

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

ACS Synth. Biol., **Just Accepted Manuscript** • DOI: 10.1021/sb5001565 • Publication Date (Web): 23 May 2014

Downloaded from <http://pubs.acs.org> on June 6, 2014

Just Accepted

“Just Accepted” manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides “Just Accepted” as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. “Just Accepted” manuscripts appear in full in PDF format accompanied by an HTML abstract. “Just Accepted” manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). “Just Accepted” is an optional service offered to authors. Therefore, the “Just Accepted” Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these “Just Accepted” manuscripts.



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^{3,4}, Morten O.A. Sommer^{1,2}, Harris H. Wang⁵

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 Genetics, Harvard Medical School, Boston, MA 02115, USA.

⁵ 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)

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 loci 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

INTRODUCTION

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 medicines¹ and chemicals². 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 sizeable 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 projects¹⁰

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 oligos¹¹ at a significantly lower cost and turnaround time compared to oligos produced by column-based synthesis¹²⁻¹⁴. Here, we describe Microarray Oligonucleotide (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 was 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 anti-consensus (TCCTCC) ribosomal binding sites, or insertion of a T7 promoter sequence (TAATACGACTCACTATAGGG)

1
2
3 upstream of operons. A summary of the genomic perturbation design is given in Table 1. In
4 all, the designed oligo library constituted 13,000 possible targeted perturbations against the
5 *E. coli* genome (see Supplementary Material).
6

7 Traditional column-based synthesis of 90mer MAGE oligos of this library size would
8 take months to years to generate at a cost >\$10 million USD, which is practically infeasible.
9 Thus, we turned to new approaches in long oligonucleotide synthesis using DNA
10 microarrays¹¹. A 130mer single-stranded DNA library was generated using the Oligo Library
11 Synthesis (OLS) platform from Agilent Technologies (Santa Clara, CA, USA). Because the
12 amount of oligos needed for MAGE is much higher than the total yield from microarray
13 synthesis (μM versus pM), we devised a PCR-based amplification protocol to generate a
14 renewable supply of MAGE-compatible oligo pool from our initial microarray-derived oligo
15 library, which we describe in greater detail below.
16
17
18
19

20 Each oligo library subpool (e.g. CDS Knockout, T7 insertion, RBS-up, RBS down)
21 was designed with a unique set of barcodes, which allowed for selective amplification of only
22 the subpool library by using specific primers (see Figure 1). A total volume of 38.4 mL of PCR
23 was first performed to ensure that we generated enough oligos for >10 MAGE cycles. One of
24 the two primers used for the PCR contained a 5'-phosphothioester bond and a 3'-uracil, while
25 the other contained a 5' phosphorylated group. The resulting amplicons contain a 5'-
26 phosphothioester bond on the MAGE-compatible sense-strand and a 5'-phosphate on the
27 reverse complement strand. This 130 bp double-stranded (ds) DNA amplicon was digested
28 with λ -exonuclease. Since λ -exonuclease has much higher activity for unphosphorylated
29 substrates¹⁶ and its activity is blocked by phosphothioester bonds, the digest results in a
30 130mer single-stranded (ss) DNA library of the MAGE-compatible sense-strand that
31 contained the phosphothioester bond at the 5' end. Subsequently, the 20 bp flanking
32 barcodes used for PCR amplification of the subpools were removed from each end of the
33 ssDNA to yield unique 90mer oligos needed for high efficiency MAGE. This was
34 accomplished by first digesting the 130mer ssDNA library, which contained an internal uracil
35 at the barcode junction from the previous PCR (see Figure 1), with an uracil DNA
36 glycosylase, endonuclease VIII¹⁷, which removes the uracil from the ssDNA strand. The uracil
37 excision effectively removes the 20mer barcode at the 5' end of the ssDNA library, yielding a
38 110mer library. The 3' barcode was designed with a DpnII restriction site placed immediately
39 after the target oligo sequence. To remove the 3' barcode, a guide primer complementary to
40 the 3' barcode including the DpnII site was used to hybridize to the 110mer oligo and was
41 digested with DpnII to yield the designed 90mer oligo library. The use of a guide primer
42 ensured that the remaining part of the single-stranded oligo is not digested by DpnII, which
43 only cuts dsDNA. The resulting oligo library contained unique 90 base single-stranded oligos.
44
45
46
47
48
49
50
51
52

53 Gel electrophoresis of the oligos was performed after each step to ensure that the
54 processing resulted in the expected product (see Figure 2a and 2b). Serial dilutions of the
55 library were further visualization on a TBE-UREA gel to estimate library concentration using
56
57
58
59
60

1
2
3 ImageJ¹⁸. We estimate that a typical amplification generates 1.4 nmols of oligos (~12 μ M in
4 115 μ L), which is sufficient for 14 MAGE cycles at 2 μ M per 50 μ L reaction per cycle.
5

