Programming cells by multiplex genome engineering and accelerated evolution

Harris H. Wang[1,2,3]*, Farren J. Isaacs[1]*, Peter A. Carr[4,5], Zachary Z. Sun[6], George Xu[6], Craig R. Forest[7] & George M. Church[1]

The breadth of genomic diversity found among organisms in nature allows populations to adapt to diverse environments[1,2]. However, genomic diversity is difficult to generate in the laboratory and new phenotypes do not easily arise on practical timescales[3]. Although *in vitro* and directed evolution methods[4–9] have created genetic variants with usefully altered phenotypes, these methods are limited to laborious and serial manipulation of single genes and are not used for parallel and continuous directed evolution of gene networks or genomes. Here, we describe multiplex automated genome engineering (MAGE) for large-scale programming and evolution of cells. MAGE simultaneously targets many locations on the chromosome for modification in a single cell or across a population of cells, thus producing combinatorial genomic diversity. Because the process is cyclical and scalable, we constructed prototype devices that automate the MAGE technology to facilitate rapid and continuous generation of a diverse set of genetic changes (mismatches, insertions, deletions). We applied MAGE to optimize the 1-deoxy-D-xylulose-5-phosphate (DXP) biosynthesis pathway in *Escherichia coli* to overproduce the industrially important isoprenoid lycopene. Twenty-four genetic components in the DXP pathway were modified simultaneously using a complex pool of synthetic DNA, creating over 4.3 billion combinatorial genomic variants per day. We isolated variants with more than fivefold increase in lycopene production within 3 days, a significant improvement over existing metabolic engineering techniques. Our multiplex approach embraces engineering in the context of evolution by expediting the design and evolution of organisms with new and improved properties.

With the advent of next-generation fluorescent DNA sequencing[10], our ability to sequence genomes has greatly outpaced our ability to modify genomes. Existing cloning-based technologies are confined to serial and inefficient introduction of single DNA constructs into cells, requiring laborious and outdated genetic engineering techniques. Whereas *in vivo* methods such as recombination-based genetic engineering (recombineering) have enabled efficient modification of single genetic targets using single-stranded DNA (ssDNA)[11–14], no such attempts have been made to modify genomes on a large and parallel scale. MAGE provides a highly efficient, inexpensive and automated solution to simultaneously modify many genomic locations (for example, genes, regulatory regions) across different length scales, from the nucleotide to the genome level (Fig. 1).

Efficiency of the MAGE process was characterized using a modified *E. coli* strain (EcNR2). Mediated by the bacteriophage λ-Red ssDNA-binding protein β, allelic replacement is achieved in EcNR2 by directing ssDNA or oligonucleotides (oligos) to the lagging strand of the replication fork during DNA replication[14]. We optimized a number of parameters (see Supplementary Information, Supplementary Fig. 2 and Supplementary Table 1) to maximize efficiency of oligo-mediated allelic replacement. To generate sequence diversity in any region of the chromosome by allelic replacement, a pool of targeting oligos is repeatedly introduced into a cell. Under optimized conditions, we can successfully introduce new genetic modifications in >30% of the cell population (Supplementary Fig. 2d) every 2–2.5 h.

Oligo-mediated allelic replacement is capable of introducing a variety of genetic modifications at high efficiency. The efficiency of generating a mismatch or insertion modification is correlated to the amount of homologous sequence between the oligo and its chromosomal target (Fig. 2a, b); the efficiency of producing a deletion modification is correlated to the size of the deletion (Fig. 2c). Figure 2d shows that the predicted two-state hybridization free energy ΔG (ref. 15) between the oligo and target chromosomal sequence is a predictor of the allelic replacement efficiency. Thus, in a pool of oligos with degenerate sequences, oligos with more homology to the target will be incorporated in the chromosome at a higher frequency than those with less homology. This feature of MAGE enables tunable generation of divergent sequences along favourable evolutionary paths.
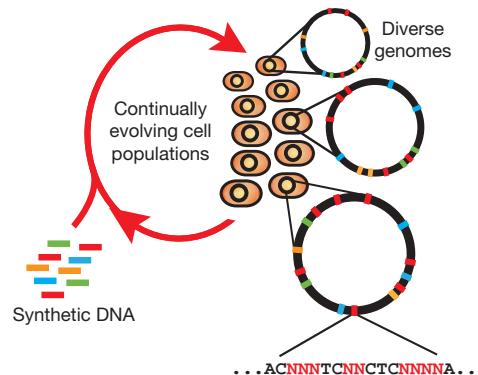
Figure 1 | Multiplex automated genome engineering enables the rapid and continuous generation of sequence diversity at many targeted chromosomal locations across a large population of cells through the repeated introduction of synthetic DNA. Each cell contains a different set of mutations, producing a heterogeneous population of rich diversity (denoted by distinct chromosomes in different cells). Degenerate oligo pools that target specific genomic positions enable the generation of a diverse set of sequences at each chromosomal location.

[1]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Program in Biophysics, Harvard University, Cambridge, Massachusetts 02138, USA. [3]Program in Medical Engineering Medical Physics, Harvard-MIT Division of Health Sciences and Technology, [4]The Center for Bits and Atoms, [5]Media Lab, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [6]Harvard College, Cambridge, Massachusetts 02138, USA. [7]George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA.
*These authors contributed equally to this work.
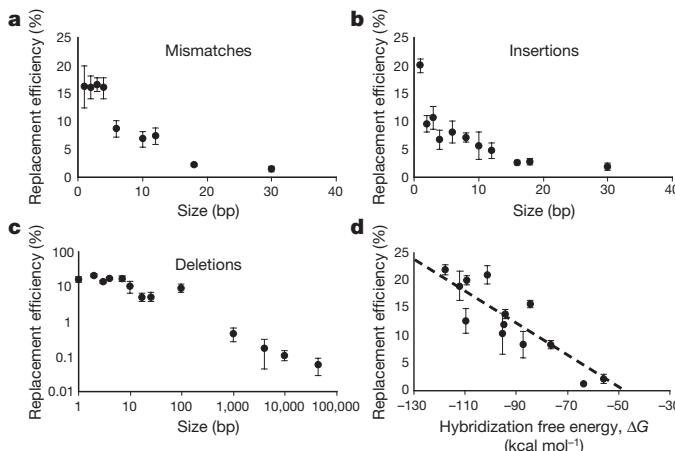
Figure 2 | Characterization of allelic replacement efficiency as a function of the type and scale of genetic modifications. a, Introducing mismatch mutations of up to 30 bp. **b**, Inserting exogenous sequences of up to 30 bp. **c**, Removing up to 45 kbp of chromosomal sequence using a single oligo. **d**, Correlation of replacement efficiency and two-state hybridization energy ΔG between the oligo and the targeted complement region in the genome. See Supplementary Fig. 1 for an illustration of oligo interaction with genomic targets and Supplementary Table 3 for a list of oligos and corresponding ΔG values. Dashed line is the linear regression correlation ($y = -0.288x - 13.7$, $R^2 = 0.799$). All oligos used were 90 bp with two phosphorothioate bonds at the 3′ and 5′ ends. All error bars indicate ± s.d.; $n = 3$.

To determine the rate at which MAGE generates sequence diversity, we used three different 90-mer oligos to produce mismatch changes in a targeted region of the *lacZ* gene in three distinct cell populations. The *cN6* and *cN30* oligos contained 6 and 30 consecutive degenerate bases, respectively; the *iN6* oligos contained 6 degenerate bases interspersed across a 30-bp region (Fig. 3). For these cell populations, the targeted *lacZ* region was sequenced in 96 random clonal isolates after MAGE cycles 2, 5, 10 and 15 that provided a snapshot of the genotypic variation in each population. Through successive cycles of MAGE, the chromosomal sequence of the *lacZ* region increasingly diverged away from wild type (Fig. 3). For *cN6*, after five MAGE cycles, we detected an average change of 3.1 bp per cell across the population (Fig. 3 inset, blue line), which equates to the generation of more than 4.3×10^9 bp
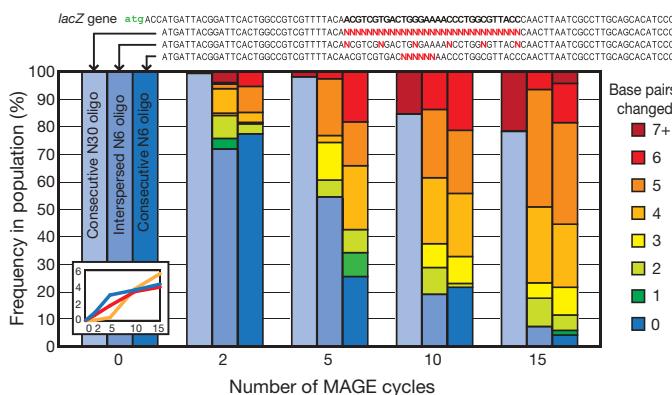
Figure 3 | Sequence diversity generated across three separate cell populations as a function of the number of MAGE cycles. Three 90-mer oligo pools were investigated: *cN30*, *iN6* and *cN6*. *cN30* contains oligos with 30 bp of consecutive degeneracy; *iN6* contains oligos with 6 bp of degeneracy spaced every 5 bp; *cN6* contains oligos with 6 bp of consecutive degeneracy. Frequency of strains in each population that contains 0 to 7+ bp of differences from the wild-type *lacZ* sequence are colour-coded. The inset shows average number of base pairs changed from wild type across the whole cell population as a function of the number of MAGE cycles using the three oligo pools *cN30* (orange line), *cN6* (blue line) and *iN6* (red line).

of variation per day (3.1 bp changes per five cycles in 7×10^8 cells at 10 cycles per day). Within 15 cycles, cell populations containing all possible N6 genotype combinations were generated using either *cN6* or *iN6* oligos. Because the replacement efficiency for a 30-bp mismatching oligo is lower (1.5% from Fig. 2a), only 21.8% of the *cN30* cell population had undergone allelic replacement after 15 cycles. We detected an average change of 5.6 bp per cell from the wild-type sequence across the whole *cN30* population (Fig. 3 inset, orange line).

The depth at which MAGE generates diversity is determined by a combination of three factors: (1) the degree of sequence variation desired at each locus; (2) the number of loci targeted; and (3) the number of MAGE cycles performed. When a single locus is targeted using a degenerate oligo pool, genetic diversity is generated across the population at that locus and is a function of the oligo pool complexity only. If more than one locus is targeted simultaneously, diversity is generated through the combinatorial arrangement of the different
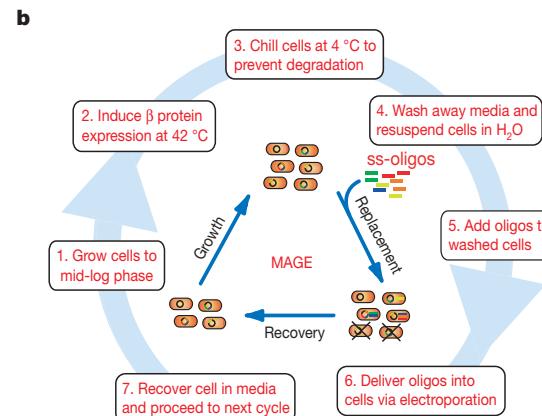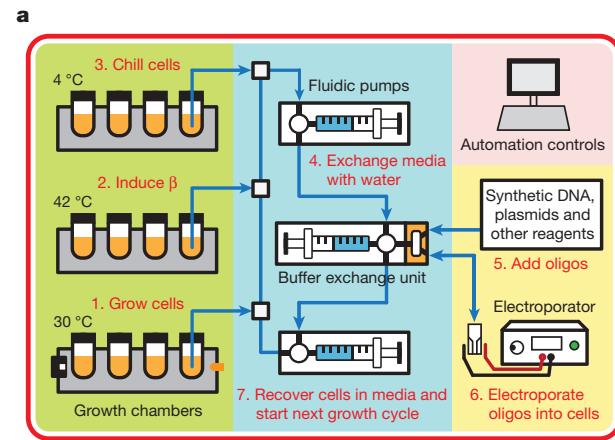
Figure 4 | MAGE automation. a, Detailed schematic diagram of MAGE prototype including climate-regulated growth chambers with real-time cell density monitors (green), anti-fouling fluidics for transfer of cells between growth chambers and exchange of media and buffers (blue), and real-time generation of competent cells for transformation with synthetic DNA (yellow). Cultures are carried through different chambers at different temperature regimes (30 °C, 42 °C, 4 °C) depending on the necessary MAGE steps (that is, cell growth, heat-shock, cooling). Cells are made electrocompetent by concentration onto a filter membrane and resuspension with wash buffer. Oligos are delivered into cells by electroporation. **b**, Step-by-step diagram of MAGE cycling steps at a total run time of 2–2.5 h per cycle. Owing to high voltage ($18\,\text{kV}\,\text{cm}^{-1}$) electroporation, ~95% of cells are killed at each cycle. Hence, the electroporation event serves to both introduce oligos into cells and to dilute the cell population, cells are then recovered and grown to mid-log phase (7×10^8 cells ml^{-1}) in liquid medium for the subsequent cycle.

2

modified loci. The frequency at which each locus is modified can be computationally predicted through a binomial distribution (Supplementary Fig. 3). Although the cell population at any cycle may reflect only a subset of all variants theoretically possible, we can cumulatively generate more variants than the actual size of the cell population ($\sim 7 \times 10^8$ cells) through successive MAGE cycles. Thus, we can generate all variants regardless of population size through computational predictions and continuous cycling by MAGE.

Given the cyclical and scalable nature of our approach, we constructed an integrated prototype device that automates the MAGE process to enable fast and reliable cellular programming. The device contains growth chambers to maintain healthy cell cultures and electroporation modules to repeatedly deliver DNA into the cells, thereby facilitating genome engineering and evolution (Fig. 4, Methods and Supplementary Information). Complex culturing conditions can be programmed into the device for growth of a diverse set of organisms and ecosystems.

