

Supplementary Materials, Methods and Data

Long-adapter single-stranded oligonucleotide probes for massively multiplexed cloning of kilobase genome regions

Lorenzo Tosi^{1*}, Viswanadham Sridhara^{1*}, Yunlong Yang^{1*}, Dongli Guan¹, Polina Shpilker¹, Nicola Segata², H. Benjamin Larman^{3†}, Biju Parekkadan^{1,4,5†}

¹Department of Surgery, Center for Surgery, Innovation, & Bioengineering, Massachusetts General Hospital, Harvard Medical School and the Shriners Hospitals for Children, Boston, Massachusetts 02114, USA

²Centre for Integrative Biology, University of Trento, Trento, Italy

³Division of Immunology, Department of Pathology, Johns Hopkins University, Baltimore, MD, USA

⁴Harvard Stem Cell Institute, Cambridge, Massachusetts 02138, USA

⁵Department of Biomedical Engineering, Rutgers University and the Department of Medicine, Rutgers Biomedical and Health Sciences, Piscataway, New Jersey 08854, USA

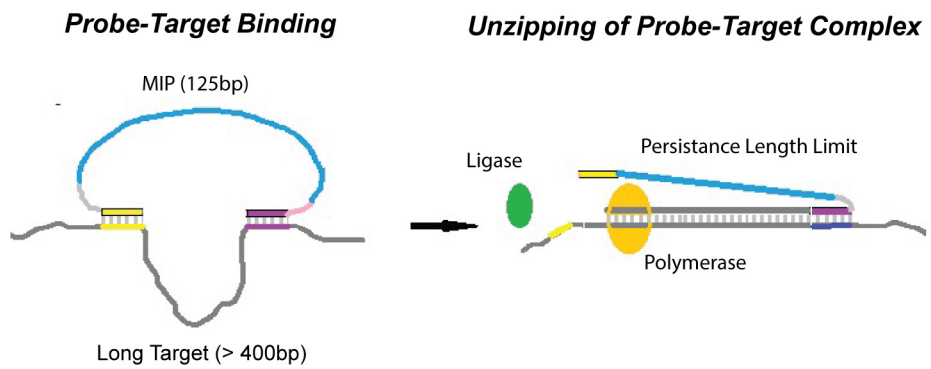
*Contributed equally

†Correspondence and requests for materials should be addressed to B.P.

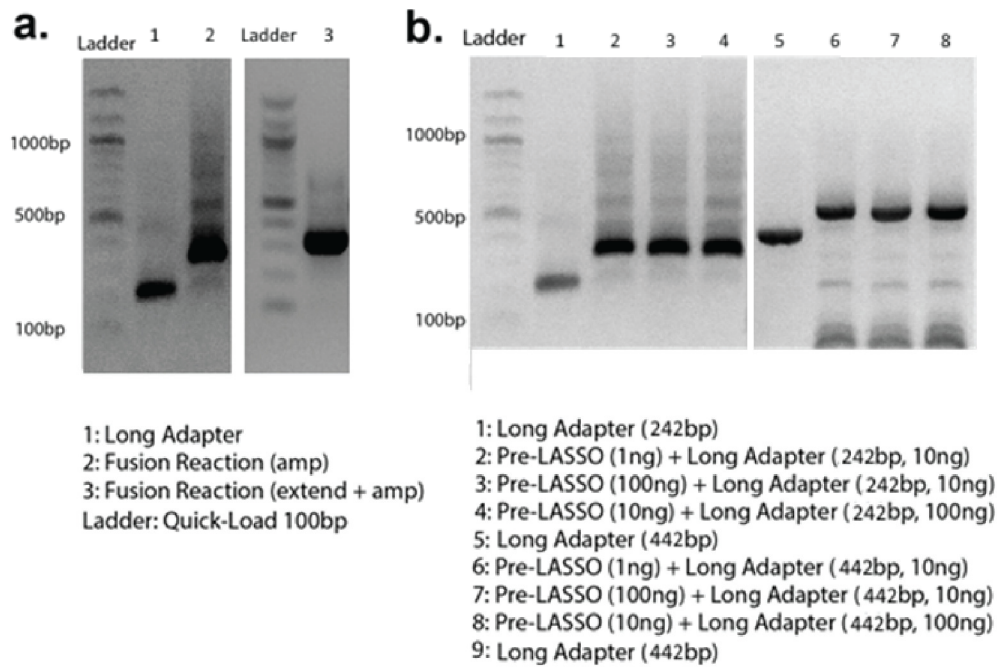
(biju_parekkadan@hms.harvard.edu) or H.B.L. (hlarman1@jhmi.edu)