6
7 We performed a side-by-side comparison of the microarray-derived oligos with those
8 obtained through standard column-synthesis from a commercial vendor (Integrated DNA
9 Technologies, Iowa, USA). A test group of 5 oligos generated from the microarray pool and
10 the commercial vendor was used for MAGE and the efficiency of oligo incorporation was
11 determined (see Figure 2c). We find that 3 out of the 5 tested oligos showed slightly lower
12 incorporation efficiencies in the microarray library compared to the column library. We
13 attribute the decreased efficiency to differences in individual oligo concentrations that may
14 result during amplification from the microarray library. Nonetheless, these results offered a
15 convincing proof-of-concept that microarray-derived oligos are compatible with MAGE
16 mutagenesis. Since the introduction of synthetic regulation to native genomic loci has been
17 an outstanding challenge in synthetic biology, we sought to further explore our sub-library that
18 generated T7 promoter insertions in the untranslated regions (UTR) upstream of each of 2585
19 operons in the *E. coli* genome. This multiplexed promoter insertion perturbation enables the
20 generation of a mutant library that contains new transcriptional regulation in the presence of
21 an inducible T7 polymerase system^{19,20}. Thus in this current work, we focused on
22 characterizing our methods in greater detail using this T7 insertion library.
23
24
25
26
27
28

29
30 The T7 promoter insertion library was designed using MODEST¹⁵ to target the UTR
31 region 40 bp upstream of all non-essential and non-transcription-factor operons in *E. coli*.
32 Downstream polycistronic open reading frames without adequate intergenic space were not
33 targeted. In all, 2585 unique insertion target sites were identified. We designed 90 bp oligos
34 by flanking a 20 bp T7 promoter (TAATACGACTCACTATAGGG) sequence with 35 bp of
35 homologous sequence at each end to the target genomic integration site. In general,
36 introduction of a 20 bp insertion is expected have lower incorporation efficiency than smaller
37 point mutations or deletions²¹. Nonetheless, these larger insertions represent a unique
38 opportunity to introduce synthetic regulation and to challenge the limits of our MAGE
39 capabilities using large oligo library pools derived from microarrays.
40
41
42

43
44 Following the generation of the 2585-oligo library, we applied this T7 promoter pool to
45 mutagenize the *E. coli* genome by MAGE for 12 cycles. The entire oligo processing pipeline
46 and MAGE mutagenesis was performed twice in parallel to generate two separate cell
47 libraries to test the reproducibility of our protocol. To assess the success of T7 insertion in
48 these combinatorial libraries, we performed deep sequencing of the final cell populations. We
49 first extracted genomic DNA of the cell populations and hybridized the genomic DNA with a
50 biotinylated oligo containing the T7 promoter sequence. Genomic regions that contained the
51 T7 promoter insertions can thus be enriched from total genomic DNA when it is applied to
52 streptavidin beads that bind to the biotinylated oligo-genomic hybrid. We then sequenced the
53 enriched genomic library by deep-sequenced to identify T7 insertion sites.
54
55
56
57
58
59
60

1
2
3 Next-generation sequencing analysis of the data showed that the two separate
4 libraries contained 87 reads and 208 reads with the T7 promoter sequence, with the reads
5 mapping to 56 and 98 targeted operons respectively in the two libraries. Only 4 of the total
6 154 targets were redundant, resulting in identification of 150 unique targets out of 2585
7 possible. The sequencing coverage (the number of times the genome has been sequenced)
8 of the libraries was 2064x and 1364x respectively. We used a Monte Carlo model to simulate
9 the insertion frequency of the total library, and found that the total cell library is expected to
10 contain between 2250 and 3500 modified genes (95 % confidence interval, see
11 Supplementary Figure 1). These results show that all or most of the targeted sites are
12 predicted to have T7 promoter insertions within the cell library.
13

14
15 We proceeded to further validate our mutagenesis results and to estimate the
16 insertion efficiency. We randomly selected 12 targets, including 8 that had not been detected
17 by population sequencing. We amplified ~200 bp PCR fragments spanning the T7 insertion
18 sites of each of the 12 genes from the cell library and sequenced the PCR products by Next
19 Generation Sequencing. For all 12 loci, we found reads containing the T7 promoter insertion
20 sequence. This provides further support that the cell library contains a majority of the 2585 T7
21 promoter insertions (See Table 1 and Figure 3).
22

23
24 The insertion frequencies for the 12 loci were estimated by comparing the number of
25 reads with T7 promoter insertions to the number of reads without the insertion. The average
26 frequency was 0.017%, and thus the average number of insertions per cell can be estimated
27 as $\mu = k \times p = 0.434$ where $p = 0.00017$ and $k = 2585$ targets. This means that 43% of the cell
28 library is expected to have at least one insertion on average. The top 1% of the population is
29 expected to have at least four (4.3) insertions based on the calculations $m = \mu +$
30 $2.326 \sqrt{k \times p(1-p)} = 4.3$ (for details about the MAGe efficiency calculations see the
31 Supporting Information as well as Wang and Church, 2011⁷). The MAGe efficiency for
32 insertions has previously been predicted based on fitting of empirically determined
33 efficiencies^{7,22}. Using such approach, the insertion efficiency (IE) is $IE = 0.15 \times e^{-0.075 \times (b-1)} =$
34 0.0361, where b is the size of the insertion in number of bases ($b = 20$). Based on this
35 insertion efficiency, the predicted average frequency of each insertion can be calculated by
36 $p_{N-predicted} = 1 - \left(1 - \frac{IE}{k}\right)^N = 1 - \left(1 - \frac{0.0361}{2585}\right)^{12} = 0.000167$, where N is the number of
37 MAGe cycles. The predicted average insertion frequency of 0.0167% matches very well with
38 the measured frequency of 0.017%, providing confidence to our predicted population
39 mutagenesis profile and its applications to analyze MAGe mutagenesis of complex oligo
40 pools. These results further highlight that that cell libraries with combinations of multiple
41 insertions per genome can be generated using this massively multiplexed approach.
42
43

