ACS Editors' Choice

ACS SyntheticBiology

Letter

pubs.acs.org/synthbio

Terms of Use

Direct Mutagenesis of Thousands of Genomic Targets Using Microarray-Derived Oligonucleotides

Mads T. Bonde,[#,†,‡] Sriram Kosuri,[#,¶] Hans J. Genee,[†,‡] Kira Sarup-Lytzen,[†,‡] George M. Church,[*,§,∥] Morten O.A. Sommer,[*,†,‡] and Harris H. Wang[*,⊥]

[†]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2970 Hørsholm, Denmark

[‡]Department of Systems Biology, Technical University of Denmark, DK-2800 Lyngby, Denmark

[¶]Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90025, United States

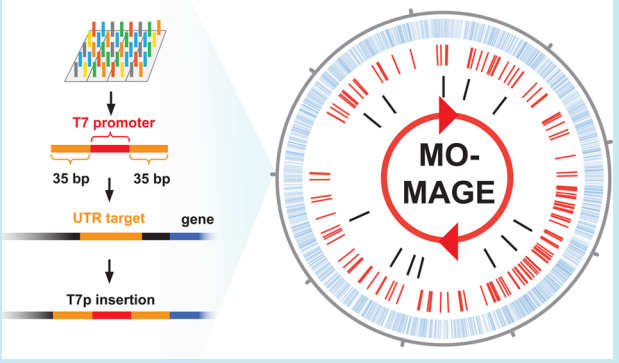
[§]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts 02115, United States

[∥]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, United States

[⊥]Department of Systems Biology, Columbia University, New York, New York 10032, United States

Supporting Information

ABSTRACT: Multiplex Automated Genome Engineering (MAGE) allows simultaneous mutagenesis of multiple target sites in bacterial genomes using short oligonucleotides. However, large-scale mutagenesis requires hundreds to thousands of unique oligos, which are costly to synthesize and impossible to scale-up by traditional phosphoramidite column-based approaches. Here, we describe a novel method to amplify oligos from microarray chips for direct use in MAGE to perturb thousands of genomic sites simultaneously. We demonstrated the feasibility of large-scale mutagenesis by inserting T7 promoters upstream of 2585 operons in *E. coli* using this method, which we call Microarray-Oligonucleotide (MO)-MAGE. The resulting mutant library was characterized by high-throughput sequencing to show that all attempted insertions were estimated to have occurred at an average frequency of 0.02% per locus with 0.4 average insertions per cell. MO-MAGE enables cost-effective large-scale targeted genome engineering that should be useful for a variety of applications in synthetic biology and metabolic engineering.



KEYWORDS: *genome engineering, MAGE, metabolic engineering, microarray, library synthesis*

A core aim of metabolic engineering and synthetic biology is to redesign and create biological systems with useful purposes, for example cell factories that produce novel medicines[1] and chemicals.[2] To achieve these goals, the need of a large set of efficient tools have spurred extensive research efforts dedicated to expanding the synthetic biology toolbox.[3,4] The alteration of gene expression and rewiring of metabolic networks are important for basic research and metabolic engineering, and methods such as transposon sequencing (Tn-seq)[5] and trackable multiplex recombineering (TRMR)[6] allow genomic perturbations into a large number of sites. However, these approaches do not allow combinations of several mutations in individual cells and are confined mostly to generating gene knockouts that require integration of a sizable selectable marker, for which there are limited options. The recent development of Multiplex Automated Genome Engineering (MAGE)[7] enables rapid and efficient targeted modification of the genome through iterative cycles of λ-Red mediated recombination using multiple oligonucleotides at once. MAGE with multiple degenerate oligos can be applied toward genome mutagenesis through a semi-rational approach, where specific targets are randomly altered to limit mutagenesis to only selected targets. Such application of MAGE has been shown to be useful for metabolic optimization to increase the biosynthetic production of lycopene and indigo in *E. coli*.[8,9] In theory, MAGE should be amenable to targeted mutagenesis using thousands of oligos all at once to target hundreds to thousands of chromosomal targets. This capability will open new possibilities for many large-scale genome engineering projects.[10]

Mutagenesis of thousands of genomic targets by MAGE requires large oligo library pools. However, synthesis of thousands of MAGE oligos by traditional column-based phosphoramidite chemistry is impractical in both time and cost. Recent developments in high-fidelity oligonucleotide microarray technologies have enabled the construction of large libraries (>55 000 oligos) of 200 bp oligos[11] at a significantly lower cost and turn-around time compared to oligos produced by column-based synthesis.[12−14] Here, we describe Microarray Oligonucleotide

Special Issue: Summer Synthetic Biology Conference 2014

Received: March 7, 2014

   dx.doi.org/10.1021/sb5001565 | *ACS Synth. Biol.* XXXX, XXX, XXX−XXX

(MO)-MAGE, a novel method to generate thousands of oligos suitable for large-scale genome engineering by MAGE and demonstrate its application for fast and robust targeted mutagenesis of the *E. coli* genome.

RESULTS AND DISCUSSION

Using a computational framework for MAGE oligo design (MODEST),[15] we first identified perturbation targets of the *Escherichia coli* genome that included most regulatory and protein coding regions (See Table 1 for details about which genes were targeted). Protein coding perturbations were made through the generation of a nonsense and frameshift mutation within the first 5% of the open-reading frame to functionally introduce a reversible gene knockout. Regulatory perturbations included up-regulation ("RBS up") using consensus (AGGAGG), down-regulation ("RBS down") using anticonsensus (TCCTCC) ribosomal binding sites, or insertion of a T7 promoter sequence (TAATACGACTCACTATAGGG) upstream of operons. A summary of the genomic perturbation design is given in Table 1. In all, the designed oligo library constituted 13 000 possible targeted perturbations against the *E. coli* genome (see Supporting Information).

Table 1. Overview of Oligos Designed and Synthesized on the Micro-Array Chip[a]

	targets	general	TFs	genes targeted
CDS knock outs	3798	3633	167	all nonessential, non-pseudo-, non-ncRNA
T7 promoters	2723	2585	138	all non-pseudo-, non-ncRNA,[b]
RBS up	3323	3172	151	all non-pseudo-, non-ncRNA
RBS down	3099	2948	151	all nonessential, non-pseudo-, non-ncRNA
total	12 943	12 338	607	

[a]Four oligo subpools were made, intended for knocking out genes by introduction of a nonsense and frameshift mutation within the first 5% of the CDS ("CDS Knock Outs"), upregulation by insertion of T7 promoters upstream of genes ("T7 promoters") and insertion of a consensus RBS sequence ("RBS up"), and down-regulation by insertion of an anti-consensus RBS sequence ("RBS down"). TFs = Transcription Factors. [b]Only genes with sufficient spacing to the next gene upstream were targeted, to ensure that the insertion of T7 promoters did not disturb upstream genes.

Traditional column-based synthesis of 90mer MAGE oligos of this library size would take months to years to generate at a cost of aproximately $500,000 USD. Thus, we turned to new approaches in long oligonucleotide synthesis using DNA microarrays.[11] A 130mer single-stranded DNA library was generated using the Oligo Library Synthesis (OLS) platform from Agilent Technologies (Santa Clara, CA, U.S.A.). Because the amount of oligos needed for MAGE is much higher than the total yield from microarray synthesis (μM versus pM), we devised a PCR-based amplification protocol to generate a renewable supply of MAGE-compatible oligo pool from our initial microarray-derived oligo library, which we describe in greater detail later.