TABLE OF CONTENTS

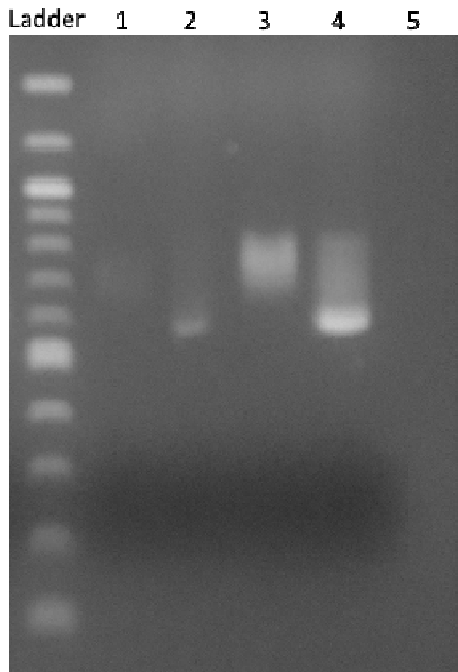
Content	Page
Supplementary Figure 1. Failure of Conventional MIPs to Capture Long DNA Fragments.	3
Supplementary Figure 2. Optimization of fusion PCR step of single LASSO probe synthesis.	4
Supplementary Figure 3. Ligation-dependent circularization of fusion PCR products. (Full blot Fig 1d)	5
Supplementary Figure 4. Optimization of circularization by ligation of fusion PCR products.	6
Supplementary Figure 5. Optimization of Gap Filling mix composition for single target capture using LASSO probes.	7
Supplementary Figure 6. Full blot of figure 2d	8
Supplementary Figure 7. Estimation of the percentage of functional captured KanR2 ORFs.	9
Supplementary Figure 8. Optimization of different parameters for ORFeome capture	10-11
Supplementary Figure 9. LASSO versus MIP Capture	12
Supplementary Figure 10. Cloning human microbiome-derived ORFs	13
Supplementary Table 1. Sequences of Long Adapters	14
Supplementary Table 2. Sequences of Primers	15
Supplementary Table 3. Sequences of Single Pre-LASSO probes	16
Supplementary References	17



Supplementary Figure 1. Failure of Conventional MIPs to Capture Long DNA Fragments. Proposed model for unsuccessful target capture. A MIP initially hybridized with a longer target is shown on the left. On the right, the complex “unzips” at the ligation arm from the hybridization site due to the stiffness of nascent dsDNA.



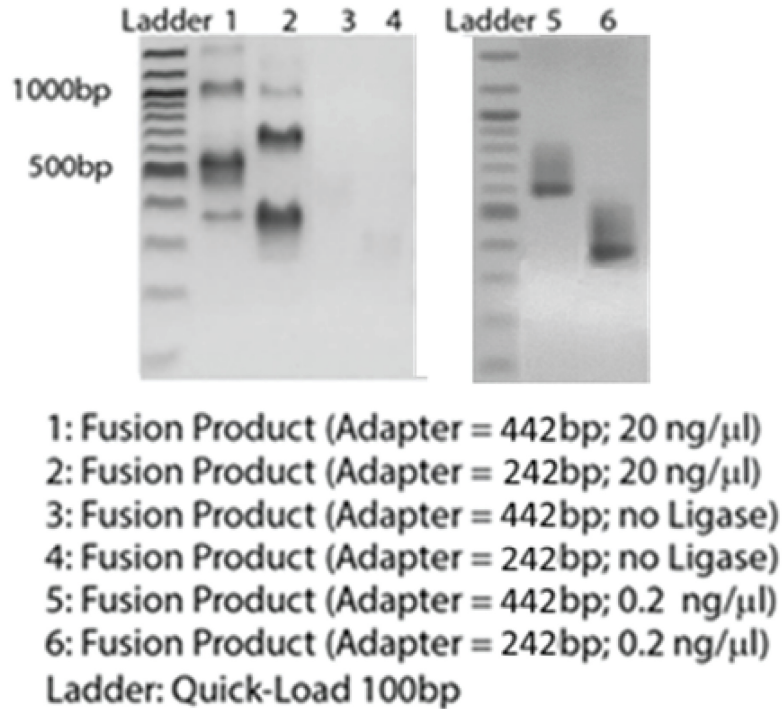
Supplementary Figure 2. Optimization of fusion PCR step of single LASSO probe synthesis. (a) Different amplification and extension conditions of the fusion reaction were tested. Lane 1: Long Adapter (242 bp). Lane 2: Fusion PCR of a pre-LASSO probe (150 bp) with a Long Adapter (242 bp) by direct PCR. Lane 3: Fusion PCR of a pre-LASSO probe (150 bp) with a Long Adapter (242 bp) obtained performing a “fusion by extension” step prior the PCR amplification. The “fusion by extension” involved subjecting the pre-LASSO probe and the Long Adapter to 10 PCR extension cycles (denaturation, annealing and extension) without the primers in the PCR master mix. After the extension, the primers were added in solution and PCR amplification performed for 30 cycles. (b) Testing different concentrations of pre-LASSO probe (150 bp) and Long Adapters (242 bp, 442 bp) in fusion PCR. As shown in lanes 2,3,4; lanes 6,7,8 the expected fusion products were obtained by using all three lengths Long Adapters with no visible differences in yield and specificity.