44
45 Here, we have presented a proof-of-concept demonstration for the generation and application
46 of oligo pools that specifically target thousands of unique chromosomal loci across a cell
47 population to introduce promoters amenable for synthetic regulation. The Microarray
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Oligonucleotide MAGE enables rapid mutagenesis of bacterial genomes using oligos derived
4 directly from microarrays without intermediate cloning or cassette selection steps, which can
5 significantly expand the combinatorial genomic diversity of the resulting cell population. A
6 resulting cell library will be useful for various screens for metabolic engineering purposes
7 such as increased production of biochemicals or tolerance towards biomass inhibitors. The
8 MO-MAGE process described here allow researchers to limit the combinatorial space to only
9 specific mutations that are expected to have a much higher chance of leading to a desired
10 phenotype than random mutations. Thus, the effect of rationally designed targets can be
11 assessed very effectively because the mutagenesis quality (*i.e.* the proportion of interesting
12 mutants to other cells) is much higher. Whereas current approaches are limited to only
13 creating few genomic changes at a time, our method could be used to target all predicted
14 beneficial mutations predicted from metabolic models, and screen for optimal phenotypes.
15
16
17
18
19

20 As the complexity of engineering tasks of synthetic biology and metabolic engineering
21 increases, the need to reduce cost of genome engineering becomes more important. MO-
22 MAGE allows the synthesis of oligos at a fraction of the cost (>1,000x) compared to traditional
23 column based oligo synthesis to make large-scale genome engineering accessible to most
24 laboratories. At a price of 36 USD per column-based oligo, MO-MAGE (2800 USD) is
25 currently cost competitive when using more than 78 oligos. This method could provide a
26 paradigm shift by making large-scale genome engineering of many thousands of targets
27 available as a standard tool for strain optimization and other projects where large-scale
28 targeted mutagenesis searches are needed. New advances in DNA and gene synthesis will
29 further foster growth in genome engineering of microbial and eukaryotic systems²³ and in
30 theory, oligo pools containing millions of oligos can be applied for MO-MAGE. However, since
31 increasing oligo diversity leads to lower replacement efficiency per genomic target, there is a
32 practical limitation in the number of oligos that can be meaningfully applied depending on the
33 required replacement efficiency per site for a given experiment.
34
35
36
37
38

39 For some projects, a higher integration frequency might be of interest, to allow more
40 combinations of insertions and a higher quality library. The amount of MAGE cycles
41 performed can be increased to higher levels, and automated solutions could increase the
42 feasible number of MAGE cycles to several 100's. For creation of complex libraries where
43 many combinations per cell is desired, single base pair substitutions and small insertions and
44 deletions can be applied to increase the replacement efficiency. For instance if 1000 oligos
45 designed for making single base pair substitutions are used for 100 cycles of MAGE, a cell
46 library with 25 average replacements per cell is predicted. This allows creation of
47 unprecedented targeted combinatorial libraries of specific chromosomal modifications.
48 Ultimately we envision MO-MAGE method will be used to make thousands of specific
49 chromosomal changes predicted to result in a desired phenotype, and combined with
50 selection or screening of the cell library for interesting mutants.
51
52
53
54
55
56
57
58
59
60

METHODS

Strains and culture conditions. We used a strain based on the EcNR2 strain²² 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 dH₂O) 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 dH₂O) 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 synthesised on the microarray chip designed to insert a T7 promoter upstream of *E. coli* genes can be found in the supplementary material.

MAGE. MAGE was performed according to the protocol provided in²¹. Briefly, the cells were grown to mid-log phase, whereafter the beta protein of the lambda-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 μ M oligos in 50 μ L and electroporated in a Bio-Rad MicroPulser, BTX ECM-830 with 1 mm gap cuvette, whereafter the cells were incubated for 2-3 hours 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. 2009²⁴ and NimbleGen SeqCap EZ Exome Library SR v2.2 protocol²⁵. A biotinylated oligo targeting the T7 promoter was incubated 66 hours 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 NEBNext® DNA Sample Prep Master Mix Set 1 kit for Illumina sequencing and manually ordered adapters kindly provided by Luhan Yang (Harvard Medical School, George Church Lab).