Each oligo library subpool (e.g., CDS Knockout, T7 insertion, RBS up, RBS down) was designed with a unique set of barcodes, which allowed for selective amplification of only the subpool library by using specific primers (see Figure 1). A total volume of 38.4 mL of PCR was first performed to ensure that we generated enough oligos for >10 MAGE cycles. One of the two primers used for the PCR contained a 5′-phosphothioester bond and a 3′-uracil, while the other contained a 5′ phosphorylated group. The resulting amplicons contain a 5′-phosphothioester bond on

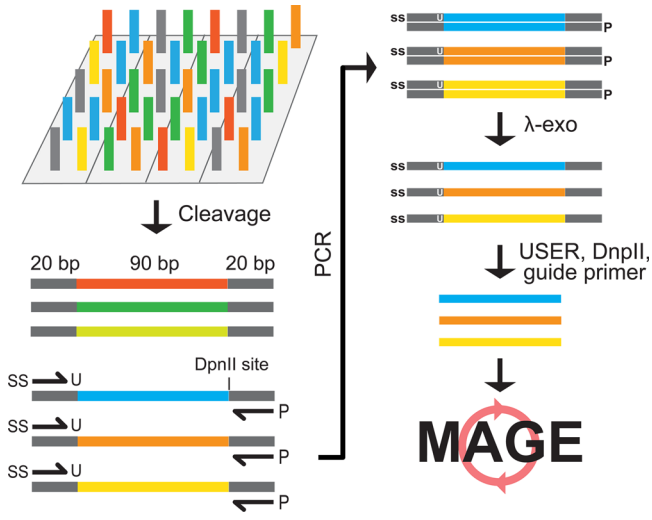


Figure 1. MO-MAGE method for targeted whole genome mutagenesis. 130 base oligonucleotides were designed and synthesized on a DNA microarray, which can be ordered from several commercial vendors. The oligos can be designed with different barcodes, which allow selective PCR amplification of a desired subpool. One of the primers are 5′ phosphorylated, which allow the degradation of only one of the strands by λ-exonuclease, resulting in single stranded oligos. The barcodes are removed by enzymatic treatment with endonuclease VIII (cutting the barcode by removing a uracil from the modified primer), DpnII and a guide primer (to make a double stranded cut site for DpnII). The final 90 bp single stranded oligos are directly applicable for MAGE.

the MAGE-compatible sense-strand and a 5′-phosphate on the reverse complement strand. This 130 bp double-stranded (ds) DNA amplicon was digested with λ-exonuclease. Since λ-exonuclease has much higher activity for unphosphorylated substrates[16] and its activity is blocked by phosphothioester bonds, the digest results in a 130mer single-stranded (ss) DNA library of the MAGE-compatible sense-strand that contained the phosphothioester bond at the 5′ end. Subsequently, the 20 bp flanking barcodes used for PCR amplification of the subpools were removed from each end of the ssDNA to yield unique 90mer oligos needed for high efficiency MAGE. This was accomplished by first digesting the 130mer ssDNA library, which contained an internal uracil at the barcode junction from the previous PCR (see Figure 1), with an uracil DNA glycosylase, endonuclease VIII,[17] which removes the uracil from the ssDNA strand. The uracil excision effectively removes the 20mer barcode at the 5′ end of the ssDNA library, yielding a 110mer library. The 3′ barcode was designed with a DpnII restriction site placed immediately after the target oligo sequence. To remove the 3′ barcode, a guide primer complementary to the 3′ barcode including the DpnII site was used to hybridize to the 110mer oligo and was digested with DpnII to yield the designed 90mer oligo library. The use of a guide primer ensured that the remaining part of the single-stranded oligo is not digested by DpnII, which only cuts dsDNA. The resulting oligo library contained unique 90 base single-stranded oligos.

Gel electrophoresis of the oligos was performed after each step to ensure that the processing resulted in the expected product (see Figure 2a and b). Serial dilutions of the library were further visualized on a TBE-UREA gel to estimate library concentration using ImageJ.[18] We estimate that a typical amplification generates 1.4 nmols of oligos (~12 μM in 115 μL), which is sufficient for 14 MAGE cycles at 2 μM per 50 μL reaction per cycle.

dx.doi.org/10.1021/sb5001565 | ACS Synth. Biol. XXXX, XXX, XXX−XXX

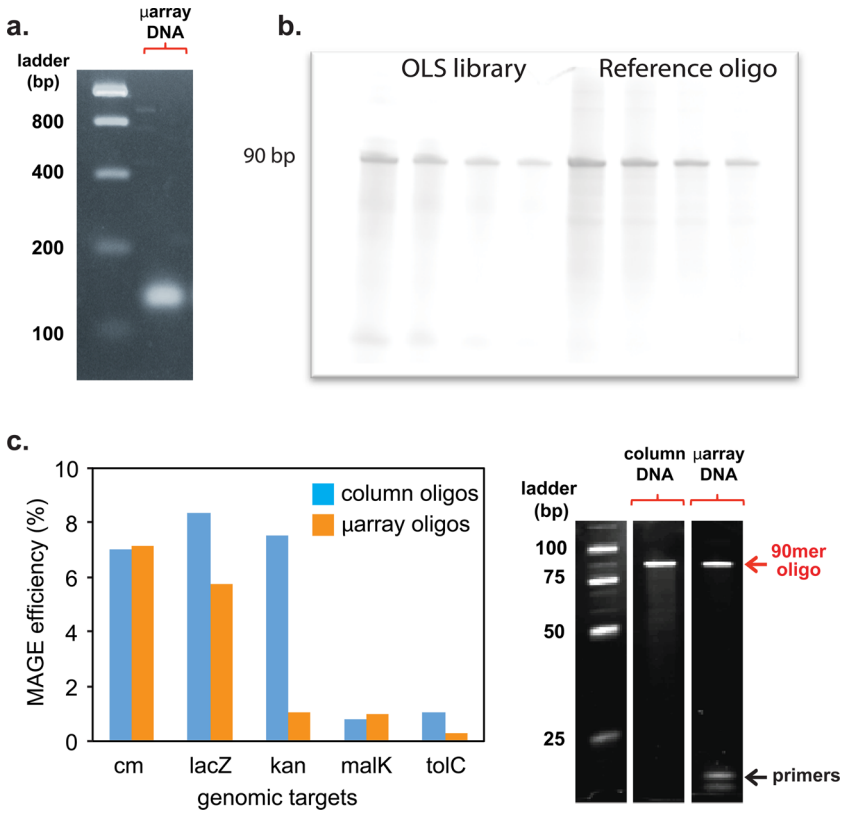


Figure 2. (a) PCR amplicons of the T7 promoter oligo pool can be seen as the strong bands around 130 bp (4% Agarose E-Gel EX with Low Range Quantitative DNA Ladder). (b) Serial dilutions of the processed single stranded T7 promoter oligo library (left 4 lanes) compared to a reference oligo of 90 bp (right 4 lanes), which indicates correct processing of the oligos from 130 bp to 90 bp oligos ready for MAGE (TBE-Urea gel 4% from Invitrogen). (c) Comparison of MAGE efficiency using column-synthesized oligos and microarray-processed oligos by MO-MAGE. Gel shows size distribution of the two processed oligo pools (TBE-Urea gel 4% from Invitrogen).

