# CapStarr-seq protocol

**Salvatore Spicuglia** ( salvatore.spicuglia@inserm.fr )
Spicuglia's lab (TAGC, Inserm U1090, Aix-Marseille University, Marseille)
**Lan T.M. Dao**
Spicuglia's lab (TAGC, Inserm U1090, Aix-Marseille University, Marseille)
**Laurent Vanhille**
Spicuglia's lab (TAGC, Inserm U1090, Aix-Marseille University, Marseille)
**Aurélien Griffon**
Spicuglia's lab (TAGC, Inserm U1090, Aix-Marseille University, Marseille)
**Nicolas Fernandez**
TGML, TAGC, Inserm U1090, Aix-Marseille University, Marseille)

**Method Article**

**Keywords:** Enhancer, high-throughput reporter assay, Starr-seq, capture, mammals

**Posted Date:** October 15th, 2015

**DOI:** https://doi.org/10.1038/protex.2015.096

# Abstract

Cell-type specific regulation of gene expression requires the activation of promoters by distal genomic elements defined as enhancers. The identification and the characterization of enhancers are challenging in mammals due to their genome complexity. Here we describe the CapStarr-Seq procedure \(Vanhille et al. Nat Commun. 2015. doi: 10.1038/ncomms7905), a high-throughput strategy to quantitatively assess enhancer activity in mammals. This approach couples capture of regions of interest to previously developed Starr-seq technique. CapStarr-Seq is applicable to a wide range of biological questions and will provide an efficient tool to better decipher mechanisms involved in gene regulation.


# Introduction

Recent advances in epigenomics have enabled genome-wide identification of enhancers in various tissues and species. However, these do not allow direct quantification of enhancer activity. Several studies have focused on the implementation of high-throughput gene reporter assays for large scale testing of enhancer activity. Among them, a technique named Starr-seq enabling genome-wide quantification of enhancer activity in Drosophila cell lines has been developed[1] and have been extensively used to interrogate enhancer function[2]. One major limitation for its application to mammalian systems resides in their genome size and complexity, rendering challenging the preparation of representative libraries and requiring very high sequencing depth. To circumvent this problem, we developed CapStarr-seq, a technique capturing genomic sequences of interest for high-throughput assessment of enhancer activity in mammals[3]. Here, we detail the step to perform CapStarr-seq as shown in Figure 1 . The provided protocol largely correspond the one described in Vanhille et al.[3], except that primers and adaptors are designed for the Illumina platform, thus allowing multiplexing sequencing. We also provide suggestions for improvement at various points during the procedure. CapStarr-Seq is a fast and cost-effective approach to assess the activity of potential enhancers for a given cell type and will be helpful in decrypting transcription regulation mechanisms applicable to a wide range of biological questions.

# Reagents

Agencourt AMPure XP DNA Purification Kit \(Beckman Coulter, A63881) NEBNext® DNA Library Prep Reagent Set for Illumina \(NEB, E6000S) NEBNext® Multiplex Oligos for Illumina® \(Index Primer Set 1) \(NEB, E7335L) KAPA HiFi HotStart Readymix DNA polymerase \(Clinisciences, KK2602) SureSelect DNA Capture Array \(Agilent Technologies), either 244k \(G4458A) or 1M \(G3358A) Cot-1 DNA \(Life Technologies) Oligo aCGH/ChIP-on-chip Hybridization Kit \(Agilent Technologies, 5188-5220) Hybridization Gasket Slide Kit - 1 microarray per slide format \(Agilent Technologies, G2534-60003) Oligo aCGH/ChIP-on-Chip Wash Buffer Kit \(Agilent Technologies, 5188-5226) Power SYBR® Green PCR Master Mix \(Life Technologies, 4368708) AgeI-HF \(NEB, R3552L) SalI-HF \(NEB, R3138S) NcoI-HF \(NEB, R3193S) In-Fusion HD Cloning Kit \(Clontech, 639648) MegaX DH10B™ T1R Electrocomp™ Cells \

(Life Technologies, C6400-03) Neon 100 ul Transfection Kit \(Life Technologies, MPK10096) QIAGEN Plasmid Maxi Kit \(QIAGEN, 12163) QIAquick Gel Extraction Kit \(QIAGEN, 28704) QIAquick PCR Purification Kit \(QIAGEN, 28104) RNeasy Plus Mini Kit \(QIAGEN, 74136) QIAshredder \(QIAGEN, 79656) RNase-Free DNase Set \(QIAGEN, 79254) DNAse Turbo \(Life Technologies Ambion, AM2238) SuperScript® III First-Strand Synthesis System \(Life Technologies, 18080051) µMACS mRNA Isolation Kit - Small Scale \(Miltenyi Biotec, 130-075-201) DNA High Sensivity Kit \(Agilent Technologies, 5067-4626) Qubit® RNA HS Assay Kit \(Life Technologies, Q32855) Qubit® ssDNA Assay Kit \(Life Technologies, Q10212) Qubit® dsDNA HS Assay Kit \(Life Technologies, Q32851) UltraPure™ Phenol:Chloroform:Isoamyl Alcohol \(25:24:1, v/v) \(Life Technologies, 15593-031) Isopropanol \(VWR, 437423R) Eppendorf DNA LoBind tubes \(DUTSCHER DOMINIQUE, 33871) Bioruptor 0.65 mL microtubes for DNA shearing \(Diagenode, C30010011) Custom primers are described in Table 1