Adaptor1	PE-A1-F	TACTACTCTTTCCCTACACGACGCTCTTCCGATCTac*T
	PE-A1-R	/5Phos/gtAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
Adaptor2	PE-A2-F	TACTACTCTTTCCCTACACGACGCTCTTCCGATCTtg*T
	PE-A2-R	/5Phos/caAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG

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*²⁶ was applied to perform the alignment of the reads to the reference genomes (see parameters in the script in supplementary materials). *Samtools*²⁷ 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 supplementary materials)

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.

ACKNOWLEDGEMENT

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 DOE (as above) and the NSF (SA5283-11210). M.T.B., H.J.G., K.S.L. and M.O.A.S. acknowledge funding from the Novo Nordisk Foundation.

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.

- 1
2
3 (13) Tian, J., Ma, K., and Saaem, I. (2009) Advancing high-throughput gene synthesis technology. *Mol. Biosyst.* 5,
4 714–22.
5
6
7 (14) Richmond, K. E., Li, M.-H., Rodesch, M. J., Patel, M., Lowe, A. M., Kim, C., Chu, L. L., Venkataramaian,
8 N., Flickinger, S. F., Kaysen, J., Belshaw, P. J., Sussman, M. R., and Cerrina, F. (2004) Amplification and
9 assembly of chip-eluted DNA (AACED): a method for high-throughput gene synthesis. *Nucleic Acids Res.* 32,
10 5011–8.
11
12 (15) Bonde, M. T., Klausen, M. S., Anderson, M. V., Annika, I. N., Wang, H. H., and Sommer, M. O. A. (2014)
13 MODEST: A Web-based Design Tool for Oligonucleotide-mediated Genome Engineering and Recombineering.
14 *Nucleic Acids Res. Webserver* .
15
16
17 (16) Mitsis, P. G., and Kwagh, J. G. (1999) Characterization of the interaction of lambda exonuclease with the
18 ends of DNA. *Nucleic Acids Res.* 27, 3057–63.
19
20
21 (17) Bitinaite, J., and Nichols, N. M. (2009) DNA cloning and engineering by uracil excision. *Curr. Protoc. Mol.*
22 *Biol. Chapter 3*, Unit 3.21.
23
24
25 (18) Abramoff, M.D., Magalhaes, P.J., Ram, S. J. (2004) Image processing with ImageJ. *Biophotonics*.
26
27
28 (19) Wang, H. H., Kim, H., Cong, L., Jeong, J., Bang, D., and Church, G. M. (2012) Genome-scale promoter
29 engineering by coselection MAGE. *Nat. Methods* 9, 591–3.
30
31
32 (20) Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F., and Voigt, C. A. (2012) Modular control of multiple
33 pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res.* 40, 8773–81.
34
35
36 (21) Wang, H. H., and Church, G. M. (2011) Multiplexed genome engineering and genotyping methods
37 applications for synthetic biology and metabolic engineering. *Methods Enzymol.* 498, 409–26.
38
39
40 (22) Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., and Church, G. M. (2009)
41 Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460, 894–8.
42
43
44 (23) Leproust, E. (2008) Agilent's Microarray Platform: How High-Fidelity DNA Synthesis Maximizes the
45 Dynamic Range of Gene Expression Measurements. *Agil. Technol.*
46
47
48 (24) Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., Fennell, T., Giannoukos,
49 G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S., and Nusbaum, C. (2009) Solution hybrid selection
50 with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* 27, 182–9.
51
52
53 (25) Nimblegen. NimbleGen SeqCap EZ Library SR User's Guide. *Library (Lond)*.
54
55
56 (26) Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009) Ultrafast and memory-efficient alignment of
57 short DNA sequences to the human genome. *Genome Biol.* 10, R25.
58
59
60 (27) Flicek, P., and Birney, E. (2009) Sense from sequence reads: methods for alignment and assembly. *Nat.*
Methods 6.

Table 1. Overview of oligos designed and synthesized on the micro-array chip. 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 downregulation by insertion of an anti-consensus RBS sequence (“RBS down”).

	Targets	General	TFs	Genes targeted
CDS Knock Outs	3,798	3,633	167	All non-essential, non-pseudo, non-ncRNA
T7 promoters	2,723	2,585	138	All non-pseudo, non-ncRNA, *
RBS up	3,323	3,172	151	All non-pseudo, non-ncRNA
RBS down	3,099	2,948	151	All non-essential, non-pseudo, non-ncRNA
Total	12,943	12,338	607	

*Only genes with sufficient spacing to the next gene upstream were targeted to only allow insertion T7 promoters without disturbing upstream genes.

Table 2. 12 genes randomly selected for deep amplicon sequencing and analysis. The number of reads with a T7 promoter insertion was compared to reads without an insertion to calculate the insertion frequency.

