Massively parallel DNA target capture using Long Adapter Single Stranded Oligonucleotide (LASSO) probes assembled through a novel DNA recombinase mediated methodology

LORENZO TOSI[1], Lamia Chkaiban[1], H. Benjamin Larman[2], Jeffrey Rosenfeld[3], and Biju Parekkadan[1]

[1]Rutgers The State University of New Jersey
[2]Johns Hopkins University
[3]Robert Wood Johnson Medical School Department of Neuroscience and Cell Biology

May 11, 2021

Abstract

In the attempt to bridge the widening gap from DNA sequence to biological function, we developed a novel methodology to assemble Long-Adapter Single-Strand Oligonucleotide (LASSO) probe libraries that enabled the massively multiplexed capture of kilobase-sized DNA fragments for downstream long read DNA sequencing or expression. This method uses short DNA oligonucleotides (pre-LASSO probes) and a plasmid vector that supplies the backbone for the mature LASSO probe through Cre-Loxp intramolecular recombination. This strategy generates high quality LASSO probes libraries (~46% of probes). We performed NGS analysis of the post-capture PCR amplification of DNA circles obtained from the LASSO capture of 3087 E.coli ORFs spanning from 400- to 4,000 bp. The median enrichment of all targeted ORFs versus untargeted ORFs was 30 times. For ORFs up to 1kb in size, targeted ORFs were enriched up to a median of 260-fold. Here, we show that LASSO probes obtained in this manner, are able to capture full-length open reading frames from total human cDNA. Furthermore, we show that the LASSO capture specificity and sensitivity is sufficient for target capture from total human genomic DNA template. This technology can be used for the preparation of long-read sequencing libraries and for massively multiplexed cloning of human sequences.

Massively parallel DNA target capture using Long Adapter Single Stranded Oligonucleotide (LASSO) probes assembled through a novel DNA recombinase mediated methodology

Lorenzo Tosi[1], Lamia Chaikban[1], Benjamin Larman[2], Jeffrey Rosenfield[3,4], Biju Parekkadan[1,3*]

[1] Department of Biomedical Engineering, Rutgers University, Piscataway, New Jersey 08854, USA

[2] Institute of Cell Engineering, Division of Immunology, Department of Pathology, Johns Hopkins University, Baltimore, MD, USA

[3] Cancer Institute of New Jersey, New Brunswick, New Jersey 08854, USA

[4] Department of Pathology, Robert Wood Johnson Medical School, New Brunswick, NJ 08903, USA

*Correspondence and requests for materials should be addressed to B.P. (biju.parekkadan@rutgers.edu; 599 Taylor Road, Piscataway, NJ 08854)

Key words : PCR, LASSO probe, Cre-Lox, DNA target capture

list of abbreviations :

LASSO, Long-Adapter Single-Strand Oligonucleotide
MIP, molecular inversion probe
NGS, next generation sequencing
ORF, open reading frame
PCR, polymerase chain reaction
pLASSO, LASSO plasmid
RPKM, Reads per kilo base per million mapped reads

Abstract

In the attempt to bridge the widening gap from DNA sequence to biological function, we developed a novel methodology to assemble Long-Adapter Single-Strand Oligonucleotide (LASSO) probe libraries that enabled the massively multiplexed capture of kilobase-sized DNA fragments for downstream long read DNA sequencing or expression. This method uses short DNA oligonucleotides (pre-LASSO probes) and a plasmid vector that supplies the backbone for the mature LASSO probe through Cre-Loxp intramolecular recombination. This strategy generates high quality LASSO probes libraries (~46% of probes). We performed NGS analysis of the post-capture PCR amplification of DNA circles obtained from the LASSO capture of 3087 E.coli ORFs spanning from 400- to 4,000 bp. The median enrichment of all targeted ORFs versus untargeted ORFs was 30 times. For ORFs up to 1kb in size, targeted ORFs were enriched up to a median of 260-fold. Here, we show that LASSO probes obtained in this manner, are able to capture full-length open reading frames from total human cDNA. Furthermore, we show that the LASSO capture specificity and sensitivity is sufficient for target capture from total human genomic DNA template. This technology can be used for the preparation of long-read sequencing libraries and for massively multiplexed cloning of human sequences.

1. Introduction

Advances in DNA sequencing have led to an exponential increase in the quantity of sequence data. Databases now contain the genomes of hundreds of plants and tens of thousands of microorganisms. Despite the availability of these data, there is still a gap in understanding the function of genes within a genome. Massively parallel technologies that enable the synthesis and cloning of long DNA sequences are thus important in linking sequence to function. [1]

The recent development of multiplexed functional assays allows for the rapid testing of thousands to millions of sequences across a wide array of biological functions. [2, 3, 4, 5, 6, 7, 8, 9]The DNA sequences of interest may be obtained by genome fragmentation[10], mutagenesis of existing sequences[11] or direct synthesis of oligonucleotides (oligos).[12] Direct oligo synthesis allows for testing of controlled hypotheses against one another without the constraints of natural variation or mutagenesis. However, individual oligos are generally shorter than 200 nucleotides (nt), limiting potential applications. [13] Gene synthesis from oligo libraries can be used to extend these lengths[14, 15], but the high cost of individual assembly and processing becomes prohibitive for large gene libraries.[16, 17] A number of alternative methods for multiplexed gene synthesis have demonstrated the assembly of hundreds to thousands of short fragments. [18, 19, 20]However, those methods are limited in achievable maximum gene length (<800 bp) and produce highly-biased libraries with constraints on sequence homology. [37]

For high throughput functional studies that involve natural DNA sequences of interest, a valuable alternative to de novo DNA chemical synthesis, is the selection of DNA targets from a natural source. PCR has been used for 30 years to select a DNA sequence of interest from a DNA template. [21] However, traditional PCR or multiplexed PCR are generally not feasible for massive parallel DNA target selection because of non-specific amplification caused by interaction between the primers. [22, 23, 24] A different approach to primer design and DNA target capture to enable greater specificity is the use of molecular inversion probes (MIP).[25] MIPs are short single-stranded DNA (ssDNA) molecules (~150 bp), that contains two annealing sites at the strand ends (the ligation and the extension arms), which are complementary to the target sequence.

Upon hybridization of the MIP to its target, the 5' and 3' DNA ends become adjacent and available for an intramolecular ligation reaction. To adapt this method to perform exon capture in combination with next generation sequencing, a DNA polymerase can be used to 'gap-fill' between target-specific MIP sequences designed to flank a full or partial exon, before ligase-driven circularization, thereby capturing a copy of the intervening sequence.[26, 27, 28] Uncircularized species are digested by exonucleases to reduce background, and circularized species are PCR amplified via primers directed at the common linker. It has been shown that MIPs are capable of massive parallel enrichment of short genomic regions. [28, 29] However, MIPs are inefficient at capturing larger target sequences because of the short length of the linker region. Increasing the length of the MIPs linker has been shown to allow the capture of longer targets. [30, 31, 32] The scalability of this long linker method was limited though, due to a requirement of a separate PCR reaction for each individual probe.