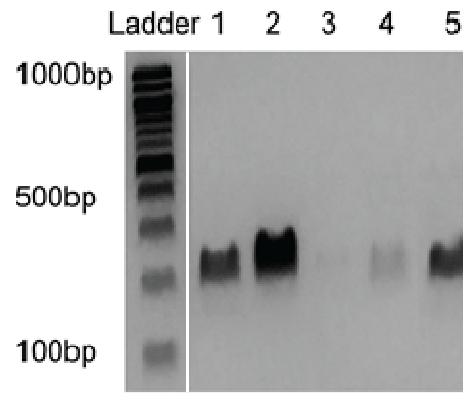
- 1: No ligase, 15' Digestion
- 2: Linearized product 550bp
- 3: Circular product 550bp
- 4: Linearized product 550bp
- 5: No ligase, 1h Digestion

Supplementary Figure 3. Ligation-dependent circularization of fusion PCR products. (Full blot Fig 1d)

A fusion PCR products of approximately 550 bp was obtained from a 150 bp pre-LASSO probe with Long Adapters of 442 bp. The fusion product (1 μ g) with sticky ends (EcoRI digested) was diluted to 0.2 ng/ μ l in 1X T4 DNA Ligase buffer and T4 ligated. After ligation, linear DNA was digested with exonucleases. DNA circles were column-purified, and run in a gel (Lane 3) in parallel with the linear non ligated and undigested fusion product (Lane 4). The circular nature of the DNA present in the bands was confirmed by the ligase negative controls where all DNA was completely digested by treating the DNA with exonucleases for 1h as expected (Lanes 5 and 6).

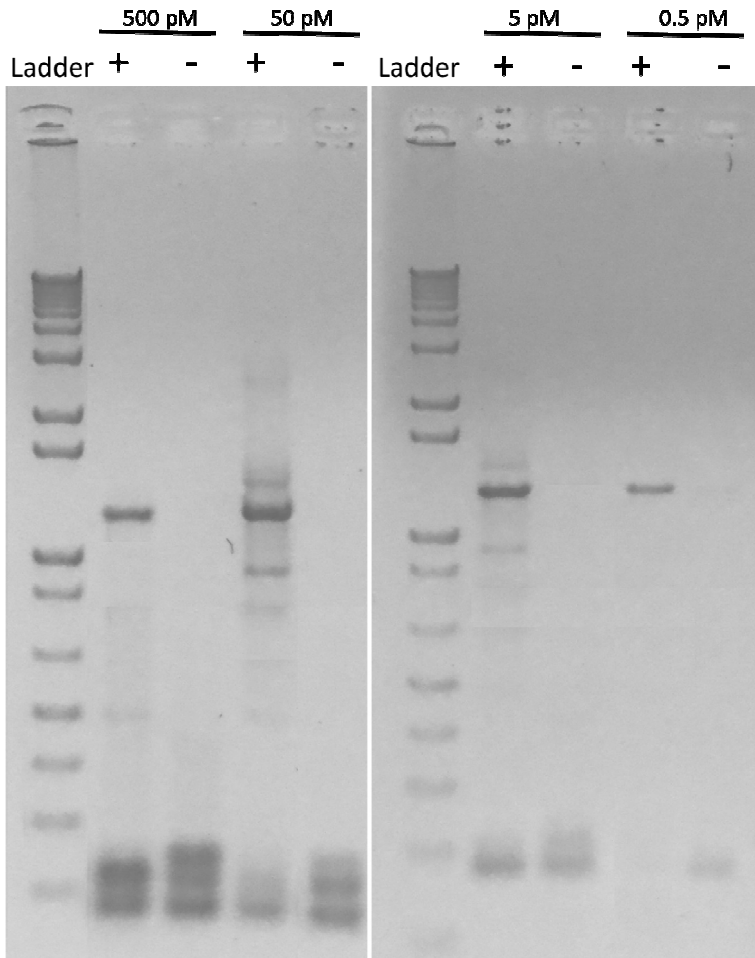


Supplementary Figure 4. Optimization of circularization by ligation of fusion PCR products. Two different length fusion PCR products of approximately 370 bp and 570 bp that were obtained from a 150 bp pre-LASSO probe with Long Adapters of 242 bp and 442 bp respectively. Fusion products (1 μg) with sticky ends (EcoRI digested) were diluted to 20 ng/μl and 0.2 ng/μl in 1X T4 DNA Ligase buffer and T4 ligated. After ligation, linear DNA was digested with exonucleases. DNA circles were column-purified, and run in a gel. In reactions containing 20 ng/μl of fusion PCR products, we observed concatemeric DNA circles (Lane 1 and 2). The circular nature of the DNA present in the bands was confirmed by the ligase negative controls where all DNA was completely digested by the exonucleases as expected (Lanes 3 and 4). No circular concatemers were visible in the gel when ligation was performed at 0.2 ng/μl (Lane 5 and 6).