# Equipment

Vortex Rotator Heat block Table centrifuge Magnetic rack for 1.5 ml tubes Diagenode™ Bioruptor® Pico Ultrasonicator Qubit 2.0 fluorometer \(Life Technologies) Nanodrop Spectrophotometer \(Thermo Scientific) Thermal Cycler for PCR Gel Imaging System Gel electrophoresis Agilent 2100 Bioanalyzer Agilent p/n G2938C Neon® Transfection System \(Life Technologies) Gene Pulser II Electroporation System \(BIO-RAD) Gene pulser/Micropulser Cuvettes 0.1 cm gap \(BIO-RAD, 165-2089) Hybridization chamber, stainless \(Agilent p/n G2534A) Hybridization oven rotator for Agilent Microarray Hybridization Chambers \(Agilent p/n G2530-60029) Hybridization oven 700 series \(SciGene) Real-time PCR system Illumina Sequencer

# Procedure

A. **Preparation of size-selected genomic DNA \(Timing: ~1 day)** 1. _Genomic DNA extraction_ Genomic DNA is extracted using standard method. DNA is then well-dissolved in water. 2. _Fragmentation of genomic DNA_ a. Dilute required amount of genomic DNA to 20 ng/µL in TE buffer. b. Aliquot 100 µL in 0.65 mL microtubes for DNA shearing. c. Keep tube on ice for 10 min, then vortex and spin briefly. d. Place tubes on the tube holder. e. Start sonication with Diagenode™ Bioruptor® Pico Ultrasonicator using the following setting: 15 seconds ON, 90 seconds OFF and 7 pulses. 3. _Size select fragmented DNA_ The original protocol 3 used agarose gel to select the desired fragmented DNA size. Here, size selection is based on the use of Agencourt AMPure XP beads \(Beckman Coulter, A63881).The following bead-to-sample ratio is used to make a library size distribution with a peak size of 300 bp. For a different library size, the bead-to-sample ratio will need to be adjusted. a. Homogenize AMPure XP beads at RT before use. b. Adjust the fragmented DNA volume to 100 µL. c. Add 0.75X volume of beads \(75 µL) to the DNA sample, mix well by pipetting and incubate at RT for 5 min. d. Place the tube on magnetic rack for 5 min. e. Carefully transfer 73 µL of the supernatant to a new 1.5 mL tube and discard the beads. f. Add 0.5X volume of beads \(50 µL) to the collected supernatant \(so total bead-to-sample ratio is 1.25X at this step), mix well by pipetting and incubate at RT for 5 min. g. Place the tube on the magnetic rack for 5 min