To overcome the scalability limitation of producing longer MIPs, we developed a method that allows the production of thousands of targeted probes, that are essentially long-linker MIPs, in the same reaction tube. [33] The assembly method of these Long-Adapter Single-Strand Oligonucleotides (LASSOs) was based on the fusion of a probe precursor (pre-LASSO that contains the ligation and extension arms) with a conserved linker sequence (Long Adapter) by PCR. The fusion PCR amplicon was subsequently circularized by intramolecular ligation and subjected to inverse PCR, so that the LASSO annealing arms were made to flank the long-adapter sequence in the final configuration. Despite the functionality of a LASSO library in cloning a near complete bacterial ORFeome, in a later study we found that the LASSO library we used for the capture was composed only by ~10% correctly assembled probes while the rest of the probes were discordant (with arms that originate from different probes).[34]

The purity (number of sequence-correct probes) and quality (higher percentage of correctly assembled probes) of mature LASSO libraries can undoubtedly impact capture efficiency of targeted genome regions. For highly complex eukaryotic genomes, including human application, high purity of a mature LASSO probe library is likely a stringent requirement. To address these issues, we developed a different highly scalable approach to assemble LASSO probes based on a cloning and recombination strategy. The probe precursor pre-LASSO are obtained as a short (160bp) DNA oligo pool, which is incorporated into a custom plasmid pLASSO and transformed into E.coli. Site specific recombination of two loxP sites oriented in the same direction in pLASSO, produces the excision of a DNA mini-circle that contains the mature LASSO precursor already in the final configuration thus avoiding the formation of discordant probes of the previous method.

We found that the excision of the mini-circle was enhanced when we used the uncoiled form of pLASSO obtained by nicking as substrate for the recombination. This novel assembly strategy produces LASSO libraries with a much higher fraction of correctly assembled LASSO probes with a consequent improvement in the capture efficiency and pave the way for LASSO probe in human applications.

2. Material and methods

All DNA oligonucleotides, pre-LASSO probes and primers sequences are available in **Supplementary Table 1**

2.1. pLASSO vector generation and linearization

pLASSO vector was derived from the linear pLox2+ (NEB). 50 ng of pLox2+ were circularized by adding 1 µl of T4 DNA ligase (NEB), 1X T4 DNA ligase buffer, nuclease-free water to 25 µl total volume and incubated overnight at 16 0C. The ligation reaction (0.5µl) was used to transform 5-alpha chemically competent *E. coli* cells (NEB).*E.coli* colonies were collected from ampicillin agar plates after overnight incubation. Single colonies (four) were inoculate in 5 ml corning tubes containing Ampicillin LB medium and shacked at 200 RPM ON at 37 0C.

The cells were pelleted and subjected to plasmid extraction by using the PureLink Quick Plasmid Miniprep Kit (Invitrogen) as described by the vendor and checked for the presence of pLox2+ by running 2µl of the plasmid miniprep in a 0.8% agarose gel with Sybr Green. The pLox2+ plasmids (500ng) was digested with

20 units of EcoRI enzyme (NEB) and 5 units of Alkaline Phosphatase (NEB) in 25µl of 1X CutSmart buffer (NEB), Incubate in the thermal cycler at 370C for 1h and heat inactivate at 800C for 10min. The digested pLox2+ (200 ng) was run on a Sybr Green agarose gel (1%), visualized at the blue light and the DNA band correspondent with the expected 2866 bp fragment was cut from the gel with a scalpel and DNA extracted using Monarch DNA Gel Extraction Kit (NEB).

100ng of the synthetic dsDNA fragment Backbone (gBlocks, IDT) were digested with 20 unit of EcoRI restriction enzyme in 25ul of 1X CutSmart buffer at 37C for one hour and purified by using "DNA Purification SPRI Magnetic Bead as described by the vendor. The EcoRI digested Backbone (10ng) were mixed with 50ng of the the 2866 bp fragment obtained from the EcoRI digestion of pLox 2+ in 25µl volume of 1X T4 DNA ligase buffer and n units of T4 DNA ligase. Ligation was performed overnight at 16°C. 0.5 µL of the ON ligation were used for transformation of 5-alpha chemically competent E. coli cells NEB and plated on an Ampicillin resistance selective agar plate. Colonies were harvested from Ampicillin agar plates and grown overnight in LB broth (5 ml). The pLASSO plasmid was extracted with PureLink Quick Plasmid Miniprep Kit as described by the vendor from pelleted broth cultures. The identity of pLASSO was verified cutting with single or combination of the SalI, BamHI, SwaI and EcoRI restriction enzymes and checking expected DNA band sizes on agarose gel according with pLASSO map (**Supplementary material 1**).

Finally, pLASSO plasmid was linearized by performing an inverted PCR in a 25µl of 1X Kapa Hi Fidelity Buffer, pLASSO (0.5ng), dNTPS (0.3 mM) and 0.5 units of Kapa Taq DNA polymerase, and Linearization primers (0.3 µM). NEB1F and NEB1R. The thermal profile was: initial denaturation 4min at 95°C, (95°C for 20 sec, 60°C for 20 sec, 2 min at 72°C) for 28 cycles, and final extension 3 min at 72°C

2.2. pre-LASSO probes

The pre-LASSO probes were 158-bp long. The E.coli pre-LASSO probe library was purchased as ssDNA oligo pool from Twist Bioscience. The positive control pre-LASSO 1kbM13 was obtained from IDT ad as dDNA (gBlock). The DNA sequences of pre-LASSO 1kbM13 is available in (**Supplementary Table 2**). The ssDNA Twist oligo pool (20 ng) was PCR amplified using selector primers Sap1F and BamH1R. The PCR was performed in a 25µl of 1X Kapa Hi Fidelity Buffer with pre-LASSO probe library (4ng), dNTPS (0.3 mM) and 0.5 units of Kapa Taq DNA polymerase, and Selector primers (0.3 µM) aF. The PCR thermal cycle was 3min at 95°C, (98°C for 20 sec, 58°C for 15 sec, 1 min at 72°C) for 8 cycles. The correct size of the amplicon (˜160bp) was verified by loading 3 µl of the PCR volume on a 2% agarose gel with ethidium bromide and using Low Molecular Weight Ladder as reference . For the single pre-LASSO probe 1kbM13 and for the 3,108 E. coli K12 ORF pre-LASSO library subpool, the design was: 5´ CAGACGACGGCCAGTGTCGAC,Ligation Arm, AACACTTCTTGCGGCGATGGTTCCTGGCTCTTCGATC, Extension Arm, GGATCCTACGGTCATTCAGC 3´

The ORFs of the E. coli K12 genome that were longer than 400 nucleotides were targeted with ligation and extension arms positioned at the beginning and end of the sequences, respectively, and extended until the desired melting temperature was reached. Specifically, the algorithm first selected the ORFs leading and trailing 32-nucleotide sequences for the two arms at melting temperatures for the ligation and extension arms of between 65 °C and 85 °C and 55 °C and 80 °C, respectively. If at least one of these conditions were not satisfied, the algorithm increased the length of the arms by one nucleotide and re-tested the conditions until they were satisfied or until the end of the ORF was reached. Because probes had non-homogeneous lengths (depending on the Tm of ligation and extension arms) up to a maximum length of 158bp, we extended all probes to 158bp by adding random oligonucleotides in between the primer selector annealing site and the extension arm.