To demonstrate an application of the MAGE process, we optimized metabolic flux through the DXP biosynthesis pathway to overproduce the isoprenoid, lycopene, in an *E. coli* strain (EcHW2) that contained the pAC-LYC plasmid that is necessary for the final steps of lycopene production. Twenty endogenous genes (*dxs, dxr, ispD, ispE, ispG, ispH, idi, ispA, appY, rpoS, crl, elbA, elbB, yjiD, purH, rnlA, yggT, ycgZ, ymgA, ariR*) documented to increase lycopene yield[16,17] were targeted to tune translation (Fig. 5a). Specifically, for each of the 20 genes, 90-mer oligos containing degenerate ribosome binding site

(RBS) sequences (DDRRRRDDDD; D = A, G, T; R = A, G) flanked by homologous regions on each side were used, with a total pool complexity of 4.7×10^5 ($3^6 \times 2^5 \times 20$). The replaced RBS regions were designed to be more similar to the canonical Shine–Dalgarno sequence (TAAGGAGGT)[18], giving rise to enhanced translation efficiency. Additionally, four genes (*ytjC, fdhF, aceE, gdhA*) from secondary pathways[19] were targeted for inactivation by oligos that introduced two nonsense mutations in the open reading frame, further improving flux through the DXP pathway. In contrast to prior strategies[20–23] that were experimentally limited by the number of genetic components that could be manipulated at once, here we optimized 24 genes simultaneously to maximize lycopene production.

As many as 15 billion genetic variants (4.3×10^8 bp variations per cycle for 35 MAGE cycles) were generated. Screening of variants was done by isolating colonies that produced intense red pigmentation on Luria–Bertani agar plates. Variants were isolated from $\sim 10^5$ colonies screened after 5–35 cycles of MAGE (see Supplementary Information), some exhibiting as much as a fivefold increase in lycopene production relative to the EcHW2 ancestral strain (Fig. 5b). Under similar experimental conditions, our highest lycopene yield of $\sim 9,000$ p.p.m. (μg per g dry cell weight) is better than documented yields[17,19]. Sequencing of six variants (EcHW2a–EcHW2f) revealed RBS convergence towards consensus-like Shine–Dalgarno sequences in genes localized at the beginning and end of the biosynthesis pathway (*dxs, dxr, idi, ispA*) as well as various gene knockouts from secondary pathways ($\Delta ytjC$, $\Delta gdhA$, $\Delta fdhF$) (Fig. 5c and Supplementary Table 2).
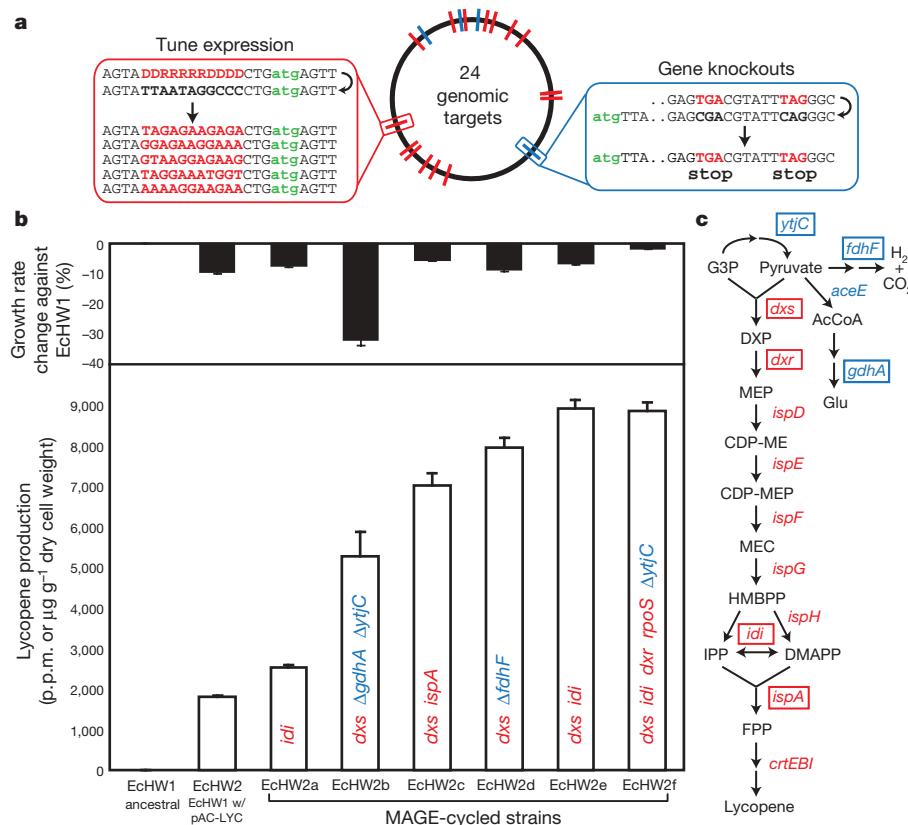
Figure 5 | Optimization of the DXP biosynthesis pathway for lycopene production. a, Genomic positions of 24 targeted genes with the RBS optimization strategy on the left (red) and gene knockout strategy on the right (blue). The gene knockout strategy involves the introduction of two nonsense mutations. All 90-mer oligos contain two phosphorothioated bases at the 3′ and 5′ termini. **b**, Black bars represent the growth rate of isolated variants (EcHW2a–f) relative to the ancestral EcHW1 strain. White bars represent lycopene production in p.p.m., which is normalized by dry cell weight in ancestral and mutant strains. Colour-coded labels in each white bar represent genetic modifications found by sequencing. All error bars indicate ± s.d.; $n = 3$. **c**, Modifications to the lycopene biosynthesis pathway of isolated variants EcHW2a–f with relevant genes highlighted by rectangular boxes. Blue labels represent knockout targets, red labels represent RBS tuning targets. AcCoA, acetyl-CoA; CDP-ME, 4-diphosphocytidyl-2-C-methyl-D-erythritol; CDP-MEP, 4-diphosphocytidyl-2C-methyl-D-erythritol-2-phosphate; DMAPP, dimethylallyl diphosphate; FPP, farnesyl diphosphate; G3P, glyceraldehyde 3-phosphate; HMBPP, (*E*)-4-hydroxy-3-methylbut-2-enyl-diphosphate; IPP, isopentenyl diphosphate; MEP, 2-C-methyl-D-erythritol-4-phosphate; MEC, 2C-methyl-D-erythritol-2,4-cyclodiphosphate.

Different tuning parameters of the lycopene pathway can be individually and combinatorially assessed in the isolated variants. For example, translation optimization of *idi* alone (EcHW2a) increased lycopene production by 40%. Whereas optimizing *dxs* and *idi* increased lycopene production by 390% (EcHW2e), additional optimization at *rpoS* and *dxr*, along with inactivation of *ytjC*, improved the growth rate of EcHW2f to that of the EcHW1 ancestor. Interestingly, *rpoS* is the alternative RNA polymerase subunit sigma factor σ^S, the master stress response regulator, and its upregulation can increase stress resistance to the accumulation of lycopene, a very hydrophobic molecule[17]. *ytjC* is an uncharacterized phosphoglycerate mutase enzyme and is thought to increase metabolic flux through the DXP pathway by increasing the accumulation of glyceraldehyde 3-phosphate intermediates[19]. Inactivation of glutamate dehydrogenase (*gdhA*) increases lycopene production, but causes a 32% decrease in growth rate of EcHW2b relative to wild type. Combinations of genetic modifications can also be assessed against each other. For example, an optimized *dxs* and *idi* strain (EcHW2e) produces 12% more lycopene than a strain with *dxs* and Δ*fdhF* (EcHW2d) and 27% more than a strain with *dxs* and *ispA* modifications (EcHW2c). The optimized DXP biosynthesis pathway presented here can be used to produce many other isoprenoid compounds of industrial and pharmaceutical relevance[24].

The diversity (that is, the degree of sequence change per target) generated by MAGE is adjustable and the specificity of targeting is always high. Oligos with defined sequences produce well-defined modification, whereas oligos with degenerate sequences produce high-diversity modifications tailored for exploring a vast sequence space. In this study, we used well-defined oligos to inactivate protein-coding sequences and high-diversity degenerate oligos to modify and sample different RBS sequences. We have also used the MAGE platform to perform whole-genome recoding of *E. coli* to enhance the incorporation of non-natural amino acids into proteins[25] and construct safer and multi-virus-resistant strains (F.J.I. *et al.*, manuscript in preparation). MAGE is thus a complementary technology to *de novo* genome synthesis[26], allowing the tuning of synthetic and natural genomes *in vivo* for various applications.

MAGE is also an accelerated evolution platform that permits the repeated introduction and maintenance of many neutral (or deleterious) mutations in the cell population. Although these mutations would normally disappear in the population via genetic drift or natural selection, MAGE accelerates the rate of their accumulation in any individual cell, thus increasing the likelihood of finding sets of mutations that may interact synergistically to produce a surprisingly beneficial phenotype. Using this technology, we could engineer or evolve cells with higher transfection efficiency (for example, harnessing natural competence systems[27]), increased allelic replacement efficiency (for example, expressing higher levels or mutants[28] of the λ-Red β protein) and perform large-scale bacterial artificial chromosome engineering. The simple allelic replacement mechanism could make this method amenable for use in other organisms, given that other ssDNA-binding protein homologues are functional[29]. Currently, 30 US dollars of commercially synthesized oligos can introduce up to 27 bp of modification at full degeneracy for a single genomic target. To target many loci, obtaining oligos from programmable DNA microchips[30] can significantly decrease the cost in comparison to traditional oligo synthesis. We envision that large-scale pipelines to program synthetic organisms and ecosystems[10] will greatly benefit from integration of hardware, software and wetware to engineer and evolve microbial, plant and animal systems.

METHODS SUMMARY

Strains. The EcNR2 strain is MG1655 with λ-prophage::*bioA*/*bioB* and *cmR*::*mutS*. The EcHW2 strain is MG1655 with λ-prophage::*bioA*/*bioB*, *kanR*::*mutS* and pAC-LYC plasmid. EcHW2a–f strains were MAGE-cycled EcHW2 derivatives containing DXP pathway modifications. Details of strains used are described in Methods.

Allelic replacement. Liquid cell cultures were inoculated from colonies obtained from a plate and grown at 30 °C to an absorbance (600 nm) of ~0.7 ($\sim 7 \times 10^8$ cells ml^{-1}) in a rotor drum at 200 r.p.m. To induce expression of the λ-Red recombination proteins (Exo, Beta and Gam), cell cultures were shifted to 42 °C for 15 min and then immediately chilled on ice. In a 4 °C environment, 1 ml of cells was centrifuged at 16,000*g* for 30 s. Supernatant medium was removed and cells were resuspended in dH2O (Gibco catalogue number 15230). This process was repeated twice with water. Supernatant water was removed and oligos suspended in water were added to the cell pellet. The oligos/cells mixture was transferred to a pre-chilled 1 mm gap electroporation cuvette (Bio-Rad) and electroporated with a Bio-Rad GenePulser electroporation system under the following parameters: 1.8 kV, 200 Ω and 25 μF. LB-min medium (1–3 ml) was immediately added to the electroporated cells. The cells were recovered from electroporation and grown at 30 °C for 2–2.5 h. Once cells reached mid-logarithmic growth they were used in additional MAGE cycles, isolated or assayed for genotype and/or phenotype analysis.

Lycopene assay. EcHW2 and derivatives (EcHW2a–f) were quantitatively assessed for lycopene production after 24 h of growth in LB-min-chloramphenicol liquid medium at 30 °C. Lycopene was extracted and quantified at an absorbance of 470 nm (see Methods). Lycopene yield was calculated by normalizing the amount of lycopene extracted to the dry cell weight.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
2. Tringe, S. G. *et al.* Comparative metagenomics of microbial communities. *Science* **308**, 554–557 (2005).
3. Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Rev. Genet.* **4**, 457–469 (2003).
4. Ellington, A. D. & Szostak, J. W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
5. Crameri, A., Raillard, S.-A., Bermudez, E., Stemmer, W. P. & C. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288–291 (1998).
6. Joo, H., Lin, Z. & Arnold, F. H. Laboratory evolution of peroxide-mediated cytochrome P450 hydroxylation. *Nature* **399**, 670–673 (1999).
7. Zhang, Y. X. *et al.* Genome shuffling leads to rapid phenotypic improvement in bacteria. *Nature* **415**, 644–646 (2002).
8. Pfleger, B. F., Pitera, D. J., Smolke, C. D. & Keasling, J. D. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nature Biotechnol.* **24**, 1027–1032 (2006).
9. Cadwell, R. C. & Joyce, G. F. Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* **2**, 28–33 (1992).
10. Shendure, J. *et al.* Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005).
11. Zhang, Y., Buchholz, F., Muyrers, J. P. & Stewart, A. F. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nature Genet.* **20**, 123–128 (1998).
12. Costantino, N. & Court, D. L. Enhanced levels of λ Red-mediated recombinants in mismatch repair mutants. *Proc. Natl Acad. Sci. USA* **100**, 15748–15753 (2003).
13. Sharan, S. K., Thomason, L. C., Kuznetsov, S. G. & Court, D. L. Recombineering: a homologous recombination-based method of genetic engineering. *Nature Protocols* **4**, 206–223 (2009).
14. Ellis, H. M., Yu, D., DiTizio, T. & Court, D. L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl Acad. Sci. USA* **98**, 6742–6746 (2001).
15. Markham, N. R. & Zuker, M. DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.* **33**, W577–W581 (2005).
16. Jin, Y. S. & Stephanopoulos, G. Multi-dimensional gene target search for improving lycopene biosynthesis in *Escherichia coli*. *Metab. Eng.* **9**, 337–347 (2007).
17. Kang, M. J. *et al.* Identification of genes affecting lycopene accumulation in *Escherichia coli* using a shot-gun method. *Biotechnol. Bioeng.* **91**, 636–642 (2005).
18. Chen, H., Bjerknes, M., Kumar, R. & Jay, E. Determination of the optimal aligned spacing between the Shine – Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Res.* **22**, 4953–4957 (1994).
19. Alper, H., Jin, Y. S., Moxley, J. F. & Stephanopoulos, G. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab. Eng.* **7**, 155–164 (2005).
20. Alper, H., Miyaoku, K. & Stephanopoulos, G. Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. *Nature Biotechnol.* **23**, 612–616 (2005).
21. Farmer, W. R. & Liao, J. C. Precursor balancing for metabolic engineering of lycopene production in *Escherichia coli*. *Biotechnol. Prog.* **17**, 57–61 (2001).

22. Kim, S. W. & Keasling, J. D. Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in *Escherichia coli* enhances lycopene production. *Biotechnol. Bioeng.* **72**, 408–415 (2001).

23. Yuan, L. Z., Rouviere, P. E., Larossa, R. A. & Suh, W. Chromosomal promoter replacement of the isoprenoid pathway for enhancing carotenoid production in *E. coli*. *Metab. Eng.* **8**, 79–90 (2006).

