-free library generation (Oxford Nanopore, UK) on a MinION Flow Cell (R9.4.1) and 3–15×10^6 150 bp paired-end reads on an Illumina NextSeq 2000 or on an Illumina NovaSeq X Plus. Quality control and adapter trimming were performed with bcl2fastq 2.20.0.445 and porechop 0.2.3_seqan2.1.1 for Illumina and ONT sequencing, respectively. Next, we performed hybrid assembly with Illumina and ONT reads using Unicycler 0.4.8 with the default parameters. Finally, the resulting single, circular contig representing the entire genome was manually inspected for errors in Geneious Prime® 2023.2.1. and annotated based on sequence homology using the BLAST function implemented in Geneious Prime® 2023.2.1. based on *E. coli* K-12 MG1655 (NCBI ID: U00096.3) as reference. Protein-coding gene annotations between the original and updated genome versions were performed using the compare_annotations.py script available from https://github.com/rrwick/Compare-annotations (developed by Ryan Wick, University of Melbourne, Melbourne, Australia). Gene essentiality in rich bacterial medium was determined based on Ref[58]. We note that the previously reported[20]. 51 bp insertion between *mrcB* and *hemL* of *E. coli* MDS42, not reported in the original AP012306.1 GenBank sequence, was confirmed by our analysis.

Segment DNA synthesis and assembly

Chromosomal segments of Ec_Syn57 were synthesized and assembled into BAC vectors by BGI Research (Shenzhen, China) and GenScript USA Inc. (Piscataway, USA, and Nanjing, China). Segments 50, 60, 62, 63, 65, 66, 67, 68, and 75 were synthesized and assembled by BGI Research, while Segments 2, 5, 6, 8, 10 A, 10B, 13, 19, 22, 35, 37, 45, 53, 55, 57, 61, 64, 73, 86, and 89 were synthesized by GenScript USA Inc. Segment synthesis at BGI Research utilized 55-80 nucleotide-long oligonucleotides as building blocks that shared 20 nucleotide terminal homology, synthesized using the standard phosphoramidite DNA synthesis procedure[116]. These 55-80-nt ssDNA oligonucleotides were subsequently assembled into 600 bp dsDNA fragments using polymerase cycling assembly (PCA). Following agarose gel electrophoresis-based size verification, assembled dsDNA products were purified and cloned into a pBR-based plasmid containing an Amp-R marker gene using isothermal assembly and transformed into *E. coli* DH10B cells. Following plasmid extraction, each plasmid was validated using Sanger sequencing. Next, the sequence-verified 600 bp dsDNA fragments were assembled into 1000–2500 bp products, cloned into a pBR-based plasmid containing a kanamycin resistance gene as a selectable marker, using isothermal assembly, and sequence-validated using Sanger sequencing. Next, two or three 1000–2500 bp fragments were assembled into 4.5–8 kb fragments using the same strategy, sequence validated, and used for subsequent bacterial artificial chromosome assembly in yeast. Segment synthesis at GenScript USA Inc. followed the same strategy but with minor modifications. We synthesized and assembled 5–8 kbp intermediate dsDNA fragments, sharing 500 bp terminal homology with the neighboring fragment or, in the case of segment ends, with the pYES2L vector. These intermediate fragments were directly constructed from chemically synthesized ssDNA fragments and cloned into either pUC57 high-copy or pCC1 low-copy BAC-based vectors. The use of low-copy vectors allowed us to mitigate frequent instability issues caused by the toxicity of endogenous *E. coli* genes in *E. coli* cloning hosts[117,118]. Following fragment assembly and sequence validation, yeast assemblies were performed in *Saccharomyces cerevisiae* W303, BY4733, or MaV203 using the intermediate dsDNA fragments, mixed at a 6:1 molar ratio of fragment to vector and assembled in yeast[119], followed by sequence validation using Illumina whole-BAC sequencing. The target BAC vector of yeast assemblies,

pYES2L (Supplementary Data 1), was de novo synthesized based on the pYES1L-URA (Addgene plasmid #84301) plasmid and contains an additional oriV origin-of-replication for potential conditional amplification in EPI300 cells (Epicenter). Following yeast assembly, segments at BGI Research were isolated from 4 ml yeast culture by lysis and then electroporated into *E. coli* MDS42 Δ*recA* cells carrying the genomically integrated Segment 69 of Ec_Syn57 and, following overnight recovery at 32 °C, plated on LBL agar plates containing 80-100 µg/ml spectinomycin. Plates were incubated at 32 °C until colony formation. Following yeast assembly and sequence validation using Illumina whole-BAC sequencing, the pYES2L-Segment assemblies were electroporated into *E. coli* MDS42 Δ*recA* cells and, following overnight recovery at 32 °C, plated on 2×YT agar plates containing 80 µg/ml spectinomycin, and the plates were incubated at 32 °C until colony formation. Following MASC PCR validation using the MASC PCR primer set described in Supplementary Data 1 and BAC extraction using the ZR BAC DNA Miniprep Kit (Zymo Research) based on the manufacturer's protocol, pYES2L segment assemblies were validated using Illumina whole-BAC sequencing. Our workflow allowed us to assemble all but one of the 87 segments smoothly. Segment 10 was challenging to synthesize, and we therefore constructed it by splitting its sequence into two fragments, a 21 kbp part A and a 24 kbp part B. Finally, we also constructed a 90.1 kbp neochromosomal region carrying genes and operons that were disrupted by DNA synthesis errors on 24 previously synthesized segments.

Clonal, DNA-synthesis-error-free DNA constructs for all debugging, genome editing, and genome construction steps were sourced from commercial DNA synthesis companies (i.e., GenScript USA, Genewiz, Twist Bioscience, Ranomics), and we relied on these companies' sequence validation to confirm DNA sequence in all cases. Plasmids for expression assays and promoter activity measurements were synthesized by GenScript using full-plasmid synthesis and delivered as error-free, clonal, sequence-validated constructs. All utilized DNA constructs are described and reported in Supplementary Data 1 and available from the corresponding authors upon request

Transcriptome and translatome analysis

We explored transcriptomic and translatomic changes by collecting cells from mid-exponential cultures using rapid filtration, then performing RNA-seq and ribosome profiling on the same samples. As we sought to explore the native transcriptome and translatome of analyzed cells, and the addition of antibiotics and cell-harvest conditions are known to influence ribosome profiling experiments[120,121], we performed cell growth and sample collection without the presence of antibiotics and carried out all cell harvest steps immediately following cell growth. To minimize changes in cell state, we implemented the rapid filtration technique described in ref. 120. Experiments were performed in three independent replicates, initiated from separate starter cultures. Cells from single colonies were inoculated into Lysogeny Broth Lennox (LBL) and grown at 37 °C, 250 rpm for 18 h, and then diluted into 500 ml 37 °C prewarmed LBL to an OD_{600} of 0.02 in a 2000 ml baffled Erlenmeyer flask with a vented cap. Cells were grown at 37 °C, 250 rpm until cultures reached OD_{600} = 0.40–0.45. Next, cells were immediately collected by pouring 400 ml culture into a Nalgene™ Rapid-Flow™ filter unit with 0.45 µm PES membrane (169-0045) and vacuum filtered. Cells were then washed off the filter membrane using 1 ml prewarmed LBL, and immediately dripped into liquid N_2. Once frozen, cell suspensions in liquid nitrogen were supplemented with 1 ml ice-cold lysis buffer (containing 10 mM $MgCl_2$, 100 mM NH_4Cl, 20 mM Tris, pH 8.0, 0.1% NP-40, 0.4% Triton X-100, 100 U/ml of RNase-free DNase I (New England Biolabs, USA), 0.5 U/µl of SUPERase·In RNase Inhibitor (Invitrogen, catalog number AM2696) and 2 mM chloramphenicol). Frozen cell samples with lysis buffer were pulverized using a liquid nitrogen pre-chilled Cell Crusher Tissue Pulverizer (Cell Crusher, 607KSL) according to the manufacturer's protocol. Samples were kept in liquid nitrogen throughout the entire procedure. Cell extracts were then centrifuged at 14,000 × g for 20 min at 4 °C to remove cell debris. The supernatants were collected, and RNA concentration in each sample was determined using the Qubit RNA High Sensitivity Assay Kit (Invitrogen, USA). Ribosome profiling was performed according to the protocol described earlier[120]. Specifically, the clarified cell lysates were digested with micrococcal nuclease from *Staphylococcus aureus* (MNase, Roche, USA) for 1 h at 25 °C under continuous agitation based on the manufacturer's protocol. Digestions were stopped by supplementing samples with EGTA (Millipore-Sigma, USA) to a final concentration of 6 mM. Next, the monosome fraction was purified by size exclusion chromatography on Amersham Micro-Spin S-400 HR columns (Cytiva, USA) according to the manufacturer's protocol. Following monosome isolation, samples were denatured by adding SDS to a final concentration of 1%, and the RNA fraction was extracted and resolved on a 15% TBE-urea denaturing gel (Invitrogen, USA) according to the manufacturer's protocol. The RNA band between 15 and 45 bp, corresponding to ribosome footprints, was excised, gel-purified, and resuspended in 15 µl of 10 mM RNase-free Tris buffer, pH 7.0. The 3′ end of isolated RNA was subsequently dephosphorylated using T4 polynucleotide kinase (New England Biolabs) at 37 °C for 1 h based on the manufacturer's protocol. Next, the RNA was purified and resuspended in 10 µl of 10 mM RNase-free Tris buffer, pH 7.0. Ribo-seq RNA libraries were prepared for sequencing using the SMARTer® smRNA-Seq Kit for Illumina (#635030) from Takara Bio USA, Inc., according to the manufacturer's protocol. Finally, sequencing-ready libraries were sequenced using Illumina's standard sequencing protocol on an Illumina NovaSeq 6000 to generate more than 30 million single-end 50 bp reads from each sample.