- 1: Ampligase Ligase + Q5 Polymerase
- 2: Ampligase Ligase + OmniKlentaq
- 3: 9N Ligase + Q5 Polymerase
- 4: 9N Ligase + OmniKlentaq
- 5: 9N Ligase + Taq Polymerase
- Ladder: Quick-Load 100bp

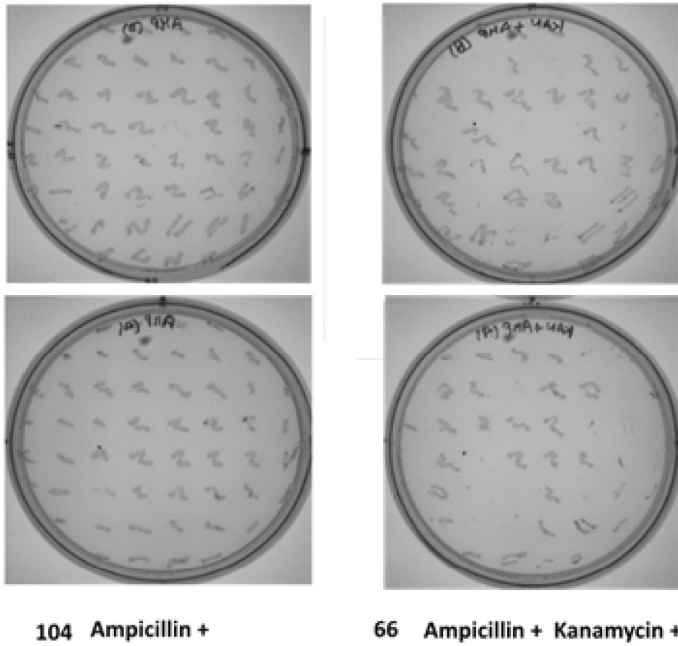
Supplementary Figure 5. Optimization of Gap Filling mix composition for single target capture using LASSO probes. The aim of this experiment was to compare different DNA polymerases and thermostable DNA ligases gap filling mix formulations in capturing a 100 bp target. Capture was performed by using a LASSO probe that was obtained by fusing a 150 bp pre-LASSO probe and a 242bp Long Adapter as described in Material and Methods.



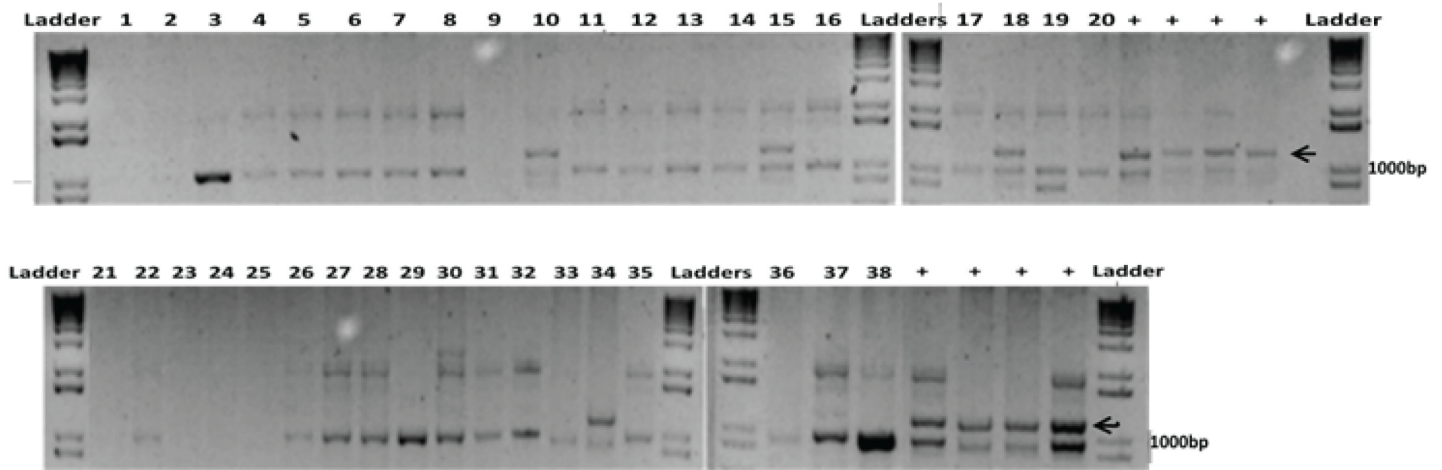
Supplementary Figure 6. Full blot of figure 2d

Post capture PCR of circles obtained by capturing a 1,038 bp target sequence within the M13Mp18 dsDNA (~500 fM) in presence of an equimolar (~500 fM) background of total genomic DNA of *E. coli*, using serial dilution of a LASSO probes. Negative controls contain sheared gDNA but no target.