We performed a side-by-side comparison of the microarray-derived oligos with those obtained through standard column-synthesis from a commercial vendor (Integrated DNA Technologies, Iowa, U.S.A.). A test group of 5 oligos generated from the microarray pool and the commercial vendor was used for MAGE and the efficiency of oligo incorporation was determined (see Figure 2c). We find that 3 out of the 5 tested oligos showed slightly lower incorporation efficiencies in the microarray library compared to the column library. We attribute the decreased efficiency to differences in individual oligo concentrations that may result during amplification from the microarray library. Nonetheless, these results offered a convincing proof-of-concept that microarray-derived oligos are compatible with MAGE mutagenesis. Since the introduction of synthetic regulation to native genomic loci has been an outstanding challenge in synthetic biology, we sought to further explore our sublibrary that generated T7 promoter insertions in the untranslated regions (UTR) upstream of each of 2585 operons in the *E. coli* genome. This multiplexed promoter insertion perturbation enables the generation of a mutant library that contains new transcriptional regulation in the presence of an inducible T7 polymerase system.[9,19] Thus, in this current work, we focused on characterizing our methods in greater detail using this T7 insertion library.

The T7 promoter insertion library was designed using MODEST[15] to target the UTR region 40 bp upstream of all nonessential and nontranscription-factor operons in *E. coli*. Downstream polycistronic open reading frames without adequate intergenic space were not targeted. In all, 2585 unique insertion target sites were identified. We designed 90 bp oligos by flanking a 20 bp T7 promoter (TAATACGACTCACTATAGGG) sequence with 35 bp of homologous sequence at each end to the target genomic integration site. In general, introduction of a 20 bp insertion is expected to have lower incorporation efficiency than smaller point mutations or deletions.[7] Nonetheless, these larger insertions represent a unique opportunity to introduce synthetic regulation and to challenge the limits of our MAGE capabilities using large oligo library pools derived from microarrays. Furthermore, effective selection of desirable mutants from a smaller genomic library with T7 promoter insertions have previously been performed successfully.[9]

Following the generation of the 2585-oligo library, we applied this T7 promoter pool to mutagenize the *E. coli* genome by MAGE for 12 cycles. The entire oligo processing pipeline and MAGE mutagenesis was performed twice in parallel to generate two separate cell libraries to test the reproducibility of our protocol. To assess the success of T7 insertion in these combinatorial libraries, we performed deep sequencing of the final cell populations. We first extracted genomic DNA of the cell populations and hybridized the genomic DNA with a biotinylated oligo containing the T7 promoter sequence. Genomic regions that contained the T7 promoter insertions can thus be enriched from total genomic DNA when it is applied to streptavidin beads that bind to the biotinylated oligo-genomic hybrid. We then sequenced the enriched genomic library by deep-sequenced to identify T7 insertion sites.

Next-generation sequencing analysis of the data showed that the two separate libraries contained 87 reads and 208 reads with

dx.doi.org/10.1021/sb5001565 | *ACS Synth. Biol.* XXXX, XXX, XXX−XXX

the T7 promoter sequence, with the reads mapping to 56 and 98 targeted operons respectively in the two libraries. Only 4 of the total 154 targets were redundant, resulting in identification of 150 unique targets out of 2585 possible. The sequencing coverage (the number of times the genome has been sequenced) of the libraries was 2064× and 1364× respectively. We used a Monte Carlo model to simulate the insertion frequency of the total library, and found that the total cell library is expected to contain between 2250 and 3500 modified genes (95% confidence interval, see Supporting Information Figure 1). These results show that all or most of the targeted sites are predicted to have T7 promoter insertions within the cell library.

We proceeded to further validate our mutagenesis results and to estimate the insertion efficiency. We randomly selected 12 targets, including 8 that had not been detected by population sequencing. We amplified ~200 bp PCR fragments spanning the T7 insertion sites of each of the 12 genes from the cell library and sequenced the PCR products by Next Generation Sequencing. For all 12 loci, we found reads containing the T7 promoter insertion sequence. This provides further support that the cell library contains a majority of the 2585 T7 promoter insertions (see Table 2 and Figure 3).

The insertion frequencies for the 12 loci were estimated by comparing the number of reads with T7 promoter insertions to the number of reads without the insertion. The average frequency was 0.017%, and thus, the average number of insertions per cell can be estimated as $\mu = k \times p = 0.434$ where $p = 0.00017$ and $k = 2585$ targets. This means that 43% of the cell library is expected to have at least one insertion on average. The top 1% of the population is expected to have at least four (4.3) insertions based on the calculations $m = \mu + 2.326(k \times p(1 - p))^{1/2} = 4.3$ (for details about the MAGE efficiency calculations see the Supporting Information as well as Wang and Church, 2011[7]). The MAGE efficiency for insertions has previously been predicted based on fitting of empirically determined efficiencies.[7,8] Using such approach, the insertion efficiency (IE) is $IE = 0.15 \times e^{-0.075 \times (b-1)} = 0.0361$, where b is the size of the insertion in number of bases ($b = 20$). Based on this insertion efficiency, the predicted average frequency of each insertion can be calculated by $p_{N-\text{predicted}} = 1 - (1 - IE/k)^N = 1 - (1 - 0.0361/2585)^{12} = 0.000167$, where N is the number of MAGE cycles. The predicted average insertion frequency of 0.0167% matches very well with the measured frequency of 0.017%, providing confidence to our predicted population mutagenesis profile and its applications to analyze MAGE mutagenesis of complex oligo pools. These results further highlight that that cell libraries with combinations of multiple insertions per genome can be generated using this massively multiplexed approach.

Table 2. Twelve Genes Randomly Selected for Deep Amplicon Sequencing and Analysis[a]

	reads with T7 insertion	reads total	insertion frequency
acrD	549	452814	0.1212%
edd	59	157788	0.0374%
osmC	13	85100	0.0153%
fryB	26	299019	0.0087%
SodB	25	450003	0.0056%
pssA	6	151518	0.0040%
secE	15	561640	0.0027%
thrL	4	197901	0.0020%
GlnD	4	263023	0.0015%
acrA	7	481259	0.0015%
mdaB	2	237727	0.0008%
hemC	1	129353	0.0008%

[a]The number of reads with a T7 promoter insertion was compared to reads without an insertion to calculate the insertion frequency.

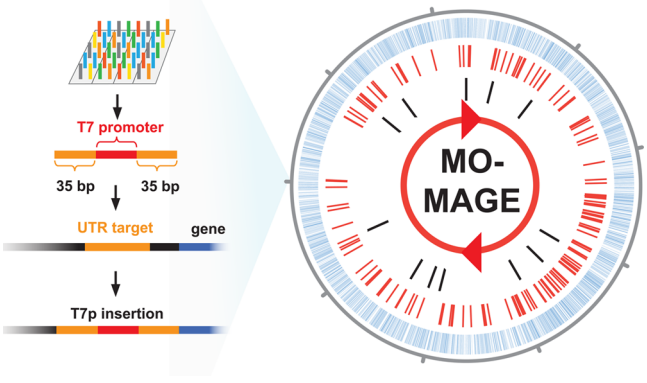


Figure 3. MO-MAGE of 2585 genomic targets corresponding to untranslated regions (UTR) upstream of genes for insertion of 20-bp T7 synthetic promoter. Designed targets are shown in blue. Mutated targets verified by whole-genome sequencing are shown in red (see Supporting Information Table 3 for complete list). Mutated targets verified by amplicon sequencing are shown in black.