We obtained matched strand-specific (stranded) RNA-seq libraries for each Ribo-Seq library by extracting total RNA from the clarified cell lysates in parallel with Ribo-seq library preparations. Sequencing libraries were prepared according to our earlier protocol[8], with minor modifications. Specifically, the rRNA-depleted and fragmented total RNA fraction was converted to sequencing libraries using the SMARTer® smRNA-Seq Kit for Illumina (Catalog number #635030) from Takara Bio USA, Inc., according to the manufacturer's protocol that generates strand-specific sequencing libraries. Library concentrations were measured using Qubit and analyzed for fragment length distribution using an Agilent 2100 Bioanalyzer. Finally, RNA-seq libraries were sequenced on an Illumina NovaSeq 6000 to generate more than 25 million single-end 50 bp reads from each sample. Demultiplexing, quality control, and adapter- and poly-A-trimming were performed with bcl-convert (version 3.9.3). cDNA reads were aligned to their corresponding reference by using Bowtie2 2.3.0[114] in --very-sensitive-local mode. Sequencing reads mapping to rRNA genes were excluded from downstream analyses, and ribosome-footprint length distributions were determined based on ribosome footprint reads mapping to protein-coding CDSes. Next, we performed metagene analysis following the protocol described in ref. 120 to determine the offset position of the ribosomal P-site within Ribo-seq sequencing reads for each sample. Following metagene analysis, ribosome-footprint P-sites were mapped to their corresponding genomic reference sequence. Finally, translation levels of all protein-coding genes were determined as the number of ribosome footprints mapped by P-site location to each protein-coding CDS region. Transcription levels were evaluated by mapping RNA-seq reads using Bowtie2 2.3.0[114] in a very-sensitive-local mode. Read counts and expression metrics were determined for each CDS, and differential expression analysis was performed using the EdgeR package (version 4.0.12, ref. 122) using the standard settings. Differential expression analyses for Ribo-seq-based multiplexed debugging were performed directly from read counts due to EdgeR's inability to calculate differential expression for genes with near-zero expression levels. To avoid excluding genes due to low ribosome footprint coverage, we selected all genes at least two-fold

differentially expressed for follow-up oligo-editing. The calculation of the sequestered ribosome fraction on Syn61Δ3's genome was performed by dividing the number of ribosome footprints mapped to genes containing forbidden codons by the number of all ribosome footprints that map to all protein-coding CDS regions on the genome. As forbidden, non-recoded codons accumulate ribosomes before the unassigned codon due to stalling and ribosome collision and result in a wide peak across non-recoded genes (i.e., *rpoS*, as illustrated in Fig. 3b), this approach allowed a better detection of stalled ribosome fraction than the analysis of only the ribosome footprint A-(-aminoacyl-)-sites mapped to non-recoded codons. The validation of proteomics-detected intragenic open reading frames was performed using a ten-fold increased Ribo-seq coverage, after resequencing select, previously constructed Ribo-seq libraries to generate more than 500 million single-end reads. Relative ribosome A-(-aminoacyl-)-site occupancy of sense codons in *E. coli* MDS42 and Syn61Δ3 was determined based on the number of Ribo-seq reads mapping to a given codon position and the frequency of the given codon across all analyzed coding sequences. Analyses of codon occupancies were based on the set of protein-coding genes (CDS) in the given genome that (1) are not annotated as pseudogenes and (2) do not include internal in-frame stop codons. Codons overlapping between neighboring genes were also excluded. Finally, we also excluded protein-coding genes with low translation levels by excluding all genes displaying a normalized ribosome occupancy level below 50.0 (*i.e.*, RPKM ≤ 50.0). We evaluated codon occupancy at the A site for ribosome footprints in each sample, following the method described by Nedialkova and Leidel (2015)[123]. Specifically, for each sample, we first determined the A-site offset within ribosome footprints using metagene analysis by selecting the P-site offset that aligned and formed a peak in coverage at the Translation Initiation Site (i.e., the AUG start codon). A-site coordinates were derived from P-site locations determined by metagene analysis. A-site fractional coverage over codon frequency values in each sample and for all sense codons was determined as the $\log_2$ ratio of the codon frequency normalized codon coverage over codon frequencies averaged at the A-site. Finally, translation efficiency was evaluated as the ratio between Ribo-seq and RNA-seq coverage for the given protein-coding gene.

To ensure accurate genome and protein-coding gene annotations, we utilized de novo assembled, reannotated, and manually inspected reference genome sequences in all transcriptomic, translatomic, and proteomic analyses. Gene annotations were transferred based on sequence homology using the BLAST function implemented in Geneious Prime® 2023.2.1. from the genome of *E. coli* K-12 MG1655 (NCBI GenBank ID U00096.3) and manually inspected for errors, as indicated in section Bacterial genome sequencing and annotation above.

Translation Initiation Rate (TIR) predictions for each protein-coding gene of *E. coli* MDS42 and Ec_Syn57 were performed using the Ribosome Binding Site Calculator Version 2.1.1 (available at www.denovodna.com)[52,53,102,103]. We predicted TIR values for each coding sequence by extracting the complete sequence of each CDS, including a 100 bp-long UTR, and TIR values were predicted at the annotated start codon of each CDS. The relative TIR for each CDS in its recoded form was determined by comparing the resulting TIR dataset from Ec_Syn57 to the TIR values obtained for the same CDS in the parental *E. coli* MDS42.

Cappable-seq sample preparation and analysis

We identified transcriptional start sites across the genome of *E. coli* MDS42 and its 57-codon, partially recoded derivative containing Segments 1 to 8, Segments 9 to 18, 36 to 44, 46 to 49, and 51 to 59 using Cappable-seq, a transcriptomic method that enriches and sequences nascent, unprocessed triphosphorylated mRNAs originating directly from bacterial promoters. mRNA samples for Cappable-seq analysis were collected by inoculating both strains from single colonies into Lysogeny Broth Lennox (LBL) and grown at 37 °C, 250 rpm for 18 h. Cultures were then diluted into 50 ml of 37 °C prewarmed LBL to an OD_{600} of 0.02 in a 300 ml baffled Erlenmeyer flask with a vented cap. Cells were grown aerobically at 37 °C, 250 rpm until cultures reached OD_{600} = 0.43–0.47, submerged in ice-water slurry for one minute with constant agitation, and cells were immediately collected by centrifugation at 5000 × g at 4 °C for 8 min. Next, following the removal of the supernatant, cell pellets were immediately dropped into liquid nitrogen. Samples were stored at −80 °C until RNA extraction. Cell samples were prepared in duplicates. Total RNA from frozen samples was extracted by using the RNeasy Midi Kit (Qiagen, USA) according to the manufacturer's instructions with in-column DNase treatment with the RNase-Free DNase Set (Qiagen, USA). Total RNA samples were then converted to Cappable-seq libraries according to New England Biolabs's published protocol[54] available from the manufacturer (New England Biolabs) at www.neb.com/en-us/protocols/2018/01/19/cappable-seq-for-prokaryotic-transcription-start-site-determination, with minor modifications. Modifications included the use of the Zymo Research RNA Clean & Concentrator-5 kit (R1015, Zymo Research, USA) for RNA purification and recovery instead of AMPure XP beads and the use of heat-denaturation (80 °C) to release streptavidin-bound 3'-Desthiobiotin-GTP (New England Biolabs, N0761) capped RNA. Following enrichment and decapping, RNA samples were converted to sequencing libraries using the NEBNext Small RNA Library Prep kit for Illumina (New England Biolabs, E7300) using the manufacturer's protocol. Sequencing libraries were then sequences on a NextSeq 2000 sequencer using the Illumina P3 100 cycle flow cell and the P2 100 cycle flow cell (Illumina, USA) to generate approximately 0.4-1.5×10^9 50 bp paired-end reads. We applied ultradeep sequencing (i.e., 1000–4000× average coverage across the entire genome, approximately tenfold higher than in prior Cappable-seq experiments[54]) to identify even weakly active transcriptional start sites (TSS) with high confidence. Non-mapping reads, formed due to adapter-dimer formation in the NEBNext Small RNA Library Prep kit, were discarded from follow-up analyses. We identified TSS across each strain's corresponding reference genome by mapping R1 reads using Bowtie2 2.3.0[114] in local mode with -L 16, implemented in Geneious Prime® 2023.2.1. Alignments were exported as.bam files and analyzed using the previously described suite of programs for the identification and filtering of TSS data. Scripts and data analysis protocols are freely available on GitHub (https://github.com/elitaone/cappable-seq) and described in references[54,95]. Following the identification of genomic TSS and their direction, each TSS's mRNA output was quantified as the TPM (transcript per million) of Cappable-seq reads originating from a given TSS. Finally, TSS coordinates were binned based on their location (intragenic or intergenic, and residing within the recoded chromosomal region or outside[54]), and each category was quantified, followed by averaging each category across the two independent replicates. We utilized the same data analysis workflow to measure antisense transcription using our stranded RNA-seq data.

We relied on previously generated Pacific Biosciences long-read SMRT-Cappable-seq data from *E. coli* K-12 MG1655 to analyze operon structures and identify their corresponding promoter on the genome of Ec_Syn57. Sequencing data (Sample GSM3290315, MG1655_rich_PacBio RS II, generated in rich bacterial medium and harvested in late exponential phase (*i.e.*, an OD_{600} = 0.55-0.6)) was downloaded from Gene Expression Omnibus dataset GSE117273[55]. Pacific Biosciences RS II circular consensus sequencing (CCS) reads were mapped to Ec_Syn57's genome using Minimap4 (version 2.24) using -t 10 -x map-pb as settings, implemented in Geneious Prime® 2023.2.1. TSS coordinates were identified as the most 5'-mapped nucleotide of each sequencing read. Synonymous recoding-impacted promoters were manually identified as sequences 0-80 bp upstream of each TSS that overlapped with intergenic mutations or codon changes on Ec_Syn57's genome. Promoter troubleshooting DIvERGE oligonucleotides based

on SMRT-Cappable-seq-identified promoter locations were designed to cover identified promoter regions and synthesized by GenScript using the standard phosphoramidite DNA synthesis procedure[68].