then carefully remove all supernatant. h. Wash the beads twice with freshly prepared 80% ethanol. After the second wash, briefly spin, place the tube on the magnetic rack and completely remove all residual ethanol. i. Dry the beads at 40 °C until the pellet starts to break. Avoid over drying the bead pellet, time varying depending on the amount of beads. j. Elute DNA in 102 µL of water and transfer exactly 100 µL of the supernatant in a new 1.5 mL tube. k. Add 0.85X volume of beads \(85 µL), mix well by pipetting and incubate at RT for 5 min. l. Place the tube on the magnetic rack for 5 min. Then carefully remove all supernatant. m. Wash the beads twice with freshly prepared 80% ethanol. After the second wash, spin briefly, place the tube on the magnetic rack and completely remove all residual ethanol. n. Dry the beads at 40 °C until the bead pellet starts to break. o. Elute DNA in 75 µL of water and transfer 73 µL of the supernatant in a new 1.5 mL tube. p. Analyze size-selected DNA on Agilent 2100 Bioanalyzer \(An example of size-selected DNA is provided in Figure 2a). NOTE: The exact bead to sample ratio is critical to select the desired DNA fragments. Therefore, it is important to be extremely careful when pipetting the beads. B. **Preparation of Pre-capture DNA library \(Timing: ~1 day)** Follow the protocol of NEBNext® DNA Library Prep Reagent Set for Illumina® \(NEB, E6000S) with some modifications. Starting material: 1 µg of size-selected DNA 1. _End repair of size-selected DNA_ a. Mix the following components in a 1.5 mL tube: Size-selected DNA: 1 – 75 µL Phosphorylation Reaction Buffer \(10X): 10 µL T4 DNA polymerase: 5 µL T4 Polynucleotide Kinase: 5 µL dNTPs: 4 µL DNA Polymerase I, Large \(Klenow): 1 µL Sterile water: up to 100 µL Total volume: 100 µL b. Incubate in a Thermocycler for 30 min at 20 °C. c. Clean up the reaction with AMPure XP beads: add 1.2X volume of beads \(120 µL) to the reaction and follow the cleanup protocol. At final step, elute DNA in 34 µL of water and carefully transfer 32 µl of the supernatant to a new 1.5 mL tube. 2. _dA-Tailing of End Repaired DNA_ a. Mix the following components in a 1.5 mL tube: End Repaired, Blunt DNA: 32 µL NEBuffer 2 \(10X): 5 µL Deoxyadenosine 5'-Triphosphate: 10 µL Klenow Fragment \(3'→5' exo-): 3 µL Total volume: 50 µL b. Incubate in a Thermocycler for 30 min at 37 °C. c. Clean up the reaction with AMPure XP beads: add 1.2X volume of bead \(60 µL) to the reaction and follow the cleanup protocol. At final step, elute DNA in 12 µL of water and carefully transfer 10 µL of the supernatant to a new 1.5 mL tube. 3. _Adaptor ligation of dA-Tailed DNA_ a. Mix the following components in a 1.5 mL tube: dA-Tailed DNA: 10 µL Quick Ligation Reaction Buffer \(2X): 25 µL NEBNext Adaptor : 10 µL Quick T4 DNA Ligase : 5 µL Total volume: 50 µL b. Incubate in a Thermocycler for 15 minutes at 20 °C. c. Add 3 µL of USER™ enzyme mix by pipetting up and down, and incubate at 37 °C for 15 min. d. Clean up the reaction with AMPure XP beads: add 1.2X volume of bead \(63.6 µL) to the reaction and follow the cleanup protocol. At final step, elute DNA in 27 µL of water and carefully transfer 25 µL of the supernatant to a new 1.5 mL tube. NOTE: Skip the Size Select Adaptor Ligated DNA in the NEB's protocol and directly proceed the next step. 4. _PCR enrichment of ligated DNA_ The PCR primers for this step are designed with 15 bp of homology arms flanking the cloning site of the Starr-seq vector1. a. Mix the following components in a PCR tube: Ligated DNA: 2.5 µL Capstarr-seq primer mix \(12.5 µM each, Table 1, No.1 & 2): 5 µL KAPA HiFi HotStart Readymix DNA Polymerase: 25 µL Sterile H2O: 17.5 µL Total volume: 50 µL b. Place the tube in Thermocycler and run the following program: 98 °C for 45 s 98 °C for 15 s, 65 °C for 30 s, 72 °C for 30 s \(7 cycles) 72 °C for 2 min Keep at 4 °C c. Clean up the PCR reaction with AMPure XP beads: add 1.2X volume of beads \(60 µL) to the reaction and follow the cleanup protocol. At final step, elute DNA in 22 µL of water and carefully transfer