Since a SwaI digestion step was used to assemble the LASSO probes, the algorithm discarded the design of pre-LASSO probes where a Swa1 restriction site was present in the ligation or extension arm. A full list of the 3089 ORFs with valid ligation and extension arms and their corresponding pre-LASSO probes is reported in the **Supplementary Table 2** .

2.3. LASSO probe assembly

Pre-LASSO cloning in pLASSO . The fusion of the pre-LASSO probe in pLASSO vector was performed by using NEBuilder HiFI DNA assembly (NEB). By adjusting with PCR grade water to 20 µl total volume the following components: Linearized pLASSO (˜50 ng), Pre-LASSO probe pool ˜16 ng, Linearized pLASSO ˜50 ng, NEBuilderHiFi DNA Assembly Master Mix 2X 10µL. The solution was Incubated in a PCR thermal cycler at 50°C for 15 minutes. The NEBulder reaction 1µl was added to 50 µl of electrocompetent 5-alpha E. coli cells (NEB) into a chilled 1 mm cuvette (BioRad) and electroporated using a Micro Pulser (BioRad). The transformed cells were recovered in 950µl optimal broth medium and were all plated on a 150 mm petri dish containing 100 µg ml^{-1} ampicillin. The numbers of single colonies were estimated by serial dilutions.

Depending on the number of colonies obtained, the electroporation was repeated n times in order to obtain approximately 10 times coverage of the E.coli library (˜30k colonies)

All colonies from the petri dishes were recovered, resuspended in 5 ml LB medium and subjected to plasmid extraction (Miniprep Kit, Qiagen). The DNA concentration was measured with a Nanodrop.

The successful cloning of pre-LASSO library in pLASSO was verified by digesting in 50µl total volume of 1X CutSmart Buffer, 500ng of the recovered pLASSO library with 20 units of SalI and 20 units BamHI. Digestion was performed for 1h at 37°C. The digestion solution (4µl) was loaded on a 2% agarose EtBr gel and subjected to electrophoresis and observed with a trans UV. If DNA band correspondent with the size of the pre-LASSO library (˜160bp) was present in the lane, the pre-LASSO library was considered successfully cloned in pLASSO.

Cre-recombination. The pLASSO library (2µg) was subjected to nicking endonuclease digestion in a PCR tube in 50µl of 1X Cut Smart Buffer with 20 units of Nt.BbvCI nicking endonuclease (NEB). The digestion was performed in a PCR thermal cycler for 1h at 37°C and heat inactivated for 10 min at 80°C. The nicked unpurified pLASSO library (250ng, correspondent to 6.25µl) were directly added into a PCR tube containing 37.75 µl of nuclease free water, 5µl of CRE recombinase buffer (NEB) and 1µl of Cre-recombinase (ABCAM) added last. The recombination reaction was performed at 37°C for 30 min and heat inactivate at 700C for 10min. The temperature was then lowered to 25°C and 1µl (20 units) of SwaI resctiction enzyme (NEB) were added directly into the recombination solution. The digestion was carried on for 1h at 25°C and heat-inactivated at 70°C for 10 min. At this point, the temperature was lowered again to 37 °C and 2µl ATP 10mM and 1µl (10 units) of Exonuclease V (NEB) were added into solution, mixed by pipetting up and down, incubated at 370C for 30min and heat-inactivate at 700C for 30min.

Inverted PCR . Inverted PCR was performed in a 50 µl total volume containing 10 µl of the unpurified Cre-recombination-digestion reaction as described above, 10 µl of 5x KAPA HiFi Fidelity Buffer, 0.3 mM each dNTP, and 0.3 µM reverse primer ThioIR and 0.3 µM forward primer Sap1F and 1 µl of KAPA HiFi HotStart DNA Polymerase. Both SapI and ThiolR anneal with opposite orientations in the inverted PCR primer annealing site central section of the pre-LASSO probe (AACACTTCTTGCGGCGATGGTTCCTG-GCTCTTCGATC). The first three bases of ThioIR were phosphorothioate bonds to prevent digestion during subsequent ExoI treatment, and a 3' -terminal uracil base for subsequent primer removal using Uracil-DNA Glycosylase (USER enzyme); Sap1F includes the SapI restriction site (Type IIS side-cutting restriction enzyme) for primer removal via digest with the isoschizomer BspQI. The PCR thermal profile was initiated for 3 min at 95 °C, followed by 18 cycles of 20 s at 98 °C, 15 s at 60 °C and 60 s at 72 °C, and then 4 min at 72 °C.

The size of the inverted PCR product was purified using AMPure beads (0.8×), washed with 80% ethanol twice and eluted with 25 µl of nuclease-free water. The concentration of purified inverted PCR product was measured by Nanodrop.

Production of mature LASSO probes . Approximately 200 ng of purified inverted PCR product was digested by adding in a total volume of 25 µl, 2.5 µl of 10× CutSmart buffer (NEB) and 1 µl of BspQI restriction enzyme (NEB). Digestion was performed at 50 °C for 1 h followed by 20 min at 80 °C. After digestion, 1 µl (five units) of Lambda Exonuclease was added directly to the BspQI digested DNA and incubated for 30 min at 37 °C followed by 10 min at 80 °C. At this point, 2 µl (1 unit $µl^{-1}$) of USER enzyme (NEB) was added

in solution and incubated for 30 min at 37 °C.

2.4. DNA templates used in capture experiments

For the LASSO probe capture experiments sensitivity test and positive controls for capture, we used a 7,249 bp ssDNA isolated from the M13mp18 phage (NEB). For the capture experiments of the E. coli ORFeome by LASSO probes, total genomic DNA of the E. coli strain K12, substrain W3110 (Migula) Castellani and Chalmers (ATCC 27325) was extracted from 500 µl of LB broth (Sigma Aldrich) overnight culture using a Charge Switch genomic DNA Mini Bacteria Kit (Life Technology). Sheared total genomic DNA of E. coli K12 was obtained by sonicating 1 µg of total DNA in a volume of 200 µl in a 1.5 ml Eppendorf tube on ice using a Branson sonifier 450 (VWR Scientific) at an output control of 2 and a duty cycle of 50% for 40 s.

2.5. E. coli ORFeome capture

The LASSO libraries were hybridized on E. coli genomic DNA. The hybridization was performed in 15 µl of 1× Ampligase DNA Ligase Buffer (Epicentre) containing 250 ng of unshared E. coli K12 total genomic DNA and 5 ng LASSO probe library. In the hybridization volume, the concentration of *E. coli* chromosomes was approximately 10 pM. The concentration of individual LASSO probes was approximately 14 pM (44 nM for the complete LASSO). The solution (15 µl) containing the MIP or LASSO probe pool and the *E. coli* DNA was denatured for 5 min at 95 °C in a PCR thermocycler (Eppendorf Mastercycler), then incubated at 65 °C for 60 min. After hybridization, 5 µl of the freshly prepared gap filling mix was added into the hybridization solution while maintaining the reaction at 65 °C in the thermal-cycler. Gap filling and ligation was performed for 30 min at 65 °C. After capture, the DNA samples were denatured for 3 min at 95 °C, and the temperature was reduced to 37 °C. Next, 4 µl of linear DNA digestion solution was added immediately. Digestion was performed for 1 h at 37 °C, followed by 20 min at 80 °C.