Promoter transcriptomics

We synthesized 12 select promoters in their wild-type parental, *E. coli* MDS42-derived and 57-codon, Ec_Syn57-derived versions, and individually cloned them into a pCC1-4k single-copy BAC, harboring the *E. coli* F plasmid origin-of-replication vector, using whole-plasmid synthesis (GenScript). The complete sequence of the 24 plasmids is available in Supplementary Data 1. We analyzed these promoter variants' mRNA output in *E. coli* MDS42 cells using coupled DNA- and mRNA-sequencing in five independent biological replicates sequenced in two technical replicates (*i.e.*, 10 replicates in total for each promoter variant). To measure RNA expression in a multiplexed manner, we first barcoded each plasmid library member downstream of the assayed promoter using a 30-bp barcode containing randomized (i.e., N) positions. Each barcode was sample-specific and contained degenerate positions to measure DNA and mRNA levels of multiple variants from the same promoter-plasmid. Consequently, every promoter variant in our screen was assayed in multiple biological and technical replicates and also multiplexed across multiple members of the plasmid library in each experimental replicate. Following plasmid synthesis and purification, plasmids to be barcoded were diluted to 2 ng/µl and PCR amplified using the primers containing the full barcode sequence using KAPA HiFi HotStart ReadyMix (KK2602, Roche, USA). Barcoded PCR products were run on a 1.2% agarose gel, and the amplified plasmid DNA was extracted using the Monarch® DNA Gel Extraction Kit (New England Biolabs, USA) according to the manufacturer's protocol. Barcoded plasmids were circularized by mixing 4 µl of the eluted, extracted PCR product with 2 µl of NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs, USA) according to the manufacturer's protocol. Following assembly and DNA cleanup, the barcoded plasmids were individually transformed into electrocompetent TransforMax EPI300 (LGC Bioresearch Technologies, USA) cells and, following 1 h of recovery, plated on LBL agar plates containing 12.5 µg/ml chloramphenicol. Next, we extracted plasmids from 16 independent colonies from each construct (16×24 in total), using the 96-Well Plate Plasmid DNA Mini-Preps Kit (Vacuum Based, B814152-0005) from Bio Basic, Inc., according to the manufacturer's protocol. Plasmids were then transformed into electrocompetent *E. coli* MDS42 cells in five independent replicates to generate the assayed promoter-plasmid library. To generate sequencing libraries, overnight cultures of the promoter library, grown in LBL broth containing 12.5 µg/ml chloramphenicol at 37 °C aerobically, were diluted 1:200 into 100 ml LBL containing 12.5 µg/ml chloramphenicol, and incubated at 37 °C under continuous shaking. Upon reaching OD_{600} of ~0.8, cultures were split into two technical replicates, quickly chilled on ice under constant agitation, and 50 ml from each culture was pelleted at 4 °C for subsequent DNA and RNA extraction. Plasmid DNA was extracted from cultures using the EZ-10 Spin Column Plasmid DNA Miniprep Kit (Bio Basic, Inc.) according to the manufacturer's protocol and the barcode region was amplified for Illumina amplicon sequencing using primers 5'-ATCGCAAACGTCACGGCTAATGGAAGCGTTCAACTAGCAG and 5'-GTGACATTTATGACGCGGGCCCGCGTCATAAATGTCACAC, followed by barcoding PCRs to multiplex samples. Total RNA from each sample was extracted using the PureLink RNA Mini Kit (Invitrogen, USA) and converted to first-strand cDNA using SensiFAST First-strand cDNA synthesis kit (BIO65054, Meridian Life Science) with 1 µg RNA as input. Amplicon RNA-seq libraries were constructed in the same way as plasmid DNA libraries outlined above, but 5 µl of the synthesized cDNA was used as the template for the amplification of the barcode region. Finally, purified amplicon DNA was sequenced with 20% PhiX control DNA spike-in on an Illumina MiSeq instrument using the MiSeq v2 chip to generate 250 bp paired-end reads. Relative expression levels were determined using the following steps. First, raw Illumina reads were processed to obtain per-barcode read coverage by merging paired-end reads using bbmerge from bbmap ver. 38.82. Next, reads were trimmed to remove all bases flanking the 30 bp barcode sequences using cutadapt ver. 3.4, and each barcode was counted across all samples. Next, barcode counts were first corrected for library sequencing depth by dividing the read count for the specific barcode by the number of reads obtained for the given DNA-seq or RNA-seq library. Next, the relative expression of each barcoded promoter was obtained by dividing the sequencing-depth-corrected RNA-seq barcode count by the sequencing-depth-corrected DNA-seq barcode count generated from the same culture and technical replicate. Finally, the mean relative expression for each specific promoter barcode was calculated across all replicates, and the resulting data were based to plot promoter relative expression levels for each promoter of interest. All utilized promoter-plasmid and barcode sequences are available in the Supplementary Data 1 of this work; the Source Data file contains the source data obtained from promoter transcriptomics experiments.

Fluorescence-based promoter activity assays

We utilized a kanamycin-resistant pUC-based sfGFP plasmid design to measure promoter activities for selected sequences and validate our transcriptomics-based findings. We individually synthesized and cloned all selected sequences in front of a strong RBS–sfGFP construct using whole-plasmid synthesis (GenScript), followed by transformation into the parental *E. coli* MDS42 Δ*recA* cells. For fluorescence measurements, we diluted the promoter-sfGFP construct-containing cultures 1:100 from overnight starters into 15 ml 2×YT in 300 ml shake flasks containing the corresponding antibiotic and cultivated for 24 h at 37 °C, 200 rpm aerobically. We then determined sfGFP expression levels in samples by pelleting and washing 1 ml of the culture with PBS (Invitrogen) and resuspending cell pellets in 150 µL BugBuster Protein Extraction Reagent (MilliporeSigma). Reactions were incubated for 15 min at room temperature and then spun down at 16,000 × g for 10 min. The fluorescence of the BugBuster-treated supernatants and the OD_{600} of the original culture were measured using the Synergy H1 Hybrid Reader (BioTek) plate reader using the bottom mode analysis with an excitation at 480 nm and emission measurement at 515 nm. Fluorescence values were normalized based on OD_{600} data. Promoter activity measurements were performed in three independent replicates.

SynOMICS-based genome assembly

We performed SynOMICS genome construction cycles in *E. coli* MDS42 Δ*recA* cells or in *E. coli* DH10B cells, containing the pRedCas2 plasmid. The pRedCas2 contains the cI857 repressor-pL promoter-driven Lambda-Red operon, consisting of *gam*, *exo*, and *bet* (from pORTMAGE-3[69], Addgene plasmid #72678; http://n2t.net/addgene:72678; RRID:Addgene_72678), a p15A origin-of-replication, a constitutively expressed chloramphenicol resistance marker and SpCas9 and tracrRNA (from pCas9[47,124], Addgene plasmid #42876). BACs containing individual recoded chromosomal regions were isolated using the ZR BAC DNA Miniprep Kit (Zymo Research) based on the manufacturer's protocol, and 500–1000 ng BAC DNA was electroporated into recipient cells using our standard electroporation protocol[8]: Specifically, cells were diluted 1:100 from an overnight starter culture into 2×YT broth containing 20 µg/ml chloramphenicol, grown aerobically at 32 °C, 250 rpm in a shaking incubator until an OD_{600} of 0.3, followed by induction at 42 °C for 15 min. Cells were then chilled on ice, washed twice using ice-cold water on ice and resuspended to 100 µl in water or 10% glycerol. Finally, cells were electroporated on a BioRad GenePulser electroporator in 1 mm cuvettes (1800V, 200 Ohm, 25 µF) and resuspended in 2×YT broth. Following electroporation, cells were allowed to recover overnight at 32 °C in a rotor drum and then plated onto 2×YT agar plates containing 80 µg/ml spectinomycin and 20 µg/