24. Khosla, C. & Keasling, J. D. Metabolic engineering for drug discovery and development. *Nature Rev. Drug Discov.* **2**, 1019–1025 (2003).

25. Cropp, T. A. & Schultz, P. G. An expanding genetic code. *Trends Genet.* **20**, 625–630 (2004).

26. Gibson, D. G. *et al.* Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215–1220 (2008).

27. Metzgar, D. *et al. Acinetobacter* sp. ADP1: an ideal model organism for genetic analysis and genome engineering. *Nucleic Acids Res.* **32**, 5780–5790 (2004).

28. Nakayama, M. & Ohara, O. Improvement of recombination efficiency by mutation of Red proteins. *Biotechniques* **38**, 917–924 (2005).

29. Datta, S., Constantino, N., Zhou, X. & Court, D. L. Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proc. Natl Acad. Sci. USA* **105**, 1626–1631 (2008).

30. Tian, J. *et al.* Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432**, 1050–1054 (2004).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.H.W. (hhwang@genetics.med.harvard.edu) or F.J.I. (farren@alumni.upenn.edu).

METHODS

Strains and culture conditions. The λ prophage was obtained from strain DY330[31], modified to include the *bla* gene and introduced into wild-type MG1655 *E. coli* by P1 transduction at the *bioA/bioB* gene locus and selected on ampicillin to yield the strain EcNR1 (λ-Red[+]). Replacement of *mutS* with the chloramphenicol resistance gene (*cmR* cassette) in EcNR1 produced EcNR2 (mutS[−], λ-Red[+]). EcNR2 was grown in low salt LB-min medium (10 g tryptone, 5 g yeast extract, 5 g NaCl in 1 l dH$_2$O) for optimal electroporation efficiency. A premature stop codon was introduced into the *cmR* gene of EcNR2 with oligo cat_fwd_stop (Supplementary Table 3) to produce EcFI5, thus inactivating the *cmR* gene. An oligo (cat_fwd_restore) containing the wild-type sequence was used to restore the CmR phenotype. The pAC-LYC plasmid[32] containing genes *crtE*, *crtB* and *crtI* was electroporated into EcNR1 to generate EcHW1, which produces lycopene at basal levels. Replacement of *mutS* with a kanamycin resistance gene in EcHW1 produced EcHW2.

Oligonucleotides and DNA sequencing. All oligonucleotides were obtained from Integrated DNA Technologies with standard purification. Oligonucleotides used in the MAGE process contained the following modifications: (1) 30–110 bp in length for optimization experiments; (2) different numbers of phosphorothioated bases; and (3) degenerate nucleotides as described elsewhere in this paper. Additional primers were purchased to amplify relevant genetic regions of the lycopene pathway to sequence strains that expressed high levels of lycopene. DNA sequencing to confirm allelic replacements was performed by Agencourt Bioscience.

LacZ and chloramphenicol replacement efficiency assays. Replacement efficiency was characterized by performing the allelic replacement protocol on EcNR2 cells using 90-mer oligos (Supplementary Table 3) that produced a premature stop codon in the chromosomal *lacZ* gene. In general, 250–500 cells were plated on LB-min containing 5-bromo-4-chloro-3-indoyl-β-D-galactoside and isopropyl-β-D-thiogalactoside (USB Biochemicals) agar plates. Efficiency of allelic replacement was calculated by taking the ratio of the number of white colonies to the total number of colonies on plates. A similar strategy was used in the *cmR* gene recovery experiments with strain EcFI5 where 30–110mer oligos were used to determine optimal oligo length for allelic replacement (Supplementary Fig. 2a). These oligos contained two phosphorothioate bonds at both the 5′ and 3′ termini. Cells were plated on LB-min-chloramphenicol and LB-min agar plates and grown overnight. Efficiency of allelic replacement was calculated by taking the ratio of the number of colonies on LB-min-chloramphenicol plates to the number of colonies on LB-min plates.

MAGE automation device. Cells were grown in sterilized 10-ml glass vials placed in thermally conductive blocks. The growth chambers were climate-regulated through temperature controllers that actuate Peltier heating/cooling elements. Cultures acclimatized quickly (<15 s) in chambers held at different temperatures due to small volumes (for example, 3 ml). Aeration of the culture was achieved through agitation of the chambers at 300 r.p.m. using an orbital shaker. Real-time monitoring of growth rates was achieved by detecting changes in light transmittance across the chamber from light-emitting diodes emitting at 600 nm. Cultures were transferred between chambers through solenoid isolation valves using syringe pumps. Cells were made electrocompetent by a filtration system, which uses a syringe pump to concentrate cells onto a filter membrane (0.22 μm pore size) and resuspend them off the membrane with appropriate electroporation buffer (for example, dH$_2$O). Single-stranded oligos (or PCR products) were electroporated into cells in a conductive cuvette with an electric pulse (18 kV cm^{-1}). After each cycle, the entire system (chambers, syringes, filters) was washed with 70% ethanol followed by dH$_2$O three times to reduce contamination and biofilm formation. All instruments were digitally controlled through software written in the LabView programming environment (National Instruments).

Cultures were initially inoculated into 3 ml of medium in the growth chamber. The device then executed repeatedly and continuously through the following steps: (1) grow cells at 30 °C to a pre-set density (that is, OD$_{600\,\text{nm}}$ of 0.7); (2) induce cells for allelic replacement via 42 °C heat shock for 15 min; (3) chill cells at 4 °C to halt cellular metabolism; (4) wash cells through 15–20 iterations of filtration and resuspension with dH$_2$O; (5) mix cell suspension and synthetic DNA; (6) deliver DNA into cells by electroporation; and (7) resuspend electroporated cells with growth media.

Colorimetric screen and assay for lycopene production. Cells from the cycled EcHW2 population were plated on LB-min-chloramphenicol agar plates and grown for 1 day at 30 °C and 2 days at room temperature to produce red colonies. The 24 gene targets were divided into three oligo pools, one containing 10 targets, a second containing 14 targets and a third containing all 24 targets, which were cycled through the MAGE process in three separate cultures. In total, 10^5 colonies with increased colour intensity by visual inspection were screened after cycles 5, 10, 15, 20, 25, 30 and 35. From 127 isolates, six strains (EcHW2a–f, with representation from each pool) were selected for direct DNA sequencing across all gene targets and quantitatively measured for lycopene production. For lycopene quantification, these isolated colonies were grown in LB-min medium for 24 h. Lycopene was extracted from 1 ml of cells as follows: centrifuged at 16,000*g* for 30 s, removal of supernatant media and resuspended with water, and re-centrifuged at 16,000*g* for 30 s. Once the supernatant was removed, the cells were resuspended in 200 μl acetone and incubated in the dark for 15 min at 55 °C with intermittent vortexing. The mixture was centrifuged at 16,000*g* for 30 s and the supernatant lycopene solution was transferred to a clean tube for quantification. Absorbance at 470 nm of the extracted lycopene solution was measured using a spectrophotometer and calibrated against a known lycopene standard (Sigma-Aldrich, catalogue number L9879) to determine the lycopene content. Dry cell weight was determined by baking a washed cell pellet at 105 °C for 24 h in the dark. Lycopene yield of different EcHW2 derivatives (EcHW2a–f) was calculated by normalizing the amount of lycopene extracted to the dry cell weight.

31. Yu, D. *et al.* An efficient recombination system for chromosome engineering in *Escherichia coli. Proc. Natl Acad. Sci. USA* **97**, 5978–5983 (2000).
32. Cunningham, F. X. Jr, Sun, Z., Chamovitz, D., Hirschberg, J. & Gantt, E. Molecular structure and enzymatic function of lycopene cyclase from the cyanobacterium *Synechococcus* sp strain PCC7942. *Plant Cell* **6**, 1107–1121 (1994).

Optimized Design Criteria for MAGE Oligonucleotides

Oligonucleotide-mediated allelic replacement was achieved in the modified *E. coli* strain EcNR2 (mutS[−], λ-Red[+]) by directing oligos to the lagging strand of the replication fork during DNA replication[1]. Targeting the lagging strand of replicating DNA with single-stranded oligonucleotides (ss-oligos) has been shown to be more efficient than targeting the leading strand in this mutS[−] strain (EcNR2)[2]. The replacement efficiency was characterized by using oligos either to inactivate the *lacZ* gene and screen for white colonies on Xgal/IPTG-containing agar plates or to fix a defective *cat* gene and select for chloramphenicol resistant colonies. In rare cases during MAGE experiments, we observed small genomic sequence changes in the oligo targeted region that are not by design, i.e., 7[+] bp mutations when only 6 bp are targeted (Fig. 3). We hypothesize that these additional mutations are likely the result of allelic replacement by faulty oligos that arise from errors during oligo synthesis. Purification of oligos may reduce instances of such cases.

Supplementary Figure 2a shows that replacement efficiency was found to be dependent on oligo length, highest at 90 basepairs (bp). We hypothesize that the 90 bp oligo has the most optimal replacement efficiency for two main reasons. First, the λ-Red single-stranded DNA-binding protein β, has been shown to require at least 30 bp to complex with oligos *in vitro*[3]. *In vivo*, shorter oligos have fewer basepairs of homology to hybridize to the targeted chromosomal site, thus decreasing the likelihood of replacement. Second, while oligos longer than 90 bp may have more regions of homology to the chromosome, they are also more likely to form secondary structures. Inhibitory secondary structures (*e.g.*, hairpin loops) can lead to dramatically lower efficiencies of replacement since reducing the number of exposed bases on the oligo will decrease the frequency of hybridization to its chromosomal target. Along these lines, we observed that oligos with computationally predicted minimal folding energies[4] of less than -12.5 kcal/mol showed significantly reduced allelic replacement frequencies experimentally

(Supplementary Fig. 2b). Phosphorothioate bonds located at the terminal bases may increase replacement efficiency by preventing *in vivo* degradation of synthetic oligonucleotide molecules by endogenous exonucleases in the cell[5]. Phosphorothioated nucleotides increased replacement efficiency by more than 2-fold when placed at the 5' terminus, but showed no effect when placed at the 3' terminus (Supplementary Fig. 2c). Increasing the number of phosphorothioated bases at the 5' terminus increased the efficiency of replacement, which saturates to its highest level at four phosphorothioated bases (Supplementary Fig. 2d). Oligonucleotides, in which all bases contained phosphorothioated bonds, did not incorporate into the chromosome (data not shown). The replacement efficiency remains high across a wide range of oligo concentrations (0.05-50 μM), thus allowing for large and highly complex oligo pools (Supplementary Fig. 2e). Allelic replacement efficiency was low when low concentrations of oligos (<0.05 μM) were used, suggesting a dilution effect. In fact, at low oligo concentrations (*i.e.*, 0.005 μM), there are on average three DNA molecules per volume of a cell (~10^{-18} m^3), leading to drastically decreased likelihood of a replacement event. Therefore, increasing the amount of oligos available for allelic replacement by either increasing the oligo concentration during electroporation or by increasing the oligo half-life inside the cell (via terminal phosphorothioated nucleotides) will lead to higher efficiencies of replacement. Interestingly, we observed chromosomal deletions of up to 45 kbp with a single 90mer oligo using the EcNR2 (recA+) strain as well as a recA- EcNR2 derivative, suggesting a recA-independent β-mediated mechanisms.

Design of MAGE Oligonucleotides for DXP Pathway

The main text of this paper described the design criteria that were implemented to optimize the DXP pathway for lycopene production. Here, we provide additional details to clarify our oligo design criteria. The design of every oligo was based on optimization experiments such that oligo length, concentration, stability, secondary structure, strand bias and modification were

optimal (Supplementary Table 1 and Supplementary Fig. 2). Two main oligo design strategies were implemented: 1) oligos with specified sequences produced specific changes by making targeted modification that knocked out the expression of target genes (*ytjC*, *fdhF*, *aceE*, *gdhA*) and 2) oligos with degenerate sequences produced diverse changes tailored for exploring a vast sequence space of RBS strengths. Importantly, in both oligo designs, the location of the genetic modification is precise and well-defined based on homology arms of the oligos. As described in the main text, the degenerate oligos were designed to mutate RBS sequences to be more similar to the canonical Shine-Dalgarno sequence (TAAGGAGGT)[6], giving rise to enhanced translation efficiency. More specifically, RBS optimization utilized 90mer oligo pools containing the DDRRRRRDDDD degeneracy at the 41-51 bp position of the oligos (D = G, A, T and R = G, A). This mutation region targeted the -4 through -14 positions from the start codon of each gene with an optimal RBS spacing of 5 bp for replacement by one of the oligos from the degenerate pool. We also calculated the cost and maximum level of degeneracy that can be introduced into a single oligo. For 30 USD we obtain 50 nmol yield of a 90mer oligo, giving us 3×10^{16} molecules, which can support full degeneracy of 27 bp.

MAGE Automation

Automation instrumentation was constructed using the following major components:

Electroporator: ECM 630	BTX Technologies Inc (MA, USA)
Digital controllers: RS-232 serial modules	Superlogics Inc. (MA, USA)
Syringe pumps: Cavro XLP600 9-port	Tecan Group Ltd. (NC, USA)
Solenoid valves: Miniature Rocker Isolation Valves	Central Distribution Sales (NH, USA)
Temperature controller: CNI-3233-C24	Omega Engineering Inc. (CT, USA)
Orbital shaker: Advanced 3500 Orbital Shaker	VWR International LLC (PA, USA)
Control system software: LabView	National Instruments (TX, USA)
Growth chamber system: custom manufactured	David Breslau Design Inc. (NH, USA)

1. Costantino, N. & Court, D.L. Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc Natl Acad Sci U S A* **100**, 15748-15753 (2003).
2. Ellis, H.M., Yu, D., DiTizio, T. & Court, D.L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc Natl Acad Sci U S A* **98**, 6742-6746 (2001).
3. Mythili, E., Kumar, K.A. & Muniyappa, K. Characterization of the DNA-binding domain of beta protein, a component of phage lambda red-pathway, by UV catalyzed cross-linking. *Gene* **182**, 81-87 (1996).
4. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-3415 (2003).
5. Wu, X.S. et al. Increased efficiency of oligonucleotide-mediated gene repair through slowing replication fork progression. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 2508-2513 (2005).
6. Chen, H., Bjerknes, M., Kumar, R. & Jay, E. Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of Escherichia coli mRNAs. *Nucleic Acids Res* **22**, 4953-4957 (1994).

Supplementary Figure 1

Supplementary Figure 1 | Diagram of the oligo-genome hybridization structure during mismatch, insertion, and deletion modifications.