The gap filling mix composition for 1ml volume of gap filling mix was as follows: 791 µl of PCR grade water, 100 µl of glycerol, 4 µl of 10 mM dNTPs, 1 µl of Ampligase DNA Ligase (100 U µ l-1), 4 µl of Omni Klentaq LA, 100 µl of 10× Ampligase DNA Ligase Buffer.. This gap filling mix could be stored at -20°C for up to three months.

The linear DNA digestion solution (volume: 48 µ l) was composed of 24 µl of nuclease-free water, 6 µl of Exonuclease I (20 units µ l^{-1}), 6 µ l of Exonuclease III (100 units µ l^{-1}), 6 µl of Lambda Exonuclease (5 units μl^{-1}) and 6 µ l of Exonuclease VII (10 units μl^{-1}) (all from NEB).

2.6. Capture of human ORFs using single LASSO probes from human cDNA .

To capture of human ORFs, was performed using 50 ng Human Reference cDNA (Zyagen) and 0.02 ng of LASSO probe β -Actin and LASSO probe GAPDH (pre- LASSO probe sequences available in **Supplementary Table 1**). Hybridization and capture were performed as described above and the sizes of the captured ORFs were verified by gel electrophoresis. To confirm the identities of the captured ORFs, the captured GAPDH and β -Actin were cloned into a pMiniT vector (NEB) using a NEB PCR cloning kit. The purified colony PCR products from the single transformants containing or RPLP0 in pMiniT were analyzed by Sanger sequencing.

2.7. Post-capture PCR

The captured ORFs were amplified using 10 µl of the capture reaction containing DNA circles in 50 µ l of PCR master mix composed of 5 ul of 10 x Klentaq DNA Polymerase Buffer, 50 µl of Omni Klentaq LA, 200 µ M of dNTPs and 0.4 µ M of primers CaptF and CaptR. The PCR thermal profile was initiated for 3 min at 95 °C, followed by 25 cycles of 20 s at 98 °C, 15 s at 60 °C and 2 min s at 72 °C, and then 4 min at 72 °C. To visualize the amplicons derived from the circles, 3 µ l of the PCR product was loaded in a 1.1% agarose gel containing Ethidium Bromide (Sigma) (0.2 µ g ml-1).

2.8. Next-generation sequencing of LASSO libraries

Paired-end NGS for inversion PCR products Non-sheared Inversion PCR products were sent for paired-end

NGS as per manufacturer's protocol (2 × 150 bp NextSeq 550 System Sequencing Illumina). Raw paired end sequencing outputs (R1 and R2 files) from Illumina NextSeq were processed and analyzed using a computational pipeline to quantitate distribution of properly formed (concordant) probes over improperly formed (discordant) probes depending on how read pairs align

to the probe library reference sequences (N = 3087). Primer sequences were removed using Trimmomatic-0.36. [35] The reads were then aligned to the LASSO design sequences using the BLAT tool. [36] The results were tallied using standard linux tools.

2.9. Next-generation sequencing of the capture libraries

Illumina library construction. Post-capture PCR products (16 µl) were enzymatically shared by adding 2 µl 10X Fragmentase Reaction Buffer v2 and 2 µl NEBNext dsDNA Fragmentase. The sheared DNA (22 µl) was subjected to end repair, 5' phosphorylation, dA-tailing and Illumina adaptor ligation using the NEBNext Ultra DNA Library Prep Kit for Illumina (NEB) as described by the vendor. PCR enrichment of adaptor ligated DNA was performed using NEBNext Multiplex Oligos (NEB) with index primers. The thermal profile was 30 s at 98 °C, followed by eight cycles of 10 s at 98 °C, 75 s at 63 °C and, 5 min at 72 °C. The PCR products were finally purified using the Agencourt AMPure XP system as described in the NEB protocol. The quality of the Illumina library was verified by checking the size distribution on an Agilent Bioanalyzer using a high-sensitivity DNA chip. The concentration of the Illumina library was measured by qPCR using the NEBNext Library Quant Kit for Illumina (NEB). DNA sequencing was performed using Illumina NextSeq instrument and standard reagents (Illumina).

Next-generation sequencing computational analyses. To check whether the sequencing and sample preparation went well, we did a quality check on the raw data using the FastQC tool (version 0.11.5). Low-quality read trimming, along with adapter clipping, was performed using Trimmomatic version 0.36. [35] The resulting fastq files from the trimmomatic output were then mapped against a reference genome sequence using Bowtie 2 version 2.4.1 (ref. 15). The reference genome used was E. coli K12 Using SAMtools. version 0.1.19-44428cd (ref. 16), we filtered the reads to include only those satisfying mapping quality scores of at least 30 and then sorted the resulting bam file. Since the probes were made for genes that satisfy the requirements of the current protocol, we considered these genes as targets. The rest of the genes, along with the intergenic regions, were considered as non-targets.

We then used bed tools version 2.24.0 (ref. 17) to estimate the depth of the regions. The depth of coverage was normalized to the length of the target to allow for accurate comparison between targets of different size. Matches between the reads and the LASSO probes were determined by using BLAT v35x1. [35]

2.10. Statistical analysis

The data are presented as mean or median ± s.e.m., as stated in the figure legends. Statistical significance was assessed using a Student's t-test for pair-wise comparison, and one-way ANOVA for comparisons between multiple ([?] 3) conditions. $P < 0.05$ was considered significant15,17.

Data availability. The authors declare that all data supporting the findings of this study are available within the paper and its Supplementary Information. Next-generation sequencing raw data of the captures performed with the LASSO are available on the NCBI Short Read Archive: https://www.ncbi.nlm.nih.gov/sra under project ID: PRJNA693554

3. Results

3.1. Design of starting components for LASSO probe synthesis

The major innovation in this new method of LASSO production was the introduction of a plasmid-mediated process to amplify and reconfigure a LASSO probe library with improved control to minimize unwanted byproducts of probe self-circularization. The assembly procedure we developed begins with the two main components: a precursor probe (pre-LASSO probe, **Fig. 1a** .) that is a ssDNA oligonucleotide (158bp) which contains the ligation and extension arms designed to hybridize with sequences that flank the targeted

region and a plasmid vector which we refer to as pLASSO (**Fig. 1b.**). The role of the pLASSO plasmid is to supply the backbone (in blue) for a mature LASSO (**Fig 1c** .) and a number of functional sites required for the assembly of a mature LASSO. The pLASSO vector was obtained in house starting from the pLox2+ linear plasmid (NEB) and a synthetic DNA fragment (backbone) as described in the material and methods. Uniquely, this plasmid was customized to have two LoxP sites (purple triangles) for Cre-recombination and a linearization primer annealing site for linearization. The plasmid also contained an ampicillin-resistance gene for selection in *E.coli.* In order to start the assembly of the LASSO probe, pLASSO needs to be linearized by PCR with linearization primers that have tails a and b identical to the a and b ends (primer selector annealing sites) of the pre-LASSO probe.