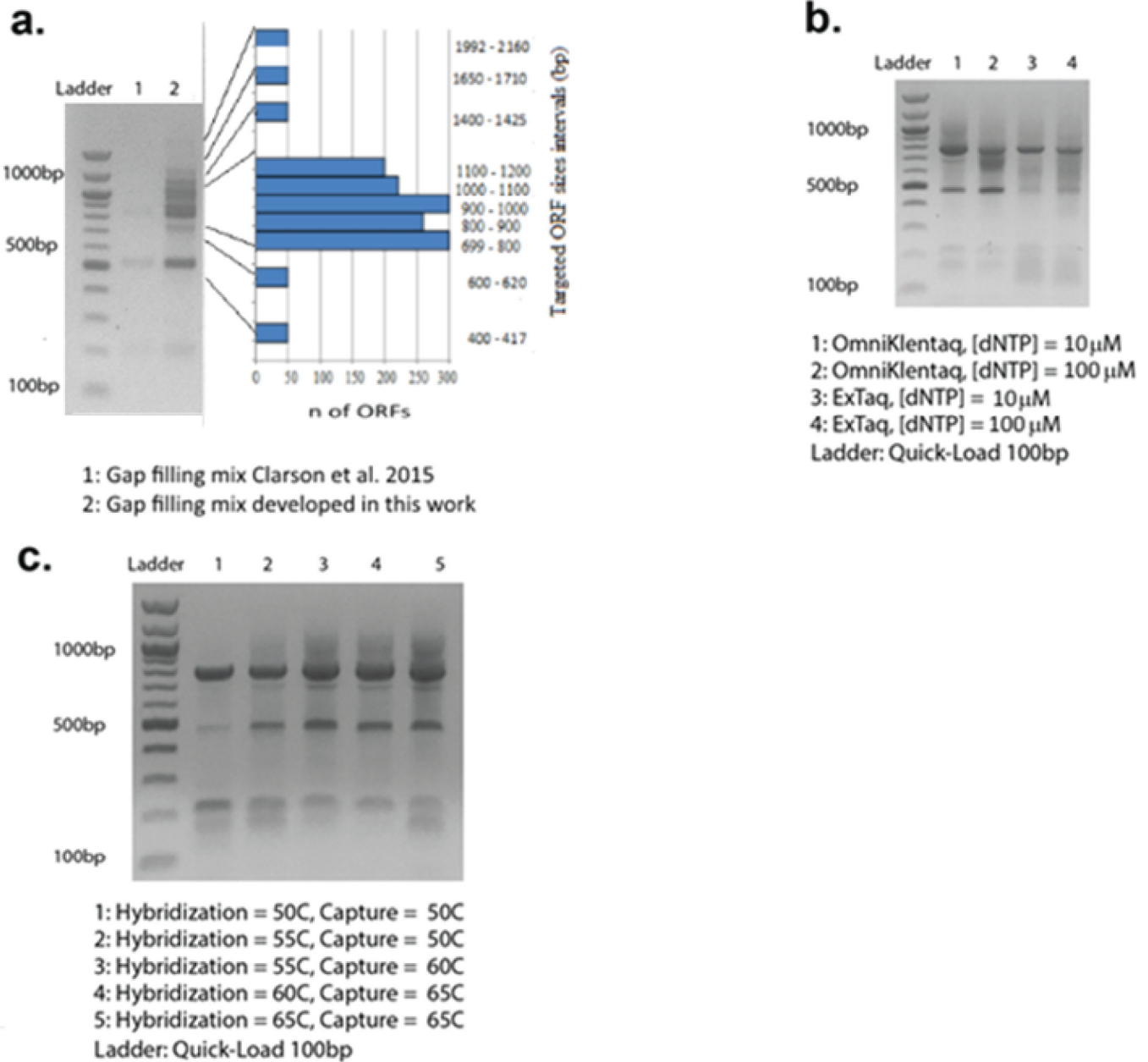
a.



b.