Supplementary Figure 2

Supplementary Figure 2 | Characterization of the allelic replacement frequency in the MAGE strain (EcNR2) and its derivative (EcFI5) by screening for the introduction of a nonsense mutation in the *lacZ* gene or recovery of the *cmR* gene. a, Replacement efficiency as a function of oligonucleotide length. Oligos contain two phosphorothioated bonds at both the 3' and 5' termini. **b,** Predicted optimal folding energy ΔG of 90-mer oligos as a function of frequency of replacement. Oligos with ΔG < -12.5 kcal/mol are considered to have significant secondary structure that hinder allelic replacement. **c,** The effect of terminal phosphorothioated bonds on replacement efficiency. Oligos of 90 bp with 0, 1(*) or 2 (**) phosphorothioated bonds between bases at the 3', 5' or both 3' and 5' termini were tested. **d,** Replacement efficiency as a function of the number of consecutive terminal 5' phosphorothioated bonds of a 90mer oligo as measured by the introduction of a nonsense mutation in the *lacZ* gene. **e,** Replacement efficiency as a function of the concentration of 90-mer oligos containing no phosphorothioated bonds (in black) or 4 phosphorothioated bonds at the 5' terminus (in red). Solid lines represent data fitted using polynomial functions. Error bars, ±SD.

Supplementary Figure 3

Supplementary Figure 3 | Predicted distribution of genetic variants in a population that has undergone simultaneous allelic manipulation at 10 different genomic locations at 30% overall replacement efficiency. In this case, 10 different genes are simultaneously targeted for inactivation. Each colored solid line represents the histographic distribution of variants containing different numbers of knockouts (KO) across the population as a function of MAGE cycle number. As MAGE cycles increase, population evolves towards acquiring all 10 gene KO's. The population is binomially distributed according to the equation:

$$P(K,N) = \sum_{j=0}^{K} \binom{K}{j} \left((1-M)^N \right)^{(K-j)} \left(1 - (1-M)^N \right)^j ,$$

where K is the number of loci simultaneously targeted, N is the number of MAGE cycles, and M is the mutation rate at any individual locus.

Supplementary Table 1

MAGE Parameters	Optimal Values
Oligo length	90 bp
Oligo concentration range	0.05 – 50 μM
Oligo stability	Four 5' phosphorothioated bases
Oligo secondary structure	> –12.5 kcal/mol
Strand bias	Target lagging strand
Size of genetic modification	Predict efficiency using hybridization energy
Cycle time	2 – 2.5 hours

Supplementary Table 1 | Optimized parameters for maximal allelic replacement efficiency.

Supplementary Table 2

EcHW2a (KO's: none)		
idi	wild-type RBS	acatgtgagaaattatg
	optimized RBS	ggaaggggatgattatg
EcHW2b (KO's: Δ*gdhA*, Δ*ytjC*)		
dxs	wild-type RBS	ttaataggcccctgatg
	optimized RBS	taggaaatggtctgatg
EcHW2c (KO's: none)		
dxs	wild-type RBS	ttaataggcccctgatg
	optimized RBS	aaaaggaagaactgatg
ispA	wild-type RBS	ccggacaatgagtaatg
	optimized RBS	ggagaagggaagtaatg
EcHW2d (KO's: Δ*fdhF*)		
dxs	wild-type RBS	ttaataggcccctgatg
	optimized RBS	gtaaggagaagctgatg
EcHW2e (KO's: none)		
dxs	wild-type RBS	ttaataggcccctgatg
	optimized RBS	ggagaaggaaactgatg
idi	wild-type RBS	acatgtgagaaattatg
	optimized RBS	tgaggaataaaattatg
EcHW2f (KO's: Δ*ytjC*)		
dxs	wild-type RBS	ttaataggcccctgatg
	optimized RBS	tagagaagagactgatg
rpoS	wild-type RBS	gtaggagccaccttatg
	optimized RBS	gagaggatggacttatg
idi	wild-type RBS	acatgtgagaaattatg
	optimized RBS	aaaagaggttgattatg
dxr	wild-type RBS	actctggatgtttcatg
	optimized RBS	ttaagggtgtattcatg

Supplementary Table 2 | Optimized RBS sequences of strains EcHW2a-f

Programming cells by multiplex genome engineering and accelerated evolution

Supplementary Table 3 - Sequences of all oligos used
List generated by H.H.Wang, 2009

(note: * indicates phosphorothiolated bond)

Fig 2a	**introduce bp mismatches**	
	lacZ_mut_1	A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGAGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATC*C*C
	lacZ_mut_2	C*A*TGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTAAGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACAT*C*C
	lacZ_mut_3	G*A*CCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTTAATGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCA*C*A
	lacZ_mut_4	A*T*GACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTTAACACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAG*C*A
	lacZ_mut_6	A*T*TCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCTCTAGATTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGC*C*A
	lacZ_mut_10	T*G*ATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGAGCTCTTCTAACCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC*C*C
	lacZ_mut_12	A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTCAATAACGGTTGCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATC*C*C
	lacZ_mut_18	G*A*AAACCCTGGCGTTACCCAACTTAATCGCCTTGCATAGGGATAACAGGGTAATAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATC*G*C
	lacZ_mut_30	A*T*GATTACGGATTCACTGGCCGTCGTTTTATAAGCTTGGCATTCTAGATATCGTTAAGCCACCCAACTTAATCGCCTTGCAGCACATC*C*C
Fig 2b	**introduce bp insertions**	
	lacZ_ins_1	C*A*TGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGAGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACA*T*C
	lacZ_ins_2	C*A*TGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTAAGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCAC*A*T
	lacZ_ins_3	G*A*CCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTTAAGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCA*G*C
	lacZ_ins_4	A*T*GACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTTAACAGTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTT*G*C
	lacZ_ins_6	C*A*CTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGTCTAGAGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTT*C*G
	lacZ_ins_8	A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGAGCTCTTCTCTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGC*A*G
	lacZ_ins_10	T*G*ATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGAGCTCTTCTAACCTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTG*C*A
	lacZ_ins_12	A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTCAATAACGGTTGGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCC*T*T
	lacZ_ins_16	G*A*TTACGGATTCACTGGCCGTCGTTTTACAACGTCGTTCTAGAGCTCTTCCTAGACTGGGAAAACCCTGGCGTTACCCAACTTAATCG*C*C
	lacZ_ins_18	G*G*CGTTACCCAACTTAATCGCCTTGCAGCACATCCCTAGGGATAACAGGGTAATCCTTTCGCCACGTGGCGTAATAGCGAAGAGGCCC*G*C
	lacZ_ins_30	A*T*GATTACGGATTCACTGGCCGTCGTTTTATAAGCTTGGCATTCTAGATATCGTTAAGCCCAACGTCGTGACTGGGAAAACCCTGGCG*T*T
Fig 2c	**introduce bp deletions**	
	lacZ_del_1	A*G*CGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCTGAATGGCGAATGGCGCTTTGCCTGGTTTCCGGCACCAGAAGC*G*G
	lacZ_del_2	A*T*GACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGACTAGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCA*C*A
	lacZ_del_3	C*C*ATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATC*C*C
	lacZ_del_4	T*G*ATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGCCCTCGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCT*T*T
	lacZ_del_7	C*A*TGACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATC*C*C
	lacZ_del_10	A*T*GACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCC*T*T
	lacZ_del_17	A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGG*C*G
	lacZ_del_26	A*T*GACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGT*A*A
	lacZ_del_100	A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGAGTTGCGCAGCCTGAATGGCGAATGGCGCTTTGCCTGGTTTCC*G*G
	lacZ_del_1000	A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGTGATTCGAGGCGTTAACCGTCACGAGCATCATCCTCTGCATGG*T*C
	lacZ_del_3000	G*T*GAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTTGTCAAAAATAATAATAACCGGGCAGGCCATGTCTGCCCGTA*T*T
	lacZ_del_10000	A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGCCGTCCTGCAGATGAATCTGCCACAGCCCCTCTTCGCCTGGTA*A*C
	lacZ_del_45000	T*T*ACCACATCAAATCATCAGGCACTTTGAAGTCGGCATACGGGTCGTATAACATCCCACATCACTATCAAGAAGATTATCTTTACCTC*T*G
Fig 2d	**introduce bp mismatches**	hyb dG (kcal/mol)
	lacZ_oligo_m1_v1(**)	G*G*AAACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTTGACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAA*T*C	-117.7
	lacZ_oligo_m1_v2(**)	G*G*AAACAGCTatgACCATGATTACGGATTCACTGGCCAGTCGTTTGACAACGTAGCTGGTGGGAAAACCCTGGCGTTACCCAACTTAA*T*C	-109.5
	lacZ_oligo_m1_v3(**)	G*G*AAACAGCTatgACCATGATTACGGATGCACTGGCCGTCGTTTTACAACGTCGTGACTGAGAAAACCCTGGCGTTACCCAACTTAA*T*C	-112.3
	lacZ_oligo_m1_v4(**)	G*G*AAACAGCTatgACCATGATTACGGATGCACTGGCCGTCGTTTGACAACGTCGTGACTGAGAAAACCCTGGCGTTACCCAACTTAA*T*C	-109.6
	lacZ_oligo_m1_v5(**)	G*G*AAACAGCTatgACCATGATTACGGATGCACTGGCCAGTCGTTTGACAACGTAGCTGACTGAGAAAACCCTGGCGTTACCCAACTTAA*T*C	-101.4
	lacZ_oligo_m1_v6(**)	G*G*AAACAGCTatgACCATGATTACGGATGCACGGGCAGTCCTTTGACACCGTAGTGCCTGAGAAAACCCTGGCGTTACCCAACTTAA*T*C	-84.5
	lacZ_oligo_m2_v1(**)	G*G*AAACAGCTatgACCATGATTACGGATTCACTGGTAATCGTTGGCCAACGGACTGACTGGGAAAACCCTGGCGTTACCCAACTTAA*T*C	-94.4
	lacZ_oligo_m2_v2*(**)	G*G*AAACAGCTatgACCATGATTACGGAGGAACTGGCCGTCGTTGGCCAACGTCGTAACTGGGAAAACCCTGGCGTTACCCAACTTAA*T*C	-95.4
	lacZ_oligo_m2_v3(**)	G*G*AAACAGCTatgACCATGATTACGGAGGAACTGGTAATCGTTGGCCAACGGACTGACTTAAAAACCCTGGCGTTACCCAACTTAA*T*C	-76.6
	lacZ_oligo_m2_v4(**)	G*G*AAACAGCTatgACCATGATTACGGAGGAATAAGTAATTAATGGCCTGAGGACTTAGTTAAAAACCCTGGCGTTACCCAACTTAA*T*C	-64.0
	lacZ_oligo_m3_v1(**)	G*G*AAACAGATatgACCAGGATTACGGCTTCACTGGACGTCGTTTGACAACGTCGTGACTGGGCAAACCCTGACGTTACCCCAACTTAA*T*C	-87.2
	lacZ_oligo_m3_v2(**)	G*G*AAACAAAGatgtACCCGAATTACGACGTCACTGTAAGTCGTTGGCCAACGTTAAGACTGGACCAACCCTTAAGTTACCTCCCTTAA*T*C	-41.2
	lacZ_oligo_m3_v3(**)	A*G*AAACAGCTaGgACCATGATTCCGGATTCAGCCGTCGTTTGGCCAACGTCGGACCTGGGAAACAACTGGCGTTTAACAACTTAC*G*A	-94.3
	lacZ_oligo_m3_v4(**)	A*T*CAACAGCTCGAACCATGATGTAGGATTCACGAACCGTCGTTGGCCAACGTCGGACCTGGGAAACAACTGGCGTTTAACAACTTAC*G*A	-55.8
Fig 3	**introduce bp mismatches (degenerate oligos)**	
	lacZ_30mer_degen (**)	A*T*GATTACGGATTCACTGGCCGTCGTTTTACANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCAACTTAATCGCCTTGCAGCACATC*C*C
	lacZ_6inters_degen (**)	A*T*GATTACGGATTCACTGGCCGTCGTTTTACANCGTCGNGACTGNGAAAANCCTGGNGTTACNCAACTTAATCGCCTTGCAGCACATC*C*C
	lacZ_6mer_degen(**)	A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACNNNNNNAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATC*C*C
Fig 3	**PCR primers for sequencing**	
	lacZ_pcr_seq_fprimer	gtaaaacgacggcaGGCAGTGAGCGCAACG
	lacZ_pcr_seq_rprimer	TTCTCCGTGGGAACAAACG
Fig 5	**introduce bp mismatches (degenerate oligos)**	
	dxs_oligo	g*c*ggactacatcatccagcgtaataaataaacaataagtaDDRRRRRDDDDctgatgagtttttgatattgccaaataccccgaccctgg*c*a
	appY_oligo	a*t*tgacagatgaaaactacggagcaaacataatccatcttHHHHHYYYYHHaaaattaaaacaagttgatagtagtcaaataacaacc*a*a
	rpoS_oligo	c*a*tttaaatcatgaactttcagcgtattctgactcataagHHHHHYYYYHHccgtgatcccttgacggaacattcaagcaaaagcctg*g*t
	crl_oligo	t*c*aatctgctcttcgggtgtccactcggtaacgtcattgcHHHHHYYYYHHgtgatgcaactgttttaccaaattggcaaaatttagc*a*a
	elbA_oligo	a*g*aaataaacatcctaatgaaccatattaaataccgtgggDDRRRRRDDDDcaaatgaagtggatagtaattgacacggtaattcaac*t*t
	elbB_oligo	c*g*ccgcatccgctcagaattacgccaattttcttcattgtHHHHHYYYYHHcgcaattgactgaaacacattaattttttttaataaaa*a*t
	yjiD_oligo	t*t*tcaggtaaaacagcctgaagtgattgtcgcatcattttHHHHHYYYYHHacgtttagcaacatatgcggcttgccctgaacaccgg*g*c
	idi_oligo	t*t*cactcttcaattatctataatgatgagtgatcagaattDDRRRRRDDDDattatgcaaacggaacacgtcattttattgaatgcac*a*g
	ispC_oligo	t*c*gagccggtcgagcccagaatggtgagttgcttcatgaaHHHHHYYYYHHtgagacagaataaaaagcaaaacgccgccagccgatc*t*g
	ispD/F_oligo	g*a*accacggcgcagaatccaaatgagtggttgccatgttHHHHYYYYHHgatttatcgattgttttgccccgcagactgtgcgcg*c*t
	purHD_oligo	t*c*gcgaaaaattcagctaacgctctctgtaatagtcaaatDDRRRRRDDDDaccatgcaacaacgtcgtccagtccgccgcgctctgc*t*c
	rnlA_oligo	t*t*caggggggtattattgtagagtttccccatatgtttctDDRRRRRDDDDgaaatgacaatcaggagttacaaaaacttaaatctgg*t*c
	yggT_oligo	t*g*caatttttggtgcgcgtgattactctaaaaaataaggaDDRRRRRDDDDgccatgaaatacgttgacttttcctgctttcaacggtca*t*t
	ispE_oligo	c*a*ccgacggtggtcaacgcatcaagttaaaaatggataacDDRRRRRDDDDataatgcggacacagtggccctctccggcaaaactta*a*t
	ispG_oligo	t*t*gattttctacgttgaattggagcctggttatgcatgaaHHHHHYYYYHHgttacccgtctgttactgcgccggtgattgttcggca*t*t
	ispH_oligo	c*a*caaaaaccacgcgggttggccaacaggatctgcatgttHHHHHYYYYHHcggatcgatttccggatcgatttcaatacaaaatgaa*c*g
	ispA_oligo	g*c*tgtctgacaatgaagacgcctctctaaccccttttacaDDRRRRRDDDDgtaatggacttttccgcagcaactcgaagcctgcgtta*a*g
	ycgZ_oligo	c*t*ccccgcagaatctaaagtcactgaattttgatgcatgctHHHHHYYYYHHaaaaatgttaatcaatgttaagtgaagtaacatagc*a*a
	ymgA_oligo	t*t*tcatattttatacgttcattatcagatgtcttcataacHHHHHYYYYHHtaacctgttaaataatcaggctgttgcattattcaaa*a*a
	ariR_oligo	c*a*gttattgcattatgaattgtagtatcttcaagcatggtHHHHHYYYYHHttacttatgaaatttttaatgtattctgtttattttct*t*a
	gdhA_oligo	C*A*TATTCTCTGGAGTCATTCCTCAACCATGTCTAAAAGCGCGACCCGAATTAAACCGAGTTCGCGCAAGCCGTTCGTGAAGTAATGAC*C*A
	aceE_oligo	G*A*GCACGCTCAACACCTTCTTCACGGATGACCGATTGAGATCGCCTGGAGTCAGTCGCGAGTTTCGATCGGATCCACGTCATTTGGGAA*A*C
	fdhF_oligo	G*C*AAAATCGTCCGGGCGGAGGCAGCGCAGGGGTAAACCAACCAGGGTACCCTGTGACTGAAGGGTTATTATGGCTGGGACTTCATTAA*C*G
	ytjC_oligo	G*C*TCACCTTTGGCGGTCAGCGGGCTGTCAGACTGGCCCTAAATACGTCACTCGGCGTTCCACTGCGTTTCACCGTGGCGGACTAGGTA*T*A
Fig 5	**forward PCR primer for Sequence**	
	dxs_f	gtaaaacgacggcaACCAGCAACTTGGTAAAAGTACC
	appY_f	gtaaaacgacggcaAAAAATGTAACTTTGAAATAAGTTAGAATAAAAAACAA
	rpoS_f	gtaaaacgacggcaCGGGAACAACAAGAAGTTAAGG
	crl_f	gtaaaacgacggcaTCAGCCCGGAAGAGGAC
	elbA_f	gtaaaacgacggcaGAAGTCAATAATTTATTGCTTTCACAAAATCT
	elbB_f	gtaaaacgacggcaGCGCCAACACCAGGG
	yjiD_f	gtaaaacgacggcaCATGACTTTTTTCCTGAAAAGTCAGT
	idi_f	gtaaaacgacggcaCTCTCTATTCCTGTCATTTCTGACTG
	ispC_f	gtaaaacgacggcaAGACGACGATCGCCGT
	ispD/F_f	gtaaaacgacggcaCGAAACTTAAAGCGCGAAACG
	purHD_f	gtaaaacgacggcaGTAATTTCTGTATTTTGCCCACGG
	rnlA_f	gtaaaacgacggcaGAGTTGCTGCTTGGTGAGAT
	yggT_f	gtaaaacgacggcaAGGCAGTTTGAAGTTGCAC
	ispE_f	gtaaaacgacggcaTTACCGGGTGATGCAACC
	ispG_f	gtaaaacgacggcaCGTAAAGACGGTAATTTGAACTTAACC
	ispH_f	gtaaaacgacggcaCCGTGAATTTATGGATGCAGG
	ispA_f	gtaaaacgacggcaCTGAGCGAGCTGGAACAG
	ycgZ_f	gtaaaacgacggcaAAAAGTTGAACACTAGTTGGCGA