Hosted file

`image1.emf` available at https://authorea.com/users/413259/articles/521655-massively-parallel-dna-target-capture-using-long-adapter-single-stranded-oligonucleotide-lasso-probes-assembled-through-a-novel-dna-recombinase-mediated-methodology

Hosted file

`image2.emf` available at https://authorea.com/users/413259/articles/521655-massively-parallel-dna-target-capture-using-long-adapter-single-stranded-oligonucleotide-lasso-probes-assembled-through-a-novel-dna-recombinase-mediated-methodology

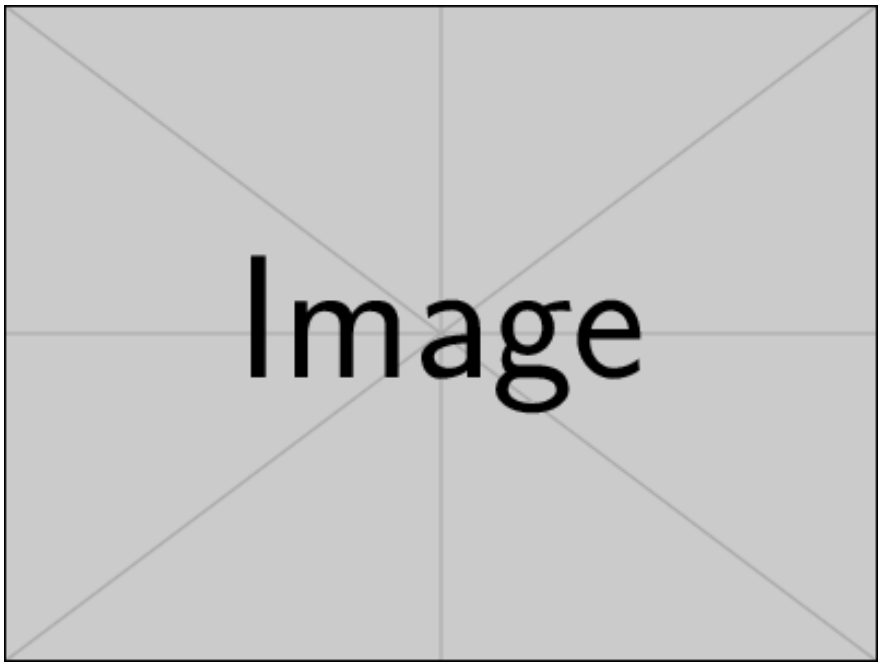

Figure 1, Schematic of starting components for DNA recombinase mediated LASSO probe synthesis. a, pre-LASSO probe that contains the ligation and extension arms (˜25bp) identical to the 5' and 3' ends of the targeted DNA sequence (ORF).**b,** pLASSO vector that contains two loxp sites oriented in the same direction, origin of replication, ampicillin resistance and primer annealing sites for linearization. **c,** Final Mature ssDNA LASSO probe. Ligation and extension arms are linked to the Backbone sequence (blue) that was obtained from pLASSO by a series of processing reactions.

3.2. Cre-recombinase mediated assembly of LASSO probe libraries

In order to assess the effectiveness of our novel DNA recombinase mediated assembly methodology in pro-

ducing complex LASSO probe libraries, we performed a comparison to our previous work. In this previous project, we had designed pre-LASSO probes to target 3164 ORFs from E.coli genomic DNA (Tosi et al 2017). We maintained the previous central inverse PCR primer annealing site while we modified the previous pre-LASSO probe DNA ends for entry in pLASSO vector (**Fig. 1 a,**a and b termini). Analogously to what was reported by Tosi et al. (2017), out of the ˜ 4000 E.coli K2 annotated ORFs we removed pre-LASSO probes corresponding to ORF targets smaller than 400 bp as a precaution to avoid potentially skewing our capture library during its subsequent PCR amplification and we also removed additional 160 probes that targeted different capture targets' lengths as negative controls. Adjusting the thresholds for target length, melting temperature or the length of the ligation / extension arms determines the number of acceptable probes to 3078. As a result of these filters, approximately 22.5% of the E. coli K12 ORFeome (900 ORFs) was thus left untargeted and is used as an internal, negative control for our experiments. The ssDNA pre-LASSO library was obtained as a pre-LASSO probe oligo pool with each oligo at a size of ˜ 160bp. The E.coli LASSO probe library was assembled as shown in **Fig. 2** . The pre-LASSO ssDNA oligo pool was converted into dsDNA format by PCR (**Fig. 2a**), and cloned in the linearized pLASSO (**Fig. 2b**). The pLASSO library obtained was expanded by transformation in E.coli cells and harvested from antibiotic selective agar plates (**Fig. 2b)** . The presence of the pre-LASSO library in the pLASSO plasmids was verified by performing double digestion with SalI and BamHI restriction enzymes (**Fig. 1b**). Gel electrophoresis (**Fig. 2c**) showed a ˜ 170bp band indicating the pre-LASSO library was successfully cloned in pLASSO.

The library was subsequently treated with Cre recombinase enzyme to create DNA mini-circles with LASSO probe precursors Interestingly, this recombination could not occur in the native supercoiled state of the plasmid library and required relaxation of the structure. Plasmids were converted in the relaxed form by digesting with the nicking endonuclease Nt.BvC1 that produces a DNA nick in correspondence of a restriction site located in the backbone (**Fig. 2 d**). The products of Cre-mediated recombination at loxP sites are DNA minicircles containing the pre-LASSO probes and 2.7 kb DNA circles containing the remaining part of pLASSO (**Fig. 2 e**). The agarose gel electrophoretic run of the recombination reaction (**Fig. 2 f** illustrates successful formation of the expected DNA minicircles (orange arrow) together with the 2.7 kb circular DNA remaining parts of pLASSO (green arrow). Since Cre-mediated recombinase enzyme requires no energy cofactors and quickly reaches equilibrium between substrate and reaction products, a DNA band corresponding to the size of the pLASSO substrate (blue arrow) was expectedly observed.