	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%

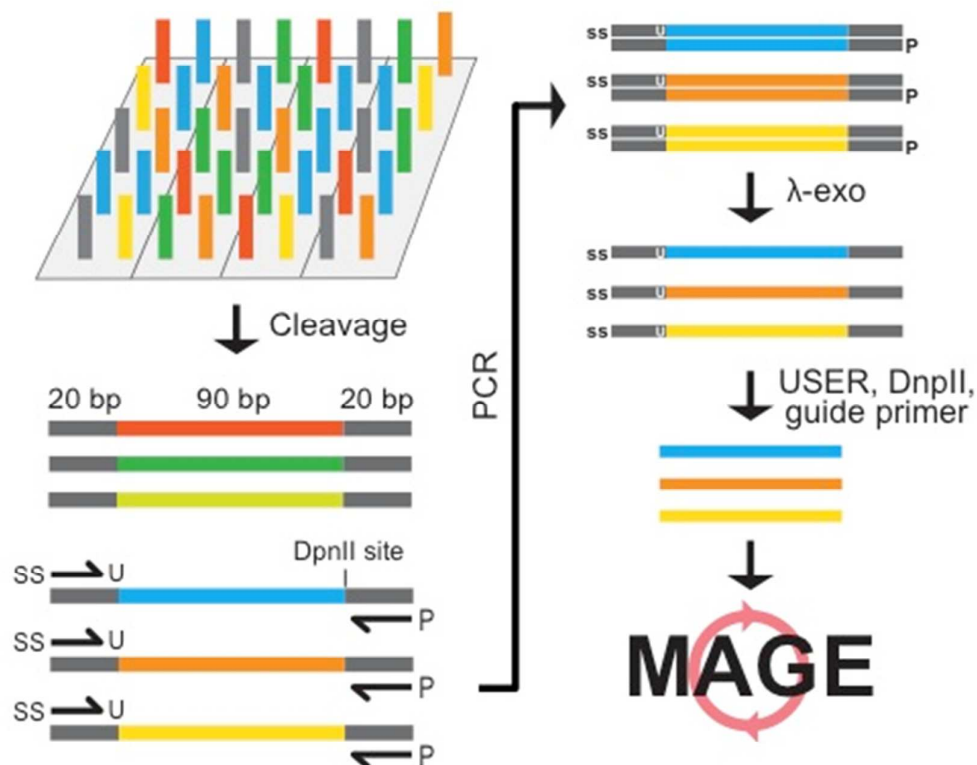


Figure 1. MO-MAGE method for targeted whole genome mutagenesis. 130 base oligonucleotides were designed and synthesized on a DNA microarray, which can be ordered from several commercial vendors. The oligos can be designed with different barcodes, which allow selective PCR amplification of a desired subpool. 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.

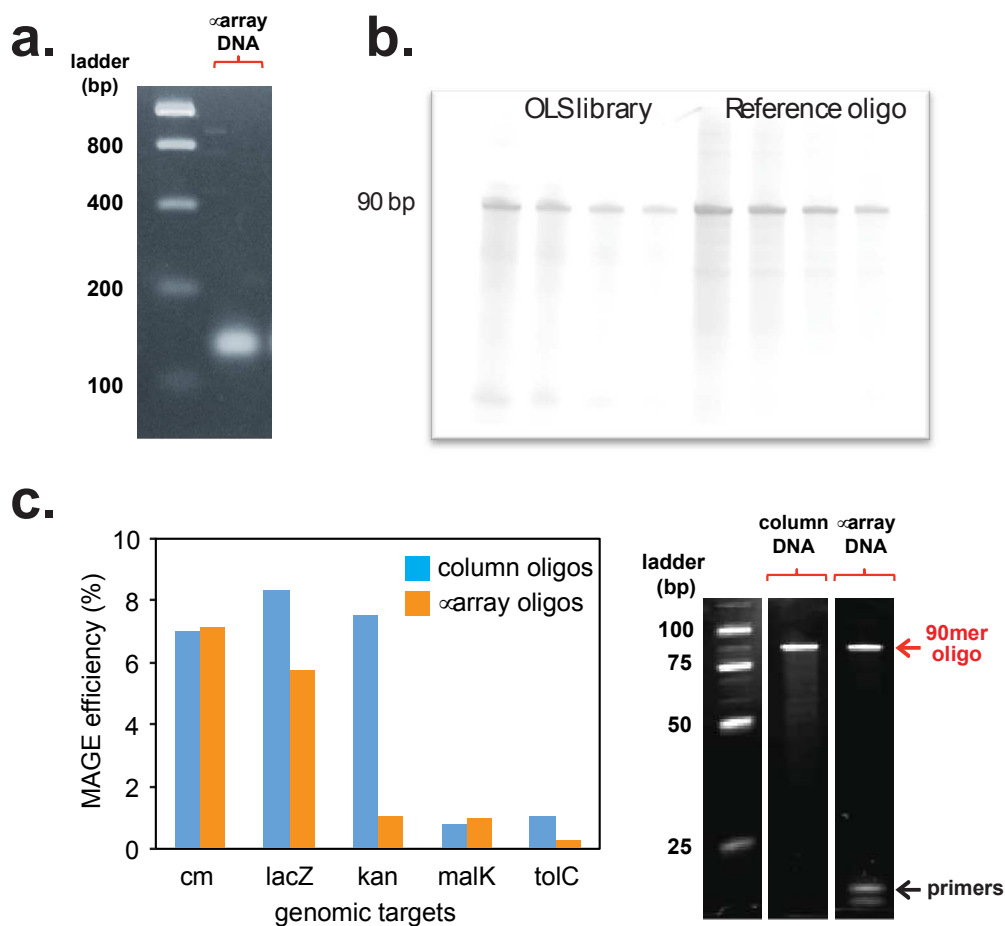


Figure 2 (a) PCR amplicons of the T7 promoter oligo pool can be seen as the strong bands around 130 bp (4 % Agarose E-Gel EX with Low Range Quantitative DNA Ladder) **(b)** Serial dilutions of the processed single stranded T7 promoter oligo library (left 4 lanes) compared to a reference oligo of 90 bp (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).