Here, we have presented a proof-of-concept demonstration for the generation and application of oligo pools that specifically target thousands of unique chromosomal loci across a cell population to introduce promoters amenable for synthetic regulation. The Microarray Oligonucleotide MAGE enables rapid mutagenesis of bacterial genomes using oligos derived directly from microarrays without intermediate cloning or cassette selection steps, which can significantly expand the combinatorial genomic diversity of the resulting cell population. A resulting cell library will be useful for various screens for metabolic engineering purposes such as increased production of biochemicals or tolerance toward biomass inhibitors. The MO-MAGE process described here allow researchers to limit the combinatorial space to only specific mutations that are expected to have a much higher chance of leading to a desired phenotype than random mutations. Thus, the effect of rationally designed targets can be assessed very effectively because the mutagenesis quality (i.e., the proportion of interesting mutants to other cells) is much higher. Whereas current approaches are limited to only creating few genomic changes at a time, our method could be used to target all predicted beneficial mutations predicted from metabolic models, and screen for optimal phenotypes.

As the complexity of engineering tasks of synthetic biology and metabolic engineering increases, the need to reduce cost of genome engineering becomes more important. MO-MAGE allows the synthesis of oligos at a fraction of the cost (>1000×) compared to traditional column based oligo synthesis to make large-scale genome engineering accessible to most laboratories. At a price of 36 USD per column-based oligo, MO-MAGE (2800 USD) is currently cost competitive when using more than 78 oligos. This method could provide a paradigm shift by making large-scale genome engineering of many thousands of targets available as a standard tool for strain optimization and other projects where large-scale targeted mutagenesis searches are needed. New advances in DNA and gene synthesis will further foster growth in genome engineering of microbial and eukaryotic systems,[20] and in theory, oligo pools containing millions of oligos can be applied for MO-MAGE. However, since increasing oligo

dx.doi.org/10.1021/sb5001565 | ACS Synth. Biol. XXXX, XXX, XXX−XXX

diversity leads to lower replacement efficiency per genomic target, there is a practical limitation in the number of oligos that can be meaningfully applied depending on the required replacement efficiency per site for a given experiment.

For some projects, a higher integration frequency might be of interest, to allow more combinations of insertions and a higher quality library. The amount of MAGE cycles performed can be increased to higher levels, and automated solutions could increase the feasible number of MAGE cycles to several hundreds. For creation of complex libraries where many combinations per cell are desired, single base pair substitutions and small insertions and deletions can be applied to increase the replacement efficiency. For instance if 1000 oligos designed for making single base pair substitutions are used for 100 cycles of MAGE, a cell library with 25 average replacements per cell is predicted. This allows creation of unprecedented targeted combinatorial libraries of specific chromosomal modifications. Ultimately, we envision MO-MAGE method will be used to make thousands of specific chromosomal changes predicted to result in a desired phenotype and combined with selection or screening of the cell library for interesting mutants.

METHODS

Strains and Culture Conditions. We used a strain based on the EcNR2[8] strain for all experiments, which is based on *E. coli* K12 MG1655. The genotype is λ-Red$^+$bla$^+$bioA$^-$/bioB$^-$ mutS$^-$zeo$^+$ Lac_T7pol$^+$Spec$^+$ LacIQ$^+$. All strains for MAGE were grown in low salt LB-min medium (10 g tryptone, 5 g yeast extract, 5 g NaCl in 1 L dH_2O) for optimal electroporation efficiency with addition of specified antibiotics. All cells for liquid cultures were grown in standard LB-min medium (10 g tryptone, 5 g yeast extract, 10 g NaCl in 1 L dH_2O) with addition of specified antibiotics.

Standard Oligonucleotides. All standard oligonucleotides were purchased from Integrated DNA Technologies with standard purification.

Oligo Library Synthesis. OLS pools were synthesized by Agilent Technologies and are available upon signing a Collaborative Technology Development agreement with Agilent. Costs of OLS pools are a function of the number of unique oligos synthesized and of the length of the oligos The OLS pool were synthesized, cleaved, and delivered as lyophilized ~1−10 picomole pools.

Oligo Library Processing. Please refer to the Supporting Information.

Oligos for T7 Cell Library Generation. A list of all the 2585 oligos synthesized on the microarray chip designed to insert a T7 promoter upstream of *E. coli* genes can be found in the Supporting Information.

MAGE. MAGE was performed according to the protocol provided in ref 7. Briefly, the cells were grown to midlog phase, whereafter the β-protein of the λ-red system was induced by growing at 42 °C for 15 min whereafter the cells were chilled to 4 °C. The culture were washed to remove salts and resuspended in cold water (<4 °C). The cells were mixed with the 2 μM oligos in 50 μL and electroporated in a Bio-Rad MicroPulser, BTX ECM-830 with 1 mm gap cuvette, whereafter the cells were incubated for 2−3 h at 30 °C. The process were repeated 12 times (12 MAGE cycles) to allow a higher frequency of insertion. After 5 and 10 MAGE cycles, the cells were grown overnight in 50 mL LB low salt medium and stored at −80 °C in a 15% v/v glycerol solution.

Enrichment for T7 Promoter Containing Sequences. The enrichment protocol was based on Gnirke et al.[21] and NimbleGen SeqCap EZ Exome Library SR v2.2 protocol.[22] A biotinylated oligo targeting the T7 promoter was incubated 66 h with the prepared sequencing library fragments. The T7 promoter containing fragments that hybridized to the biotinylated oligo was enriched by multiple rounds of binding and washing with Invitrogen Streptavidin M-270 Dynabeads and Invitrogen binding and wash buffer.

Illumina DNA Sequencing and Analysis. Samples for sequencing of the 12 individual strains were processed with Illumina TruSeq v2 sample prep kit and standard Illumina adapters. Samples for sequencing of the cell libraries were prepared with the NEBNext DNA Sample Prep Master Mix Set 1 kit for Illumina sequencing and manually ordered adaptors kindly provided by Luhan Yang (Harvard Medical School, George Church Lab).

adaptor 1	PE-A1-F	TACACTCTTTCCCTCACGACGCTCTTCGATCTac*T
	PE-A1-R	/5Phos/gtAGATCGGAGAGCGGTTCAGCGGAATGCCGAG
adaptor 2	PE-A2-F	TACACTCTTTCCCTCACGACGCTCTTCGATCTtg*T
	PE-A2-R	/5Phos/caAGATCGGAGAGCGGTTCAGCGGAATGCCGAG

The libraries were sequenced in two separate lanes, whereas the 12 isolated strains were sequenced in one lane with multiplexing barcodes. The 12 prepared individual strains were sent to the Harvard Biopolymers facility (genome.med.harvard.edu) for Illumina sequencing and downloaded to the cbs.dtu.dk server and processed here. Sequences containing the T7 promoter sequence were extracted with the "grep" command. *BLASTn* was performed in CLC Bio main Workbench 6.0 with standard settings: "word size" 11, "match" 1, "mismatch" −3, "gap cost existence" 5, "gap cost extension" 2. *Bowtie*[23] was applied to perform the alignment of the reads to the reference genomes (see parameters in the script in Supporting Information). *Samtools*[24] was applied to make a consensus reference genome and indexed BAM-file (for visual inspection of read alignment) from the bowtie output (see parameters in the script in Supporting Information)

PCR and freq-seq of 12 Loci. PCR primers were designed for amplification of 12 genomic regions of around 200 bp with the T7 promoter insertion site in the middle. The amplicons were sequenced on an Illumina MiSeq, and a script created to extract all WT and mutant sequences and report the numbers.