Hosted file

`image3.emf` available at https://authorea.com/users/413259/articles/521655-massively-parallel-dna-target-capture-using-long-adapter-single-stranded-oligonucleotide-lasso-probes-assembled-through-a-novel-dna-recombinase-mediated-methodology

Figure 2 , **Schematics of DNA recombinase mediated LASSO probe library assembly** . **a** , A ssDNA pre-LASSO oligonucleotide library is converted to double stranded DNA form by PCR using primer selectors. **b** , The selected pre-LASSO library in shuttled in the linearized pLASSO vector and used for transformation in *E.coli* .**c** , Gel electrophoresis of BamHI and SalI digested library, lane 1 illustrate presence of a DNA band correspondent with the pre-LASSO insert indicating successful cloning of the pre-LASSO library. lane 2, pLASSO alone. **d** , The native supercoiled pLASSO library is digested nicking with endonuclease and subjected to Cre recombination that generates a DNA minicircle containing the pre-LASSO and a circular 2.7 kb DNA circle. The 2.7kbDNA circle, together with the unreacted plasmids and bigger DNA circles generated by inter-plasmid recombination (not shown in the drawing) are eliminated by restriction followed by exonuclease digestion. **f** , Gel electrophoresis illustrates in lane 1 successful formation of DNA minicircles (orange arrow) together with the 2.7 kb circular DNA remaining parts of pLASSO (green arrow), the unreacted plasmid (blue arrow). The approximately 6kb band (yellow arrow) correspond to the recombination of two different plasmids (inter-plasmid recombination). In lane 2, When using an un-nicked pLASSO library for Cre recombination the DNA band correspondent to DNA minicircle was absent. **e** , inverse PCR. **g** , maturation of the LASSO probe library by removal of primer annealing sites and digestion of a DNA strand. **h,** Gel electrophoresis, in lane 1 inverted PCR amplicon correspondent to the linearized

minicircle. In lane 2 negative control for Cre-recombination. **Legend.** Sal1, BamH1 and BspQ1 indicate restriction enzyme sites, nick indicates nicking endonuclease site NtBspQ1 the * indicates phosphorothioate bonds, U indicate a deoxyuracil moiety. L1 indicates 1 kb DNA Ladder (NEB), L2 indicates Low Molecular Weight DNA Ladder (NEB)

An approximately 6kb DNA band (yellow arrow) was also seen in the final product and likely corresponded to the size of pLASSO concatemers. The formation of circular concatemers was likely caused by to the recombination of loxP sites located in different pLASSO molecules (inter-molecular recombination). In order to eliminate the recombination products other than the DNA minicircles, we performed a digestion with SwaI restriction enzyme (SwaI recognition sequence present in the 2.7 kb circular DNA, pLASSO substrate and pLASSO concatemers) followed by Exonuclease V digestion (**Fig. 2d**). We also performed the Cre-recombination reaction using the un-nicked pLASSO library; in this case the DNA band correspondent to DNA minicircle (red arrow in**Fig. 2 f** lane 1) was not visible (**Fig. 2 f,** lane 2). This observation suggests that the uncoiled form of pLASSO plasmid, induced by the DNA nick, was a better substrate for Cre-recombinase that the natural supercoiled un-nicked form. The minicircles were subsequently subjected to inverse PCR (**Fig. 2 e**) using primers that anneal on the inverse PCR primer-annealing site. The expected size of the inverted PCR product was verified in agarose gel (**Fig. 2 h**). The negative control for Cre recombination showed no inverted PCR amplicon was present (gel lane 2) indicating that DNA minicircles were not formed in absence of Cre recombinase and that pLASSO was completely digested.

The inverted PCR product is in the final mature LASSO probe configuration with the annealing arms flanking the conserved region (blue) derived from pLASSO (**Fig. 2 g**). The external primer sites were then removed by digesting with BspQ1 restriction enzyme followed by exonuclease digestion and treatment with USER enzyme as described in Material and Methods. The final mature LASSO probe library was then ready for massively parallel target capture reactions.

To assess quality and uniformity of the LASSO library produced we performed NextSeq 150bp paired end sequencing of the E.coli LASSO library at the inverted PCR stage (**Fig 2 g**) in which ligation and extension arms are already coupled with the conserved DNA backbone in the final configuration.

Analysis of NGS data reveal that the majority of LASSO probes were composed of ˜45% correctly paired probe arms versus total read sequences per probe type (**Fig. 3 a)** . The mean for correctly paired probes as a ratio of concordant probes vs discordant probes calculated for all LASSO probes was 46 %. Syukri S. and coworkers (2019) when assessing the quality of the E.coli LASSO library reported only 10% of concordant probes when using our previous assembly methodology (Tosi *et al.*,2017). As shown in **Fig. 3b** , the majority of the probes were present within two tenfold the normalized abundance of the median indicating a relatively uniform representation of LASSO probes in the LASSO library.

Hosted file

`image4.emf` available at https://authorea.com/users/413259/articles/521655-massively-parallel-dna-target-capture-using-long-adapter-single-stranded-oligonucleotide-lasso-probes-assembled-through-a-novel-dna-recombinase-mediated-methodology

Figure 3 LASSO probe library analysis. a . Box Plot "Arm Concordancy" indicates the percentage of correctly paired probe arms versus total read sequences per probe type (black dots). **b** . Plot of absolute counts of concordant LASSO probes types (gray dots).

3. 3. Multiplex LASSO cloning of the E. coli ORFeome

We next evaluated the ability of the new LASSO probes to capture a library of kilobase-sized ORFs from *E. coli* genomic DNA using the same capture parameters described by Tosi L. and coworkers (2017). According to the workflow in **Fig 4a** , LASSO probes were assembled as described above and hybridized with total genomic DNA of E.coli K12, targeting the 3078 ORFs in a single reaction volume. Circles containing ORFs were PCR amplified using primers that hybridize to the conserved adapter region on each LASSO probe.

Hosted file

`image5.emf` available at https://authorea.com/users/413259/articles/521655-massively-parallel-dna-target-capture-using-long-adapter-single-stranded-oligonucleotide-lasso-probes-assembled-through-a-novel-dna-recombinase-mediated-methodology

Figure 4. ORFeome capture using LASSO probes. a, Schematic of the workflow**. b,** Post-capture PCR of circles obtained from the capture of 3,078 ORFs of E. coli K12 performed using the LASSO probe library. The inset is a histogram denoting the size distribution of the targeted ORFs split into bin sizes of 40 bp. Targeted ORFs have an increase in 140bp of residual LASSO sequences once captured and run on a gel. **c.** Median RPKM enrichment ratios of targeted ORFs versus non-targeted genetic elements ratios of a LASSO probe library obtained by using the DNA Recombinase Mediated Assembly (blue) and the assembly method developed by Tosi L. and coworkers in 2017 (red). **d,** Bee swarm plot combined with boxplot Average depth of sequencing per kilobase for each targeted ORF (n=3087) and non targeted ORF (n=905). Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots. n = 3057, 1004 sample points. **e.**Normalized read depth of targeted ORFs as a function of the length of the ORF

Post capture PCR of circles obtained from the capture of 3078 ORFs of E.coli K12 was run in a 1.2% agarose gel and is shown in **Fig. 4 b.** and their apparent size distribution corresponded well with that of the targeted ORFs. Post capture PCR amplicon was enzymatically fragmented and sequenced on an Illumina NextSeq instrument to obtain 150 nucleotide paired end reads.

For reads mapping to the E. coli genome, we calculated target enrichment factors, which we defined as the reads per kilobase of genetic element per million reads (RPKM), which were mapped to the targeted ORFs versus non-targeted ORFs. Furthermore, RPKM targeted/non-targeted ratios were analyzed for different length genetic elements by binning **Fig. 4 c** In this experiment, LASSO targeted ORFs were enriched in all bins (up to ˜250 × for ORFs < 1kb) representing 8 times improvement in comparison to enrichment previously measured by Tosi and coworkers (2017).