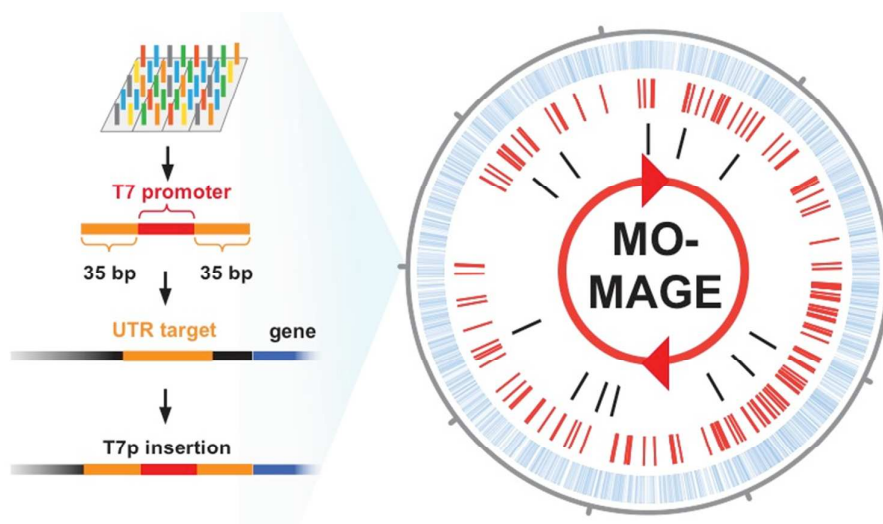
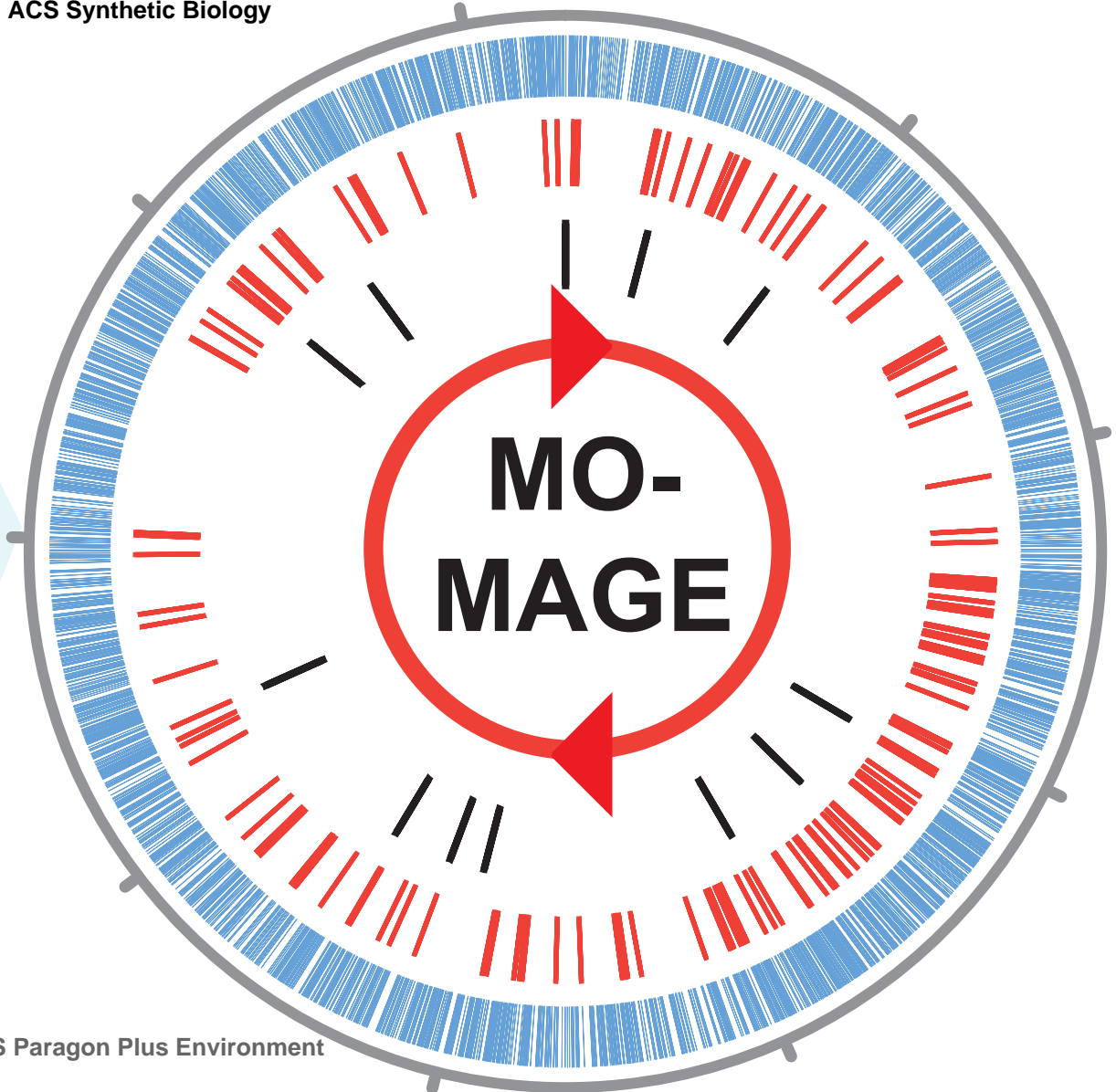