ASSOCIATED CONTENT

Supporting Information

This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*Email: msom@bio.dtu.dk.
*Email: hw2429@columbia.edu.
*Email: gchurch@genetics.med.harvard.edu.

Author Contributions

#M.T.B. and S.K. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

H.H.W. acknowledges funding from the National Institutes of Health Director's Early Independence Award (1DP5OD009172-01).

G.M.C acknowledges funding from the Department of Energy (DOE) (as above) and the National Science Foundation (NSF) (SA5283-11210). M.T.B., H.J.G., K.S.L. and M.O.A.S. acknowledge funding from the Novo Nordisk Foundation and the European Union Seventh Framework Programme (FP7-KBBE-2013-7-single-stage) under grant agreement no. 613745, Promys.

REFERENCES

(1) Weber, W., and Fussenegger, M. (2012) Emerging biomedical applications of synthetic biology. *Nat. Rev. Genet. 13*, 21−35.
(2) Serrano, L. (2007) Synthetic biology: Promises and challenges. *Mol. Syst. Biol. 3*, 158.
(3) Tyo, K. E., Alper, H. S., and Stephanopoulos, G. N. (2007) Expanding the metabolic engineering toolbox: More options to engineer cells. *Trends Biotechnol. 25*, 132−7.
(4) Khalil, A. S., and Collins, J. J. (2010) Synthetic biology: Applications come of age. *Nat. Rev. Genet. 11*, 367−79.
(5) Van Opijnen, T., and Camilli, A. (2013) Transposon insertion sequencing: A new tool for systems-level analysis of microorganisms. *Nat. Rev. Microbiol. 11*, 435−42.
(6) Warner, J. R., Reeder, P. J., Karimpour-Fard, A., Woodruff, L. B. a, and Gill, R. T. (2010) Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat. Biotechnol. 28*, 856−62.
(7) Wang, H. H., and Church, G. M. (2011) Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering. *Methods Enzymol. 498*, 409−26.
(8) Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., and Church, G. M. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature 460*, 894−8.
(9) Wang, H. H., Kim, H., Cong, L., Jeong, J., Bang, D., and Church, G. M. (2012) Genome-scale promoter engineering by coselection MAGE. *Nat. Methods 9*, 591−3.
(10) Esvelt, K. M., and Wang, H. H. (2013) Genome-scale engineering for systems and synthetic biology. *Mol. Syst. Biol. 9*, 641.
(11) Kosuri, S., Eroshenko, N., Leproust, E. M., Super, M., Way, J., Li, J. B., and Church, G. M. (2010) Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat. Biotechnol. 28*, 1295−9.
(12) Tian, J., Gong, H., Sheng, N., Zhou, X., Gulari, E., Gao, X., and Church, G. (2004) Accurate multiplex gene synthesis from programmable DNA microchips. *Nature 432*, 1050−1054.
(13) Tian, J., Ma, K., and Saaem, I. (2009) Advancing high-throughput gene synthesis technology. *Mol. Biosyst. 5*, 714−22.
(14) Richmond, K. E., Li, M.-H., Rodesch, M. J., Patel, M., Lowe, A. M., Kim, C., Chu, L. L., Venkataramaian, N., Flickinger, S. F., Kaysen, J., Belshaw, P. J., Sussman, M. R., and Cerrina, F. (2004) Amplification and assembly of chip-eluted DNA (AACED): A method for high-throughput gene synthesis. *Nucleic Acids Res. 32*, 5011−8.
(15) Bonde, M. T., Klausen, M. S., Anderson, M. V., Wallin, A. I. N., and Sommer, M. O. A. (2014) MODEST : A web-based design tool for oligonucleotide-mediated genome engineering and recombineering. *Nucleic Acids Res.*, DOI: 10.1093/nar/gku428.
(16) Mitsis, P. G., and Kwagh, J. G. (1999) Characterization of the interaction of lambda exonuclease with the ends of DNA. *Nucleic Acids Res. 27*, 3057−63.
(17) Bitinaite, J., and Nichols, N. M. (2009) DNA cloning and engineering by uracil excision. *Curr. Protoc. Mol. Biol. Chapter 3*, DOI: 10.1002/0471142727.mb0321s86.
(18) Abramoff, M. D., Magalhaes, P. J., and Ram, S. J. (2004) Image processing with ImageJ. *Biophotonics Int. 11*, 36−42.
(19) Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F., and Voigt, C. A. (2012) Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res. 40*, 8773−81.
(20) LeProust, E. (2008) Agilent's microarray platform: How high-fidelity DNA synthesis maximizes the dynamic range of gene expression measurements. Agilent Technologies, Santa Clara, CA.
(21) Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S., and Nusbaum, C. (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol. 27*, 182−9.
(22) Nimblegen (2013) *NimbleGen SeqCap EZ Library SR User's Guide*; Roche Nimblegen, Madison, WI.
(23) Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol. 10*, R25.
(24) Flicek, P., and Birney, E. (2009) Sense from sequence reads: Methods for alignment and assembly. *Nat. Methods 6*, S6−S12, DOI: 10.1038/nmeth.1376.

SUPPORTING INFORMATION

Direct Mutagenesis of Thousands of Genomic Targets using Microarray-derived Oligonucleotides

Mads T. Bonde*[1,2], Sriram Kosuri*[3], Hans J. Genee[1,2], Kira Sarup-Lytzen[1,2], George M. Church[4,5], Morten O.A. Sommer[1,2], Harris H. Wang[4]

[1] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2970 Hørsholm, Denmark

[2] Department of Systems Biology, Technical University of Denmark, DK-2800 Lyngby, Denmark

[3] Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA.

[4] Department of Systems Biology, Columbia University, New York, NY 10027, USA

* These authors have contributed equally to this work
Correspondence should be addressed to M.O.A.S (msom@bio.dtu.dk), H.H. W. (hw2429@columbia.edu), or G.M.C. (gchurch@genetics.med.harvard.edu)

OLIGO PROCESSING

We obtained the oligos from an Agilent Technologies inkjet printed DNA microarray chip containing 13,000 oligos including the 2587 uniquely barcoded T7 oligos used in this experiment. A detailed protocol is included at the end of the Supporting Information.