Fig. 4d. illustrates box plots of average depth of sequencing per kilobase for each targeted and for each untargeted ORF. The targeted ORFs were significantly enriched compared with the non-targeted ORFs (by Welch two-sample t-test). The mean and the median RPKM of the targets was 2476 and 264 for the targets respectively while the mean and the median RPKM of the Non Targets was 31 and 1.26 respectively. Fold-enrichment of targets was calculated to be between 60- and 200-fold (by the median or mean of the target RPKM, respectively, over the mean non-target RPKM). At a cutoff of three times the median non-target RPKM, around 70% of the targeted ORFs were successfully captured. The normalized abundance of each target ORF was negatively correlated with the ORF length; (**Fig. 4e**). This length bias was previously reported (Tosi et al. 2017) and it reflects target length-dependent capture efficiency, post-capture PCR bias or a combination of the two effects.

3.4. LASSO probes sensitivity assessment

In order to model the feasibility of a highly sensitive and selective capture of a single gene in a background of the human genome, we performed a series of capture reactions where we spiked consecutive tenfold dilutions of the 7250 bp M13mp18 viral DNA in a constant human genomic DNA background (1.6 fM, correspondent to 50 ng/µl) to a final 0.34 fM correspondent to $25*10^{-6}$ ng/µl. The captures were performed by using tenfold dilutions of the new LASSO M13 probe that was designed to capture 1kb target sequence within the M13mp18 genome. As shown in the **Fig. 5 a** , The expected capture band was observed even when testing the lowest 0.35 fM target molar concentration that was ˜4 times lower than the Human DNA molar background (˜1.6 fM) in the capture reaction.

3.5. LASSO capture of Human ORFs

Finally, we evaluated the ability of these manufactured LASSO probes purified from a Cre-mediated plasmid

production process to capture two individual full-length ORFs from a cDNA derived from different human tissues. The ˜ 1 kb sequences of β -actin and Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) genes were successfully captured in this manner (**Fig. 5b**), as confirmed by Sanger sequencing. The evolved method to produce LASSO probes did not impact the ability of these probes to capture human cDNAs as was previously shown.

Hosted file

`image6.emf` available at https://authorea.com/users/413259/articles/521655-massively-parallel-dna-target-capture-using-long-adapter-single-stranded-oligonucleotide-lasso-probes-assembled-through-a-novel-dna-recombinase-mediated-methodology

Figure 5. LASSO capture sensitivity assessment, a . Gel electrophoresis of post capture PCR amplicons obtained by capturing a single 1kb target sequence within the M13mp18 viral DNA (red) that was spiked according to five tenfold dilutions (3.5 pM, 350fM, 35fM, 3.5fM, 0.35fM) in a constant 1.6fM total Human genomic DNA background (green). Captures displayed the 15 lanes were performed by testing three tenfold dilutions of the LASSO probe (blue) of 72pM, 7.2pM and 720 fM respectively. **-** indicates negative control for the filling mix in the capture reaction. **b** . Post capture PCR of circles obtained from the capture of β -actin and Glyceraldehyde 3-phosphate dehydrogenase (GAPDH). Lanes 6 capture of g-Actin, 7 capture of GAPDH, 8 and 9 negative control same as 6 and 7 but no LASSO library in the capture, 10 Positive control for capture reaction (1kb target within M13mp18 ssDNA. **L** indicates 1kb plus DNA ladder (NEB).

4. Discussion

LASSO technology enables massively multiplexed capture of large DNA fragments for sequencing and/or expression cloning. A key factor for LASSO capture efficiency is the quality of the LASSO probe library in terms of probe sequence identity and probe representation. A previous assembly protocol we developed presented a number of drawbacks that resulted in a poor quality of the LASSO library with only ˜10% of correctly assembled LASSO probes. The critical phase was the self-circularization by ligation of the LASSO probe precursor, wherein EcoRI digested probe ends were intramolecularly ligated to each other. Shukor S et al. [34] noted that when attempting to generate thousands of probes in a single reaction by this manner, there exists a strong possibility that intermolecular ligations would manifest as mismatched probe arms on a mature LASSO probe. The presence of the discordant probes in the mature LASSO libraries was responsible for a reduction of the efficiency, likely due to the depletion of PCR reactants used up for the amplification of low molecular weight unspecific DNA amplicons arising in the post capture PCR (Supplementary figure 1 table a).

To improve the quality of the LASSO probe libraries we developed a different LASSO assembly methodology that leads to the same mature LASSO probe configuration but uses the cre recombination of a custom plasmid (pLASSO) to supply the linker of the mature LASSO probe. The assembly process herein described starts with the cloning of the pre-LASSO library in the pLASSO vector and E.coli transformation. The multiplication of a pLASSO library in E.coli, reduces the possible skewing of the different LASSO probes in the library. The plasmid library was then subjected to Cre recombination of the two loxp sites oriented head-to tail in pLASSO resulting in the excision of a DNA minicircle containing the LASSO precursor in its final configuration (Figure 2).

The LASSO library we assembled targeted the same E. coli ORFs of our previous work [33] and the LASSO probes had identical design but displayed superior capture performance. This observation is in agreement with the higher percentage (˜46%) of concordant probes present in the E.coli LASSO library. The median RPKM for targeted ORFs versus untargeted ORFs was much higher than produced with the previous LASSO assembly methodology - especially for shorter ORFs (˜8 times higher). This finding suggests that a better quality of the LASSO probe library results in a higher capture efficiency and in the reduction of undesirable low molecular weight amplicons in post capture PCR (Supplementary Figure 1)

With a model system, we showed that the sensitivity of LASSO capture potentially allows for the massive

parallel capture of DNA targets at the whole human genome scale. We also evaluated the ability of the new LASSO probes to capture two individual full-length ORFs from a total human cDNA. The genes β-actin and glyceraldehyde 3-phosphate dehydrogenase were captured thereby verifying that this new LASSO production method provides comparable capture efficacy to the first-generation method using a self-circularization reaction. Future evaluation of this method is necessary to evaluate the breadth of this LASSO technology for the creation of human protein libraries.

As novel long-read sequencing technologies emerge, there is an increasing need for novel target enrichment methods that allow highly multiplexed selection of kilobase-sized DNA. We expect that LASSO probes can find immediate applications for targeted construction of long-read sequencing libraries. LASSO probes can also be used for the rapid and inexpensive production of pooled ORFeome libraries that can be expressed using standard vectors for functional screening applications.

Acknowledgement

This research was conducted with support under Grant Nos. R01GM127353 (B.P., B.L.), R01EB012521 (B.P.), and R01GM20861 (W.K.O.) awarded by the National Institutes of Health.

Author contributions

Conceptualization, B.P., H.B.L., and L.T.; Validation, J.L., S.S, S.L, A.T., L.T.; Formal Analysis, J.L., L.T., B.L, V.N; Writing-Original Draft Preparation, J.L. and S.S; Writing-Review & Editing, J.L., S.S, S.L, A.T., L.T., H.B.L., V.N, WKO, B.P.; Funding Acquisition, B.P. and H.B.L.

Conflicts of interest: The authors declare no conflict of interest.