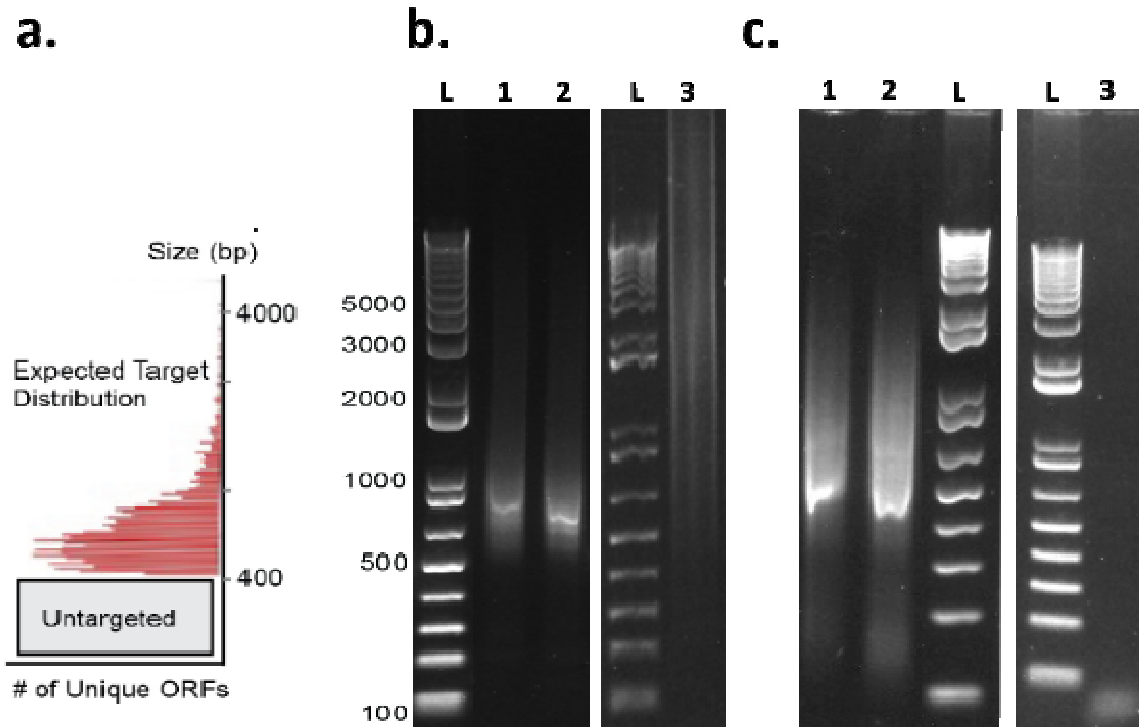
Supplementary Figure 7. Estimation of the percentage of functional captured KanR2 ORFs. A pET-21(+) expression vector (ampicillin resistance for selection) was linearized by PCR using tailed-primers with tails identical to the sequence of the primers we used in post capture PCR amplification. Post capture PCR of KanR2 was cloned in pET- 21(+) via Gibson Assembly. Transformation of BL21 kanamycin susceptible BL21 *E. coli* cells was performed by electroporation. (a) 104 *E.coli* transformant colonies were replica plated in ampicillin (100 µg/ml) selection agar plates and ampicillin (100 µg/ml) plus kanamycin (50 µg/ml) selection agar plates. 66 colonies were ampicillin and kanamycin resistant while 38 were ampicillin resistant and kanamycin susceptible. (b) Colony PCR of the 38 colonies to evaluate the presence of KanR2. Only 4 clones (Lanes 10, 15, 18, 34) contained the KanR2 inserts. Therefore the 34 empty clones were not considered in the estimation of the percentage of functional clones. In total 66 clones were kanamycin resistant, out of the 70 clones that contained the insert. 94% of the captured KanR2 ORFs were therefore considered functional.



Supplementary Figure 8. Optimization of different parameters for ORFeome capture

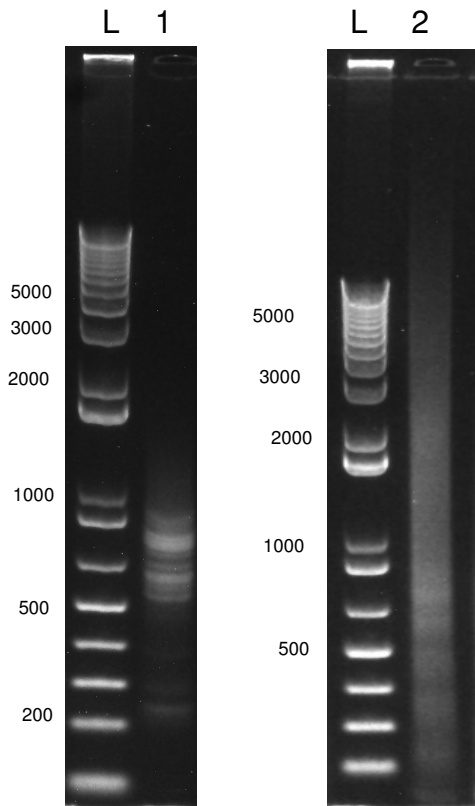
Capture experiments were performed by using a LASSO probe library assembled with Long Adapter 442 bp to capture 1,530 ORFs of *E.coli* K12 (pre-LASSO probe list available in **Supplementary File 1**). The 1,530 ORFs were selected according different intervals of ORF lengths. (a) The gap filling mix composition we developed for single target capture was tested together with the gap filling mix composition described by Carlson and coworkers (2015)⁴. Post capture PCR band pattern is in agreement with the expected ORF size distribution (Lane 2 and histogram). (b) Comparison of different post capture PCRs performed by testing Omni Klentaq (Enzymatics) or ExTaq Polymerase (TaKaRA) at different dNTP concentrations in the gap filling mix. NB. In 2015 Omni Kleantaq was discontinued by Enzymatics. We started purchasing the same enzyme from DNA Polymerase Technology, Inc. with the name of Omni Kleantaq LA. We found that we were able to obtain the same capture results by diluting it before adding it to the gap filling mix as described in Material and Methods. Our gap filling mix is composed of 0.025U/ μ l of Ampligase DNA Ligase in final capture volume. Different authors used much higher concentrations of Ampligase DNA Ligase in the final capture volume:

Brian J. O’Roak et al. (2012)³ 1 U/μl, Carlson et al. (2015)⁴ 3 U/μl, Jin Billy Li et al. (2009)⁵ 0.16 U/μl, Peidong Shen (2011)⁶ 0.25 U/μl. We investigated whether increasing the concentration of the Ampligase DNA Ligase up to 1 U/μl (maintaining Omni KlenTaq at 0.042 U/μl and dNTPs 10μM) could improve the capture efficiency. We noticed no differences in yield or band pattern (data not shown) indicating that 0.025 U/μl of Ampligase DNA Ligase in final capture volume was sufficient for capture. (c) PCRs post capture performed by testing different temperatures for hybridization and capture. The best patterns were obtained when both hybridization and gap filling were performed at 65 C.



Supplementary Figure 9. LASSO versus MIP Capture

(a) Histogram denoting the size distribution of the targeted *E.coli* K12 ORFs split into bin size of 40 bp. Reproduced from main text for convenience. (b) Gel electrophoresis of the capture of 3,108 ORFs of *E. coli* K12 performed by using LASSO-242bp, LASSO-442bp and MIP probe libraries. It should be noted that targeted ORFs will have an increase in 140bp of residual LASSO sequences once captured and run on a gel. Lane 1: LASSO-442bp library, post capture PCR. Lane 2: LASSO-242bp library, post capture PCR. Lane 3: MIP library, post capture PCR. (c) Gel electrophoresis of attB-capture amplicons for pDONR221 (Gateway) cloning of LASSO-242bp, LASSO-442bp and the MIP captures. Amplicons were obtained by re-amplification of the capture amplicons shown in lanes 1, 2, and 3 in (b) using attB-tailed primers. Lane 1: Re-amplification of LASSO-442bp post capture PCR using attB-tailed primers. Lane 2: Re-amplification of LASSO-242bp. post capture PCR using attB-tailed primers. Lane 3: Re-amplification of MIP post capture PCR using attB-tailed primers. L : TrackIt 1 Kb Plus DNA Ladder (Thermo Fisher)



Lane 1 Capture performed with LASSO-242bp library.

Lane 2 Capture performed with the MIP library.

L : TrackIt 1 Kb Plus DNA Ladder (Thermo Fisher)

Supplementary Figure 10. Cloning human microbiome-derived ORFs

Gel electrophoresis of amplicons obtained from LASSO-242bp and MIP captures when using as template for capture total DNA extracted from a healthy human donor's stool.

Supplementary Table 1. Sequences of Long Adapters
(Underlined sequences represent annealing positions for primers in PCR)

Oligo Name	Sequence
Long Adapter 242 bp	5'AGAGAAGTCCTAGCACGGTAACCTccgaggatgtcatcaaagagtttaagagtttatgagatttaag gtcaagatggaggaagcgtcaacggacacgagttcgagattgagggagaaggagaaggccggccttacgaggg cacacaaaccgctaagctcaaggtcacaaaaggaggaccctccccttctcctgggatattctgagccctcagttcca gtacggaagcGAATTCCAGCTT-3'
Long Adapter 442 bp	5'AGAGAAGTCCTAGCACGGTAACCTccgaggatgtcatcaaagagtttaagagtttatgagatttaag gtcaagatggaggaagcgtcaacggacacgagttcgagattgagggagaaggagaaggccggccttacgaggg cacacaaaccgctaagctcaaggtcacaaaaggaggaccctccccttctcctgggatattctgagccctcagttcca gtacggaagcacaagcctatgttaaacaccctgcccacatccctgactatctgaagctcctcctgaaggctcaagt gggagagattcatgaactcgaggacggagggcgtgacagtcacacaagatagcaccctcaggacggagaggt tattataaggtgaaactcagaggaaccaactcccctccgatggccctgtcatGAATTCCAGCTT-3'