	ymgA_f	gtaaaacgacggcaAGGAAACTCTCGGGGAAATTG
	ariR_f	gtaaaacgacggcaTAAGCAATTACTCGTAATGAAATCGG
	gdhA_f	gtaaaacgacggcaGCTTACCGTCACATTCTTGATG
	aceE_f	gtaaaacgacggcaGCAACTAAACGTAGAACCTGTCT
	fdhF_f	gtaaaacgacggcaCCGTAATATCAGGGAATGACCC
	ytjC_f	gtaaaacgacggcaTGGCCTGAATTTTAGCGGG
Fig 5	**reverse PCR primer for sequencing**	
	dxs_r	CGATTTTGTCGCGGCG
	appY_r	CAAACAGAGAAAAAACAAGCGATAATG
	rpoS_r	CGATCATCCGGCGGC
	crl_r	GTCGTTCAACCACTTCAGTGT
	elbA_r	TATTATGTAAACATTTATTTGTTTATTTTGCTTTCTCC
	elbB_r	CAAACCCCCCCGGC
	yjiD_r	CCAGAGTGTGGCAGTACG
	idi_r	CAGGAGGCGGTAATTTCCACG
	ispC_r	GATCAACATCCTCAAGCGCT
	ispD/F_r	CGGCTTTCAGACCTGCC
	purHD_r	GGATCTGATGTTCTTCCATAATGGC
	rnlA_r	CAGCAGGATTGATGGTTTCAAAAA
	yggT_r	GTAAACCGGCAATCCAGATGA
	ispE_r	TGTCAATGCTGATATTCGCACC
	ispG_r	TCAGCGCAATGCGATAGTC
	ispH_r	AGGTGGCATCAAACACCG
	ispA_r	GACGCAGATCGTCATCATCC
	ycgZ_r	TCTTCATAACGACTTCATAGCTAACC
	ymgA_r	TACTTTATGAAATTTTAATGTATTCTGTTTATTTTCTTACC
	ariR_r	TATACTTAATAATTAGAAGTTACATATCATCAGCTG
	gdhA_r	CCTTTTACCACCGCCCATC
	aceE_r	CGTGCACGGAAGAAGTGG
	fdhF_r	CCAATAACGGCGCGC
	ytjC_r	ACGCGATCGCTGAGC
Suppl Fig 2a	**introduce bp mismatches**	
	cat_fwd_stop	GCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAA<u>TGA</u>ACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAA
	cat_fwd_restore	GCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAA<u>TGT</u>ACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAA
	cat_fwd_re110*	T*A*TCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAA*A*G
	cat_fwd_re90*	G*C*ATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTA*A*A
	cat_fwd_re70*	G*A*ACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTAC*G*G
	cat_fwd_re50*	A*G*GCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGC*T*G
	cat_fwd_re30*	G*T*CAGTTGCTCAATGTACCTATAACCAG*A*C
Suppl Fig 2b	**introduce 1 bp mismatch**	
	argO_oligo	a*g*gcaagcttagcgcctctgttttatttttccatcagatagcgcTtaactgaacaaggcttgtgcatgagcaataccgtctctcgcca*g*c
	atpE_oligo	t*a*gttaacgttctgatattgctctttaaataaaagcaacgcttaTtacgcgacagcgaacatcacgtacagacccagacctacagcga*t*c
	b1146_oligo	c*g*attagcctcatcgttctgtggttaaaaattgaaagtgttctgTtaatctttcggatagatatccggtcttaagtcagatttcgtaa*t*t
	b1228_oligo	c*a*atcaaatattctgcccatgcggggaaggatgagaagcttcgaTtaaggttcggctcgagcgtcagcgagagagcgttgccgcaggc*a*a
	b1578_oligo	c*g*ggcttgtaatgggttaagtgataacagatgtctggaaatataAgggcaaatccacaggtgaactacgctcctcctgtttgttacgca*a*t
	b4273_oligo	g*t*ataaaatatccgtattcatatcagcacaaggtggatttgcccTtatatttccagacatctgttatcacttaacccattacaagccc*g*c
	coaD_oligo	c*t*gccggagaatgtccatcaggcgctgatggcgaagttagcgtaAcgtttatgccggatggtatgccatccggcggcgcatgaattact*t*c
	fabH_oligo	c*c*agggaacacaaatgcaaattgcgtcatgttttaatccttatcTtagaaacgaaccagcgcggagcccaggtgaatccaccgccaa*a*g
	fliN_oligo	g*a*tatcattactccgtctgagcgaatgcgccgcctgagccgttaAtgatgaataaccacgctactgtgcaatcttccgcgccggtttc*t*g
	frlD_oligo	a*c*ggcgtgcgcggcgaaaaccattcagtaccacggtgcctggtaAgtataacgttggcgtgagcatcttcacgccaacgtgctgttac*t*t
	frmR_oligo	t*c*cgttgacgacactattgaactggttcgtgcctatcttaaataActgaatctattaccatattgaggaagcggagagatgaaatca*c*g
	hda_oligo	t*t*cggataaggcgttcgcgccgcatccgacaataaacaccttatTtacaacttcagaatttctttcacaaacggaatggtcagcttac*g*t
	hycl_oligo	t*c*accccatcaagaacatccctgtcctgattccttaatgaaaaaTtactcttcttccaccgctaactgcgcgaagccgccatttcctt*c*c
	hypB_oligo	a*t*ggaccagtggctgaactggctggagacacagcgattgcataAgcgttcccggccagatccgcaccattgacggcaaccaggcgaa*a*g
	ilvA_oligo	g*c*gcgggtaggcctgataagcgaagcgctatcaggcattttccTtaaccggccaaaaagaacctgaacgccgggttattggtttcgt*c*g
	kgtP_oligo	g*c*cggtgcaagcaccggctataccgtctggcaactgacccgtcaTtaaagacgcatcccctccctttgcgatgtagcatcaaagaaa*c*c
	lolA_oligo	t*a*ttatccgaaaaatcgagcgacagattgctcactcaggtgcctTtacttacgttgatcatctaccgtgacgccttgcggcggggtga*a*g
	lon_oligo	a*g*cctgccagccctgtttttattagtgcattttgcgcgaggtcaTtattttgcagtcacaacctgcatacagacggttcattttgca*g*c
	lplA_oligo	g*a*gctacgggagttatcggcatggatggcgggggctgtaaagtaAttacccgcccatgcgggcaactttctcttcgatttgccggatt*t*t
	lpxK_oligo	a*g*cgacattcatgactccatcaatcgaacgctgccgccggcgtaaTTagttgccagaagccagcaaggttagttgcgtaagcagtttcg*c*t
	marA_oligo	g*g*cgaatcgcgctttttacatccattaaatcattacaacagctaAttgaaaacgtgacaacgtcactgaggcaatcatgaaaccactt*t*c
	mreC_oligo	a*g*agagccagattacccagcgtccctggctacgatagctcgccaTtattgccctcccggcgcacgcgcaggcggttgagcaccacttt*g*c
	murF_oligo	g*t*taaagccggaataatatttgaccaaatgttcggccagccaaaTtaacatgtcccattctcctgtaaagcgcgtactaccttcttcca*t*g
	paaX_oligo	c*g*gatggacaaaactctcttctggcacaaccggagtcagaccgtTtatctgataaattggcataacgcctcctgttcaatattcaagc*c*g
	pbl_oligo	t*a*tgcccaaaatatatattgtatataccaggcttaatgaactaAacaatcgtaaggcttcagcaaaatgagcctaaagcctcttttt*t*t
	pgpA_oligo	c*c*cgatgagacgtgacaagcgtcacatcaggcatcggtgcacaaTtacgacagaatacccagcggccagtgacgataaaatacca*g*g
	plsX_oligo	a*t*atcccgtcacttgcaaactgcgagttcgctggcagcgtcctgTtaccgcagagttccgcttttgccaccgtccagcagctcaaaac*c*a
	priB_oligo	t*g*aaacggcagaacttgcgacgacggaaataacgtgccatatggTtagtctccagaatctatcaattcaatctgctcggcatgcaaaa*c*c
	priC_oligo	g*c*gcggaaaaaatcgaaaacaggttagcgcgtttaacccgctAcaatggagagaaatatgtcactggaaaatgccctgacgatgtc*a*a
	ptsN_oligo	t*a*tcaaatcattacggataccgaaggtactccggatgaagcgtaAttattcggtaatgtctcttttagacgttgtgagagaaaacagt*a*c
	relA_oligo	a*c*agtatatatcaatctacattgtagatacgagcaaatttcggcTtaatcccgtgcaaccgacgcgcgtcgatacatccggcacct*g*g
	speG_oligo	c*a*caaaacaccgggtcagactctcctgaagccgaccgcacaataAcattaataataatcgatcgtattttttgatggtgtaaacccgtt*c*g
	sucB_oligo	g*a*taatgccttatccggtctacagtgcaggtgaaacttaaactaTtacacgtccagcagcagacgcgtcggatcttccagcaactctt*t*g
	sufA_oligo	t*t*tgacatcgtcaagttgcttcagtattacgagacatagtaccgcTtaatacccaaagctttcgccacagccacattcattctgggctt*t*a
	yaiS_oligo	t*a*tttcacagtcaaacttgctgcattaaaaaaacataaaagctaAgaacgacgcgattatatgcgtcatgatcgcctgcgtgcagttg*c*a
	ybaA_oligo	t*g*aatttttgggctgacggcacgccgcacagcccgtcacgctacTtattcgtcgatgattgactcgaatccgccatagatcattcgct*t*g
	ybcF_oligo	a*c*gcgtcaagcgtcgcatcaggcacaaatgtctaatgcctacgaTtacagcgaaatacaggtccccgcttcgccaggttcgtctctt*c*a
	ybjK_oligo	t*c*ggttcaaggttgatgggtttttttgttatctaaaacttatctaTtaccctgcaaccctctcaaccatcctcaaaatctcctcgcgcg*a*t
	ybjR_oligo	g*a*gaaaactgccgcctggctacgcgaccacggaaaaactgccgcgTtaatcctgcccgtatttctccagcaatgcttcggcaatcgcct*g*a
	ycbK_oligo	a*a*taatacgatagttcatactgcccctgtttcgttaagcaattaTtaccagtgccgtgctggcccggtatcaatatgcacaaagttac*t*a
	ycgX_oligo	c*t*tttatatattgaagatgtagaaaataataaaccattaatctaActaaagttggatacttaagaaatgcttcataattcagtaaggc*a*t
	yejE_oligo	g*g*tgaagccgtccgcgacgcatttgatcctaataaggcggtgtaAcatgacgcaaactctgttagcgattgaaaatttgtcggtgggt*t*t
	yffB_oligo	g*g*tttcagtgattccagttatcagcaattttttccatgaggtgtaAtctatgtcgtgcccggttattgagctgacacaacagctttattc*g*c
	yfiA_oligo	t*c*ggtgaaaagcagcgccaacttcgtcgaagaagttgaagaagagtaAtcctttatattgagtgtatcgccaacgcgccttcgggcgcgtt*t*t
	ygaU_oligo	t*g*acgccaattcccattatccagcaggcgatggctggcaattaaTtactcttccggaatacgcaaacacttgcccccggataaatttta*t*c
	ygbA_oligo	t*g*atttaactgcaaattgccggacagatctgcctgtccggcataTtattcatgaggttttttcggacgatatttttccggcagttctg*g*c
	ygiZ_oligo	t*t*tctctgtctatgagagccgttaaaacgactctcatagatttaTtaatagcaaaatataaaccgtccccaaaaaagccaccaaccac*a*a
	yhcO_oligo	c*t*tgtccgacgatatgcccccggttcggggggcttttttttgcgTtaatgacgaacattaaaaacgtataccccagcctcttcct*c*t
	yibA_oligo	a*a*tgaaattataacttccgctattgataagctgaagcgttcataAcgtgagttgcctatgcacagtgggggattcccgccggcacggt*g*c
	yjjV_oligo	a*t*tgtgcgccaactgccggatgcggcgtgaacgccttatccggcTtacggcacgttaaacaacgtatacgtgttattaagcaacgctt*g*c
	ynfA_oligo	g*c*tgacacttcctcttatcttattgataaaatggatttatgttcTtacgtgcgccccagcccgcaacaatgatcaacatgccgcaaa*g*c
	ypjC_oligo	a*a*accaacatgaaaatcacttttttcttatattatcgaaaaggtTtattcatttctttagcgcattcaaaaaactgatcggcattat*t*t
	yqiB_oligo	a*g*ggttaacaggctttccaaatggtgtccttaggtttcacgacgTtaataaaccggaatcgccatcgctccatgtgctaaacagtatc*g*c
Suppl Fig 2c&d	**introduce bp mismatches**	
	lacZ_oligo_m1_v1(no *)	GGAAACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC
	lacZ_oligo_m1_v1(3'*)	GGAAACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAAT*C
	lacZ_oligo_m1_v1(5'*)	G*GAAACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC
	lacZ_oligo_m1_v1(*)	G*GAAACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAAT*C
	lacZ_oligo_m1_v1(3'**)	GGAAACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAA*T*C
	lacZ_oligo_m1_v1(5'**)	G*G*AAACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC
	lacZ_oligo_m1_v1(**)	G*G*AAACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAA*T*C
	lacZ_oligo_m1_v1(3'***)	G*G*A*AACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC
	lacZ_oligo_m1_v1(5'****)	G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC
	lacZ_oligo_m1_v1(5'****[8])	G*G*A*A*A*C*A*G*CTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC
Suppl Fig 2e	**introduce bp mismatches**	
	lacZ_oligo_m1_v1(no *)	GGAAACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC
	lacZ_oligo_m1_v1(5'****)	G*G*A*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC

Optimized Design Criteria for MAGE Oligonucleotides

Oligonucleotide-mediated allelic replacement was achieved in the modified *E. coli* strain EcNR2 (mutS$^-$, λ-Red$^+$) by directing oligos to the lagging strand of the replication fork during DNA replication[1]. Targeting the lagging strand of replicating DNA with single-stranded oligonucleotides (ss-oligos) has been shown to be more efficient than targeting the leading strand in this mutS$^-$ strain (EcNR2)[2]. The replacement efficiency was characterized by using oligos either to inactivate the *lacZ* gene and screen for white colonies on Xgal/IPTG-containing agar plates or to fix a defective *cat* gene and select for chloramphenicol resistant colonies. In rare cases during MAGE experiments, we observed small genomic sequence changes in the oligo targeted region that are not by design, i.e., 7$^+$ bp mutations when only 6 bp are targeted (Fig. 3). We hypothesize that these additional mutations are likely the result of allelic replacement by faulty oligos that arise from errors during oligo synthesis. Purification of oligos may reduce instances of such cases.

Supplementary Figure 2a shows that replacement efficiency was found to be dependent on oligo length, highest at 90 basepairs (bp). We hypothesize that the 90 bp oligo has the most optimal replacement efficiency for two main reasons. First, the λ-Red single-stranded DNA-binding protein β, has been shown to require at least 30 bp to complex with oligos *in vitro*[3]. *In vivo*, shorter oligos have fewer basepairs of homology to hybridize to the targeted chromosomal site, thus decreasing the likelihood of replacement. Second, while oligos longer than 90 bp may have more regions of homology to the chromosome, they are also more likely to form secondary structures. Inhibitory secondary structures (*e.g.*, hairpin loops) can lead to dramatically lower efficiencies of replacement since reducing the number of exposed bases on the oligo will decrease the frequency of hybridization to its chromosomal target. Along these lines, we observed that oligos with computationally predicted minimal folding energies[4] of less than -12.5 kcal/mol showed significantly reduced allelic replacement frequencies experimentally

(Supplementary Fig. 2b). Phosphorothioate bonds located at the terminal bases may increase replacement efficiency by preventing *in vivo* degradation of synthetic oligonucleotide molecules by endogenous exonucleases in the cell[5]. Phosphorothioated nucleotides increased replacement efficiency by more than 2-fold when placed at the 5' terminus, but showed no effect when placed at the 3' terminus (Supplementary Fig. 2c). Increasing the number of phosphorothioated bases at the 5' terminus increased the efficiency of replacement, which saturates to its highest level at four phosphorothioated bases (Supplementary Fig. 2d). Oligonucleotides, in which all bases contained phosphorothioated bonds, did not incorporate into the chromosome (data not shown). The replacement efficiency remains high across a wide range of oligo concentrations (0.05-50 μM), thus allowing for large and highly complex oligo pools (Supplementary Fig. 2e). Allelic replacement efficiency was low when low concentrations of oligos (<0.05 μM) were used, suggesting a dilution effect. In fact, at low oligo concentrations (*i.e.*, 0.005 μM), there are on average three DNA molecules per volume of a cell (~10^{-18} m^3), leading to drastically decreased likelihood of a replacement event. Therefore, increasing the amount of oligos available for allelic replacement by either increasing the oligo concentration during electroporation or by increasing the oligo half-life inside the cell (via terminal phosphorothioated nucleotides) will lead to higher efficiencies of replacement. Interestingly, we observed chromosomal deletions of up to 45 kbp with a single 90mer oligo using the EcNR2 (recA+) strain as well as a recA- EcNR2 derivative, suggesting a recA-independent β-mediated mechanisms.

Design of MAGE Oligonucleotides for DXP Pathway

The main text of this paper described the design criteria that were implemented to optimize the DXP pathway for lycopene production. Here, we provide additional details to clarify our oligo design criteria. The design of every oligo was based on optimization experiments such that oligo length, concentration, stability, secondary structure, strand bias and modification were

optimal (Supplementary Table 1 and Supplementary Fig. 2). Two main oligo design strategies were implemented: 1) oligos with specified sequences produced specific changes by making targeted modification that knocked out the expression of target genes (*ytjC*, *fdhF*, *aceE*, *gdhA*) and 2) oligos with degenerate sequences produced diverse changes tailored for exploring a vast sequence space of RBS strengths. Importantly, in both oligo designs, the location of the genetic modification is precise and well-defined based on homology arms of the oligos. As described in the main text, the degenerate oligos were designed to mutate RBS sequences to be more similar to the canonical Shine-Dalgarno sequence (TAAGGAGGT)[6], giving rise to enhanced translation efficiency. More specifically, RBS optimization utilized 90mer oligo pools containing the DDRRRRRDDDD degeneracy at the 41-51 bp position of the oligos (D = G, A, T and R = G, A). This mutation region targeted the -4 through -14 positions from the start codon of each gene with an optimal RBS spacing of 5 bp for replacement by one of the oligos from the degenerate pool. We also calculated the cost and maximum level of degeneracy that can be introduced into a single oligo. For 30 USD we obtain 50 nmol yield of a 90mer oligo, giving us 3×10^{16} molecules, which can support full degeneracy of 27 bp.

MAGE Automation

Automation instrumentation was constructed using the following major components:

Electroporator: ECM 630	BTX Technologies Inc (MA, USA)
Digital controllers: RS-232 serial modules	Superlogics Inc. (MA, USA)
Syringe pumps: Cavro XLP600 9-port	Tecan Group Ltd. (NC, USA)
Solenoid valves: Miniature Rocker Isolation Valves	Central Distribution Sales (NH, USA)
Temperature controller: CNI-3233-C24	Omega Engineering Inc. (CT, USA)
Orbital shaker: Advanced 3500 Orbital Shaker	VWR International LLC (PA, USA)
Control system software: LabView	National Instruments (TX, USA)
Growth chamber system: custom manufactured	David Breslau Design Inc. (NH, USA)

1. Costantino, N. & Court, D.L. Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc Natl Acad Sci U S A* **100**, 15748-15753 (2003).
2. Ellis, H.M., Yu, D., DiTizio, T. & Court, D.L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc Natl Acad Sci U S A* **98**, 6742-6746 (2001).
3. Mythili, E., Kumar, K.A. & Muniyappa, K. Characterization of the DNA-binding domain of beta protein, a component of phage lambda red-pathway, by UV catalyzed cross-linking. *Gene* **182**, 81-87 (1996).
4. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-3415 (2003).
5. Wu, X.S. et al. Increased efficiency of oligonucleotide-mediated gene repair through slowing replication fork progression. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 2508-2513 (2005).
6. Chen, H., Bjerknes, M., Kumar, R. & Jay, E. Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of Escherichia coli mRNAs. *Nucleic Acids Res* **22**, 4953-4957 (1994).

Supplementary Figure 1

Supplementary Figure 1 | Diagram of the oligo-genome hybridization structure during mismatch, insertion, and deletion modifications.

Supplementary Figure 2

Supplementary Figure 2 | Characterization of the allelic replacement frequency in the MAGE strain (EcNR2) and its derivative (EcFI5) by screening for the introduction of a nonsense mutation in the *lacZ* gene or recovery of the *cmR* gene. a, Replacement efficiency as a function of oligonucleotide length. Oligos contain two phosphorothioated bonds at both the 3' and 5' termini. **b,** Predicted optimal folding energy ΔG of 90-mer oligos as a function of frequency of replacement. Oligos with ΔG < -12.5 kcal/mol are considered to have significant secondary structure that hinder allelic replacement. **c,** The effect of terminal phosphorothioated bonds on replacement efficiency. Oligos of 90 bp with 0, 1(*) or 2 (**) phosphorothioated bonds between bases at the 3', 5' or both 3' and 5' termini were tested. **d,** Replacement efficiency as a function of the number of consecutive terminal 5' phosphorothioated bonds of a 90mer oligo as measured by the introduction of a nonsense mutation in the *lacZ* gene. **e,** Replacement efficiency as a function of the concentration of 90-mer oligos containing no phosphorothioated bonds (in black) or 4 phosphorothioated bonds at the 5' terminus (in red). Solid lines represent data fitted using polynomial functions. Error bars, ±SD.

Supplementary Figure 3

Supplementary Figure 3 | Predicted distribution of genetic variants in a population that has undergone simultaneous allelic manipulation at 10 different genomic locations at 30% overall replacement efficiency. In this case, 10 different genes are simultaneously targeted for inactivation. Each colored solid line represents the histographic distribution of variants containing different numbers of knockouts (KO) across the population as a function of MAGE cycle number. As MAGE cycles increase, population evolves towards acquiring all 10 gene KO's. The population is binomially distributed according to the equation:

$$P(K,N) = \sum_{j=0}^{K} \binom{K}{j}\left((1-M)^N\right)^{(K-j)}\left(1-(1-M)^N\right)^j,$$

where K is the number of loci simultaneously targeted, N is the number of MAGE cycles, and M is the mutation rate at any individual locus.

Supplementary Table 1

MAGE Parameters	Optimal Values
Oligo length	90 bp
Oligo concentration range	0.05 – 50 μM
Oligo stability	Four 5' phosphorothioated bases
Oligo secondary structure	> –12.5 kcal/mol
Strand bias	Target lagging strand
Size of genetic modification	Predict efficiency using hybridization energy
Cycle time	2 – 2.5 hours

Supplementary Table 1 | Optimized parameters for maximal allelic replacement efficiency.