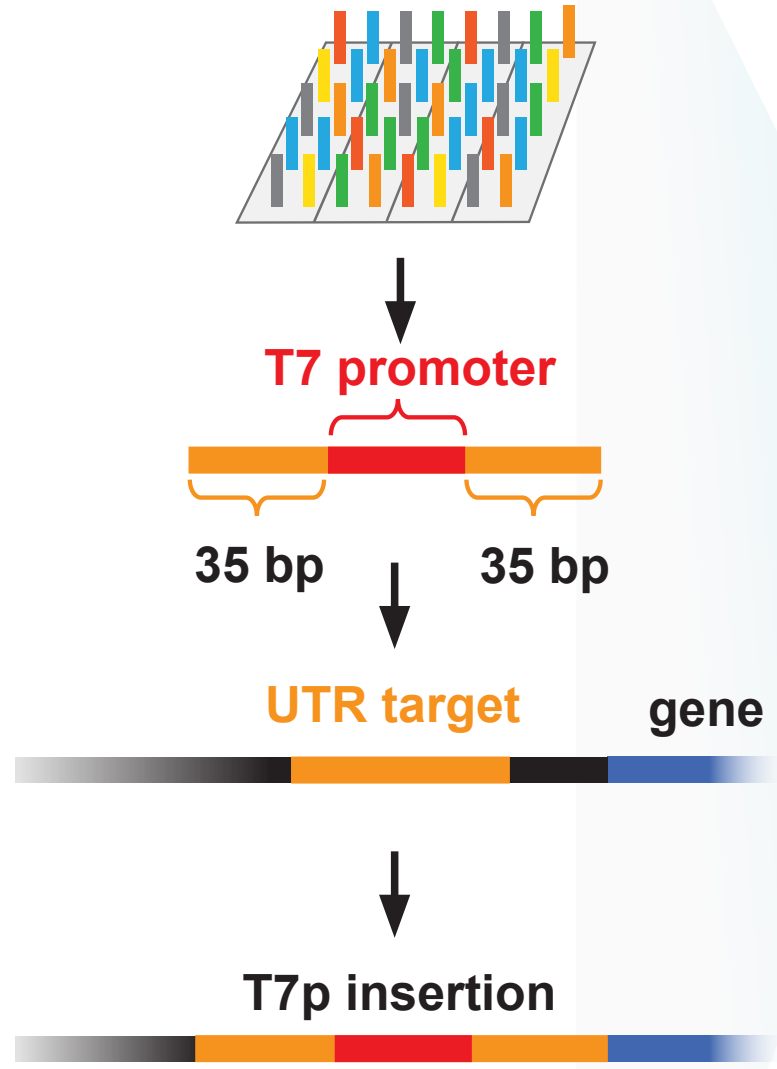
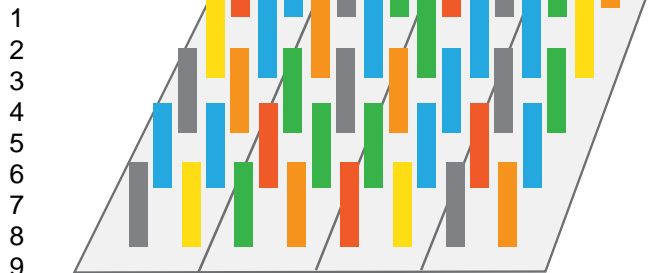