Reference list

1. Lubock, N., Kosuri, S. Genetic engineering: Lassoing genomic libraries. Nat Biomed Eng 1, 0098 (2017).
2. Inoue F, Ahituv N. Decoding enhancers using massively parallel reporter assays. Genomics. 2015 Sep; 106(3):159-164
3. Gasperini M, Starita L, Shendure J. The power of multiplexed functional analysis of genetic variants. Nat Protoc. 2016 Oct;11(10):1782-7. doi: 10.1038/nprot.2016.135. Epub 2016 Sep 1. PMID: 27583640; PMCID: PMC6690347.
4. Hietpas,R.T., Jensen,J.D. and Bolon,D.N.A. (2011) Experimental illumination of a fitness landscape. Proc. Natl Acad. Sci. U.S.A., 108, 7896–7901.
5. Kinney,J.B., Murugan,A., Callan,C.G. Jr and Cox,E.C. (2010) Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. Proc. Natl Acad. Sci. U.S.A., 107, 9158–9163.
6. Sharon,E., Kalma,Y., Sharp,A., Raveh-Sadka,T., Levo,M., Zeevi,D., Keren,L., Yakhini,Z., Weinberger,A. and Segal,E. (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nat. Biotechnol., 30, 521–530.
7. Starita,L.M., Young,D.L., Islam,M., Kitzman,J.O., Gullingsrud,J., Hause,R.J., Fowler,D.M., Parvin,J.D., Shendure,J. and Fields,S. (2015) Massively parallel functional analysis of BRCA1 RING domain variants. Genetics, 200, 413–422.
8. Doolan,K.M. and Colby,D.W. (2015) Conformation-dependent epitopes recognized by prion protein antibodies probed using mutational scanning and deep sequencing. J. Mol. Biol., 427, 328–340.
9. Patwardhan, R.P., Lee, C., Litvin, O., Young, D.L., Pe'er, D. and Shendure,J. (2009) High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nat. Biotechnol., 27, 1173–1175
10. Arnold,C.D., Gerlach,D., Stelzer,C., Bory´ n,Ł.M., Rath,M. and Stark,A. (2013) Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science, 339, 1074–1077
11. Sarkisyan KS et al., Local fitness landscape of the green fluorescent protein. Nature. 533, 397–401 (2016).

12. Rocklin GJ et al., Global analysis of protein folding using massively parallel design, synthesis, and testing. Science. 357, 168–175 (2017)
13. Angus M Sidore, Calin Plesa, Joyce A Samson, Nathan B Lubock, Sriram Kosuri, DropSynth 2.0: high-fidelity multiplexed gene synthesis in emulsions, Nucleic Acids Research, Volume 48, Issue 16, 18 September 2020
14. Quan,J., Saaem,I., Tang,N., Ma,S., Negre,N., Gong,H., White,K.P. and Tian,J. (2011) Parallel on-chip gene synthesis and application to optimization of protein expression. Nat. Biotechnol., 29, 449–452.
15. Kosuri,S., Eroshenko,N., Leproust,E.M., Super,M., Way,J., Li,J.B. and Church,G.M. (2010) Scalable gene synthesis by selective
16. Kosuri S, Church GM, Large-scale de novo DNA synthesis: technologies and applications. Nat. Methods 11, 499–507 (2014)
17. Hughes RA, Ellington AD, Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology. Cold Spring Harb. Perspect. Biol 9 (2017)
18. Klein, J.C., Lajoie,M.J., Schwartz,J.J., Strauch,E.M., Nelson,J., Baker,D. and Shendure,J. (2015) Multiplex pairwise assembly of array-derived DNA oligonucleotides. Nucleic Acids Res., 44, e43.
19. Kim,H., Han,H., Ahn,J., Lee,J., Cho,N., Jang,H., Kim,H., Kwon,S.and Bang,D. (2012) 'Shotgun DNA synthesis' for the high-throughput construction of large DNA molecules. Nucleic Acids Res., 40, e40
20. Hsiau,T.H.-C., Sukovich,D., Elms,P., Prince,R.N., Stritmatter,T.,Ruan,P., Curry,B., Anderson,P., Sampson,J. and Christopher Anderson,J. (2015) A method for multiplex gene synthesis employing error correction based on expression. PLoS One, 10, e0119927.
21. Saiki, R.K. et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239, 487–491 (1988). This paper was the first description of PCR, which, coupled to electrophoretic sequencing, is the primary conventional method for targeted variation analysis.
22. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ. Target-enrichment strategies for next-generation sequencing. Nat Methods. 2010 Feb;7(2):111-8
23. Cho, R.J. et al. Genome-wide mapping with biallelic markers in Arabidopsis thaliana. Nat. Genet. 23, 203–207 (1999).
24. Wang, D.G. et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280, 1077–1082 (1998).
25. Nilsson, M. et al. Padlock probes: circularizing oligonucleotides for localized DNA detection. Science 265, 2085–2088 (1994).
26. Landegren, U. et al. Molecular tools for a molecular medicine: analyzing genes, transcripts and proteins using padlock and proximity probes. J. Mol. Recognit. 17, 194–197 (2004).
27. Porreca, G.J. et al. Multiplex amplification of large sets of human exons. Nat. Methods 4, 931–936 (2007).
28. Turner, E.H., Lee, C., Ng, S.B., Nickerson, D.A. & Shendure, J. Massively parallel exon capture and library-free resequencing across 16 genomes. Nat. Methods 6, 315–316 (2009).
29. O'Roak BJ, Stessman HA, Boyle EA, et al. Recurrent de novo mutations implicate novel genes underlying simplex autism risk. Nat Commun. 2014;5(5595):1–6
30. Krishnakumar, S. et al. A comprehensive assay for targeted multiplex amplification of human DNA sequences. Proc. Natl Acad. Sci. USA 105, 9296–9301 (2008).
31. Shen, P. et al. Multiplex target capture with double-stranded DNA probes. Genome Med. 5, 50 (2013).
32. Shen, P. et al. High-quality DNA sequence capture of 524 disease candidate genes. Proc. Natl Acad. Sci. USA 108, 6549–6554 (2011).
33. Tosi L, Sridhara V, Yang Y, Guan D, Shpilker P, Segata N, Larman HB, Parekkadan B. Long-adapter single-strand oligonucleotide probes for the massively multiplexed cloning of kilobase genome regions. Nat Biomed Eng. 2017;1:0092.
34. Shukor S, Tamayo A, Tosi L, Larman HB, Parekkadan B. Quantitative assessment of LASSO probe assembly and long-read multiplexed cloning. BMC Biotechnol. 2019 Jul 24;19(1):50
35. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioin-

formatics. 2014;30(15):2114-2120. doi:10.1093/bioinformatics/btu170

36. Kent WJ. BLAT–the BLAST-like alignment tool. Genome Res. 2002 Apr;12(4):656-64. doi: 10.1101/gr.229202. PMID: 11932250; PMCID: PMC187518.
37. Calin Plesa, Angus M. Sidore, Nathan B. Lubock, Di Zhang, and Sriram Kosuri. (2018) Multiplexed Gene Synthesis in Emulsions for Exploring Protein Functional Landscapes Science. 2018 January 19; 359(6373): 343–347