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.

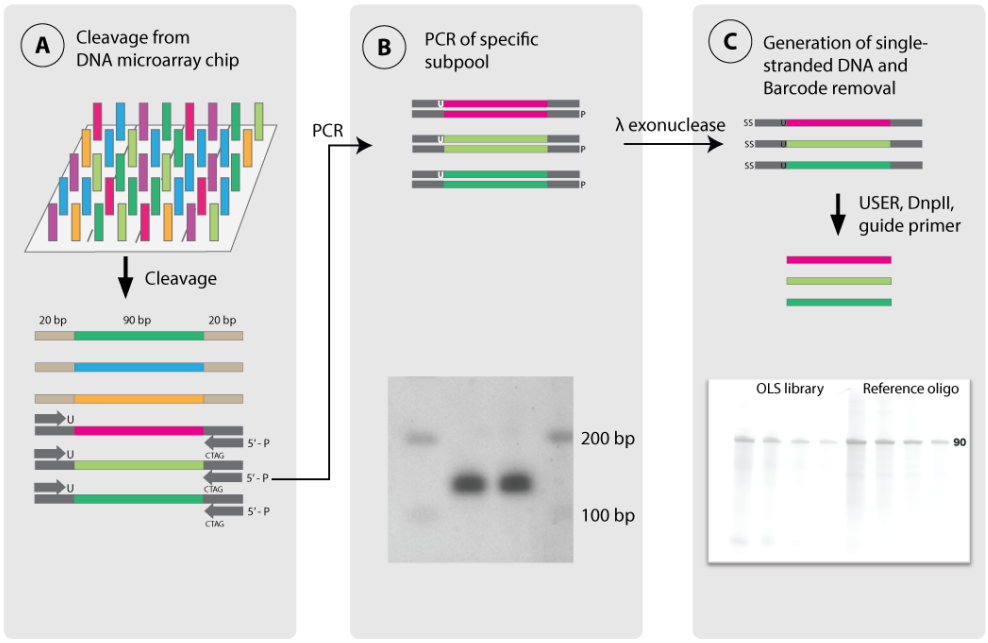
We designed each subpool with a unique barcode, which allows a selective amplification of only that library by using specific primers (see Supplementary Figure 1). We ordered the oligos from Agilent Technologies laboratories, and PCR amplifiied with a total volume of 38.4 mL, to ensure enough oligos 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

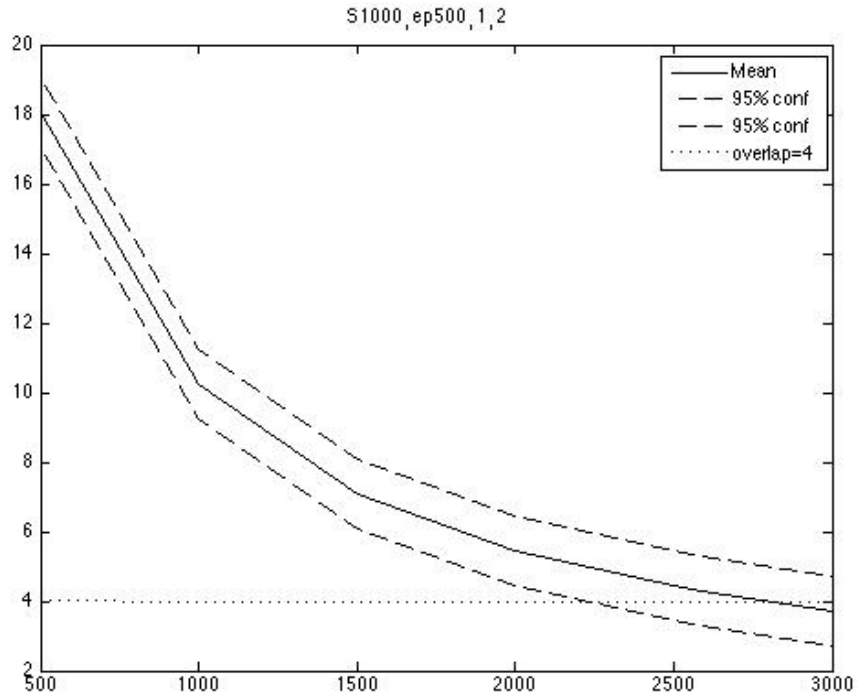
$$p_{N-predicted} = 1 - (1 - RE_{av})^N = 1 - \left(1 - \frac{0.0361}{2587}\right)^{12} = 0.000167$$

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.