ml chloramphenicol and incubated at 32 °C until colony formation (i.e., 2–10 days). In rare cases when electroporation did not yield colonies (i.e., in the case of Segment 11 and 32), we performed F-plasmid-mediated conjugation utilizing the oriT sequence present on our pYES2L plasmid (for a detailed protocol, see Supplementary Method 1.). Next, we confirmed the presence of the recoded chromosomal region, first using MASC PCR (using MASC-PCR primers listed in Supplementary Data 1) in KAPA2G Fast Multiplex MasterMix (Roche, USA) and then by validating its sequence using Illumina whole genome sequencing. To perform genomic deletions, cells were diluted 1:100 into 50 ml 2×YT broth containing 50 µg/ml spectinomycin and 20 µg/ml chloramphenicol and grown at 32 °C, 250 rpm until the early exponential phase. We eliminated the parental copy of the extrachromosomal BAC-carried recoded segment by co-electroporating in two to three 40 µl replicates, 6–10 µg of the corresponding genome-targeting pCRISPR plasmid[47,124]-derived constitutive crRNA expression plasmid and 8–10 µg of the corresponding SynOMICS deletion cassette (for the complete list of crRNA plasmids and deletion cassettes, see Supplementary Data 1). This step is initiated in cells with constitutive Cas9 expression by transiently overexpressing recombineering proteins and delivering a genome-targeting CRISPR-RNA-expressing plasmid with an antibiotic marker cassette. The marker cassette serves three purposes: 1.) it deletes—in combination with a Cas9 cut and recombineering—the parental genomic region, 2.) prepares the deleted genomic locus for the integration of the synthetic recoded segment based on short terminal homologies that match the recoded part, and 3.) allows the antibiotic-resistance-based identification of clones that lost the complete parental locus. Before electroporation, the expression of *gam*, *exo*, and *bet* was induced by a brief, 15-minute heat shock at 42 °C, 250 rpm. Following electroporation, cells were recovered overnight in 2×YT broth at 32 °C under constant agitation and then plated in 0.5–50 ml batches onto 2×YT agar plates containing 11 µg/ml gentamicin, 20 µg/ml chloramphenicol, and 50 µg/ml kanamycin. Plates were incubated at 32 °C until colony formation (i.e., 2–12 days). Plated recovery cell culture volumes were determined and adjusted based on the recovery culture's optical density prior to plating. We confirmed the deletion of the parental chromosomal copy using colony PCR with genome-specific, external primers, hybridizing 200–500 bp outside the flanking homology of the SynOMICS deletion cassette. All colony PCRs were performed using Platinum™ II Hot-Start Green PCR Master Mix (Thermo Fischer Scientific, USA). The sequence of all PCR primers is provided in Supplementary Data 1. Colonies displaying the correct PCR amplicon size (i.e., ~ 2700–3000 bp) were restreaked to 2×YT agar plates containing 8 µg/ml gentamicin, 20 µg/ml chloramphenicol, and 50 µg/ml spectinomycin and incubated at 32 °C until colony formation (*i.e.*, 2-12 days). Finally, the loss of the parental segment copy was confirmed using MASC PCR and Illumina whole-genome sequencing. We integrated the extrachromosomal, BAC-carried segment by diluting cells from an overnight starter culture into 50 ml 2×YT broth, containing 1% D-glucose and 20 µg/ml chloramphenicol, and growing cells at 32 °C, 250 rpm until the early exponential phase. As the selection stringency of CRISPR/Cas9-based selection is primarily limited by intramolecular recombination due to sequence similarity between neighboring sgRNA constructs[125,126] (Supplementary Fig. 4), we maximized SynOMICS's efficiency by utilizing a recently developed set of nonrepetitive sgRNA sequences and eliminated all unstable motifs from pINTsg[66,126–128] (Supplementary Fig. 4). The elimination of unstable sites boosted SynOMICS's integration efficiency, frequently approaching 100% (Fig. 2d, e). Following electrocompetent cell preparation[47,124] as described above, we electroporated 6 µg pINTsg plasmid. Cells were allowed to recover overnight at 32 °C in a rotor drum in 2×YT broth containing 1% D-glucose, washed once using 2×YT broth containing 1% D-glucose, and plated in 5 µl, 200 µl, and 500 µl volumes onto 2×YT agar plates containing 50 µg/ml carbenicillin, 20 µg/ml chloramphenicol, and 1% D-glucose, and incubated at 32 °C until colony formation (i.e., 2–12 days). Cells from all SynOMICS deletion and integration experiments were plated on agar plates in 14.5 cm diameter Greiner Bio-One Petri dishes (VWR Catalog No. 82050-600). Following colony formation, segment integrations were confirmed using colony PCRs checking the terminal junctions between the segment and the neighboring genomic end. Colonies displaying the correct PCR amplicons were then restreaked to 2×YT agar plates containing 50 µg/ml carbenicillin, 20 µg/ml chloramphenicol, and 1% D-glucose, and the integrity of the recoded chromosomal region was validated using MASC PCR and whole genome sequencing. Finally, cells were made suitable for the next SynOMICS integration cycle by growing them in 2×YT broth containing 20 µg/ml chloramphenicol and 0.6% L-rhamnose (CAS number: 10030-85-0) at 32 °C, inducing the expression of the pRham (*rhaB*) promoter-driven *hok* bacterial toxin that resulted in the rapid loss of the pINTsg plasmid. Given the rapid loss of the pINTsg plasmid in the presence of L-rhamnose, we could directly proceed with the next SynOMICS cycle by electroporating the next segment.

crRNA and sgRNA plasmids used in SynOMICS experiments were synthesized by GenScript USA as research-grade, primarily supercoiled plasmids, delivered freeze-dried, and resuspended to 1 µg/µl in nuclease-free H_2O. During SynOMICS cycles, we initiated DIvERGE-, MAGE-, and/or ALE-based troubleshooting after the deletion or integration steps, depending on cell fitness. The details of troubleshooting for each segment are summarized in Supplementary Method 1. All primer sequences used during genome assembly are listed in Supplementary Data 1. Based on the stringent selection posed by CRISPR/Cas9 and the multi-kbp payload capacity of the marker cassette, we expected that SynOMICS would also allow us to simultaneously correct design flaws and DNA synthesis errors within recoded segments, while testing the functionality of the recoded design. We demonstrated that the deletion cassette of SynOMICS is extendable to correct design and synthesis errors within BAC-carried segments. We utilized this feature to insert and replace genes and operons up to 6.7 kbp, correct DNA synthesis errors in eight segments, and complement lethal design flaws at four additional loci (Supplementary Method 1). This feature of SynOMICS is especially important, as it allows both the replacement of essential genomic motifs (e.g., oriC) and the parallelized correction of errors by varying only the sequence of a short double-stranded (ds)DNA deletion cassette. Finally, we utilized the same pCRISPR-assisted recombineering workflow to delete *rnc* in MDS42 and our most recoded strain, Strain 5. To delete *rnc*, we engineered an in-frame deletion that deletes *rnc* while preserving only the last 11 codons to maintain *era* expression using an ssDNA MAGE oligo. CRISPR-counterselection against the parental *rnc* sequence was performed using a 5'-TTGAAGCCGATTAATTACGA sgRNA. Following deletion using our standard CRISPR/Cas9-assisted ssDNA deletion step procedure as described above, cells were recovered overnight, and 12 ml saturated culture was plated in 3 ml batches on agar plates in 14.5 cm diameter Greiner Bio-One Petri dishes (VWR Catalog No. 82050-600). Following colony formation, deletions were confirmed using external colony PCRs. In the case of *rnc* deletion in Strain 5, plates were incubated up to 5 days due to a slower growth rate and all outgrowing colonies were screened with external PCR primers, but no successful deletants were observed. In comparison, MDS42 yielded nearly 100% deletion efficiency.

Genome editing-based debugging

Slow-growing variants from SynOMICS cycles were transformed with pORTMAGE plasmids[69], selected based on the antibiotic resistance phenotype of the given strain. Strains originating from SynOMICS deletion steps (*i.e.*, Gentamicin-R and Kanamycin-R) were made electrocompetent using our standard protocol[8] and electroporated with pORTMAGE-2 (Addgene plasmid #72677; http://n2t.net/addgene:

72677; RRID:Addgene_72677) or alternatively, pORTMAGE203D, a derivative of pORTMAGE-2 expressing CspRecT[129] instead of *gam*, *exo*, and *bet*. Strains originating from the integration step of SynOMICS were transformed with pORTMAGE-Ec1[129], unless noted otherwise in the Supplementary Method 1. pORTMAGE plasmid containing cells were induced and made electrocompetent as described earlier[68,130] and in our general electrocompetent cell protocol in Methods: SynOMICS-based genome assembly, and 40 µl freshly made electrocompetent cells were electroporated with 2 µl of 500 µM mixture of MAGE and DIvERGE ssDNA oligonucleotides in two replicates, targeting potential promoter-recoding- and DNA-synthesis-derived errors in the given strain. For a detailed list of all oligonucleotides, see Supplementary Data 1. All oligonucleotides were chemically synthesized either by GenScript, USA, or by Integrated DNA Technologies, USA, and resuspended in IDTE buffer, pH 8.0 (Integrated DNA Technologies, USA) to a 500 µM final concentration. Following oligo-delivery, cells were allowed to recover overnight in SOB or 2×YT broth, and electroporation cycles were repeated as described in Supplementary Method 1 for each segment. Following electroporation cycles, cells were diluted twice 1:100 in 50 ml SOB or 2×YT broth in a 300 ml Erlenmeyer flask with a vented cap, grown aerobically between each dilution to saturation aerobically, and then plated onto 2×YT agar plates and incubated until colonies formed. Finally, individual fast-growing colonies from each experiment were isolated—based on colony size, following colony formation on 2×YT agar plates—and subjected to whole-genome sequencing. Tn1000 and Insertion Sequence (IS) deletions were performed similarly, but the corresponding MAGE oligonucleotides, targeting the complete deletion of the given mobile genetic elements, were co-electroporated with a pCRISPR plasmid[47,124], carrying a crRNA guide sequence to cleave the genomic or BAC-carried Tn1000 or IS element (Supplementary Data 1). For a detailed description of the utilized crRNA sequences, mobile genetic element removal experiments, see Supplementary Method 1. The deletion of the genomic *recA* was performed as described in Supplementary Method 1. Ribo-seq-based multiplexed genome editing experiments were performed similarly. Differentially translated genes based on ribosome profiling results were selected as targets for multiplexed editing. We inserted a 1024-member constitutive promoter library (5′-YTKAYRGCTAGCTCAGYCCTWGGKAYWRTGCTAGC) in front of all downregulated operons. To identify operon structures before oligo design, we relied on our SMRT-Cappable-seq analysis to identify the principal promoter and the most 5′ TSS of each operon. Editing oligonucleotides were designed to insert the compressed promoter library downstream of each selected TSS, without disrupting the native promoter and regulatory sequences. Upregulated genes based on Ribo-seq data were targeted for debugging using the RBS Library Calculator[103,131] (version 2.0) of DeNovoDNA (available at https://www.denovodna.com/software/design_rbs_library_calculator) in genome editing mode. The RBS sequence of each target gene was randomized to generate a 64-member RBS library spanning a TIR range between 100 and 100,000 using at maximum 10 bp changes compared to the parental RBS sequence to ensure high editing efficiency and compensate for inaccuracies in the RBS calculator algorithm. If needed, overlapping genes were manually disentangled to accommodate newly inserted promoters or RBS sequences. To increase library size, we performed electrocompetent cell preparation in 50 ml 2×YT broth with the appropriate antibiotics and electroporated four parallel replicates for each edited line (See our general electrocompetent cell preparation and editing protocol in Methods: SynOMICS-based genome assembly). Following oligo editing, we relied on 2–3 consecutive selection cycles using our terascale selection protocol (see Methods: Adaptive laboratory evolution-based troubleshooting for details) to enrich high-fitness variants in 500 ml 2×YT broth at 37 °C before the next oligo-editing cycle. The list of MAGE oligonucleotides in included in Supplementary Data 1.