Figure 3. MO-MAGE of 2585 genomic targets corresponding to untranslated regions (UTR) upstream of genes for insertion of 20-bp T7 synthetic promoter. Designed targets are shown in blue. Mutated targets verified by whole-genome sequencing are shown in red (see supplementary table 3 for complete list).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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37





↓ Cleavage

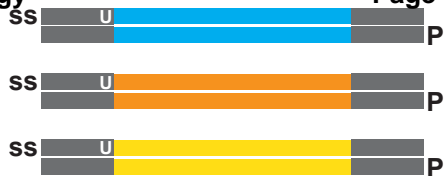
20 bp 90 bp 20 bp



DpnII site



PCR



↓ λ-exo



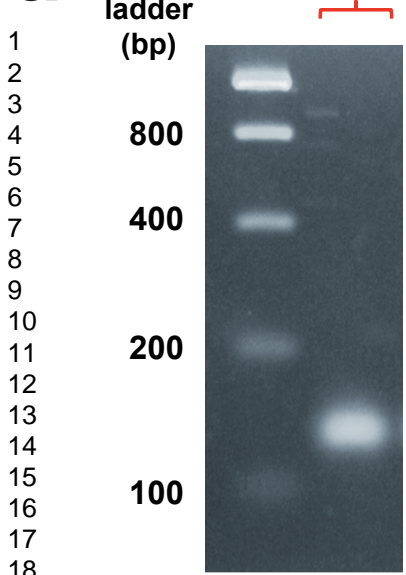
↓ USER, DpnII,
guide primer



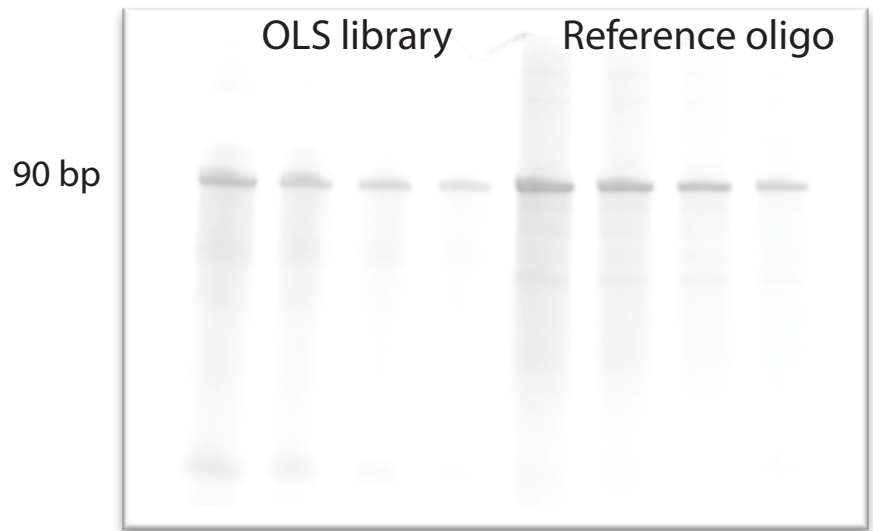
↓

MAGE

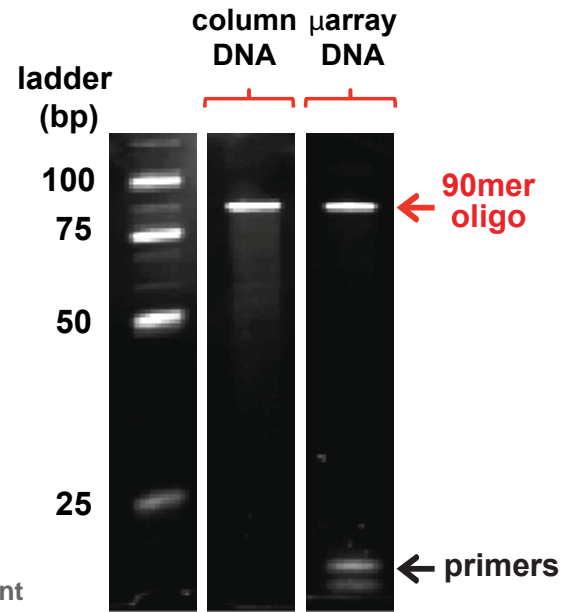
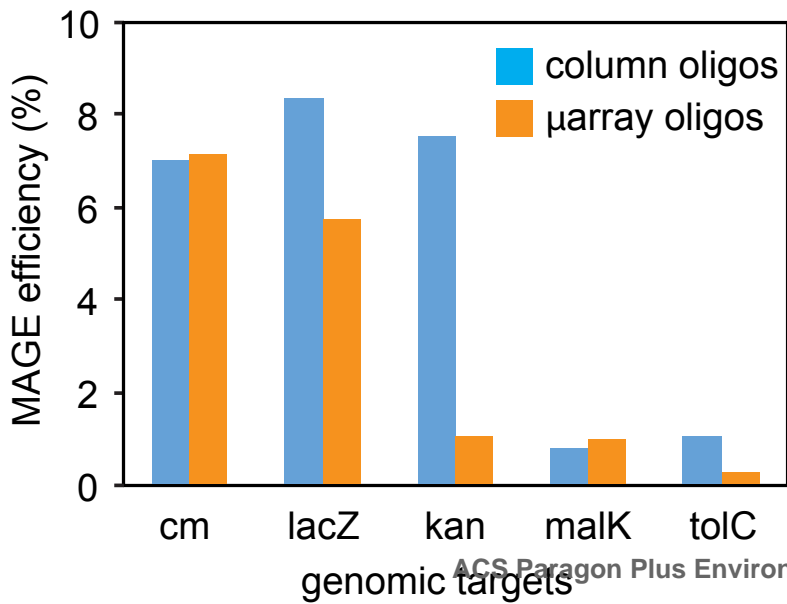
a.



b. ACS Synthetic Biology



c.



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37