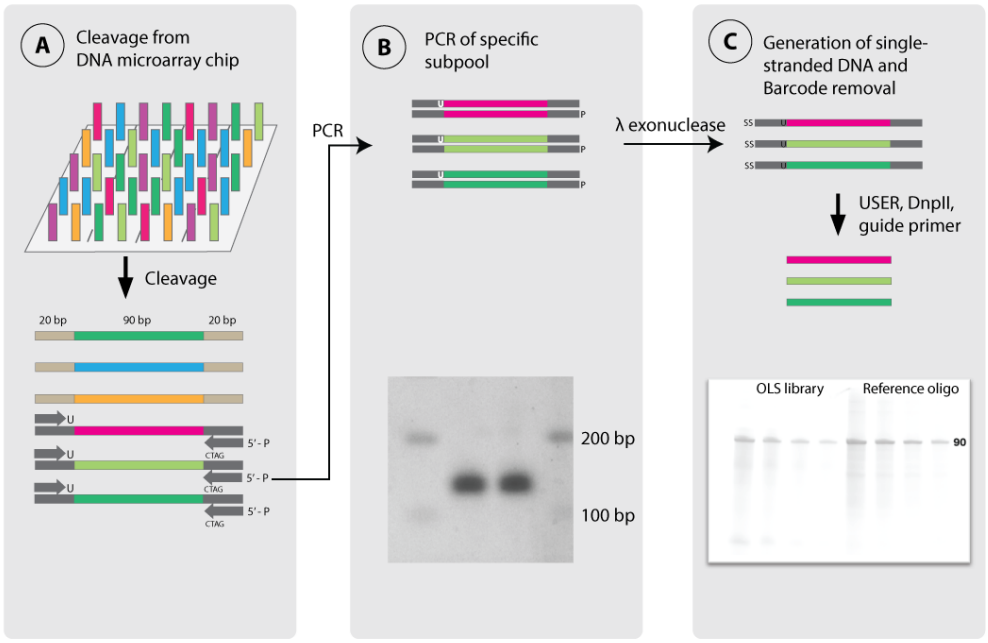
We designed each subpool with a unique barcode, which allows a selective amplification of only that library by using specific primers (see Supplementary Figure 1). We ordered the oligos from Agilent Technologies laboratories, and PCR amplifiied with a total volume of 38.4 mL, to ensure enough oligos for at least 10 MAGE cycles. We processed the double stranded DNA with Lambda exonuclease because recombineering efficiency is significantly higher when single stranded DNA oligos are applied [1]. We used a 5' phosphorylated primer to facilitate the breakdown of only one strand, as λ-exonuclease has much higher activity for unphosphorylated substrates [2]. The other primer containes thioesterbonds in the 5' end, which protects from degradation by Lambda Exonuclease.

Barcodes used for amplification of the pools need to be removed to allow efficient recombineering. We used a primer containing a uracil base for the second PCR reaction to facilitate the removal of the 5' barcodes by uracil DNA glycosylase, endonuclease VIII [3] that remove uracil from a DNA strand. This leads to a break in the single stranded DNA, effectively removing the barcode at the 5' end of the oligos. (see **1**).

The 3' barcode was designed with a DpnII restriction site, placed immediately after the target oligo sequence. A guide primer, complementary to a sequence including the 3' barcode and DpnII site was added, which allowed the removal of the 3' barcode by DpnII treatment. Using a guide primer while leaving the remaining part of the oligo single-stranded ensures that DpnII does not cut the oligo, because DpnII only cuts double stranded DNA. The

resulting oligo pool contains 2587 unique 90 bp single stranded oligos with 35 bp flanking regions complimentary to a region upstream of *E. coli* genes, and the middle part containing the T7 promoter.

Gel electrophoresis of the oligos was performed after each PCR reaction and after the final treatment, to validate the presence of the oligos and ensure that the processing resulted in the correct lengths. The oligos are ~130 bp after both PCR reactions, and 90 bp after treatment and cleavage as expected. Serial dilutions of library and reference (highest conc. to the left) facilitated the calculation of the OLS library concentration. ImageJ[4] was used to determine the library concentration to ~11.7 µM in 115 µL –enough for 13 MAGE cycles at 2 µM in 50 µL pr. cycle.



Supplementary Figure 1 Amplification and processing of MAGE-oligos from a DNA microchip A) Oligos were synthesized on a DNA microarray chip and then cleaved off the chip. We designed the oligos with multiple subpools with different barcodes. B) We performed PCR of a specific subpool containing 2587 oligos. PCR amplicons of the expected size can be seen as the strong bands around 130 bp (4 % Agarose E-Gel EX with Low Range Quantitative DNA Ladder) C) The double stranded oligos were treated with Lambda Exonuclease to create single stranded DNA oligos. Barcodes were removed by treatment with uracil DNA glycosylase, endonuclease VIII, DpnII and a guide primer making the DNA double stranded at the DpnII site. The gel shows serial dilutions of the final oligo library (left 4 lanes) compared to a reference oligo of 90 bp (right 4 lanes), which indicates correct processing of the oligos from 130 bp to 90 bp oligos ready for MAGE (TBE-Urea gel 4 % from Invitrogen).

REPLACEMENT EFFICIENCY CALCULATIONS

We used the data from the frequency sequencing of 12 genes to calculate various characteristics of the library based on formulas from Wang and Church (2011)[5].

Supplementary Table 1

	Reads with T7 insertion	Reads total	Frequency
acrD	549	452814	0.1212%
edd	59	157788	0.0374%
osmC	13	85100	0.0153%
fryB	26	299019	0.0087%
SodB	25	450003	0.0056%
pssA	6	151518	0.0040%
secE	15	561640	0.0027%
thrL	4	197901	0.0020%
GlnD	4	263023	0.0015%
acrA	7	481259	0.0015%
mdaB	2	237727	0.0008%
hemC	1	129353	0.0008%

The average frequency is p = 0.0001678 ≈ 0.017%, and thus the average number of insertions per cell can be estimated as

$$\mu = k \times p = 0.4342$$

where k = 2587 targets, meaning that 43 % of the cell library is estimated to have an insertion. The variance can be calculated as

$$\sigma = \sqrt{k \times p(1-p)} = 1.215$$

The top 1% clones is estimated to have at least m number of mutations:

$$m = \mu + 2.326\sigma = 4.3$$

Allelic replacement efficiency for insertions has previously been predicted by the following function based on fitting of empirically determined efficiencies from Wang et al. (2009a) (Wang and Church, 2011) [1,5]:

$$RE = 0.15 \times e^{-0.075 \times (b-1)} = 0.0361$$

where b is the number of basepairs in the insertion (b=20). The predicted average frequency of each insertion can be calculated by

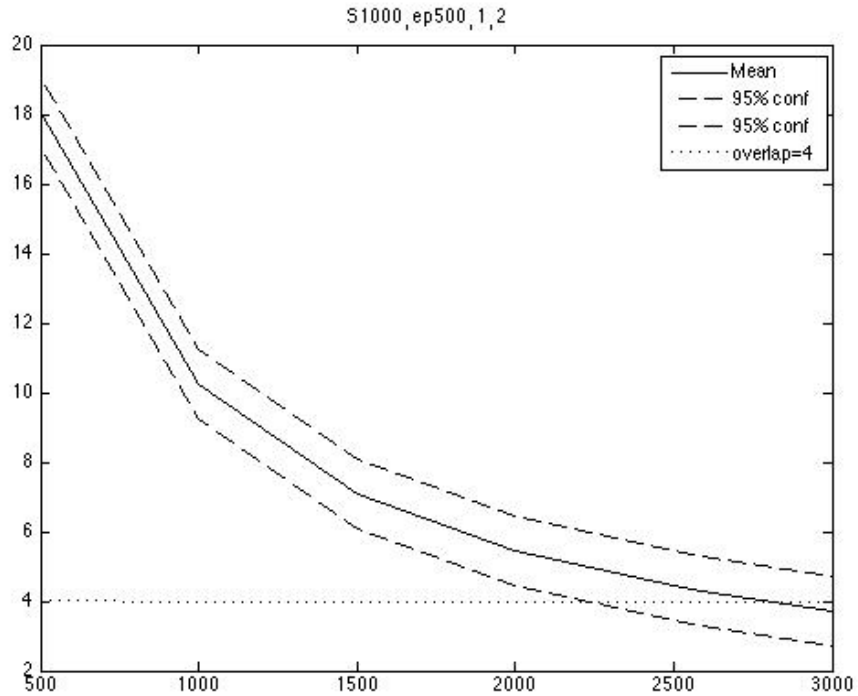
$$p_{N-predicted} = 1 - (1 - RE_{av})^N = 1 - \left(1 - \frac{0.0361}{2587}\right)^{12} = 0.000167$$