Adaptive laboratory evolution-based troubleshooting

We performed adaptive laboratory evolution experiments in rich bacterial medium on the course of SynOMICS genome construction cycles to evolve variants with increased fitness. At each daily transfer step, ~5×10^9 to 10^{10} cells were transferred into 500 ml Lysogeny Broth Lennox (LBL) supplemented with 1.5 g/l Tris/Tris-HCL as buffer and incubated aerobically for 24 hours at 37 °C, 250 rpm in a 2000 ml baffled Erlenmeyer flask with a vented cap. Adaptive laboratory evolution experiments were performed in two replicates, starting from the same seed culture or two isogenic clones of the starting strain. During ALE experiments, cells were plated onto 2×YT agar plates following every five transfers and incubated at 37 °C to assess growth rate based on the speed of colony formation. Finally, evolution experiments were terminated by spreading bacterial cells onto 2×YT agar plates, incubated at 37 °C, and an individual, fast-growing colony from each experiment was isolated and subjected to whole-genome sequencing. We progressed to the next genome construction step using this selected clone without additionally selecting variants based on the genotype of whole-genome sequencing results. As expected, based on the low mutation rate applied during evolution, whole-genome sequencing of the ALE-evolved variants indicated the accumulation of only a few mutations within the recoded region. The only exception was the strain carrying Segments 82–0, likely due to the emergence of a variant with an elevated mutation rate. The duration of each ALE experiment is listed in Supplementary Method 1.

Directed evolution of deleterious codons

To study codon preference at the deleterious *yceD–holB* locus in Segment 21 of Ec_Syn57, we relied on MAGE to randomize all recoded codon positions to compressed codons representing the same sense forbidden codons. In this recoding scheme, all recoded serine codons were randomized to AGY, all leucine codons were mutated to TTR, and all arginine codons were mutated to AGR, while stop codons were targeted with a reversion to TAG. We designed 96 MAGE oligonucleotides targeting all 172 recoded codon positions. Next, we performed three genome editing cycles on *E. coli* MDS42 carrying the recoded Segment 21 using pORTMAGE503B, followed by growth-based selection using our terascale selection protocol at 37 °C in 2×YT broth. Strain fitness improved rapidly. Following 27 generations of growth-based selection, we isolated bulk genomic DNA and sequenced the evolved population using deep Illumina whole-genome sequencing. Sequencing libraries were prepared using the Illumina DNA Prep kit and IDT 10 bp UDI indices with a target insert size of 280 bp and sequenced on an Illumina NovaSeq X Plus, producing 150 bp paired-end reads. Demultiplexing, quality control, and adapter trimming were performed with bcl-convert (v4.2.4). Reads were then trimmed to Q28 using BBDuk from BBTools and aligned to *E. coli* MDS42 carrying the recoded Segment 21 as a reference using Bowtie2 2.3.0[114] in –very-sensitive-local mode. Single-nucleotide polymorphisms (SNPs) and their frequency at each target codon position were determined using the SNP calling function of Geneious Prime® 2025.1.2.

Doubling time and growth measurements

To determine growth parameters under standard laboratory conditions, saturated overnight cultures, initiated from isogenic, single colonies, were diluted 1:100 into 100 µl 2×YT or M9 + 2% D-glucose in a 96-well flat-bottom transparent microtiter plate (Corning Inc., USA). Overnight starter cultures were grown in the same media as the downstream conditions. To assess growth kinetics, diluted cultures in ten replicates were incubated aerobically at 37 °C, 800 rpm, and 1-mm orbital shake in a BioTek Synergy H1 Multimode plate reader (Agilent Technologies, Inc., USA). Optical density at 600 nm (OD_{600}) measurements were taken every 9 min until the stationary phase was reached. We calculated the doubling time using the open-source GrowthRates package, version 4.4, according to the method described in Ref. 132.

Long-term cultivation in M9 + 2% D-glucose broth was performed in three replicates for strains that were unable to form distinct colonies on M9 + 1% glucose agar plates (Teknova, M1201) after 7 days of incubation at 37 °C. For long-term cultivations, cells were grown in M9YG broth (M6018, Teknova) and then diluted 1:100 into fresh M9 + 2% D-glucose and incubated for 14 days at 37 °C in a rotor drum.

Biolog Phenotype measurements

Biolog Phenotype microarray measurements were performed on PM Phenotype MicroArrays™ (Biolog) plates PM01, PM02 (carbon sources) and PM04 (phosphorus and sulfur sources), PM06 (nitrogen sources) and PM09 (osmotic and ionic stressors), according to the previously described protocol[77]. Specifically, strains were grown overnight at 37 °C in Tryptic Soy Broth from single colonies and diluted 1:500 into 15 ml Biolog inoculation fluid (IF-0, Biolog). To compensate for metabolic auxotrophies, we supplemented the inoculation fluid with 3 mg/ml of SC powder (Sunrise Scientific Products, 1300-030) for PM 1, 9, and phosphorus tests in PM 4, while we applied 3 mg/ml of SC powder without cysteine and methionine (MP Biomedicals, 114420322) in sulfur tests (PM 4). Plates PM 4, 6 and 9 utilized D-glucose as a carbon source. PM plates were inoculated with 100 µl cell suspension per well and incubated at 37 °C in an Odin plate reader (Biolog) for 48 h with kinetic OD monitoring at 590 nm every 20 minutes. Measurements were performed in two independent replicates, and mean OD_{600} values based on $n = 2$ were compared to *E. coli* MDS42.

Total proteome analysis and the detection of translated genes

We analyzed the proteome of *E. coli* MDS42 as wild-type and Syn61 (Addgene #174513) as a 61-codon variant of *E. coli* MDS42 by performing tandem mass spectrometry-based proteome analysis. 10 ml mid-exponential phase cultures ($OD_{600} = 0.45$) of each strain, growing at 37 °C, 250 rpm in 2×YT broth in 300 ml baffled flasks, were spun down at room temperature and washed four times in 1 ml phosphate-buffered saline (Invitrogen, USA). Cell samples were immediately frozen in liquid N_2 and stored at −80 °C until total protein extraction and proteome analysis. Proteome analysis was performed in one replicate. Next, frozen samples were thawed on ice and digested using the S-Trap™ micro columns (ProtiFi, NY, USA) digest procedure according to the manufacturer's protocol. Specifically, samples were loaded to a 10 kDa filter (Pall, TX) and washed three times with 0.15 ml ultrapure water. Next, samples were reduced followed by spinning down chemicals each time in a centrifuge (Eppendorf, Germany) for 2 min at 100 xg. Finally, proteins were dissolved on the top of the 10 kDa filter in 100 µL 50 mM TEAB (triethyl ammonium bicarbonate) buffer, followed by trypsin (sequencing-grade, from Promega, WA) digests for 3 h at 38 °C. Digested samples were spun down through the filter, dried down to 20 µl and half of the entire sample was submitted for LC-MS/MS analysis. For analysis, samples were solubilized in aqueous 0.1% formic acid for subsequent analysis by tandem mass spectrometry. LC-MS/MS analysis of digested samples was performed on a Q Exactive Orbitrap Mass Spectrometer equipped with a UNO nano-HPLC (both from Thermo-Scientific, USA). The Q Exactive instrument was run in WWA mode (Wide Window isolation) with MS1 data-dependent acquisition. Peptides were separated on a 150 µm inner diameter microcapillary trapping column packed first with 1.9 cm of C18 PepSep column (Bruker, MA). Separation was achieved by applying a gradient from 5 to 27% acetonitrile in 0.1% formic acid over 45 min at 200 nl/minute. Electrospray ionization was performed by applying a voltage of 2000 V using a custom electrode junction at the end of the microcapillary column and sprayed from metal tips (PepSep, Denmark). The mass spectrometry survey scan was performed in the Orbitrap in the range of 400–550, 550–750 and 750–1250 m/z as gas-phase-fractionation-based analysis at a resolution of $6{\times}10^4$, followed by the selection of the ten most intense ions for fragmentation using Collision Induced Dissociation (CID) in the second MS step (CID-MS2 fragmentation) in the ion trap using a precursor isolation width window of 10 m/z, AGC (automatic gain control) setting of 10,000 and a maximum ion accumulation of 150 ms. Singly charged ion species were excluded from CID fragmentation. The normalized collision energy was set to 30 V, and an activation time of 10 ms was utilized. Ions in a 10 m/z window around ions selected for MS/MS as wide window acquisition mode were excluded from further selection for fragmentation for 60 seconds.

The proteome reference database was generated by identifying all potential Open Reading Frames (ORFs) longer than 20 amino acids using the canonical genetic code and TTG, ATC, ATG, CTG, ATT, GTG, and ATA[59] as potential start codons on each genome sequence (i.e., *E. coli* MDS42 and Syn61) using Geneious Prime® 2023.2.1. The raw data was analyzed using Proteome Discoverer 3.1 (Thermo Fisher Scientific, USA) and Chimerys 2.0 (MSAID, Germany). Assignment of MS/MS spectra was performed using the Sequest HT algorithm by searching the data against a manually constructed protein sequence database which included all ORF entries from the sample's corresponding reference genome and other known contaminants, such as human keratins and common lab contaminants. Peptide amounts in each sample were quantified by LFQ (label-free quantitation). Sequest HT searches were performed using a 10-ppm precursor ion tolerance and requiring each peptide's N-/C termini to adhere to trypsin protease specificity while allowing up to two missed cleavages. Methionine oxidation (+ 15.99492 Da) was set as a variable modification. All cysteines were set to permanent no modification due to no alkylation procedure. An overall false discovery rate (FDR) of 1% on both protein and peptide levels was achieved by performing a target-decoy database search using Percolator[133]. Next, peptides originating from previously not annotated cryptic ORF sequences were identified by screening all identified peptides from a given sample using ProteoMapper version 5.2.0[134] to filter out all peptides matching known proteins on the sample's reference genome. Finally, all identified peptides were mapped back to genomic ORFs using the BLAST function implemented in Geneious Prime® 2023.2.1. and inspected manually. All identified cryptic ORFs and ORF-derived peptides are listed in Supplementary Data 6.