Supplementary Table 2. Sequences of Primers

Primer Name	Sequence (5' to 3')
FusionBlaF	AGAGAAGTCCTAGCACGGTAACCTCCGAGGATGTCATCAAAGAG
RFP200EcoR1	AAGCTGGAATTCGCTTCCGTAAGGAACTGAGGGC
RFP400EcoR1	AAGCTGGAATTCATGACAGGGCCATCGGAGGGG
BLAF	GAGTATTACCGCGGCGAATTC
SubPool2kbF	CGGTGCTGACGATGCCGAATTC
Fusion2kbF	TGCTGCTTGGATGCGTTAAATGGTCCGAGGATGTCATCAAAGAG
TiolNew	A*T*C*GCCGCAAGAAGTGTU (* indicates phosphorothioate bonds, U indicate a deoxyuracil moiety.)
SapINew	GGTTCCTGGCTCTTCGATC
ICeul200CaptF	CTCCCCTTCTCCTGGGATATTCTG
PCR1kbCaptF400	GTGAAACTCAGAGGAACCAACTTCC
PCR1kbCaptR200	CGCTTCCCTCCATCTTGACCTTAAATCTCA
attB1Capt200F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAAGGAGATACCCAGTTCCAGTA CGGAAGCGAAT
attB1Capt400F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAAGGAGATACCCTCCGATGGCC CTGTCATGAAT
attB2Capt200R	GGGGACCACTTTGTACAAGAAAGCTGGGTTTAGGAGGTTACCGTGCTAGGACTT CTC
pET21RGibson	TCCTCTGAGTTTCACCGGATCCGCGACCCATTTGC
pET21FGibson	TCAAGATGGAGGGAAGCGAATTCGAGCTCCGTCGACAA

Supplementary Table 3. Sequences of Single Pre-LASSO probes

<i>Oligo Name</i>	<i>Sequence</i>
Pre-LASSO KanR2	GAGTATTACCGCGGCGAATTCATGAGCCATATTCAACGGGAAACGTCTTGCTCTA GGAACACTTCTTGCGGCGATAGAAGGTTCTGGCTCTTCGATCGCAGTTTCATTT GATGCTCGATGAGTTTTTCTAAAGAGAAGTCCTAGCACGGTAACC
Pre-LASSO 100bp	GAGTATTACCGCGGCGAATTCCTCAACGGCAGCAGCGGATCCGTGAACACTTCTT GCGGCGATAGAAGGTTCTGGCTCTTCGATCTGATTTATGGTCATTCTCGTTTTT AGAGAAGTCCTAGCACGGTAACC
Pre-LASSO 620bp	GAGTATTACCGCGGCGAATTCCTGGAGTTTGCTTCCGGTCTGGTTCGCAACACTT CTTGCGGCGATAGAAGGTTCTGGCTCTTCGATCGATTTGGGTAATGAATATCC GGTCTTGTC AAGAGAGAAGTCCTAGCACGGTAACC
Pre-LASSO 1kb	GAGTATTACCGCGGCGAATTCCTGGAGTTTGCTTCCGGTCTGGTTCGCAACACTT CTTGCGGCGATAGAAGGTTCTGGCTCTTCGATCGCCGTTGCTACCCTCGTTCC GATGCAGAGAAGTCCTAGCACGGTAACC
Pre-LASSO 2kb	GAGTATTACCGCGGCGAATTCCTGGAGTTTGCTTCCGGTCTGGTTCGCAACACTT CTTGCGGCGATAGAAGGTTCTGGCTCTTCGATCGGCTCTGAGGGTGGCGGTT CTGAGGAGAGAAGTCCTAGCACGGTAACC
Pre-LASSO 4kb	GAGTATTACCGCGGCGAATTCCTGGAGTTTGCTTCCGGTCTGGTTCGCAACACTT CTTGCGGCGATGGTTCCTGGCTCTTCGATCGGCGAATCCGTTATTGTTTCTCCC GATGTAAGAGAAGTCCTAGCACGGTAACC
pre-LASSO RPLPO	GAGTATTACCGCGGCGAATTCCTCACATTCCCCCGGATATGAGGCAACACTTCT TGCGGCGATGGTTCCTGGCTCTTCGATCACGATGTCACCTCCACGAGGACGCAG AGAAGTCCTAGCACGGTAACC
pre-LASSO TP53	GAGTATTACCGCGGCGAATTCgtctgagtcaggcccttctgtcttgaacatgagAACACTTCTTGC GGCGATGGTTCCTGGCTCTTCGATCcacagggcaggctctggccagttggcaaacatAGAGAA GTCCTAGCACGGTAACC

SUPPLEMENTARY REFERENCES

1. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
2. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
3. O'Roak B.J. et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **21**, 338 (2012)
4. Carlson KD, et al. MIPSTR: a method for multiplex genotyping of germline and somatic STR variation across many individuals. *Genome Res* **5**, 750-761 (2015)
5. Li JB et al. Multiplex padlock targeted sequencing reveals human hypermutable CpG variations. *Genome Res* **19**, 1606-15. (2009)
6. Shen P, et al High-quality DNA sequence capture of 524 disease candidate genes. *Proc Natl Acad Sci U S A.* **108**, 6549-54 (2011)