where $RE_{av} = RE/k$, where k is the number of target sites (k=2587) and N is the number of MAGE cycles. The expected average number of insertions per cell is predicted to be

$$\mu_{predicted} = k \times p_N = 0.432029$$

LIBRARY CHARACTERIZATION AND MODELLING

Even though a large number of reads were obtained (95 and 63 million respectively), all genes were not expected to be covered, because of the very small size of a 20 bp insertion compared to the full genome making up most of the sequencing output. A simulation was developed to simulate the experiment *in silico* to investigate the probability of having x number of overlapping genes (genes that are identified in the both of the two generated cell libraries) depending on the number of modified sites. A Monte Carlo simulation of the experiment was developed, using the number of reads with a T7 promoter divided by the number of reads total to estimate the probability of drawing a read with a modified promoter for each library. The simulation was run in 7 rounds, assuming a different total number of modified genes (500, 1000, 1500, 2000, 2500, 2587 and 3000). Reads are drawn from both libraries, and with the probability calculated as described above, a gene from the total gene pool (e.g. 500 or 1500) is assigned to the read. The genes drawn from the 2 pools are compared, and the number of unique gene overlaps is calculated. Each round is repeated 1000 times per total number of modified genes, and the mean number of overlapping genes is calculated and plotted. The simulation showed that the probability of getting only 4 overlapping genes is within a 95 % confidence interval, if the total number of genes is between 2250 and 3500 genes.



Supplementary Figure 2. Visualization of a Monte Carlo simulation of the performed experiment. The experiment is simulated with varying number of modified genes in the cell library that was sequenced (shown on the x-axis). 1000 simulations were performed for 500-3000 genes (interval of 500). The mean number of overlapping genes resulting from these simulations is plotted at the y-axis along with 95% confidence intervals.

COST ESTIMATIONS

The cost comparison is based on comparing a 12k oligo chip from CustomArray with the standard price of 36 USD from Integrated DNA Technologies (IDT). The cost of oligo processing is estimated to 800 USD based on the use of appropriate chemicals and enzymes and the oligo chip is estimated to 2,000 USD based on a 12k chip from CustomArray.

Supplementary Table 2 Cost comparison (USD) of standard columns synthesis to micro-array synthesis and processing

Number of Oligos in Mix	Standard column synthesis	Total for chip-oligos	Ratio: Standard vs. micro-array
78	2808	2,800	1
100	3,600	2,800	1.3
1000	36,000	2,800	13
5000	180,000	2,800	64
10000	360,000	2,800	129

OLIGOS

PCR amplification of genomic loci upstream up 12 genes

MTB_72_thrL_FW	TGGTTACCTGCCGTGAGT
MTB_73_acrD_FW	AATGCCTCCTACTGACCAAA
MTB_74_edd_FW	CTTGTTCTATCCGGGCGA
MTB_75_osmC_FW	AGGGATTGTGATTGGTATGA
MTB_76_AcrA_FW	CTTGTTGGGCCTGTTTGT
MTB_77_MdaB_FW	TATCCTGCATCGGTGAGT
MTB_78_GlnD_FW	TTAATTCATCACGGGGCCA
MTB_79_SodB_FW	CCCCAAAAACACTTCGCT
MTB_80_FryB_FW	TTCTACCGCCGCTTCTTC
MTB_81_pssA_FW	AGCTCGGGTTTAACGTTG
MTB_82_hemC_FW	CTCGCCATCAACTTGTCT
MTB_83_secE_FW	CCCTTTTTGCACGCTTTC
MTB_84_thrL_RV	GCGGGCTTTTTTCTGTGTT
MTB_85_acrD_RV	CAACAGGATTGCCAGCAC
MTB_86_edd_RV	CCCACGAGGCTTTTTTTATTAC
MTB_87_osmC_RV	TTCCCTTCCCGCGTTT
MTB_88_AcrA_RV	GGTTTTTCGTGCCATATGTT
MTB_89_MdaB_RV	TCAGGGTGTCGTTCAGTT
MTB_90_GlnD_RV	CTGATAACGGCTGCGAAA
MTB_91_SodB_RV	ATACTCGATGGTTTCCGCA
MTB_92_FryB_RV	CAATTTTTCTTTGTGTCCCCTC
MTB_93_pssA_RV	GAGAAATCTTGGGTAGTTGGG
MTB_94_hemC_RV	ATGCGCTACATACAAGTG
MTB_95_secE_RV	CCCACTTCATCGCTTCCA

MO-MAGE PROTOCOL

Day 1

1. OLS template: add 500ul IDTE to lyophilized OLS library
2. Make 20uM dilution of 150uM primer stock: add 10ul to 65ul dH_2O
    b. Promoter-std-f/r/: B1/B2 i. Forward sequence
        1. /5Biog/C*C*TTGAATCGACACTGCAG/3deoxyU/
    ii. Reverse sequence
        1. /5Phos/CGAACTCGCCAAGGTAGATC
3. Real time PCR setup
    a. Total volume: 100ul
    b. Split into 2 tubes with 50ul each c. Do 4 reactions total:
        i. T7prom+template
        ii. T7prom-template (controls)
    d. PCR reagent setup:
        i. 50ul 2x Kappa SybrFAST kit
        ii. 2.5ul 20uM primer forward (500nM final concentration)
        iii. 2.5ul 20 uM primer forward (500nM final concentration)
        iv. 1.0uL template v. 44ul dH_2O
    e. PCR cycle setup:
        i. Step 1: 95C, 60sec
        ii. Step 2: 95C, 10sec
        iii. Step 3: 62C, 30sec
        iv. Step 4: goto Step 2, 39x
        v. Step 6: 72C, 30sec
        vi. Step 7: end
    f. Stop PCR reaction after it starts to slope off (first step)
        i. Skip to Step 6 after last cycle at 62C run down to 10sec, finish with polish PCR at 72C for 30sec
4. PCR purify the rt-PCR product
    a. Pool 50ul into 100ul, use Qiagen PCR purification kit
    b. Need to add 10ul Sodium Acetate to adjust pH during PCR cleanup c. Elute in 50ul EB
    d. Store in 4C if necessary, run out on gel to verify product (1ul or 0.5ul)
5. Second PCR amplification
    a. PCR reagent setup: make 2 full PCR plates (100ul per well) for each template. 96*2*100 = 19200uL mix
        i. 17.4mL dH_2O (20-0.62-2)
        ii. 2ml 10x PCR buffer
        iii. 100uL 150uM primer forward
        iv. 100uL 150uM primer reverse
        v. 20uL template (from step 4, 1/1000 effective dilution)
        vi. 160ul 25mM dNTPs
        vii. 400U Polymerase (80uL) Enzymatics Taq Polymerase
        viii. Polymerase not HotStart, so need to keep cool
        ix. Aliquot into 50ml Falcon tube first, then into solution basin, then into chilled 96-well plates; seal (microseal "A" film, BioRad)
    b. PCR cycle setup:
        i. Step 1: 94C, 180sec
        ii. Step 2: 94C, 10sec
        iii. Step 3: 62C, 60sec
        iv. Step 4: goto Step 2, 34x
        v. Step 6: 68C, 60sec
        vi. Step 7: 4C, forever

vii. Step 8: end

6. Purification of second PCR product
   a. Pool 2 plates of PCR products into 50 ml falcon tubes
   b. Filter with 50ml MWCO centrifugal filter units (Millipore Amicon Ultracel 10K MWCO)
      i. Put in 10 ml first, spin 4000xg for 5 min, discard eluent ii. Put in rest of 10 ml, spin for 4000xg for 15min, discard eluent
      iii. Put in 12 ml of TE (from IDTE), spin for 4000xg for 15min, discard eluent