Statistics & reproducibility

No statistical calculations were done to predetermine sample size. Sample sizes were chosen as large as possible while remaining feasible for sample handling and data collection. Three or more replicates were performed in almost all cases, with a small standard deviation across measurements. We used this strategy and sample size because the variation across assays was small and we were interested in large effects. Measurements were performed in independent replicates, and only slight variation was observed between them. The exact number of replicates is stated in the corresponding figure legend and methods section. All measurements were reproducible, and no data were excluded from the analyses. The experiments were not randomized. Sequencing and mass spectrometry raw data generation were blinded. All other forms of experiments and data generation were not blinded.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw sequencing data have been deposited to the Sequence Read Archive (SRA) under BioProject ID PRJNA1088510. The annotated genome of Ec_Syn57 is available in the Source Data file of this paper as Ec_Syn57.gb. All other sequences utilized in this study are listed in Supplementary Data 1. The previously generated Pacific Biosciences long-read SMRT-Cappable-seq data analyzed in this study is available from the Sequence Read Archive (SRA) under BioProject ID PRJNA481586. Mass spectrometry and proteomic datasets are available on MassIVE, under MSV000094380 [https://doi.org/10.25345/

C59G5GR15]. To aid future recoding efforts and the investigation of the effects described in this work, we have deposited the current set of 57-codon Ec_Syn57 genome regions on Addgene, making it freely available together with the datasets generated in this paper. Source data are provided with this paper.

Code availability

The comparison of protein-coding gene annotations between the original and updated genome versions was performed using the compare_annotations.py script available from GitHub [https://github.com/rrwick/Compare-annotations], developed by Ryan Wick, Peter Doherty Institute for Infection and Immunity of the University of Melbourne, Melbourne, Australia. Cappable-seq data analysis scripts and protocols are available on GitHub [https://github.com/elitaone/cappable-seq], developed by Bo Yan from New England Biolabs, Ipswich, MA, USA, and described in references [54,95]. The genome design scripts used for the initial design[19] of Ec_Syn57's genome are available on Zenodo[135] [https://doi.org/10.5281/zenodo.19682030].

References

1. Crick, F. H. C., Barnett, L., Brenner, S. & Watts-Tobin, R. J. General nature of the genetic code for proteins. *Nature* **192**, 1227–1232 (1961).
2. Nirenberg, M. et al. The RNA code and protein synthesis. *Cold Spring Harb. Symp. Quant. Biol.* **31**, 11–24 (1966).
3. Quax, T. E. F., Claassens, N. J., Söll, D. & van der Oost, J. Codon bias as a means to fine-tune gene expression. *Mol. Cell* **59**, 149–161 (2015).
4. Deng, Y., Kalfon, J., Chu, D. & Haar, T. von der. Hidden patterns of codon usage bias across kingdoms. *bioRxiv* 478016 (2018) https://doi.org/10.1101/478016.
5. Springstein, B. L. et al. Systematic analysis of nonprogrammed frameshift suppression in E. coli via translational tiling proteomics. *Proc. Natl. Acad. Sci. USA* **121**, e2317453121 (2024).
6. Verma, S. C., Qian, Z. & Adhya, S. L. Architecture of the Escherichia coli nucleoid. *PLOS Genet.* **15**, e1008456 (2019).
7. Lajoie, M. J. et al. Genomically recoded organisms expand biological functions. *Science* **342**, 357–360 (2013).
8. Nyerges, A. et al. A swapped genetic code prevents viral infections and gene transfer. *Nature* 1–8 https://doi.org/10.1038/s41586-023-05824-z(2023).
9. Robertson, W. E. et al. Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science* **372**, 1057–1062 (2021).
10. Zürcher, J. F. et al. Refactored genetic codes enable bidirectional genetic isolation. *Science* **378**, 516–523 (2022).
11. Dunkelmann, D. L. et al. Adding α,α-disubstituted and β-linked monomers to the genetic code of an organism. *Nature* 1–8 https://doi.org/10.1038/s41586-023-06897-6(2024).
12. Ma, N. J. & Isaacs, F. J. Genomic recoding broadly obstructs the propagation of horizontally transferred genetic elements. *Cells* **3**, 199–207 (2016).
13. Grome, M. W. et al. Engineering a genomically recoded organism with one stop codon. *Nature* 1–10 https://doi.org/10.1038/s41586-024-08501-x(2025).
14. Isaacs, F. et al. Genomically recoded Escherichia coli with optimized functional phenotypes. 2024.08.29.610322 Preprint at https://doi.org/10.1101/2024.08.29.610322 (2024).
15. Isaacs, F. J. et al. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**, 348–353 (2011).
16. Lajoie, M. J. et al. Probing the limits of genetic recoding in essential genes. *Science* **342**, 361–363 (2013).
17. Napolitano, M. G. et al. Emergent rules for codon choice elucidated by editing rare arginine codons in Escherichia coli. *PNAS* **113**, E5588–E5597 (2016).
18. Wang, K. et al. Defining synonymous codon compression schemes by genome recoding. *Nature* **539**, 59–64 (2016).
19. Ostrov, N. et al. Design, synthesis, and testing toward a 57-codon genome. *Science* **353**, 819–822 (2016).
20. Fredens, J. et al. Total synthesis of Escherichia coli with a recoded genome. *Nature* 1 https://doi.org/10.1038/s41586-019-1192-5(2019).
21. Robertson, W. E. et al. Escherichia coli with a 57-codon genetic code. *Science* **0**, eady4368 (2025).
22. Chan, L. Y., Kosuri, S. & Endy, D. Refactoring bacteriophage T7. *Mol. Syst. Biol.* **1**, 2005.0018 (2005).
23. Mukai, T. et al. Overcoming the challenges of megabase-sized plasmid construction in Escherichia coli. *ACS Synth. Biol*. https://doi.org/10.1021/acssynbio.0c00008(2020).
24. Wang, K., Torre, D., de la, Robertson, W. E. & Chin, J. W. Programmed chromosome fission and fusion enable precise large-scale genome rearrangement and assembly. *Science* **365**, 922–926 (2019).
25. Ye, Y. et al. Genomic iterative replacements of large synthetic DNA fragments in corynebacterium glutamicum. *ACS Synth. Biol*. https://doi.org/10.1021/acssynbio.1c00644(2022).
26. Hutchison, C. A. et al. Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253 (2016).
27. Zürcher, J. F. et al. Continuous synthesis of E. coli genome sections and Mb-scale human DNA assembly. *Nature* 1–8 https://doi.org/10.1038/s41586-023-06268-1(2023).
28. Itaya, M., Tsuge, K., Koizumi, M. & Fujita, K. Combining two genomes in one cell: stable cloning of the Synechocystis PCC6803 genome in the Bacillus subtilis 168 genome. *PNAS* **102**, 15971–15976 (2005).
29. Venetz, J. E. et al. Chemical synthesis rewriting of a bacterial genome to achieve design flexibility and biological functionality. *PNAS* 201818259 https://doi.org/10.1073/pnas.1818259116(2019).
30. Foo, J. L. et al. Establishing chromosomal design-build-test-learn through a synthetic chromosome and its combinatorial reconfiguration. *Cell Genomics* **3**, 100435 (2023).
31. Williams, T. C. et al. Parallel laboratory evolution and rational debugging reveal genomic plasticity to S. cerevisiae synthetic chromosome XIV defects. *Cell Genomics* **3**, 100379 (2023).
32. Richardson, S. M. et al. Design of a synthetic yeast genome. *Science* **355**, 1040–1044 (2017).
33. Zhao, Y. et al. Debugging and consolidating multiple synthetic chromosomes reveals combinatorial genetic interactions. *Cell* https://doi.org/10.1016/j.cell.2023.09.025 (2023).
34. Chen, L.-G. et al. A designer synthetic chromosome fragment functions in moss. *Nat. Plants* 1–12 https://doi.org/10.1038/s41477-023-01595-7(2024).
35. Gibson, D. G. et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56 (2010).
36. Camellato, B. R., Brosh, R., Ashe, H. J., Maurano, M. T. & Boeke, J. D. Synthetic reversed sequences reveal default genomic states. *Nature* 1–8 https://doi.org/10.1038/s41586-024-07128-2(2024).
37. Trolle, J. et al. Resurrecting essential amino acid biosynthesis in mammalian cells. *eLife* **11**, e72847 (2022).
38. van Kooten, M. J. F. M., Scheidegger, C. A., Christen, M. & Christen, B. The transcriptional landscape of a rewritten bacterial genome reveals control elements and genome design principles. *Nat. Commun.* **12**, 3053 (2021).
39. James, J. S., Dai, J., Chew, W. L. & Cai, Y. The design and engineering of synthetic genomes. *Nat Rev Genet* 1–22 https://doi.org/10.1038/s41576-024-00786-y(2024).
40. Mandell, D. J. et al. Biocontainment of genetically modified organisms by synthetic protein design. *Nature* **518**, 55–60 (2015).
41. Rovner, A. J. et al. Recoded organisms engineered to depend on synthetic amino acids. *Nature* **518**, 89–93 (2015).
42. Rugbjerg, P., Brask Dyerberg, A. S., Quainoo, S., Munck, C. & Sommer, A. lexander M. O. Short and long-read ultra-deep