20 µL of the supernatant to a new 1.5 mL tube. d. Measure the concentration of DNA with a Qubit dsDNA Kit. e. Assess quality of DNA library with Agilent 2100 Bioanalyzer using the DNA High Sensitive Kit. A shift of 100 bp in the size distribution should be observed after adding adaptors to the library \(Figure 2b). f. Perform the required number of PCR reactions \(steps 4a-4c) in order to obtain the desired amount of DNA for the capture \(10-20 µg of DNA is required for a single capture). If DNA yield is not enough, starts again library preparation from the beginning \(steps B1 to step B4). C. **Capture enrichment \(Timing: ~3 days, Hands-on: 2 hours)** 1. _Microarray design_ For target enrichment, a home-designed 3 bp resolution oligonucleotide microarray covering the desired genomic regions is constructed using the SureSelect Capture Array technology \(Agilent) and the eArray tool default setting \(https://earray.chem.agilent.com/earray/). Depending on the number of region of interest it is possible to choose either the 244k or 1M format. Note: It should be noted that the size of tested regions depends solely on the size distribution of the Pre-capture DNA library \(Steps A & B), and not on the size of the defined regions of interest. 2. _Microarray processing_ Follow the protocol of Agilent SureSelect DNA Capture array from Section 3, step 1 to step 5 \(Manual part number G4458-90000, version 1.0, July 2009). The hybridization sample mixture is as following: Sample DNA \(10 µg to 20 µg in nuclease-free water): 138 µL Blocking oligo 1 \(200 µM): Capstarr-seq_Forward \(No.1): 5 µL Blocking oligo 2 \(200 µM): Capstarr-seq_Reverse \(No.2): 5 µL Blocking oligo 3 \(200 µM): Capstarr-seq_Forward_Rev \(No.3): 5 µL Blocking oligo 4 \(200 µM): Capstarr-seq_Reverse_Rev \(No.4): 5 µL Cot-1 DNA \(1 mg/mL)#: 50 µL Agilent 10X Blocking Agent: 52 µL Agilent 2X Hi-RPM Hybridization Buffer : 260 µL Total volume: 520 µL \(#) use Cot-1 DNA from the appropriate species. D. **Amplification of eluted post-capture DNA \(Timing: ~1 days)** 1. _Purify eluted post-capture DNA with the AMPure XP beads_ Add 1X volume of beads \(490 µL) to the eluted post-capture DNA and follow the cleanup protocol. At final step, elute DNA in 52 µL of water and carefully transfer 50 µL of the supernatant to a new 1.5 mL tube. 2. _PCR amplification_ a. Set up a PCR reaction as following: Eluted post-capture DNA: 5 µL Capstarr-seq primer mix \(10 µM each, No.1 & 2): 5 µL KAPA HiFi HotStart Readymix DNA Polymerase: 25 µL Sterile H2O: 15 µL Total volume: 50 µL b. Place the tubes in Thermocycler and run the following program: 98 °C for 45 s 98 °C for 15 s, 65 °C for 30 s, 72 °C for 30 s \(20 - 24 cycles#) 72 °C for 2 min Keep at 4 °C \(#) number of cycle is depending on the efficiency of the capture. For the first time, we recommend running with 20 cycles. 3. _PCR Clean up with the AMPure XP beads_ a. Add 1X volume of bead \(60 µL) to the reaction and follow the clean up protocol. At final step, elute DNA in 42 µL of water and carefully transfer 40 µL of the supernatant to a new 1.5 mL tube. b. Measure the concentration of DNA with the Qubit dsDNA HS Assay Kit. c. \(Optional) Analyze post-capture DNA library on Agilent 2100 bioanalyzer using DNA High Sensitive Kit. The amplified post-capture library should have the same size distribution as the pre-capture library with the peak size of 400 bp \(Figure 2c). d. Perform additional number of PCR reactions \(step D2 and step D3) to reach to required amount of DNA for the next cloning step \(at least 2 µg of DNA). 4. _Evaluation of capture efficiency_ We recommend selecting a set of at least 3 targeted regions and 3 non-targeted regions for qPCR assay in order to having accurate confirmation. a. Set up two series of qPCR reactions using as template either pre-capture amplified DNA library \(step B4) or post-capture amplified library \(step D3). DNA: 0.5 ng Primer mix \(10 µM each): 2 µL Power SYBR: 10 µL Sterile H2O: up to 20 µL Total volume: 20 µL b. Quantify target enrichment Evaluate the target enrichment by ΔCt values. The difference

between pre-capture and post-capture Ct values indicates the level of enrichment and the capture is considered to be efficient when ΔCt ≥ 5 \(Figure 3). E. **Library cloning \(Timing: ~4 days)** 1. _Prepare linearized Starr-seq vector_ STARR-Seq screening vector should be obtained from Stark's lab \(ref3). The large amount of plasmid is prepared by Maxiprep \(QIAGENE). Vector is then linearized by restriction enzyme digestion as following: a. AgeI digestion - Prepare digestion mixture as follow: Plasmid DNA: 5 µg Cutsmart buffer \(10X): 10 µL AgeI-HF \(20 U/µL): 2.5 µL Sterile H2O: up to 100 µL Total volume: 100 µL - Incubate the digestion mixture at 37 °C for 4 – 6 hours. - Purify digested plasmid with QIAquick PCR Purification Kit and elute the product in 50 µL of water. b. SalI digestion - Prepare digestion mixture as follow: Plasmid DNA: 5 µg Cutsmart buffer \(10X): 10 µL SalI-HF \(20 U/µL): 2.5 µL Sterile H2O: up to 100 µL Total volume: 100 µL - Incubate the digestion mixture at 37 °C for 6 – 8 hours. c. Gel electrophoresis - Prepare agarose gel at 0.8% in a long tray. - Load digested plasmid on gel \(avoid overload plasmid DNA). - Run at low voltage until the bands are well separated. - Excise the desired band \(3.5kb) under UV 365 mm \(process as fast as possible to avoid DNA damage). - Purify linearized plasmid with QIAquick Gel Extraction Kit and elute in 50 µL of water. d. NcoI digestion \(optional) NcoI cut at a site between AgeI and SalI. This step helps to avoid any contamination of circular vector. - Digestion mixture as follow: Plasmid DNA: 5 µg Cutsmart buffer \(10X): 10 µL NcoI-HF \(20 U/µL): 2.5 µL Sterile H2O: up to 100 µL Total volume: 100 µL - Purify digested plasmid with QIAquick PCR Purification Kit and elute in 50 µL of water. 2. _In-Fusion cloning reaction_ Before using the linearized vector for cloning, it is recommended to perform a transformation test with linearized vector only to evaluate the background. Typically, the transformation of 100 ng of linearized vector should give few hundred colonies \(acceptable range). a. Perform 12 In-Fusion reactions using In-Fusion HD Cloning Kit. For 1 reaction, combine the following reagents in a PCR tube: Linearized vector: 200 ng Library: 50 ng In-Fusion enzyme mix: 2 µL H2O: up to 10 µL b. Incubate reactions for 15 min at 50 °C, and then transfer tubes to ice. c. Pool all 12 In-Fusion reactions in 1.5 mL tubes. d. Purify In-Fusion reaction with AMPure beads: Add 1X volume of bead \(100 µL) to the reaction and follow the clean-up protocol. At final step, elute recombinant plasmid in 22 µL of water and carefully transfer 20 µL of the supernatant to a new 1.5 mL tube. 3. _Transformation of recombinant library into E.coli_ Perform 12 transformations in total a. Thaw MegaX DH10B™ T1R Electrocomp™ Cells on wet ice. b. When cells are thawed, mix cells by tapping gently. Add 220 µL of cells to chilled 1.5 mL tube containing 20 µL of purified recombinant plasmid. c. Pipette 20 µL of cell/DNA mixture into a chilled 0.1 cm cuvette and electroporate. If you are using the BTXR ECMR 630 or Bio-Rad GenePulserR II electroporator, we recommend using the following electroporation conditions: 2.0 kV, 200 Ω, 25 µF. If you are using a different electroporator, you may need to vary these electroporation conditions to achieve optimal transformation efficiency. The time constant should be between 4,8 – 5,2. d. Add 1 mL of Recovery Medium to the cells in the cuvette and transfer the solution to a 15 mL culture tube. e. Perform the remaining 11 transformations and pool altogether in the same tube. f. Close the tube lid and shake at 225 rpm 37 °C for 1 hour. g. To estimate total number of clone in the library, spread 0.002%, 0.004% and 0.006% of the total transformation mix on pre-warmed LB plates containing 100 µg/mL ampicillin. Incubate plates overnight at 37 °C. h. The remaining transformation mix is directly transferred to 1 L of LB containing 100 µg/mL ampicillin and grow overnight at 37 °C with shaking at 250 rpm. i. The next day, count the number of colony on plates and

estimate the number of clones in the whole transformation reaction. Usually a total of 6-8 million colonies is achieved. 4. _Maxiprep \(QIAGEN Plasmid Maxi Kit)_ a. Collect cultured bacteria after 14-16 hours and perform 10 Maxipreps \(100 mL of bacteria per one Maxiprep) following the manufacture's protocol. b. Purified plasmid library is then well diluted in water at high concentration \(~5 μg/μL). F. **Library transfection \(Timing: ~2 days)** To ensure sufficient representation of each target region, highly efficient transfection of the CapStarr-seq library might be achieved. In total, 1.25 mg of plasmid library is transfected into 50 million cells using the cell-line specific conditions. As an example, we demonstrate here detailed protocol for K562 cell transfection follow Neon 100 uL transfection kit protocol. 1. One day before transfection, plate the cells at a density at which cells can reach to 70% - 80% confluent the next day. 2. Prepare required culture flasks, fill with pre-warmed culture medium. 3. Count the cell number. 4. Take an aliquot of 50 x 106 cells and wash in pre-warmed PBS. 5. Pellet the cells by centrifugation and completely remove PBS. 6. Resuspend cells in 1 mL Buffer R. 7. Add plasmid library: 25 μg of plasmid library per million cells, so total 1250 μg plasmid library is needed for 50 million cells. 8. Perform 10 transfections: 5 million cells per transfection with conditions as following: Pulse voltage: 1450 Pulse width: 10 Pulse number: 3 9. Spread the transfected cells in prepared culture flasks. Return cells into 37 °C, 5% CO¬¬2 incubator. 10. Optional: the overall enhancer activity of the library can be evaluated by FACS analysis with an aliquot of cells after 24 hours of transfection. Typically, a shift of GFP expression as compared to cells transfected with the empty STARR-seq vector is observed \(Figure 4). G. **Preparation of targeted cDNA \(Timing: ~1 days)** 1. _RNA extraction_ 24h after transfection, cells are collected for RNA isolation using the RNeasy Mini Plus Kit with some remarks: - Homogenize samples using QIAshredder \(QIAGEN, 79656). - Apply the DNase-on-column \(QIAGEN, 79254). - Total RNA amount is measured with Nanodrop Spectrometer. 2. _mRNA purification_ - Follow the manual instruction of μMACS mRNA Isolation Kit. Depending on the amount of total RNA, using the required number of columns \(do not overload the columns). - Purified mRNA concentration is measured by Qubit RNA HS Assay kit. Normally, mRNA account for 1% - 5% of total RNA population. 3. _DNase I treatment_ a. Prepare the following mixture in a PCR tube: mRNA: all from above step 10X Buffer: 5 μL Dnase I: 1 μL H2O: up to 50 μL Total volume: 50 μL b. Incubate the mixture at 37 °C for 30 min then place on ice. c. RNA clean-up following the Rneasy Mini Plus Kit. d. Quantify final mRNA concentration by Qubit RNA HS Assay Kit. 4. _Targeted cDNA synthesis_ Perform 10 reverse transcription reactions following the SuperScript® III First-Strand Synthesis System protocol. a. Prepare the RNA/primer mixture \(this following mixture is for 1 reaction): mRNA: 300 ng RT primer \(10 μM, Table 1, No.5): 1 μL dNTP \(10 μM): 1 μL Sterile H2O: up to 10 μL Total volume: 10 μL b. Incubate the tubes at 65 °C for 5 min, then place on ice for at least 1 min. c. Prepare the following cDNA synthesis mix \(this following mixture is for 1 reaction): 10X RT buffer: 2 μL 25 mM MgCl2: 4 μL 0.1M DTT: 2 μL Rnase OUT \(40U/ μL): 1 μL Superscript III RT \(200U/ μL): 1 μL Total volume: 10 μL d. Add 10 μL of cDNA synthesis mix to each RNA/primer mixture, mix gently and collect by brief centrifugation. Incubate at 50 °C for 50 min and terminate the reaction at 85 °C for 5 min. Chill on ice. e. Add 1 μL of Rnase H to each tube and incubate for 20 min at 37 °C. f. Purify the reaction with QIAquick PCR Purification Kit and elute in 50 μL of water. g. Measure single strand cDNA concentration by Qubit® ssDNA Assay Kit. H. **Preparation of Libraries for Sequencing \(Timing: ~4 hours)** 1. _Targeted PCR_ a. Prepare 10 PCR reactions in PCR tubes. For cDNA

Library, the following mixture is for 1 reaction: cDNA: 5 ng Targeted library PCR primer mix \(12.5 µM each, Table 1, No.6 & 8): 5 µL KAPA HiFi HotStart Readymix DNA Polymerase : 25 µL Sterile H2O: up to 50 µL Total volume: 50 µL For Input Library, the following mixture is for 1 reaction: Plasmid DNA \(from maxiprep): 0.2 ng Targeted Input PCR primer mix \(12.5 µM each, Table 1, No.7 & 8): 5 µL KAPA HiFi HotStart Readymix DNA Polymerase: 25 µL Sterile H2O: up to 50 µL Total volume: 50 µL b. Place the tubes in Thermocycler and run the following program: 98 °C for 2 min 98 °C for 20 s, 65 °C for 20 s, 72 °C for 1 min \(15 cycles) 72 °C for 2 min Keep at 4 °C c. Purify the reaction with QIAquick PCR Purification Kit and elute in 50 µL of water. d. Analyze DNA libraries on Agilent 2100 Bioanalyzer with DNA High Sensitive Kit. The expected size for amplified Input library is the total size of insert plus ~1 kb of the vector. The expected size for amplified cDNA library is the total size of the insert plus ~800 bp of the vector because of the splicing \(Figure 5). NOTE: Depending on the size distribution of the obtained DNA profile, it could be necessary to purify the expected DNA fragments after running on an agarose gel as described in Vanhille et al. 2015 2. _Index PCR_ Perform a PCR reaction to add sequencing indexes to the libraries using NEBNext® Multiplex Oligos for Illumina® \(Index Primer Set 1). a. Mix the following components: DNA library/Input: 2 ng Universal primer \(25 µM): 2.5 µL Index primer \(25 µM)# : 2.5 µL KAPA HiFi HotStart Readymix DNA Polymerase: 25 µL Sterile H2O: up to 50 µL Total volume: 50 µL \(#) If you are using the NEBNext Multiplex Oligos for Illumina \(#E7335, #E7500), for each reaction, only one of the 12 PCR indexes is used for the PCR. b. Place the tubes in Thermocycler and run the following program: 98 °C for 2 min 98 °C for 15s, 65 °C for 30 s, 72 °C for 30s \(10 cycles) 72 °C for 2 min Keep at 4 °C c. PCR Clean up with AMPure XP beads: add 1X volume of beads \(50 µL) to the PCR product and follow the cleanup protocol. At final step, elute DNA in 52 µL of water and carefully transfer 50 µL of the supernatant to a new 1.5 mL tube. d. Quantify DNA library and Input library on Agilent 2100 Bioanalyzer using DNA High Sensitive Kit. The distribution of libraries should peak at 450 bp \(Figure 5). I. **Library sequencing and processing \(Timing: 1-2 days)** 1. Perform sequencing using an Illumina high-throughput sequencing platform and standard single-end sequencing. We usually aim for coverage of 1k reads per captured regions. 2. Align sequenced reads to the adequate genome using standard tools. Do not remove multiple reads starting at the same coordinates. 3. Elongate each read accordingly to the average size of the library. J. **Quantification of enhancer activity** 1. For each captured regions quantify the number of Fragment Per Kilobase per Million mapped reads \(FPKM) in both Input and CapStarr-seq experiments. 2. Remove regions where FPKM in the Input sample is below 1. 3. For each region compute the FPKM ratio between CapStarr-seq and Input samples. 4. Determine à threshold for enhancer activity. If the experiments contains a set negative controls or randomly chosen regions, then the threshold can be estimated using a fold discovery rate approach. If not, threshold can be estimated by computing the inflexion point of the ranked FPKM ratios.

# References

1 Arnold, C. D. et al. Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science 339, 1074-1077, doi:10.1126/science.1232542 \(2013). 2 Muerdter, F., Boryn, L. M. & Arnold, C. D. STARR-seq - Principles and applications. Genomics, doi:10.1016/j.ygeno.2015.06.001 \(2015). 3 Vanhille, L. et

al. High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. Nature Communication 6, doi:10.1038/ncomms7905 \(2015).

# Figures

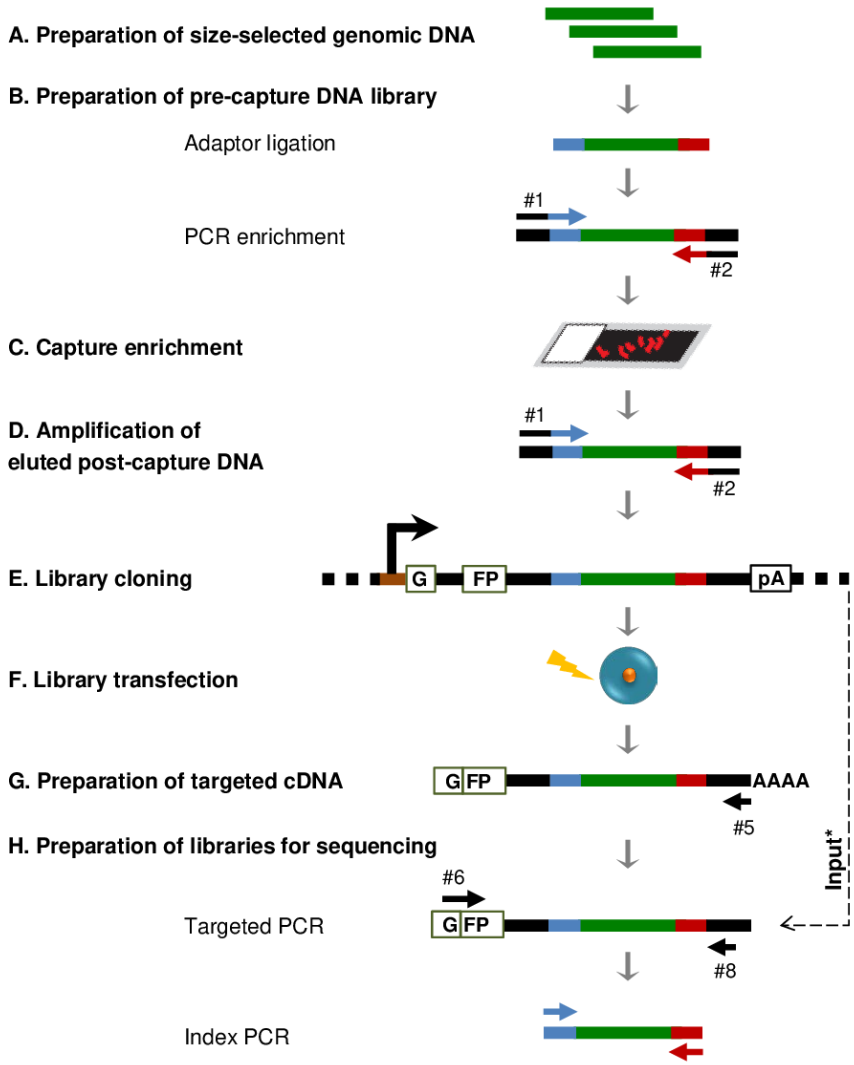


**Figure 1: Overview of the CapStarr-seq procedure.** See main text for details. Primers are listed in table 1. *For input amplification primer #6 is replaced by #7. The blue and red boxes indicate the Illumina adaptors.

## Figure 1

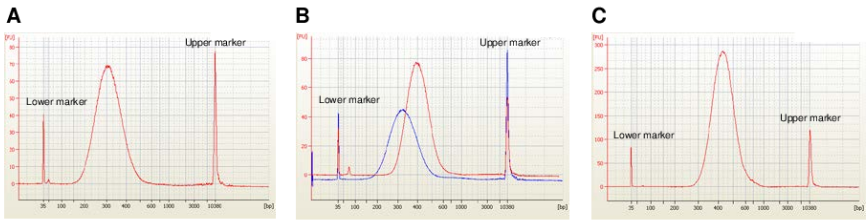


**Figure 2. Analysis of DNA size distribution.** DNA libraries were analyzed by Agilent 2100 Bioanalyzer using DNA  High Sensitive Kit. A) Size selected genomic DNA. B) Comparison of pre-capture library size distribution before (blue line) and after (red line) adding adaptors. A shift of 100 bp is observed. C) Amplified post-capture library.

## Figure 2

Analysis of DNA size distribution

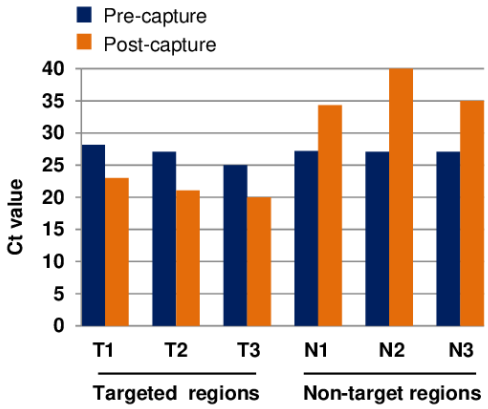


**Figure 3. qPCR-based QC of capture efficiency.** qPCR analyses of 3 targeted and 3 non-targeted regions using same amount of DNA from pre-capture and post-capture libraries.

## Figure 3

qPCR-based QC of capture efficiency

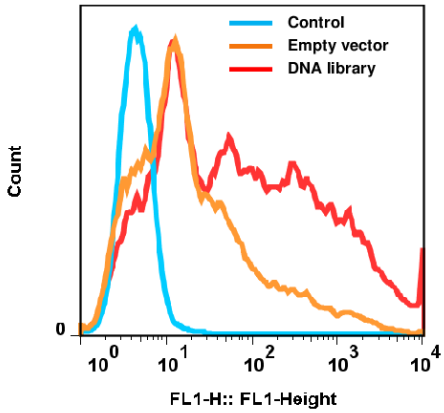


**Figure 4. FACS analysis.** Example of GFP expression in K562 cells. Cells were transfected with a CapStarr-seq library or the STARR-Seq empty vector. FACS analysis was performed after 24h of transfection. The increase of GFP expression indicated the potential activity of enhancers in the human promoter library. The shift of GFP expression in the empty vector transfected cells as compared with the non transfected cells (control) indicate basal promoter activity.

## Figure 4

FACS analysis

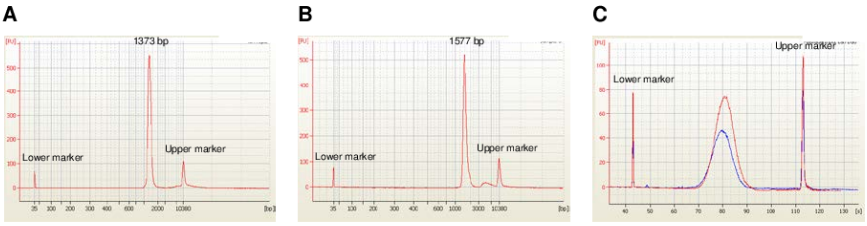


**Figure 5. Preparation of libraries for sequencing.** DNA libraries were analyzed by Agilent 2100 Bioanalyzer using DNA High Sensitive Kit. (A-B) Targeted PCR for cDNA library (A) and input library (B). A difference of ~200 bp in the average size distribution between the two libraries is observed. (C) Index PCR adding sequencing indexes for cDNA library (blue line) and input library (red line).

### Figure 5

Preparation of libraries for sequencing

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement0.pdf