7. Protease digest to remove polymerase
   a. Put all samples from filter collection into microcentrifuge tube
   b. Add 50 ul TE (from IDTE) into filter collection to collect residual samples and spin for 1 min at 4000xg, put residual sample (~50ul) into same microcentrofuge tube
   c. Add 40ul of Qiagen Protease (stock conc) into sample, sample should turn cloudy, incubate at 37C for 40 minutes in shaking Thermomixer(Eppendorf)

8. Bind sample using protein resin
   a. Add 70 ul per sample reaction (Rapid Clean, Advansta)
   b. Vortex for 15 sec
   c. Spin down in 1.5ml Centrifuge column (Pierce Thermo, Prod #89868)
      i. 1min, 2000xg
      ii. Should yield about 150ug in 400ul of sample
      iii. Measure with nanodrop
      iv. Sample now in TE (slight blue tinge), store at 4C
      v. Assay by running on gel (E-gel low range quantitative DNAladder

9. Wash sample with dH2O
   a. Spin down sample in 2ml MWCO centrifugal filter unit (Millipore Amicon Ultracel 10K MWCO, smaller version of what we used in step 6b)
   b. 14,000xg, 8min, save TE eluent in case something goes wrong
   c. Resuspend sample in the filter unit with 400ul Ambion nuclease-free dH_2O
   d. 14,000xg, 8min, save eluent in case something goes wrong
   e. Add 100 ul of dH_2O (there should be ~40ul in the filter unit), invert unit and transfer to new microcentrifuge tube
   f. 1,000xg for 2 min, sample should be in 140ul at the bottom of the microcentrifuge tube.
   g. Save 2ul of each sample in -20C to run quantitative gel against post-exo digest.

10. λ-exo digest, chew up phosphate ending strand a. KO sample: Example: 30ug total dsDNA in 140ul
   b. T7 sample: Example: 22ug total dsDNA in 140ul
   c. 20 ng/unit λ-exo is optimal, so use 600 units of λ-exo (from Enzymatics)
   d. Stock λ-exo enzyme conc. of 5 units/ul, so we need 120ul of λ-exo e. We want 10% enzyme concentration, so the total reaction volume is 1.2ml
   f. Reaction setup (for each sample)

i. 140ul dsDNA sample (max 30ug total), it's ok if there is excess λ-exo.
ii. 120ul λ-exo (5units/ul)
iii. 120ul 10x buffer
iv. 820ul Ambion dH_2O
v. Total reaction volume: 1200ul, 10% enzyme concentration, which is desired.

g. Run reaction in ThermoMixer for 4hr at 37C at 750rpm h. Heat inactivate λ-exo for 15 min at 75C
i. Freeze tube in -20C

11. Wash step
a. Spin down sample in 2ml MWCO centrifugal filter unit (Millipore Amicon Ultracel 10K MWCO, same as Step 9)
b. Add 100 ul Ambion nuclease-free dH_2O
c. Store at 4C (total volume of 150ul)
d. Run on denaturing and nondenaturing gel for assay/quantification

12. DpnII and USER digest of primer ends from ssOligos a. Full reaction: 1.5ml in a 2ml tube
i. 150ul template
ii. 30ul DpnOII (1500 units)
iii. 100ul USER (100units)
iv. 150ul DpnII Buffer
v. 100uL 150nM 15mer guide primer (15nmole primer)
vi. 970uL H_2O

b. Method:
i. Ramping with only template, buffer, guide primer and water
1. Ramp down 95Cåå60C at 0.1C/s (~6min)
2. Hold at 60C for 3min
3. Ramp down 60Cåå50C at 0.1C/s (~2min)
4. Hold at 50C for 3min
5. Ramp down 50Cåå37C at 0.1C/s (~2min)
6. Hold at 37C for 3min
ii. Add enzymes (DpnII andUSER)
iii. Incubate at 37 for 2 hours
iv. Heat inactivate DpnII (and USER) 65C for 20minutes

15. Final desalt
a. Adding 400ul of water into sample and spin sample in 3kD column at 14,000xg for 20 minutes, save supernatant just in case
b. Add 50 ul of Ambion water into about 20ul of sample, invert tube, and spin again into a fresh tube

16. Quantify amount of oligos on TBE-UREA gel

VERIFIED T7 INSERTIONS UPSTREAM OF GENES IN THE CONSTRUCTED CELL LIBRARIES

Supplementary Table 3 List of the 150 genes where a T7 sequence was identified upstream in the deep sequencing.

yagU	flhB	ynfF	ycgF	yfjR	sohB
yagK	dnaJ	ydiN	yohO	btuE	panE
mdaB	ybaA	mcbR	psrO	ibpA	uxaB
yedQ	ydiU	spy	yadN	pepB	rluB
yoaI	yebB	tolB	mltA	yjdI	ygaD
slyB	yhcC	rybB	cspB	yhaO	ydaL
pbpG	ycgJ	pgl	bglH	yqhA	ligT
gpmA	yeaP	ykgC	tesB	yqeK	ydcN
pmbA	yidE	rhtC	yfbU	ynfN	yaeF
yifK	acrD	osmC	rhaB	ygcF	ymgJ
ydcR	sroB	clcB	rsmG	ycgK	mdtH
dhaK	yccA	ydbK	yciV	hemA	yihO
norV	ubiG	mhpT	yfiE	pinE	yafN
damX	abgR	fliF	yedR	ymgA	sodC
edd	ydfO	gsp	ybcL	slyX	amyA
yohC	ydhX	sra	rnb	ileS	dcuD
ycgE	yhfZ	mppA	yciI	lexA	udp
chbC	yeeU	dsrA	sbcD	ydaN	sapA
ydhB	yddV	ccmC	yjdK	nuoC	
rumA	mutY	fadI	dgt	nuoG	
entD	cyaA	ymiB	ycgL	nadB	
hrpB	nudG	lipA	yciY	fliA	
rem	ybiC	intS	yehD	yjjK	
ydeI	artP	cpxP	xseA	cmr	
dpiB	rdlC	ldrA	ykgH	alr	
chaB	cspF	yjbB	rsmC	yceQ	

REFERENCES

(1) Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., and Church, G. M. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature 460*, 894–8.

(2) Mitsis, P. G., and Kwagh, J. G. (1999) Characterization of the interaction of lambda exonuclease with the ends of DNA. *Nucleic Acids Res. 27*, 3057–63.

(3) Bitinaite, J., and Nichols, N. M. (2009) DNA cloning and engineering by uracil excision. *Curr. Protoc. Mol. Biol. Chapter 3*, Unit 3.21.

(4) Abramoff, M.D., Magalhaes, P.J., Ram, S. J. (2004) Image processing with ImageJ. *Biophotonics*.

(5) Wang, H. H., and Church, G. M. (2011) Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering. *Methods Enzymol. 498*, 409–26.