profiles emerging heterogeneity across five platform Escherichia coli strains. *Metab. Eng.* https://doi.org/10.1016/j.ymben.2020.11.006 (2020).
43. Nyerges, Á. et al. CRISPR-interference-based modulation of mobile genetic elements in bacteria. *Synth Biol* **4**, ysz008 (2019).
44. Pósfai, G. et al. Emergent properties of reduced-genome Escherichia coli. *Science* **312**, 1044–1046 (2006).
45. Durfee, T. et al. The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse. *J. Bacteriol.* **190**, 2597–2606 (2008).
46. Broom, J. E. et al. Sequence of a transposon identified as Tn1000 (γδ). *DNA Sequence* **5**, 185–189 (1995).
47. Umenhoffer, K. et al. Genome-wide abolishment of mobile genetic elements using genome shuffling and CRISPR/Cas-assisted MAGE allows the efficient stabilization of a bacterial chassis. *ACS Synth. Biol.* **6**, 1471–1483 (2017).
48. Wang, H. H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
49. Lau, Y. H. et al. Large-scale recoding of a bacterial genome by iterative recombineering of synthetic DNA. *Nucleic Acids Res.* **45**, 6971–6980 (2017).
50. Robertson, W. E. et al. Escherichia coli with a 57-codon genetic code. *Science* **390**, eady4368 (2025).
51. Cui, L. & Bikard, D. Consequences of Cas9 cleavage in the chromosome of Escherichia coli. *Nucl. Acids Res.* gkw223 (2016) https://doi.org/10.1093/nar/gkw223.
52. Espah Borujeni, A. et al. Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences. *Nucleic Acids Res.* **45**, 5437–5448 (2017).
53. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**, 946–950 (2009).
54. Ettwiller, L., Buswell, J., Yigit, E. & Schildkraut, I. A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. *BMC Genomics* **17**, 199 (2016).
55. Yan, B., Boitano, M., Clark, T. A. & Ettwiller, L. SMRT-cappable-seq reveals complex operon variants in bacteria. *Nat. Commun.* **9**, 1–11 (2018).
56. Wang, H. H. et al. Genome-scale promoter engineering by coselection MAGE. *Nat. Methods* **9**, 591–593 (2012).
57. Ma, N. J., Hemez, C. F., Barber, K. W., Rinehart, J. & Isaacs, F. J. Organisms with alternative genetic codes resolve unassigned codons via mistranslation and ribosomal rescue. *eLife 2018* **7**, e34878 (2018).
58. Goodall, E. C. A. et al. The essential genome of Escherichia coli K-12. *mBio* **9**, e02096-17 (2018).
59. Hecht, A. et al. Measurements of translation initiation from all 64 codons in E. coli. *Nucleic Acids Res.* **45**, 3615–3626 (2017).
60. Meydan, S. et al. Retapamulin-assisted ribosome profiling reveals the alternative bacterial proteome. *Mol. Cell* **74**, 481–493.e6 (2019).
61. Stringer, A., Smith, C., Mangano, K. & Wade, J. T. Identification of novel translated small open reading frames in escherichia coli using complementary ribosome profiling approaches. *J. Bacteriol.* **204**, e00352–21 (2022).
62. Fesenko, I., Sahakyan, H., Shabalina, S. A. & Koonin, E. V. The cryptic bacterial microproteome. 2024.02.17.580829 Preprint at https://doi.org/10.1101/2024.02.17.580829 (2024).
63. Nakahigashi, K. et al. Comprehensive identification of translation start sites by tetracycline-inhibited ribosome profiling. *DNA Res.* **23**, 193–201 (2016).
64. Weaver, J., Mohammad, F., Buskirk, A. R. & Storz, G. Identifying small proteins by ribosome profiling with stalled initiation complexes. *mBio* **10**, (2019).
65. Goodall, E. C. A. et al. Identification of new genes on a whole genome scale using saturated reporter transposon mutagenesis. 2024.09.06.611592 Preprint at https://doi.org/10.1101/2024.09.06.611592 (2024).
66. Hossain, A. et al. Automated design of thousands of nonrepetitive parts for engineering stable genetic systems. *Nat. Biotechnol.* **38**, 1466–1475 (2020).
67. Promoters/Catalog/Anderson - parts.igem.org. https://parts.igem.org/Promoters/Catalog/Anderson.
68. Nyerges, Á. et al. Directed evolution of multiple genomic loci allows the prediction of antibiotic resistance. *PNAS* **115**, E5726–E5735 (2018).
69. Nyerges, Á. et al. A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *PNAS* 201520040 (2016) https://doi.org/10.1073/pnas.1520040113.
70. Barrick, J. E. & Lenski, R. E. Genome dynamics during experimental evolution. *Nat. Rev. Genet* **14**, 827–839 (2013).
71. Garoff, L. et al. Population bottlenecks strongly influence the evolutionary trajectory to fluoroquinolone resistance in Escherichia coli. *Mol. Biol. Evol.* **37**, 1637–1646 (2020).
72. Lee, H., Popodi, E., Tang, H. & Foster, P. L. Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. *PNAS* **109**, E2774–E2783 (2012).
73. Shendure, J. et al. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005).
74. Bell, G. & MacLean, C. The search for 'Evolution-Proof' antibiotics. *Trends Microbiol.* **26**, 471–483 (2017).
75. Lin, Y., Zou, X., Zheng, Y., Cai, Y. & Dai, J. Improving chromosome synthesis with a semiquantitative phenotypic assay and refined assembly strategy. *ACS Synth. Biol.* **8**, 2203–2211 (2019).
76. Karcagi, I. et al. Indispensability of horizontally transferred genes and its impact on bacterial genome streamlining. *Mol. Biol. Evol.* **33**, 1257–1269 (2016).
77. Bochner, B. R., Gadzinski, P. & Panomitros, E. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res.* **11**, 1246–1255 (2001).
78. Lasa, I. et al. Genome-wide antisense transcription drives mRNA processing in bacteria. *Proc. Natl. Acad. Sci. USA* **108**, 20172–20177 (2011).
79. Dornenburg, J. E., DeVita, A. M., Palumbo, M. J. & Wade, J. T. Widespread Antisense Transcription in Escherichia coli. *mBio* **1**, https://doi.org/10.1128/mbio.00024-10 (2010).
80. Lipońska, A. et al. Base composition at the start of the coding sequence controls the balance between translation initiation and mRNA degradation in E. coli. *Nucleic Acids Res.* **53**, 1262 (2025).
81. Weeks, R. & Ostermeier, M. Fitness and Functional Landscapes of the E. coli RNase III Gene rnc. *Mol. Biol. Evol.* **40**, msad047 (2023).
82. Magistro, G., Magistro, C., Stief, C. G. & Schubert, S. A simple and highly efficient method for gene silencing in Escherichia coli. *J. Microbiol. Methods* **154**, 25–32 (2018).
83. Yona, A. H., Alm, E. J. & Gore, J. Random sequences rapidly evolve into de novo promoters. *Nat. Commun.* **9**, 1–10 (2018).
84. Radrizzani, S., Kudla, G., Izsvák, Z. & Hurst, L. D. Selection on synonymous sites: the unwanted transcript hypothesis. *Nat Rev Genet* 1–18 https://doi.org/10.1038/s41576-023-00686-7(2024)
85. Fuqua, T. & Wagner, A. De-novo promoters emerge more readily from random DNA than from genomic DNA. 2025.08.25.672121 Preprint at https://doi.org/10.1101/2025.08.25.672121 (2025).

86. Urtecho, G. et al. Genome-wide functional characterization of Escherichia coli promoters and sequence elements encoding their regulation. *eLife* **12**, RP92558 (2023).
87. Umenhoffer, K. et al. Reduced evolvability of Escherichia coli MDS42, an IS-less cellular chassis for molecular and synthetic biology applications. *Microbial Cell Factories* **9**, 38 (2010).
88. Gibson, B., Wilson, D. J., Feil, E. & Eyre-Walker, A. The distribution of bacterial doubling times in the wild. *Proc. R. Soc. B Biol. Sci.* **285**, 20180789 (2018).
89. Forrest, D., Warman, E. A., Erkelens, A. M., Dame, R. T. & Grainger, D. C. Xenogeneic silencing strategies in bacteria are dictated by RNA polymerase promiscuity. *Nat. Commun.* **13**, 1149 (2022).
90. Myers, K. S., Noguera, D. R. & Donohue, T. J. Promoter architecture differences among alphaproteobacteria and other bacterial taxa. *mSystems* **6**, https://doi.org/10.1128/msystems.00526-21 (2021).
91. Cui, Z., Stein, V., Tnimov, Z., Mureev, S. & Alexandrov, K. Semi-synthetic tRNA complement mediates in vitro protein synthesis. *J. Am. Chem. Soc.* **137**, 4404–4413 (2015).
92. Takai, K., Okumura, S., Hosono, K., Yokoyama, S. & Takaku, H. A single uridine modification at the wobble position of an artificial tRNA enhances wobbling in an Escherichia coli cell-free translation system. *FEBS Lett.* **447**, 1–4 (1999).
93. Zaher, H. S. & Green, R. Hyperaccurate and error-prone ribosomes exploit distinct mechanisms during tRNA selection. *Mol. Cell* **39**, 110–120 (2010).
94. Tzani, I. et al. Detection of host cell microprotein impurities in antibody drug products. *Nat. Commun.* **15**, 8605 (2024).
95. Yan, B., Tzertzinis, G., Schildkraut, I. & Ettwiller, L. Comprehensive determination of transcription start sites derived from all RNA polymerases using ReCappable-seq. *Genome Res.* gr.275784.121 (2021) https://doi.org/10.1101/gr.275784.121.
96. d'Oelsnitz, S., Stofel, S. K., Love, J. D. & Ellington, A. D. Snowprint: a predictive tool for genetic biosensor discovery. *Commun. Biol.* **7**, 1–9 (2024).
97. Ireland, W. T. et al. Deciphering the regulatory genome of Escherichia coli, one hundred promoters at a time. *arXiv:2001.07396 [q-bio]* http://arxiv.org/abs/2001.07396 (2020).
98. Coppens, L. & Ledesma-Amaro, R. PromoterAtlas: decoding regulatory sequences across Gammaproteobacteria using a transformer model. https://doi.org/10.1038/s41467-026-72837-3 (2026).
99. Choe, D. et al. Adaptive laboratory evolution of a genome-reduced Escherichia coli. *Nat. Commun.* **10**, 935 (2019).
100. Nguyen, E. et al. Sequence modeling and design from molecular to genome scale with Evo. *Science* **386**, eado9336 (2024).
101. Nyerges, A. et al. Synthetic genomes unveil the effects of synonymous recoding. 2024.06.16.599206 Preprint at https://doi.org/10.1101/2024.06.16.599206 (2024).
102. Cetnar, D. P. & Salis, H. M. *Systematic Quantification of Sequence and Structural Determinants Controlling mRNA Stability in Bacterial Operons*. (2020) https://doi.org/10.1101/2020.07.22.216051.
103. Reis, A. C. & Salis, H. M. An automated model test system for systematic development and improvement of gene expression models. *ACS Synth. Biol.* (2020) https://doi.org/10.1021/acssynbio.0c00394.
104. Fleur, T. L., Hossain, A. & Salis, H. M. Automated model-predictive design of synthetic promoters to control transcriptional profiles in bacteria. *Nat Commun* **13**, 5159 (2022).
105. de Boer, C. G. et al. Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nat. Biotechnol.* **38**, 56–65 (2020).
106. Schlusser, N., Gonzalez, A., Pandey, M. & Zavolan, M. Current limitations in predicting mRNA translation with deep learning models. 2024.01.18.576214 Preprint at https://doi.org/10.1101/2024.01.18.576214 (2024).
107. Shao, B. & Yan, J. A long-context language model for deciphering and generating bacteriophage genomes. *Nat Commun* **15**, 9392 (2024).
108. Logel, D. Y., Trofimova, E. & Jaschke, P. R. Codon-restrained method for both eliminating and creating intragenic bacterial promoters. *ACS Synthetic Biology* (2022) https://doi.org/10.1021/acssynbio.1c00359.
109. King, S. H. et al. Generative design of novel bacteriophages with genome language models. 2025.09.12.675911 Preprint at https://doi.org/10.1101/2025.09.12.675911 (2025).
110. Bartley, B. A., Beal, J., Karr, J. R. & Strychalski, E. A. Organizing genome engineering for the gigabase scale. *Nat. Commun.* **11**, 1–9 (2020).
111. Dai, J., Boeke, J. D., Luo, Z., Jiang, S. & Cai, Y. Sc3.0: revamping and minimizing the yeast genome. *Genome Biol.* **21**, 205 (2020).
112. Pampuch, M., Walker, E. J. L. & Karas, B. J. Towards synthetic diatoms: the Phaeodactylum tricornutum Pt-syn 1.0 project. *Curr. Opin. Green. Sustain. Chem.* **35**, 100611 (2022).
113. Chen, Y. et al. Multiplex base editing to convert TAG into TAA codons in the human genome. *Nat. Commun.* **13**, 4482 (2022).
114. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
115. Deatherage, D. E. & Barrick, J. E. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol. Biol.* **1151**, 165–188 (2014).
116. Shen, Y. et al. Dissecting aneuploidy phenotypes by constructing Sc2.0 chromosome VII and SCRaMbLEing synthetic disomic yeast. *Cell Genomics* **3**, 100364 (2023).
117. Sorek, R. et al. Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* **318**, 1449–1452 (2007).
118. Kimelman, A. et al. A vast collection of microbial genes that are toxic to bacteria. *Genome Res.* **22**, 802–809 (2012).
119. Shen, Y. et al. Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science* **355**, eaaf4791 (2017).
120. Oh, E. et al. Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell* **147**, 1295–1308 (2011).
121. Mohammad, F., Green, R. & Buskirk, A. R. A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. *eLife* **8**, e42591 (2019).
122. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
123. Nedialkova, D. D. & Leidel, S. A. Optimization of codon translation rates via tRNA modifications maintains proteome integrity. *Cell* **161**, 1606–1618 (2015).
124. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotech.* **31**, 233–239 (2013).
125. Ding, T. et al. Reversed paired-gRNA plasmid cloning strategy for efficient genome editing in Escherichia coli. *Microb. Cell Factories* **19**, 63 (2020).
126. Reis, A. C. et al. Simultaneous repression of multiple bacterial genes using nonrepetitive extra-long sgRNA arrays. *Nat. Biotechnol.* **37**, 1294–1301 (2019).
127. Jack, B. R. et al. Predicting the genetic stability of engineered DNA sequences with the EFM calculator. *ACS Synth. Biol.* **4**, 939–943 (2015).
128. Halper, S. M., Hossain, A. & Salis, H. M. *The Synthesis Success Calculator: Predicting the Rapid Synthesis of DNA Fragments with Machine Learning*.https://doi.org/10.1101/2020.06.05.136820 (2020).

129. Wannier, T. M. et al. Improved bacterial recombineering by parallelized protein discovery. *PNAS* (2020) https://doi.org/10.1073/pnas.2001588117.
130. Szili, et al. Rapid evolution of reduced susceptibility against a balanced dual-targeting antibiotic through stepping-stone mutations. *Antimicrobial Agents and Chemotherapy* AAC.00207-00219 (2019) https://doi.org/10.1128/AAC.00207-19.
131. Farasat, I. et al. Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Mol. Syst. Biol.* **10**, 731–731 (2014).
132. Hall, B. G., Acar, H., Nandipati, A. & Barlow, M. Growth rates made easy. *Mol. Biol. Evol.* **31**, 232–238 (2014).
133. Käll, L., Storey, J. D. & Noble, W. S. Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. *Bioinformatics* **24**, i42–i48 (2008).
134. Mendoza, L. et al. Flexible and fast mapping of peptides to a proteome with proteomapper. *J. Proteome Res.* **17**, 4337–4344 (2018).
135. Gleb, K. Probing the limits of genetic recoding using multi-omics-guided evolution. Zenodo https://doi.org/10.5281/zenodo.19682030 (2026).

Acknowledgements

We thank György Pósfai (Biological Research Center, Hungary) for sharing MDS42 and Jason W. Chin's team (Medical Research Council Laboratory of Molecular Biology, UK) for sharing Syn61 and Syn61Δ3 via Addgene. The authors thank GenScript USA Inc. for their DNA synthesis support, Silvia Tornaletti and Luciano Brocchieri at TB-Seq, Inc., San Francisco, CA, for their support with Ribo-seq and RNA-seq experiments and data analysis; Ranomics, Inc., Canada, for experimental and DNA synthesis support; and Dan Snyder, Katrina Harris, and all members of SeqCenter, LLC, Pittsburgh, PA, for their support with DNA and RNA sequencing experiments. We thank Bo Yan from New England Biolabs, Ipswich, MA for support with Cappable-seq experiments; and Tan Chen, Yun Wang, Fei Huang, Xianping Deng, Jiahui Zhang at BGI Research for their help with DNA synthesis for the project.

Author contributions

A.N. developed the project, performed or directed all experiments and analyses, contributed to funding acquisition, and wrote the manuscript with input from all authors. A.N. and G.M.C. supervised research. B.B. carried out LC-MS/MS analyses. E.R. and R.F. performed genome editing, laboratory-evolution-based troubleshooting experiments, and growth rate analyses. A.C-P. removed transposons; computational analyses, including genome reannotations and analysis of impacted start codons, promoters, RBSes, and RNA levels; obtaining and troubleshooting: RNA-seq, adaptive laboratory evolution, growth of strains; DNA for BAC genome sequencing; construction, troubleshooting, and screening of recoded genome segments; corrected segments; and contributed to funding acquisition. M.B.T., S.Y., J.A., N.O., A.R., and D.C. assisted in experiments at the early stage of the project. S.Y., M.L., M.W., Q.Z., K.C., F.H., I.B., Y.Y., M.H-S., A.T., A.J., Z.L., Y.S., and M.H. performed DNA synthesis. E.R., V.A., and O.S. performed media preparation and DNA extraction for whole-genome sequencing. B.K., M.S., and V.S. provided reagents for the project. B.H. assisted in data visualization. All authors reviewed the manuscript.

Funding

Funding for this research was provided by the US Department of Energy (DOE) under grant DE-FG02-02ER63445 (to G.M.C.), by the National Science Foundation (NSF) Award number: 2123243 (to G.M.C.), by a JGI DNA Synthesis Program grant (award number 508712) (to G.M.C.), by the National Health Institute (NIH) Award number 1K99EB035165-01 (to A.N), 5K99EB035165-02 (to A.N), and 4R00EB035165-03 (to A.N), and by the EMBO LTF 160-2019 Long-Term fellowship (to A.N). A.C.P. acknowledges funding from the Swiss National Science Foundation (award number P2ELP2_181884 to A.C.P.). DNA synthesis for troubleshooting experiments in this work (proposal: https://doi.org/10.46936/10.25585/60008479) was also conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, supported by the Office of Science of the U.S. Department of Energy, operated under Contract No. DE-AC02-05CH11231 (awarded to The Regents of the University of California and the Lawrence Berkeley National Laboratory).

Competing interests

The authors declare the following competing interests: A.N. is an inventor on a granted patent related to directed evolution with random genomic mutations (DIvERGE) used in this study (US10669537B2: Mutagenizing Intracellular Nucleic Acids, Patent applicant: Biological Research Center) that has been outlicensed. Harvard Medical School has filed a provisional patent application related to the construction of Ec_Syn57 in this work (US provisional application 63/560,144, Patent applicant: President and Fellows of Harvard College), on which A.N. and G.M.C. are listed as inventors. Q.Z., M.W., M.L., A.J., K.C., Z.L., and F.H. are employed by GenScript USA Inc., but the company had no role in designing or executing experiments. G.M.C. is a founder of GRO Biosciences and EnEvolv (now part of Ginkgo Bioworks), in which he has related financial interests. Other potentially relevant financial and non-financial interests of G.M.C. are listed at http://arep.med.harvard.edu/gmc/tech.html. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-74300-9.

Correspondence and requests for materials should be addressed to Akos Nyerges or George M. Church.

Peer review information *Nature Communications* thanks Yi Wu, Kaihang Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.