Supplementary Table 2

EcHW2a (KO's: none)		
idi	wild-type RBS	acatgtgagaaatt**atg**
	optimized RBS	ggaaggggatgatt**atg**
EcHW2b (KO's: Δ*gdhA*, Δ*ytjC*)		
dxs	wild-type RBS	ttaataggcccctg**atg**
	optimized RBS	taggaaatggtctg**atg**
EcHW2c (KO's: none)		
dxs	wild-type RBS	ttaataggcccctg**atg**
	optimized RBS	aaaaggaagaactg**atg**
ispA	wild-type RBS	ccggacaatgagta**atg**
	optimized RBS	ggagaagggaagta**atg**
EcHW2d (KO's: Δ*fdhF*)		
dxs	wild-type RBS	ttaataggcccctg**atg**
	optimized RBS	gtaaggagaagctg**atg**
EcHW2e (KO's: none)		
dxs	wild-type RBS	ttaataggcccctg**atg**
	optimized RBS	ggagaaggaaactg**atg**
idi	wild-type RBS	acatgtgagaaatt**atg**
	optimized RBS	tgaggaataaaatt**atg**
EcHW2f (KO's: Δ*ytjC*)		
dxs	wild-type RBS	ttaataggcccctg**atg**
	optimized RBS	tagagaagagactg**atg**
rpoS	wild-type RBS	gtaggagccacctt**atg**
	optimized RBS	gagaggatggactt**atg**
idi	wild-type RBS	acatgtgagaaatt**atg**
	optimized RBS	aaaagaggttgatt**atg**
dxr	wild-type RBS	actctggatgtttc**atg**
	optimized RBS	ttaagggtgtattc**atg**

Supplementary Table 2 | Optimized RBS sequences of strains EcHW2a-f

Programming cells by multiplex genome engineering and accelerated evolution

Supplementary Table 3 - Sequences of all oligos used
List generated by H.H.Wang, 2009

(note: * indicates phosphorothiolated bond)

Fig 2a   introduce bp mismatches

Name	Sequence
lacZ_mut_1	A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGAGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATC*C*C
lacZ_mut_2	C*A*TGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTAAGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACAT*C*C
lacZ_mut_3	G*A*CCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTTAATGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCA*C*A
lacZ_mut_4	A*T*GACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTTAACACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAG*C*A
lacZ_mut_6	A*T*TCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCTCTAGATTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGC*C*A
lacZ_mut_10	T*G*ATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGAGCTCTTCTAACCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC*C*C
lacZ_mut_12	A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTCAATAACGGTTGCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATC*C*C
lacZ_mut_18	G*A*AAACCCTGGCGTTACCCAACTTAATCGCCTTGCATAGGGATAACAGGGTAATAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATC*G*C
lacZ_mut_30	A*T*GATTACGGATTCACTGGCCGTCGTTTTATAAGCTTGGCATTCTAGATATCGTTAAGCCACCCAACTTAATCGCCTTGCAGCACATC*C*C

Fig 2b   introduce bp insertions

Name	Sequence
lacZ_ins_1	C*A*TGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGAGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACA*T*C
lacZ_ins_2	C*A*TGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTAAGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCAC*A*T
lacZ_ins_3	G*A*CCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTTAAGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCA*G*C
lacZ_ins_4	A*T*GACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTTAACACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTT*C*G
lacZ_ins_6	C*A*CTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGTCTAGAGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTT*C*G
lacZ_ins_8	A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGAGCTCTTCTCTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGC*A*G
lacZ_ins_10	T*G*ATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGAGCTCTTCTAACCTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTG*C*A
lacZ_ins_12	A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTCAATAACGGTTGGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCC*T*T
lacZ_ins_16	G*A*TTACGGATTCACTGGCCGTCGTTTTACAACGTCGTTCTAGAGCTCTTCCTAGACTGGGAAAACCCTGGCGTTACCCAACTTAATCG*C*C
lacZ_ins_18	G*G*CGTTACCCAACTTAATCGCCTTGCAGCACATCCCTAGGGATAACAGGGTAATCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCC*G*C
lacZ_ins_30	A*T*GATTACGGATTCACTGGCCGTCGTTTTATAAGCTTGGCATTCTAGATATCGTTAAGCCCAACGTCGTGACTGGGAAAACCCTGGCG*T*T

Fig 2c   introduce bp deletions

Name	Sequence
lacZ_del_1	A*G*CGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCTGAATGGCGAATGGCGCTTTGCCTGGTTTCCGGCACCAGAAGC*G*G
lacZ_del_2	A*T*GACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGACTAGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCA*C*A
lacZ_del_3	C*C*ATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATC*C*C
lacZ_del_4	T*G*ATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGCCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCT*T*T
lacZ_del_7	C*A*TGACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATC*C*C
lacZ_del_10	A*T*GACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCC*T*T
lacZ_del_17	A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGG*C*G
lacZ_del_26	A*T*GACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGT*A*A
lacZ_del_100	A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGAGTTGCGCAGCCTGAATGGCGAATGGCGCTTTGCCTGGTTTCC*G*G
lacZ_del_1000	A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGTGATTCGAGGCGTTAACCGTCACGAGCATCATCCTCTGCATGG*T*C
lacZ_del_3000	G*T*GAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTTGTCAAAAATAATAATAACCGGGCAGGCCATGTCTGCCCGTA*T*T
lacZ_del_10000	A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGCCGTCCTGCAGATGAATCTGCCACAGCCCCTCTTCGCCTGGTA*A*C
lacZ_del_45000	T*T*ACCACATCAAATCATCAGGCACTTTGAAGTCGGCATACGGGTCGTATAACATCCCACATCACTATCAAGAAGATTATCTTTACCTC*T*G

Fig 2d   introduce bp mismatches

Name	Sequence	hyb dG (kcal/mol)
lacZ_oligo_m1_v1(**)	G*G*AAACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTTGACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAA*T*C	-117.7
lacZ_oligo_m1_v2(**)	G*G*AAACAGCTatgACCATGATTACGGATTCACTGGCCAGTCGTTTGACAACGTAGTGACTGG GAAAACCCTGGCGTTACCCAACTTAA*T*C	-109.5
lacZ_oligo_m1_v3(**)	G*G*AAACAGCTatgACCATGATTACGGATGCACTGGCCGTCGTTTTACAACGTCGTGACTGAGAAAACCCTGGCGTTACCCAACTTAA*T*C	-112.3
lacZ_oligo_m1_v4(**)	G*G*AAACAGCTatgACCATGATTACGGATGCACTGGCCGTCGTTTGACAACGTCGTGACTGAGAAAACCCTGGCGTTACCCAACTTAA*T*C	-109.6
lacZ_oligo_m1_v5(**)	G*G*AAACAGCTatgACCATGATTACGGATGCACTGGCCAGTCGTTTGACAACGTAGTGACTGAGAAAACCCTGGCGTTACCCAACTTAA*T*C	-101.4
lacZ_oligo_m1_v6(**)	G*G*AAACAGCTatgACCATGATTACGGATGCACGGGCAGTCCTTTGACACCGTAGTGCCTGAGAAAACCCTGGCGTTACCCAACTTAA*T*C	-84.5
lacZ_oligo_m2_v1(**)	G*G*AAACAGCTatgACCATGATTACGGATTCACTGGTAATCGTTGGCCAACGGACTGACTGGGAAAACCCTGGCGTTACCCAACTTAA*T*C	-94.4
lacZ_oligo_m2_v2*(**)	G*G*AAACAGCTatgACCATGATTACGGAGGAACTGGCCGTCGTTGGCCAACGTCGTGACTGGCCGTTACCCAACTTAA*T*C AATAAAACCCTGGCGTTACCCAACTTAA*T*C	-95.4
lacZ_oligo_m2_v3(**)	G*G*AAACAGCTatgACCATGATTACGGAGGAACTGGTAATCGTTGGCCAACGGACTGACTTAAAAACCCTGGCGTTACCCAACTTAA*T*C	-76.6
lacZ_oligo_m2_v4(**)	G*G*AAACAGCTatgACCATGATTACGGAGGAATAAGTAATTAATGGCCTGAGGACTTAGTTAAAAACCCTGGCGTTACCCAACTTAA*T*C	-64.0
lacZ_oligo_m3_v1(**)	G*G*AAACAGATatgACCAGGATTACGGCTTCACTGGACGTCGTTTGACAACGTCGTGACTGGGCAAACCCTGACGTTACCCGACTTAA*T*C	-87.2
lacZ_oligo_m3_v2(**)	G*G*AAACAAAGatgACCCGAATTACGACGTCACTGTAAGTCGTTGGCCAACGTTAAGACTGGACCAACCCTTAAGTTACCTCCCTTAA*T*C	-41.2
lacZ_oligo_m3_v3(**)	A*G*AAACAGCTaGgACCATGATTCCGGATTCACTAGCCGTCGTTTGGCCAACGTCGGACCTGGGAAACAACTGGCGTTTAACAACTTAC*G*A	-94.3
lacZ_oligo_m3_v4(**)	A*T*CAACAGCTCGAACCATGATGTAGGATTCACGAACCGTCGTTGGCCAACGTCGGACCTGGGAAACAACTGGCGTTTAACAACTTAC*G*A	-55.8

Fig 3   introduce bp mismatches (degenerate oligos)

Name	Sequence
lacZ_30mer_degen (**)	A*T*GATTACGGATTCACTGGCCGTCGTTTTACA NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN CAACTTAATCGCCTTGCAGCACATC*C*C
lacZ_6inters_degen (**)	A*T*GATTACGGATTCACTGGCCGTCGTTTTACA NCGTCGNGACTGNGAAAANCCTGGNGTTACNCAACTTAATCGCCTTGCAGCACATC*C*C
lacZ_6mer_degen(**)	A*T*GATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGAC NNNNNNAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATC*C*C

Fig 3   PCR primers for sequencing

Name	Sequence
lacZ_pcr_seq_fprimer	gtaaaacgacggcaGGCAGTGAGCGCAACG
lacZ_pcr_seq_rprimer	TTCTCCGTGGGAACAAACG

Fig 5   introduce bp mismatches (degenerate oligos)

Name	Sequence
dxs_oligo	g*c*ggactacatcatccagcgtaataaataaacaataagtaDDRRRRRDDDctgatgagttttgatattgccaaatacccgaccctgg*c*a
appY_oligo	a*t*tgacagatgaaaactacggagcaaacataatccatcttHHHHYYYYYHHaaaattaaaacaagttgatagtagtcaaataacaacc*a*a
rpoS_oligo	c*a*tttaaatcatgaactttcagcgtattctgactcataagHHHHHYYYYYHHccgtgatcccttgacggaacattcaagcaaaagcctg*g*t
crl_oligo	t*c*aatctgctcttcgggtgtccactcggtaacgtcattgcHHHHHYYYYYHHgtgatgcaactgttttaccaaattggcaaaatttagc*a*a
elbA_oligo	a*g*aaataaacatcctaatgaaccatattaaataccgtgggDDRRRRRDDDcaaatgaagtggatagtaattgacacggtaattcaac*c*t
elbB_oligo	c*g*ccgcatccgctcagaattacgccaattttcttcattgtHHHHHYYYYYHHcgcaattgactgaaacacattaattttttttaataaaa*a*t
yjiD_oligo	t*t*tcaggtaaaacagcctgaagtgattgtcgcatcattttHHHHHYYYYYHHacgttttagcaacatatgcggcttgccctgaacaccgg*g*c
idi_oligo	t*t*cactcttcaattatctataatgatgagtgatcagaattDDRRRRRDDDattatgcaaacggaacacgtcattttattgaatgcac*a*g
ispC_oligo	t*c*gagccggtcgagcccagaatggtgagttgcttcatgaaHHHHHYYYYYHHtgagacagaataaaaagcaaaacgccgccagccgatc*c*g
ispD/F_oligo	g*a*accacgcgcaaacatcccaaatgagtggttgccatgttHHHHYYYYYHHgatttatcgattgttttgccccgcagactgtgcgcg*c*t
purHD_oligo	t*c*gcgaaaaattcagctaacgctctctgtaatagtcaaatDDRRRRRDDDaccatgcaacaacgtcgtccagtccgccgcgctctgc*t*c
rnlA_oligo	t*t*caggggggtattattgtagagttttccccatatgtttctDDRRRRRDDDgaaatgacaatcaggagttacaaaaacttaaatctgg*t*c
yggT_oligo	t*g*caattttttggtgcgcgtgattactctaaaaaataaggaDDRRRRRDDDccatgaatacgttgacttttcctgctttcaacggtca*t*t
ispE_oligo	c*a*ccgacggtggtcaacgcatcaagttaaaaatggataacDDRRRRRDDDataatgcggacacagtggccctctccggcaaaactta*a*t
ispG_oligo	t*t*tgattttctacgttgaattggagcctggttatgcatgaaHHHHHYYYYYHHgttacccgtctgttactgcgccggtgattgttcggca*t*t
ispH_oligo	c*a*tcaaaaaccacgcgggttggccaacaggatctgcatgttHHHHHYYYYYHHgcggatcgatttccggatcgatttccaataatcaaaatgaa*c*g
ispA_oligo	g*c*tgtctgacaatgaagacgcctctctaacccctttttacaDDRRRRRDDDgtaatggactttccgcagcaactcgaagcctgcgtta*a*g
ycgZ_oligo	c*t*cccgcagaatctaaagtcactgaattttgatgcatgctHHHHYYYYYHHaaaaatgttaatcaatgttaagtgaagtaacaatagc*a*a
ymgA_oligo	t*t*tcatattttatacgttcattatcagatgtcttcataacHHHHYYYYYHHtaacctgttaaataatcaggctgttgcattattcaaa*a*a
ariR_oligo	c*a*gttattgcattatgaattgtagtatcttcaagcatggtHHHHHYYYYYHHttacttatgaaattttaatgtattctgtttatttttct*t*a
gdhA_oligo	C*A*TATTCTCTGGAGTCATTCCTCAACCATGTC TAAAAGCGCGACCCGAAT TAAACCGAGTTCGCGCAAGCCGTTCGTGAAGTAATGAC*C*A
aceE_oligo	G*A*GCACGCTCAACACCTTCTTCACGGATGACCGA TTAGATCGCCTGGAGTCAGTCGCGAGTTTCGATCGGATCCACGTCATTTGGGAA*A*C
fdhF_oligo	G*C*AAAATCGTCCGGGCGGAGGCAGCGCAGGGG TAAACCAACCAGGGTACCCTG TGACTGAAGGGTTATTATGGCTGGGACTTCATTAA*C*G
ytjC_oligo	G*C*TCACCTTTGGCGGTCAGCGGGCTGTCAGACTGGCC CTAAATACGTCACTCGGCGTTCCACTGCGGTTTCACCGTGGCGGACTAGGTA*T*A

Fig 5   forward PCR primer for Sequence

Name	Sequence
dxs_f	gtaaaacgacggcaACCAGCAACTTGGTAAAAGTACC
appY_f	gtaaaacgacggcaAAAAATGTAACTTTGAAATAAGTTAGAATAAAAAACAA
rpoS_f	gtaaaacgacggcaCGGGAACAACAAGAAGTTAAGG
crl_f	gtaaaacgacggcaTCAGCCCGGAAGAGGAC
elbA_f	gtaaaacgacggcaGAAGTCAATAATTTATTGCTTTCACAAAATCT
elbB_f	gtaaaacgacggcaGCGCCAACACCAGGG
yjiD_f	gtaaaacgacggcaCATGACTTTTTTCCTGAAAAGTCAGT
idi_f	gtaaaacgacggcaCTCTCTATTCCTGTCATTTCTGACTG
ispC_f	gtaaaacgacggcaAGACGACGATCGCCGT
ispD/F_f	gtaaaacgacggcaCGAAACTTAAAGCGCGAAACG
purHD_f	gtaaaacgacggcaGTAATTTCTGTATTTTGCCCACGG
rnlA_f	gtaaaacgacggcaGAGTTGCTGCTTGGTGAGAT
yggT_f	gtaaaacgacggcaAGGCAGTTTGAAGTTGCAC
ispE_f	gtaaaacgacggcaTTACCGGGTGATGCAACC
ispG_f	gtaaaacgacggcaCGTAAAGACGGTAATTTGAACTTAACC
ispH_f	gtaaaacgacggcaCCGTGAATTTATGGATGCAGG
ispA_f	gtaaaacgacggcaCTGAGCGAGCTGGAACAG
ycgZ_f	gtaaaacgacggcaAAAAGTTGAACACTAGTTGGCGA

ymgA_f	gtaaaacgacggcaAGGAAACTCTCGGGGAAATTG
ariR_f	gtaaaacgacggcaTAAGCAATTACTCGTAATGAAATCGG
gdhA_f	gtaaaacgacggcaGCTTACCGTCACATTCTTGATG
aceE_f	gtaaaacgacggcaGCAACTAAACGTAGAACCTGTCT
fdhF_f	gtaaaacgacggcaCCGTAATATCAGGGAATGACCC
ytjC_f	gtaaaacgacggcaTGGCCTGAATTTTAGCGGG

Fig 5   **reverse PCR primer for sequencing**

dxs_r	CGATTTTGTCGCGGCG
appY_r	CAAACAGAGAAAAAACAAGCGATAATG
rpoS_r	CGATCATCCGGCGGC
crl_r	GTCGTTCAACCACTTCAGTGT
elbA_r	TATTATGTAAACATTTATTTGTTTATTTTGCTTTCTCC
elbB_r	CAAACCCCCCCGGC
yjiD_r	CCAGAGTGTGGCAGTACG
idi_r	CAGGAGGCGGTAATTTCCACG
ispC_r	GATCAACATCCTCAAGCGCT
ispD/F_r	CGGCTTTCAGACCTGCC
purHD_r	GGATCTGATGTTCTTCCATAATGGC
rnlA_r	CAGCAGGATTGATGGTTTCAAAAA
yggT_r	GTAAACCGGCAATCCAGATGA
ispE_r	TGTCAATGCTGATATTCGCACC
ispG_r	TCAGCGCAATGCGATAGTC
ispH_r	AGGTGGCATCAAACACCG
ispA_r	GACGCAGATCGTCATCATCC
ycgZ_r	TCTTCATAACGACTTCATAGCTAACC
ymgA_r	TACTTTATGAAATTTTAATGTATTCTGTTTATTTTCTTACC
ariR_r	TATACTTAATAATTAGAAGTTACATATCATCAGCTG
gdhA_r	CCTTTACCACCGCCCATC
aceE_r	CGTGCACGGAAGAAGTGG
fdhF_r	CCAATAACGGCGCGC
ytjC_r	ACGCGATCGCTGAGC

Suppl Fig 2a   **introduce bp mismatches**

cat_fwd_stop	GCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAA<u>TGA</u>ACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAA
cat_fwd_restore	GCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAA<u>TGT</u>ACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAA
cat_fwd_re110*	T*A*TCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAA*A*G
cat_fwd_re90*	G*C*ATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTA*A*A
cat_fwd_re70*	G*A*ACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTAC*G*G
cat_fwd_re50*	A*G*GCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGC*T*G
cat_fwd_re30*	G*T*CAGTTGCTCAATGTACCTATAACCAG*A*C

Suppl Fig 2b   **introduce 1 bp mismatch**

argO_oligo	a*g*gcaagcttagcgcctctgttttattttttccatcagatagcgcTtaactgaacaaggcttgtgcatgagcaataccgtctctcgcca*g*c
atpE_oligo	t*a*gttaacgttctgatattgctctttaaataaaagcaacgcttaTtacgcgacagcgaacatcacgtacagaccccagacctacagcga*t*c
b1146_oligo	c*g*attagcctcatcgttctgtggttaaaaattgaaagtgttctgTtaatctttcggatagatatccggtcttaagtcagatttcgtaa*t*t
b1228_oligo	c*a*atcaaatattctgcccatgcggggaaggatgagaagcttcgaTtaaggttcggctcgagcgtcagcgagagacgttgccgcaggc*a*a
b1578_oligo	c*g*ggcttgtaatgggttaagtgataacagatgtctggaaatataAgggcaaatccacaggtgaactacgctcctctgtttgttacgca*a*t
b4273_oligo	g*t*ataaaatatccgtattcatatcagcacaaggtggatttgcccTtatatttccagacatctgttatcacttaacccattacaagccc*g*c
coaD_oligo	c*t*gccggagaatgtccatcaggcgctgatggcgaagttagcgtaAcgtttatgccggatggtatgccatccggccgcgcatgaattact*t*c
fabH_oligo	c*c*agggaacacaaatgcaaattgcgtcatgttttaatccttatcTtagaaacgaaccagcgcggagcccaggtgaatccaccgccaa*a*g
fliN_oligo	g*a*tatcattactccgtctgagcgaatgcgccgcctgagccgttaAtgatgaataaccacgctactgtgcaatcttccgcgccggtttc*t*g
frlD_oligo	a*c*ggcgtgcgcggcgaaaaccattcagtaccacggtgcctggtaAgtataacgttggcgtgagcatcttcacgccaacgtgctgttac*t*t
frmR_oligo	t*c*cgttgacgacactattgaactggttcgtgcctatcttaaataActgaatctattaccatattgaggaagagcgagagatgaaatca*c*g
hda_oligo	t*t*cggataaggcgttcgcgccgcatccgacaataaacaccttatTtacaacttcagaatttctttcacaaacggaatggtcagcttac*g*t
hycI_oligo	t*c*accccatcaagaacatccctgtcctgattccttaatgaaaaaTtactcttcttccaccgctaactgcgcgaagccgccatttcctt*c*c
hypB_oligo	a*t*ggaccagtggctgaactggctggagacacagcgattgcataAgcgttcccggccagatccgcaccattgacggcaaccaggcgaa*a*g
ilvA_oligo	g*c*gcgggtaggcctgataagcgaagcgctatcaggcatttttccTtaacccgccaaaaagaacctgaacgccgggttattggtttcgt*c*g
kgtP_oligo	g*c*cggtgcaagcaccggctataccgtctggcaactgacccgtcaTtaaagacgcatcccctttcccttttgcgatgtagcatcaaagaaa*c*c
lolA_oligo	t*a*ttatccgaaaaatcgagcgacagattgctcactcaggtgcctTtacttacgttgatcatctaccgtgacgccttgcggcggggtga*a*g
lon_oligo	a*g*cctgccagccctgtttttattagtgcattttgcgcgaggtcaTtattttgcagtcacaacctgcataccagacggttcattttgca*g*c
lplA_oligo	g*a*gctacgggagttatcggcatggatggcgggggctgtaaagtaAttacccgcccatgcgggcaactttctcttcgatttgccggatt*t*t
lpxK_oligo	a*g*cgacattcatgactccatcaatcgaacgctgccgccggcgtaaTtagttgccagaagccagcaaggttagttgcgtaagcagtttcg*c*t
marA_oligo	g*g*cgaatcgcgcttttttacatccattaaatcattacaacagctaAttgaaaacgtgacaacgtcactgaggcaatcatgaaaccactt*t*c
mreC_oligo	a*g*agagccagattacccagcgtccctggctacgatagctcgccaTtattgccctcccggcgcacgcgcaggcggttgagcaccacttt*g*c
murF_oligo	g*t*taaagccggaataatatttgacccaaatgttcggccagccaaaTtaacatgtcccattctcctgtaaagcgcgtactacctcttcca*t*g
paaX_oligo	c*g*gatggacaaaactctcttctggcacaaccggagtcagaccgtTtatctgataaattggcataacgcctcctgttcaatattcaagc*c*g
pbl_oligo	t*a*tgcccaaaatatatattgtatataccaggcttaatgaactaAacaatcgtaaggcttcagcaaaatgagcctaaagcctctttttt*t*t
pgpA_oligo	c*c*tgatgagacgtgacaagcgtcacatcaggcatcggtgcacaaTtacgacagaatacccagcggccagtggagacccgataaaata ca*g*g
plsX_oligo	a*t*t*ataccgtcacttgcaaactgcgagttcgctggcagcgtcctgTtaccgcagagttccgcttttgccaccgtccagcagctcaaaac*c*a
priB_oligo	t*g*aaacggcagaacttgcgacgacggaaataacgtgccatatggTtagtctccagaatctatcaattcaatctgctcggcatgcaaaa*c*c
priC_oligo	g*c*gctggaaaaaatcgaaaacaggttagcgcgtttaacccgctAcaatggagagaatatgtcactggaaaatgccctgacgatgtc*a*a
ptsN_oligo	t*a*tcaaatcattacggataccgaaggtactcccggatgaagcgtaAttattcggtaatgtctcttttagacgttgtgagagaaacagt*a*c
relA_oligo	a*c*agtatatatcaatctacattgtagatacgagcaaatttcggcTtaactcccgtgcaaccgacgcgcgtcgataacatccggcacct*g*g
speG_oligo	c*a*caaaacaccgggtcagactctcctgaagccgaccgcacaataAcattaataataatcgatcgtattttttgatggtgtaaacccgtt*c*g
sucB_oligo	g*a*taatgccttatccggtctacagtgcaggtgaaacttaaactaTtacacgtccagcagcagacgcgtcggatcttccagcaactctt*t*g
sufA_oligo	t*t*tgacatcgtcaagttgcttcagtattacgagacatagtaccgcTtataccccaaagctttcgccacagccacattcattctgggctt*t*a
yaiS_oligo	t*a*tttcacagtcaaacttgctgcattaaaaaaacataaaagctaAgaacgacgcgattatatgcgtcatgatcgcctgcgtgcagttg*c*a
ybaA_oligo	t*g*aattttttgggctgacggcacgccgcacagcccgtcacgctacTtattcgtcgatgattgactcgaatccgccatagatcattcgct*t*g
ybcF_oligo	a*c*gcgtcaagcgtcgcatcaggcacaaatgtctaatgcctacgaTtacagcgaaatacaggtccccgcttcgccgacggctctctt*c*a
ybjK_oligo	t*c*ggttcaaggttgatgggttttttgttatctaaaacttatctaTtaccctgcaaccctctcaaccatcctcaaaatctcctcgcgcg*a*t
ybjR_oligo	g*a*gaaaactgccgcctggctacgcgaccacggaaaaactgccgcgTtaatcctgcccgtatttctccagcaatgcttcggcaatcgcct*g*a
ycbK_oligo	a*a*taatacgatagttcatactgcccctgtttcgttaagcaattaTtaccagtgccgtgctggcccggtatcaatatgcacaaagttac*t*a
ycgX_oligo	c*t*tttatatattgaagatgtagaaaataataaaccattaatctaActaaagttggatacttaagaaatgcttcataattcagtaaggc*a*t
yejE_oligo	g*g*tgaagccgtccgcgacgcatttgatcctaataaggcggtgtaAcatgacgcaaactctgttagcgattgaaaatttgtcggtgggt*t*t
yffB_oligo	g*g*tttcagtgattccagttatcagcaattttttccatgaggtgtaAtctatgtcgtgcccggttattgagctgacacaacagctttattc*g*c
yfiA_oligo	t*c*ggtgaaagacgccaacttcgtcgaagaagttgaagaagagtaAtcctttatattgagtgtatcgccaacgcgccttcgggcgcgtt*t*t
ygaU_oligo	t*g*acgccaattcccattatccagcaggcgatggctggcaattaaTtactcttccggaatacgcaacacttgcccggataaattttat*c*c
ygbA_oligo	t*g*atttaactgcaaattgccggacagatctgcctgtccggcataTtattcatgaggtttttcggacgatattttttccggcagttctg*g*c
ygiZ_oligo	t*t*ctctgtctatgagagccgttaaaacgactctcatagattttaTtaatagcaaaatataaaccgtccccaaaaaagccaccaaccac*a*a
yhcO_oligo	c*t*tgtccgacgatatgccccggttcgggggcttttttttgcgTtaatgacgaacattaaaacagaaacattaaaacgcaaatgccctctttcct*c*t
yibA_oligo	a*a*tgaaattataacttccgctattgataagctgaagcgttcataAcgtgagttgcctatgcacagtgggggattcccgccggcacggt*g*c
yjjV_oligo	a*t*tgtgcgccaactgccggatgcggcgtgaacgcctatccggcTtacggcacgttaaacaacgtatacgtgttattaagcaacgctt*g*c
ynfA_oligo	g*c*tgacacttcctcttatcttattgataaaatggatttatgttcTtacgtgcgccccccagcccgcaacaatgatcaacatgccgcaaa*g*c
ypjC_oligo	a*a*accaacatgaaaatcacttttttcttatattatcgaaaaggtTtattcatttctttttagcgcattcaaaaaaactgatcggcattat*t*t
yqiB_oligo	a*g*ggttaacaggctttccaaatggtgtccttaggtttcacgacgTtaatataaccggaatcgccatcgctccatgtgctaaacagtatc*g*c

Suppl Fig 2c&d   **introduce bp mismatches**

lacZ_oligo_m1_v1(no *)	GGAAACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC
lacZ_oligo_m1_v1(3'*)	GGAAACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAAT*C
lacZ_oligo_m1_v1(5'*)	G*GAAACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC
lacZ_oligo_m1_v1(*)	G*GAAACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAAT*C
lacZ_oligo_m1_v1(3'**)	GGAAACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAA*T*C
lacZ_oligo_m1_v1(5'**)	G*G*AAACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC
lacZ_oligo_m1_v1(**)	G*G*AAACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAA*T*C
lacZ_oligo_m1_v1(5'***)	G*G*A*AACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC
lacZ_oligo_m1_v1(5'****)	G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC
lacZ_oligo_m1_v1(5'**[8])	G*G*A*A*A*C*A*G*CTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC

Suppl Fig 2e   **introduce bp mismatches**

lacZ_oligo_m1_v1(no *)	GGAAACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC
lacZ_oligo_m1_v1(5'****)	G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTT<u>TGA</u>CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC