Supplemental Materials

Challenges and considerations for reproducibility of STARR-seq assays

Maitreya Das[1, 3, 5*], Ayaan Hossain[4, 5], Deepro Banerjee[4, 5], Craig Alan Praul[5] and Santhosh Girirajan[1, 2, 3, 4, 5*]

1. Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA 16802
2. Department of Anthropology, Pennsylvania State University, University Park, PA 16802
3. Molecular and Cellular Integrative Biosciences Graduate program, Pennsylvania State University Park, PA 16802
4. Bioinformatics and Genomics Graduate program, Pennsylvania State University, University Park, PA 16802
5. Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA 16802

Table of Contents

Supplemental Text 3
Supplemental Methods 4
1. Read deduplication, alignment, and filtering 4
2. Data quality control 4
3. Peak calling 5
4. Activity comparison between peaks and exonic regions 5
5. Assessing reproducibility of published STARR-seq datasets 5
Limitations of the Perspective 6
Supplemental Figures 7
6. Supplemental Fig. S1 (A-E) 7
7. Supplemental Fig. S2 12
8. Supplemental Fig. S3 (A-C) 13
9. Supplemental Fig. S4 16
10. Supplemental Fig. S5 (A-C) 17
11. Supplemental Fig. S6 18
12. Supplemental Fig. S7 19
Supplemental Tables 20
13. Supplemental Table S1 20
14. Supplemental Table S2 22
15. Supplemental Table S3 (A-C) 24
Supplemental Protocol 27

Supplemental Text

STARR-seq experiment

We built a STARR-seq plasmid library spanning approximately 33 Mbp of the human genome and performed multiple STARR-seq runs on HEK293T cells. Our target regions were shortlisted from existing ChIP-seq data on HEK293 or HEK293T cells available on ENCODE (The ENCODE Project Consortium 2007). In brief, we first overlapped ChIP-seq sites binding to the histone modifications H3K27Ac (for active enhancers) and H3K4Me1 (for active or poised enhancers). Next, we intersected these regions with all TF-ChIP-seq sites on HEK293 (from ENCODE) to obtain a comprehensive enhancer catalog spanning 46,010 ChIP-seq sites. The selected regions comprise TF-binding sites that overlap with recognized enhancer marks. We captured our target library from commercially available human whole genome DNA using hybridization and capture probes and cloned the captured fragments into the human STARR-seq vector to build a STARR-seq plasmid library (see **Supplemental Fig. S1A-E**, **Supplemental Protocol**). To assess the impact of different genomic mutations on enhancer activity, we transfected the library into seven different mutant HEK293T lines and one wild-type line in three biological replicates and isolated reporter specific mRNA to build 24 STARR-seq output screening libraries. We directly amplified the STARR-seq plasmid library in three replicates for the input. We sequenced 24 output screening libraries and 3 input libraries using a NextSeq2000 sequencer with approximately 45 million reads per sample.

While conducting the assays we came across various design and protocol inconsistencies that enabled us to generate a list of STARR-seq design considerations, best practice guidelines and quality control checkpoints (see **Supplemental Tables**). Our analysis pipeline is provided as **Supplemental code** and on GitHub (links provided in **Supplemental Methods**). We used our data to demonstrate the effects of read filtering, significance of read depth, and compared different peak callers. All sequence data generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA879724. We also reanalyzed datasets from existing STARR-seq studies and assessed them for their data quality control steps and compared them to our data (see **Supplemental methods**). We also evaluated 24 STARR-seq studies and scored important assay features for each study based on the number of details reported by the authors (see **Supplemental Table S3A-C**).

Supplemental Methods

Read deduplication, alignment, and filtering

We used custom scripts to remove all PCR duplicates from raw FASTQ files of STARR-seq input and output libraries obtained after demultiplexing. We then aligned the paired-end reads to GRCh38 human genome assembly using BWA-MEM (Li 2013) with default settings. Next, we removed reads that were *(i)* unaligned, *(ii)* low quality (mapping quality score<30), *(iii)* multi-mapped (reads that mapped to multiple locations with equal confidence), and *(iv)* off-target using SAMtools (Li et al. 2009) with the following parameters: -F 2828 -f 2 -q 30.

To compare read loss across published datasets with our data, we reanalyzed ATAC-STARR-seq data from Wang and colleagues and human whole-genome STARR-seq data from Johnson and colleagues using our computational pipeline with modifications to the deduplication strategies (Johnson et al. 2018; Wang et al. 2018). We first aligned paired-end reads using BWA-MEM. Next, we filtered PCR duplicates using Picard (Broad Institute. 2019). Finally, we removed (i) unaligned, (ii) low quality (mapping quality score<30), (iii) multi-mapped (reads that mapped to multiple locations with equal confidence) reads and compared reads across the three datasets **(Supplemental Fig. S2)**. Analysis pipelines are posted on GitHub: https://github.com/deeprob/starrseq_dedup_align_filter.

Data quality control

We assessed data quality and replicability by calculating correlation of filtered read counts between input and output library replicates. Additionally, we calculated correlation of Reads per Million (RPM) normalized output-over-input fold changes between replicates **(Supplemental Fig. S3A-C)**. The output sequencing libraries were generated from the wild-type control HEK293T line. To compare quality of published datasets with our data, we assessed the correlation of filtered read counts between input and output replicates for our STARR-seq libraries and the reanalyzed datasets (Johnson et al. 2018; Wang et al. 2018) **(Supplemental Fig. S4)**. We also performed Principal Component Analysis (PCA) and visualized the first two principal components of input and output replicate filtered read counts for all reanalyzed datasets including our own **(Supplemental Fig. S5A-C)**.

Peak calling

We called peaks using previously published tools, MACS2 (Zhang et al. 2008), STARRPeaker (Lee et al. 2020), CRADLE (Kim et al. 2021) and DESeq2 (Love et al. 2014) using default settings. We merged the input and output replicates before peak calling using MACS2 and STARRPeaker. For CRADLE and DESeq2, the replicates were kept separate since they both utilize the variance between replicates to adjust p-value estimates of peaks. Before peak calling with DESeq2, we fragmented all regions using a sliding window of size 500 bp and a stride of 50 bp. Next, we calculated the input and output library coverage for each of these windows using BEDTools (Quinlan and Hall 2010). The input and output replicate-wise library coverage for each window was used by DESeq2 to identify differentially active regions. We used BEDTools to intersect CRADLE- and DESeq2-called "active" peaks with peaks called by other callers.

For comparing the number of peaks by varying library coverage, we randomly subsampled the input and output control libraries from 10% to 90% with increments of 5% using SAMtools (Li et al. 2009). For each subsample, we also created three replicates by changing random seed parameter. We called peaks in the subsampled libraries using STARRPeaker with the default settings. Our peak calling pipeline is posted on GitHub: https://github.com/deeprob/starrseq_peak_call.

Activity comparison between peaks and exonic regions

To compare activity between peaks and exonic regions, we first identified regions in our library which overlapped with known exonic regions from the reference human genome (GRCh38). Next, we calculated the RPKM normalized coverage of filtered reads fold changes between output and input libraries for both STARRPeaker called peaks and the identified exonic regions. Finally, we compared the fold change distributions between peaks and exonic regions using t-test **(Supplemental Fig. S6)**.

Assessing reproducibility of published STARR-seq datasets

To compare variation in enhancer activity for a given fragment between different STARR-seq studies, we used processed bigwig files from Johnson and colleagues and Lee and colleagues for their respective whole genome studies (Johnson et al. 2018; Lee et al. 2020). We used output over input fold change signals from the two whole genome libraries reported by Lee and

colleagues, submitted as bigwig files. Johnson and colleagues reported separate bigwig files for input and output read signals for their whole genome library. In this regard, we first normalized the input signal by Z-score normalization method. Next, we calculated fold change of output over normalized input. Finally, we measured Pearson and Spearman's correlation of fold changes between the three libraries **(Supplemental Fig. S7)**.

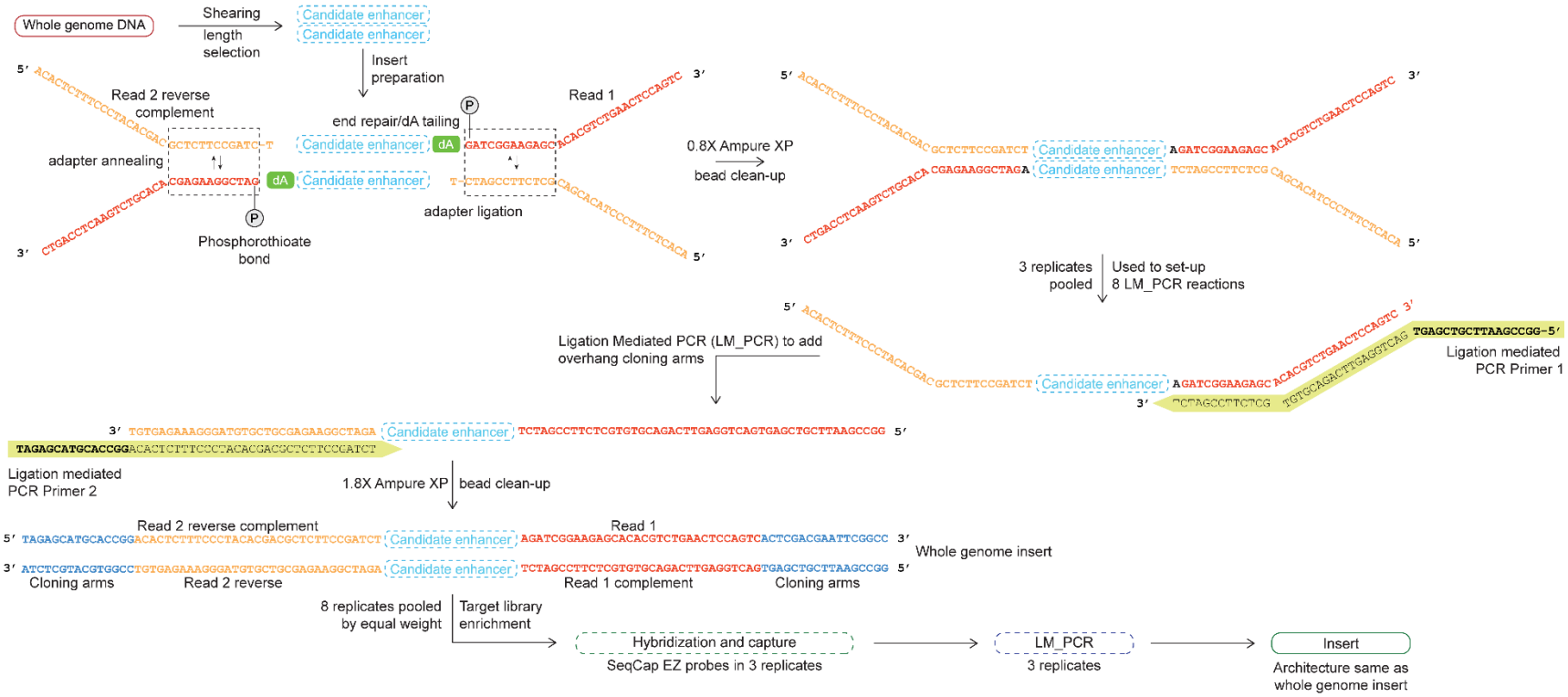
Limitations of the Perspective

We emphasize that the published and new data used in this manuscript is to demonstrate the various nuances associated with STARR-seq analysis. We excluded any biological result, interpretations or conclusions drawn from the data, as it is beyond the scope of this manuscript. However, we provide a brief description of the newly generated data, quality control assessments as well as detailed descriptions of all analyses conducted in this study, in addition to all raw and processed files for independent review.

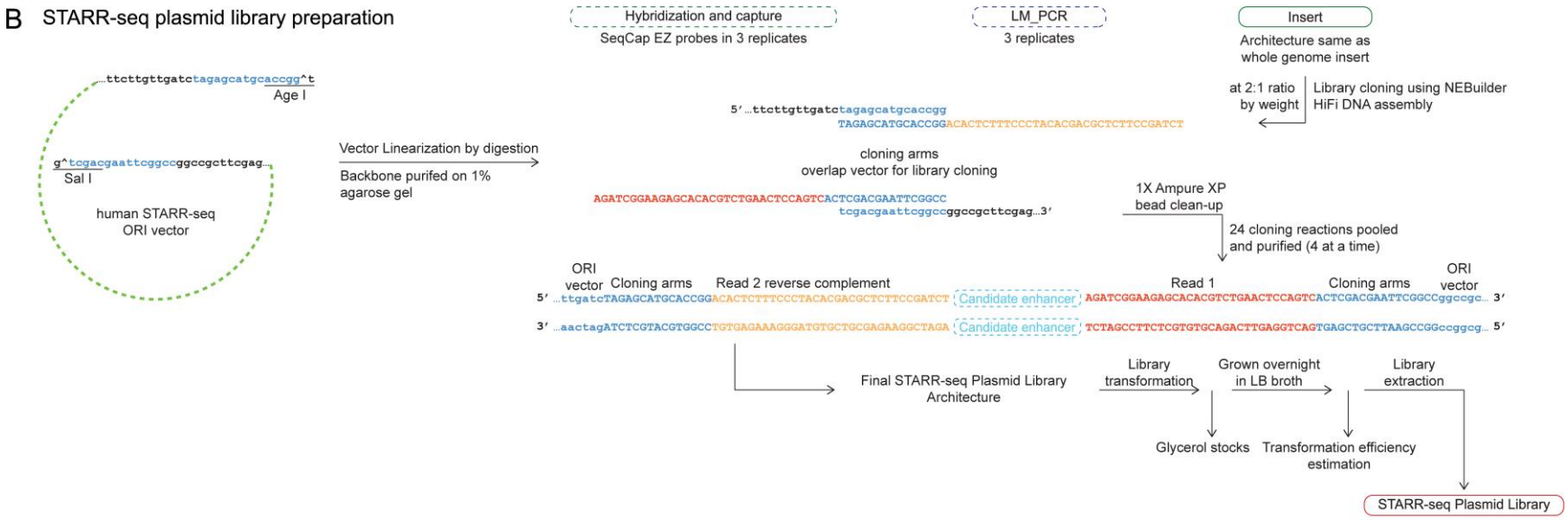
Supplemental Figures

Supplemental Fig. S1 (A-E): Schematic representation of a STARR-seq protocol highlighting all sequence information.

A Insert preparation



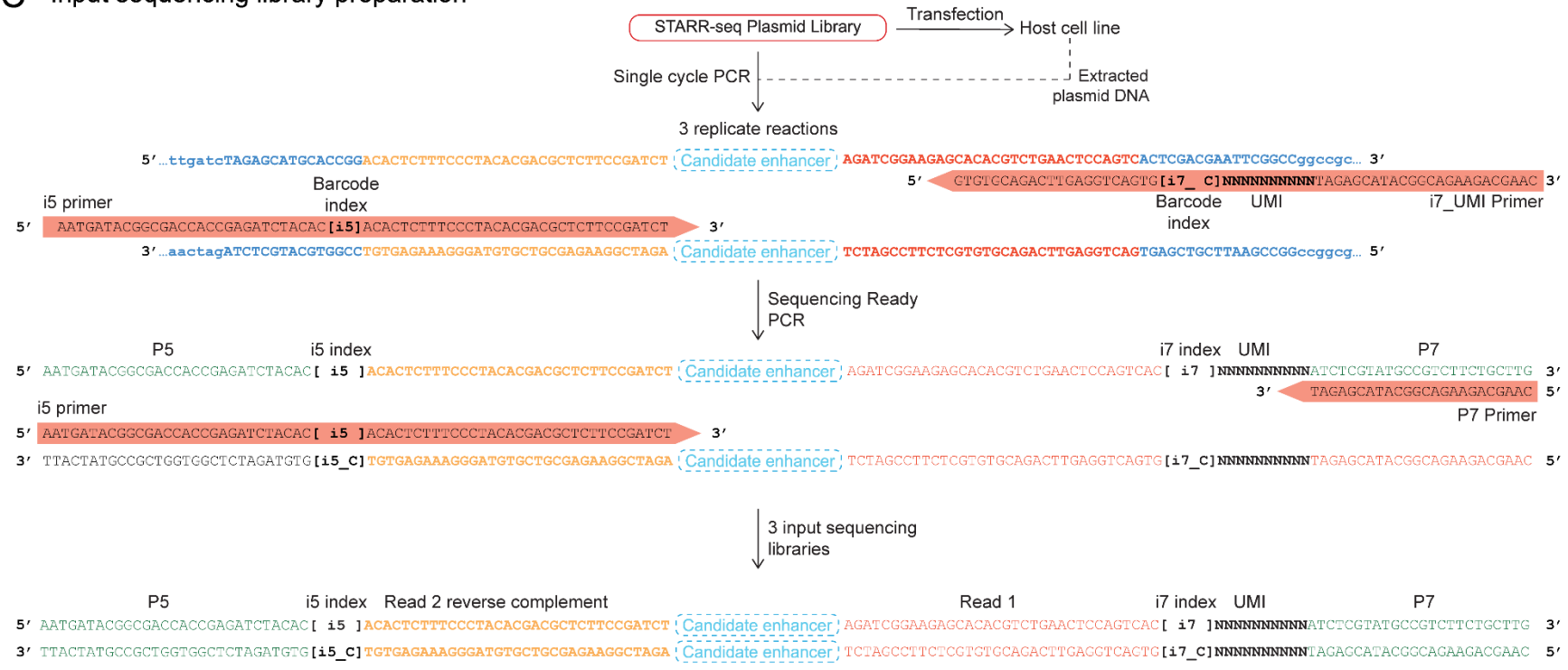
151


B STARR-seq plasmid library preparation
…ttcttgttgatctagagcatgcaccgg^t
Age I
g^tcgacgaattcggccggccgcttcgag…
Sal I
human STARR-seq
ORI vector
Vector Linearization by digestion
Backbone purifed on 1%
agarose gel
Hybridization and capture
SeqCap EZ probes in 3 replicates
LM_PCR
3 replicates
Insert
Architecture same as
whole genome insert
at 2:1 ratio
by weight
Library cloning using NEBuilder
HiFi DNA assembly
5′…ttcttgttgatctagagcatgcaccgg
TAGAGCATGCACCGGACACTCTTTCCCTACACGACGCTCTTCCGATCT
cloning arms
overlap vector for library cloning
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTCGACGAATTCGGCC
tcgacgaattcggccggccgcttcgag…3′
1X Ampure XP
bead clean-up
24 cloning reactions pooled
and purified (4 at a time)
ORI
vector
Cloning arms
Read 2 reverse complement
Read 1
Cloning arms
ORI
vector
5′ …ttgatcTAGAGCATGCACCGGACACTCTTTCCCTACACGACGCTCTTCCGATCT Candidate enhancer AGATCGGAAGAGCACACGTCTGAACTCCAGTCACTCGACGAATTCGGCCggccgc… 3′
3′ …aactagATCTCGTACGTGGCCTGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA Candidate enhancer TCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTGAGCTGCTTAAGCCGGccggcg… 5′
Final STARR-seq Plasmid Library
Architecture
Library
transformation
Grown overnight
in LB broth
Library
extraction
Glycerol stocks
Transformation efficiency
estimation
STARR-seq Plasmid Library

152

153

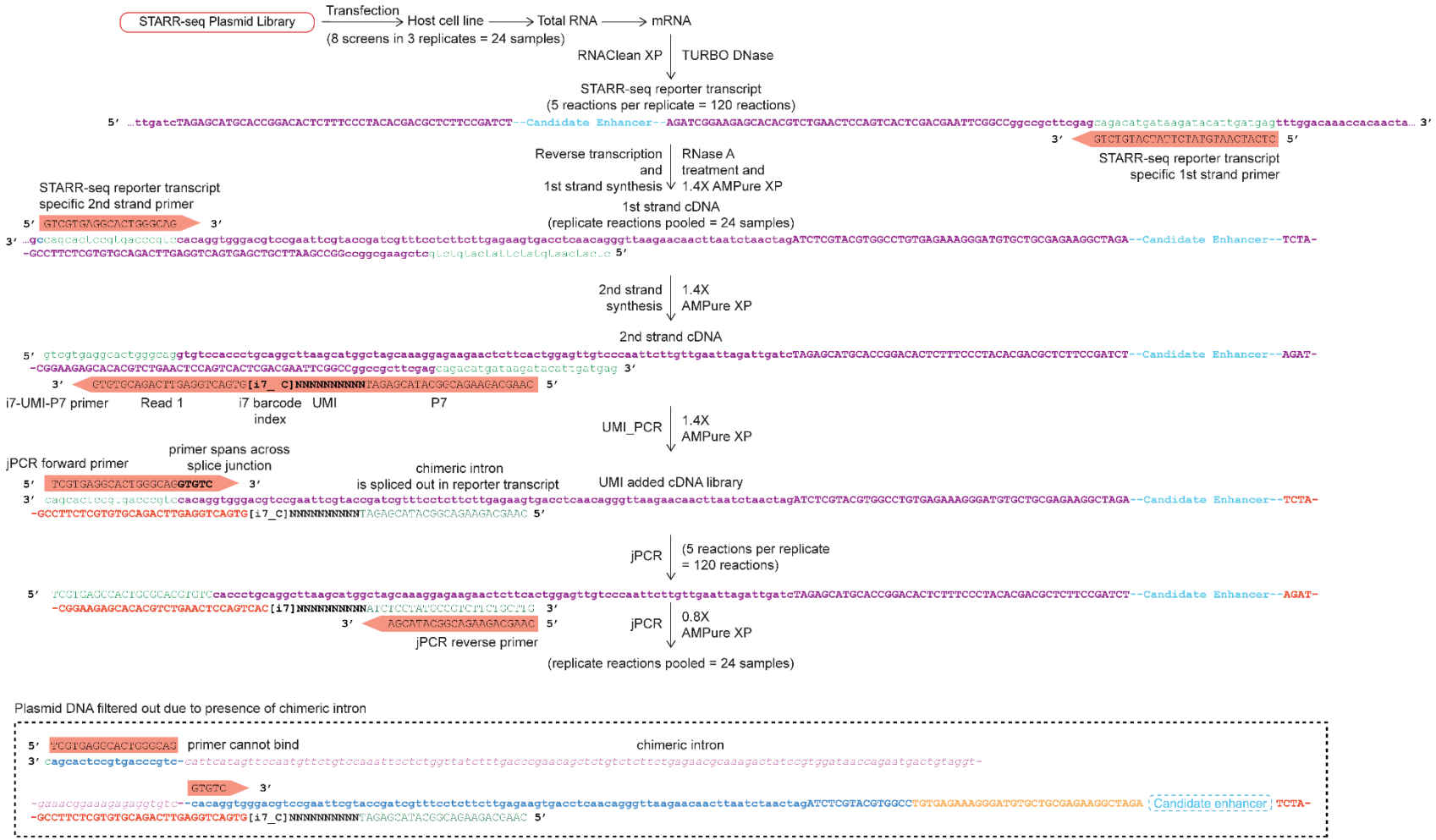
C Input sequencing library preparation



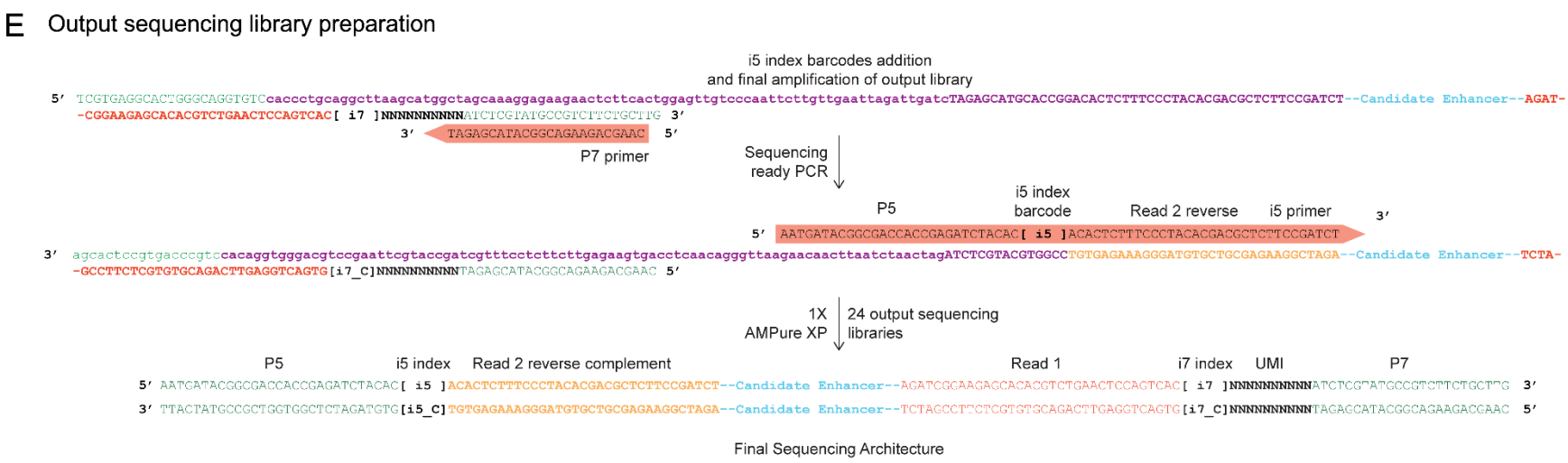
154 Final sequencing architecture

155

D cDNA library generation and UMI addition

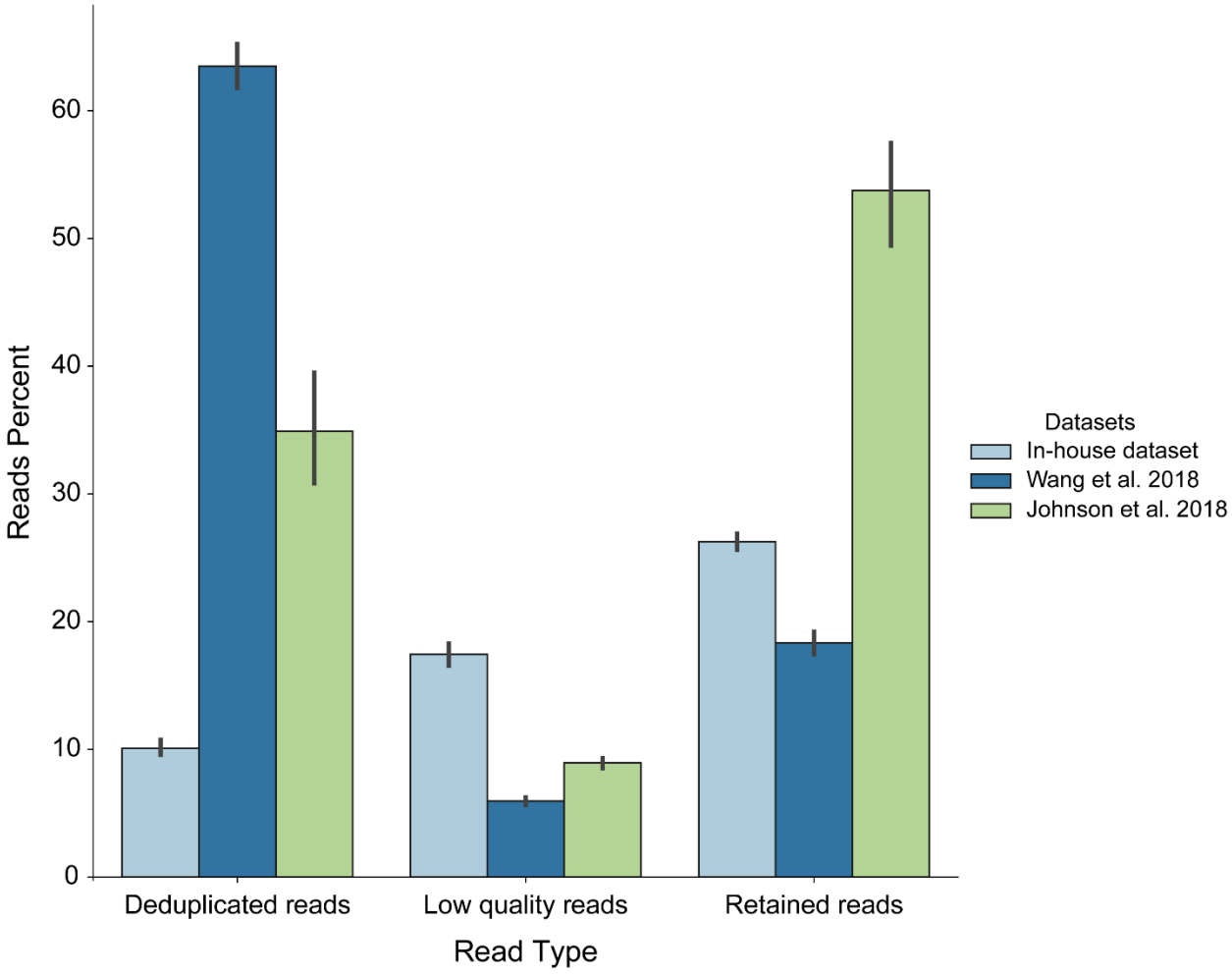


156



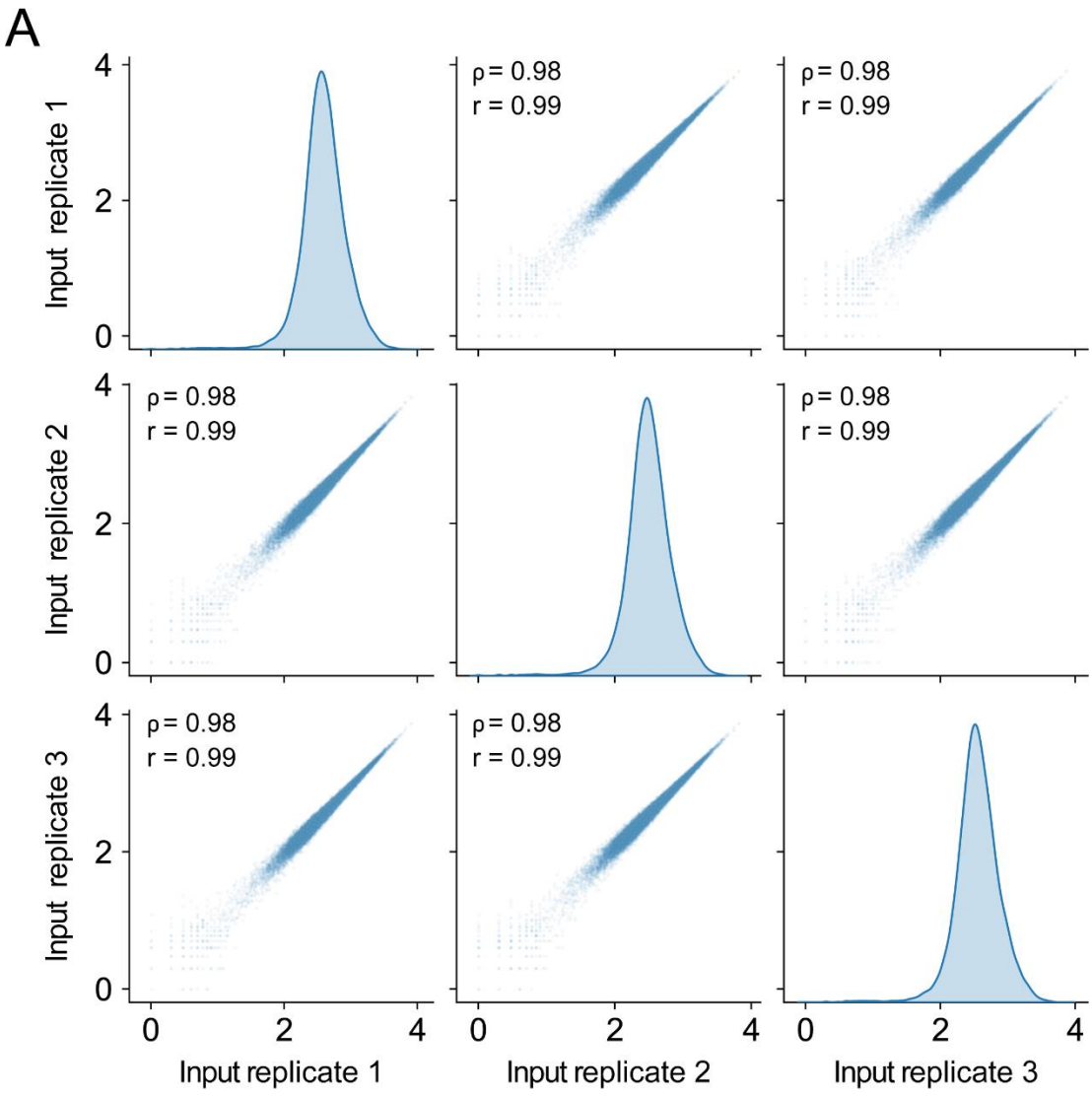
Supplemental Fig. S1: A schematic of STARR-seq experimental protocol and complete library sequence information are shown. **(A)** Insert preparation including addition of sequencing adapters and cloning arms to library fragments to facilitate sequencing and library cloning. **(B)** STARR-seq plasmid library preparation including vector linearization and cloning of library inserts into human STARR-seq vector (Muerdter et al. 2018) followed by library amplification through transformation. **(C)** Preparation of 'input' sequencing library either directly from plasmid library or from DNA extracted from library-transfected host cells. This step adds on UMIs and index barcodes to the library prior to sequencing to sort for PCR duplicates and to sequence multiple libraries on the same sequencing lane (multiplexing). **(D)** cDNA library generation for STARR-seq screening includes transfection of the plasmid library into a host cell and reverse transcription of self-transcribed reporter transcripts. This step also involves adding UMIs for detecting and removing PCR duplicates. **(E)** Preparation of 'output' sequencing library involves adding sequencing barcode indexes to the screening library before sequencing to enable multiplexing. Both 'input' and 'output' libraries are pooled and sequenced in parallel for enhancer screening.

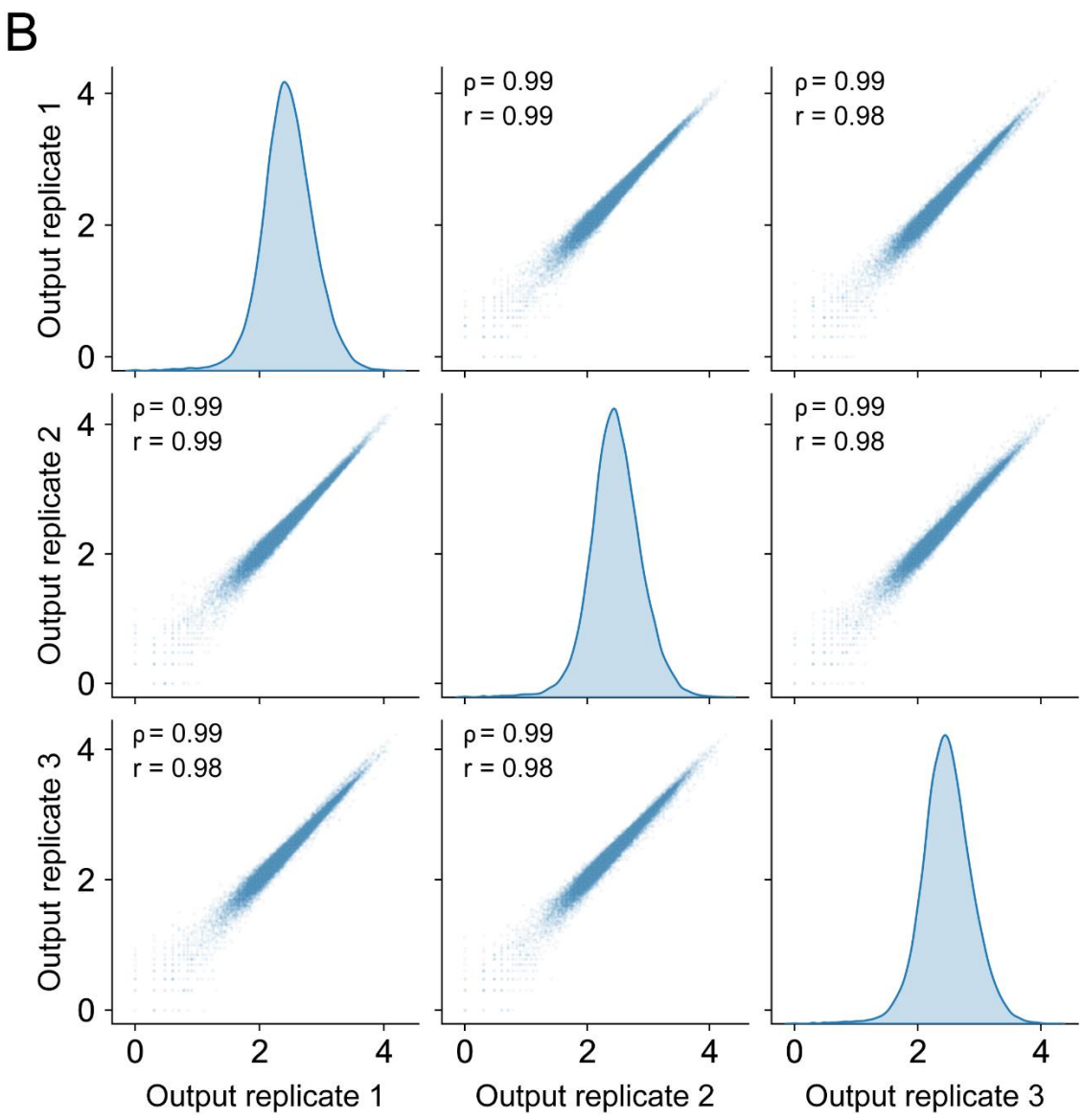
Supplemental Fig. S2: Read loss across published STARR-seq datasets.



Supplemental Fig. S2: Bar plots illustrating read loss observed across published STARR-seq datasets compared to our in-house STARR-seq assay are shown. Study by (Wang et al. 2018) and in-house STARR-seq includes focused assays, while (Johnson et al. 2018) includes data from a whole genome STARR-seq assay.

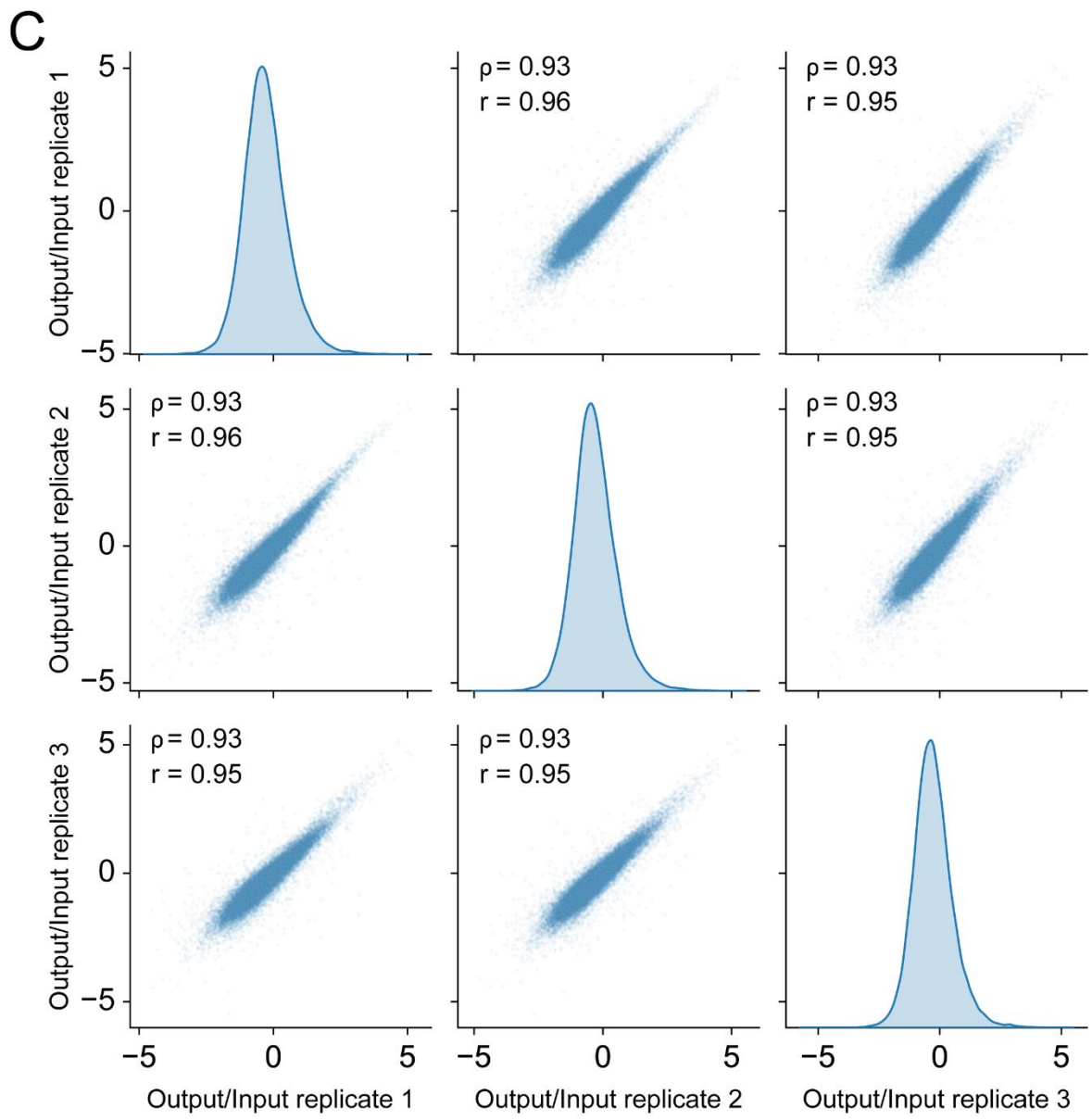
Supplemental Fig. S3 (A-C): Correlation across libraries for in-house input and output STARR-seq libraries are shown. Please note that this figure has three multi-figure panels – each in a page.




B
Output replicate 1
Output replicate 2
Output replicate 3
ρ = 0.99
r = 0.99
ρ = 0.99
r = 0.98
ρ = 0.99
r = 0.99
ρ = 0.99
r = 0.98
ρ = 0.99
r = 0.98
ρ = 0.99
r = 0.98
0
2
4

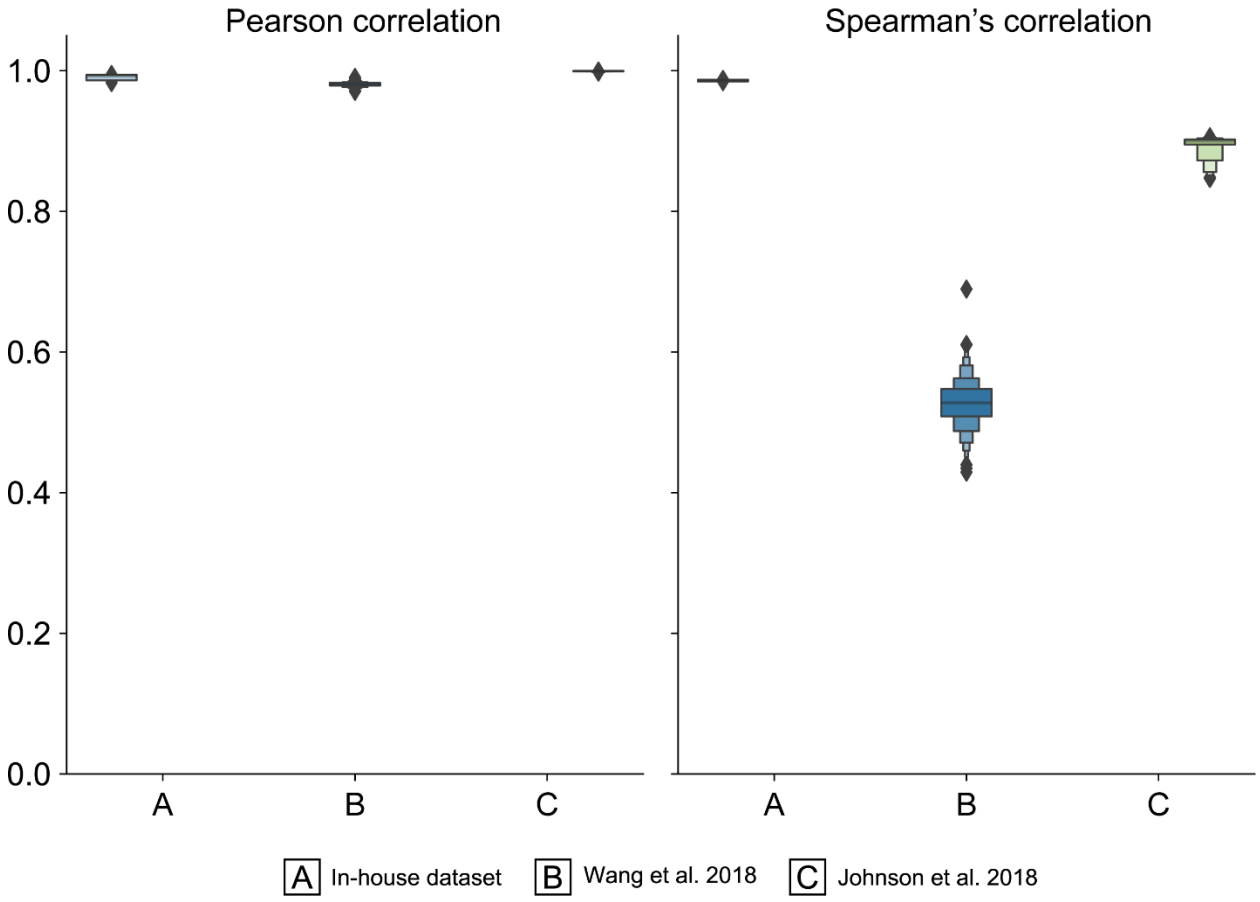
187

188



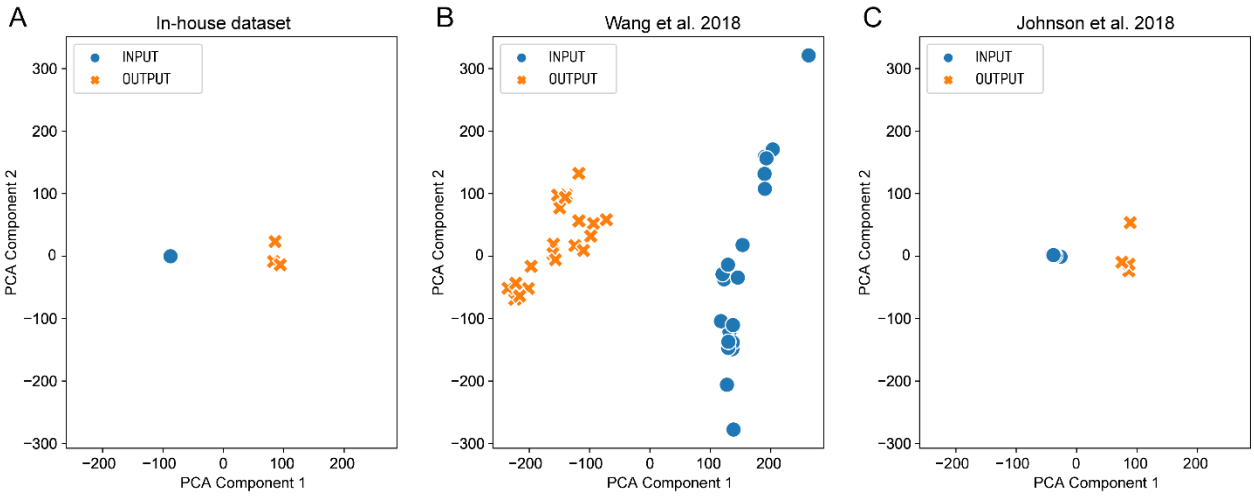
Supplemental Fig. S3: Scatter plots and correlation (Spearman's and Pearson co-efficient values indicated) across in-house STARR-seq libraries. **(A)** Scatterplots comparing log10 transformed raw read counts across three replicates of input sequencing libraries prepared from direct amplification of in-house STARR-seq plasmid library along with the observed correlation is shown. **(B)** Scatterplots comparing log10 transformed raw read counts across three replicates of output sequencing libraries along with the observed correlation is shown. The output library consists of the STARR-seq experiment conducted on control HEK293T cell line. **(C)** Scatterplots comparing $\log_2$ transformed output over input fold changes along with their observed correlation is shown. Input replicates assigned randomly to output replicates 1, 2 and 3.

Supplemental Fig. S4: Correlations across STARR-seq libraries



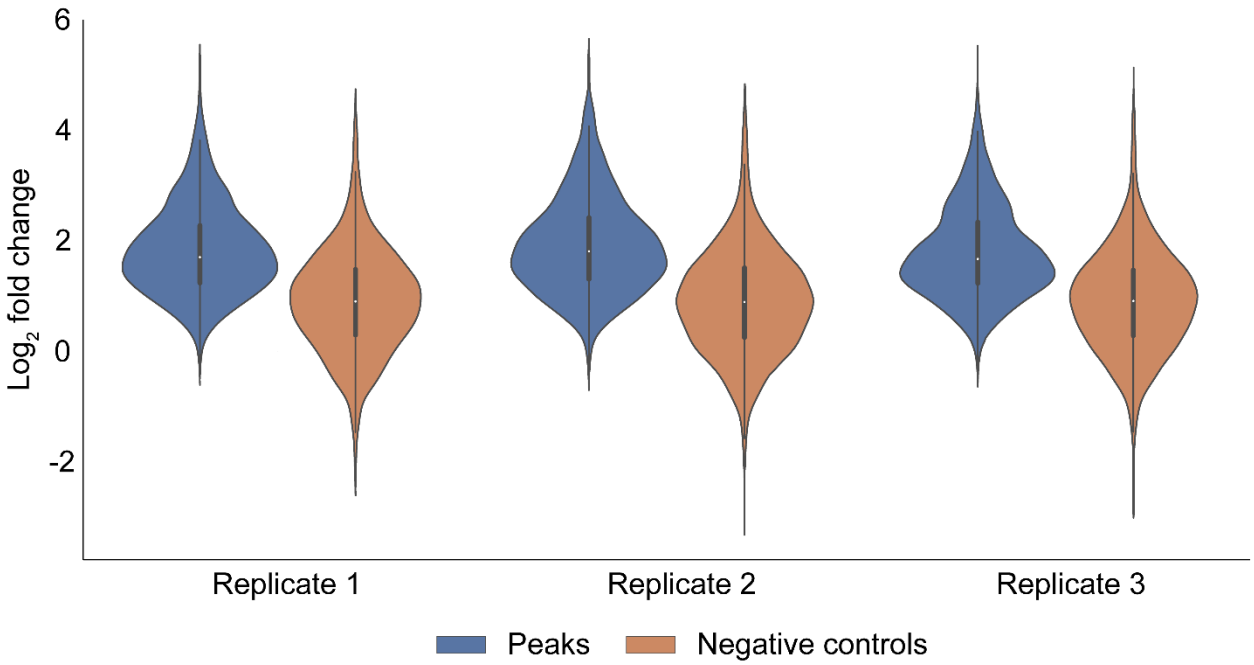
Supplemental Fig. S4: Comparison of Pearson and Spearman's correlation of input and library read counts across published studies (Wang et al. 2018; Johnson et al. 2018) and in-house STARR-seq assay is shown. Study by Wang and colleagues and in-house dataset are focused assays while Johnson and colleagues report a whole genome STARR-seq assay.

Supplemental Fig. S5 (A-C): Principal Component Analysis (PCA) for STARR-seq library replicates



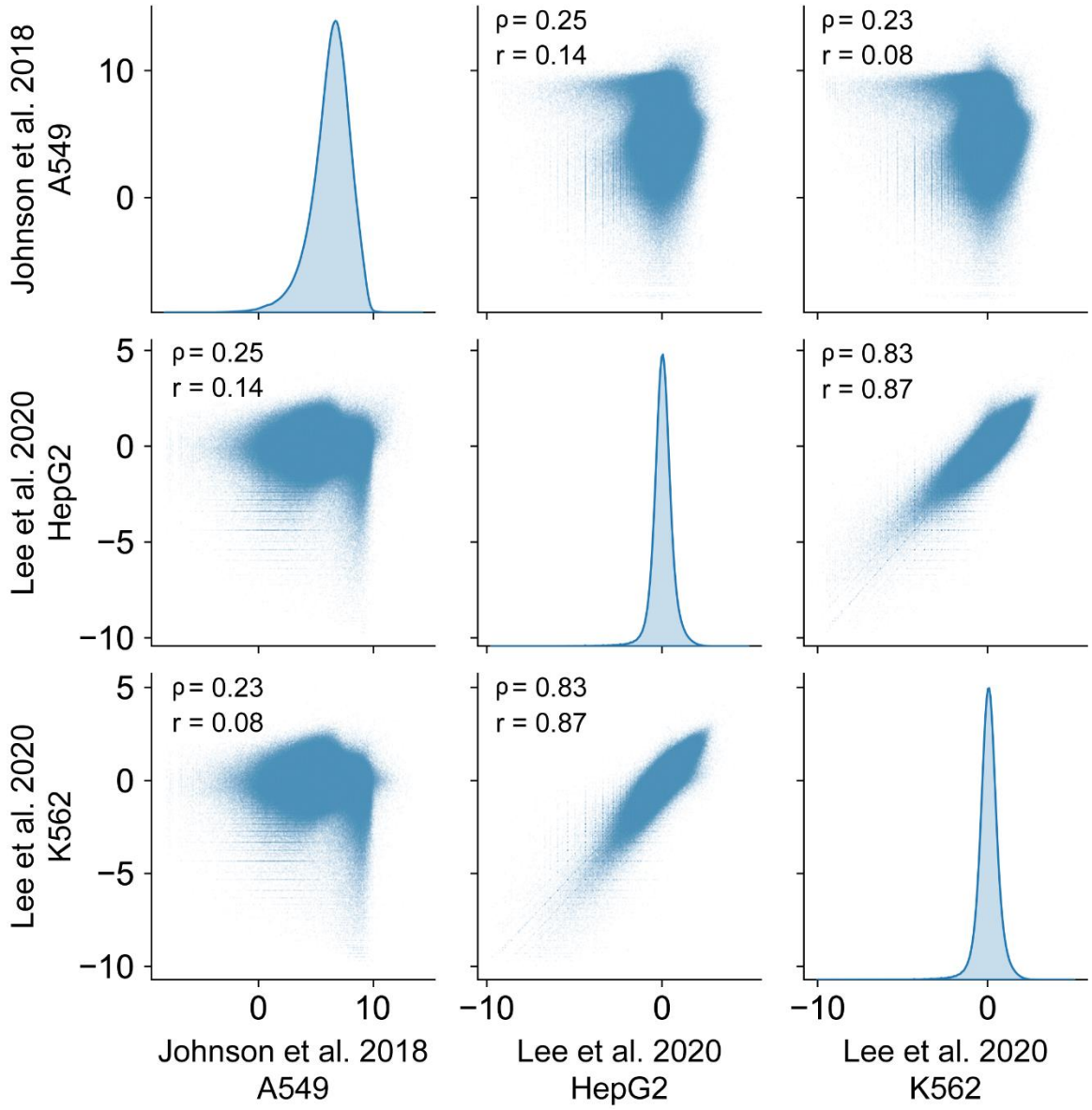
Supplemental Fig. S5: PCA plots demonstrating successful clustering of input and output STARR-seq library replicates for **(A)** Our in-house STARR-seq input and control library **(B)** Dataset from (Wang et al. 2018) **(C)** Dataset from (Johnson et al. 2018). Please note in (A) all three input library replicates (blue dots) cluster together.

Supplemental Fig. S6: Controls for STARR-seq activity



Supplemental Fig. S6: Violin plots comparing the log_2 fold change of normalized reads (Output over Input Reads Per Kilobase Million) observed between the STARRPeaker-called peaks and exonic regions from our in-house STARR-seq dataset as negative controls are shown. STARR-seq activity of exons was significantly lower than that of the peaks from enhancer regions (t-test statistic: 53.14; p-value: 0.0).

Supplemental Fig. S7: Correlation across published STARR-seq assays.



Supplemental Fig. S7: Spearman's and Pearson correlations for output over input $\log_2$ transformed fold changes for three whole genome STARR-seq assays are shown. Two assays were conducted for the same study (Lee et al. 2020) in K562 and HepG2 lines using the modified human STARR-seq vector containing the *ORI* promoter (Muerdter et al 2018). The third assay was conducted for a different study (Johnson et al. 2018) on A549 cells using the original STARR-seq vector (Arnold et al 2013).

Supplemental Tables

Supplemental Table S1: STARR-seq assay scaling

Library size	Fragment length range	Minimum number of fragments to cover target region	Number of transformation reactions	Reads for sufficient depth
Length of target region (in base pairs)	Determined by library size and biological question.	(Library size/frag length) Eg: Assuming upper limit of library size and lower limit of fragment length.	t = (n × c)/e t = number of transformations n = number of minimum fragments c = number of copies per fragment e = transformation efficiency (i.e., transformants per reaction)	r = n × x × z r = total number of reads x = minimum number of reads required per variant z = presumptive library dynamic range
Up to 100 kb	200-1200 bp	100 kb/200 bp = 500 fragments minimum.	t = (500×1000)/10^6 = 0.5 Hence a single transformation reaction would yield sufficient colonies	For both libraries assuming z = 100, x = 10. r = 500×10×100 = 500,000 reads for output library
100 kb-1 Mb	200-1200 bp	5000 fragments minimum	t = (5000×1000)/10^6 = 5 reactions	r = 5000×10×100 = 5,000,000 for output library
1-100 Mb	300-1200 bp	333,333 fragments minimum	t = (333,333×100)/10^6 = 33 reactions	r = 333,333×10×100 = 333,333,000 reads for output library
100-1000 Mb or Whole genome	400-1200 bp	2.5 million fragments for 400 bp Therefore, by increasing length, this can be reduced. Eg: 1000 bp fragments would require 1 million unique fragments.	t = (2.5×10^6×100)/10^6 = 250 reactions This can be further optimized to use cells with higher transformation efficiency. Example: a transformation efficiency of 10^7 would decrease the number of reactions to 25.	r = 2.5×10^6×10×100 = 2,500,000,000 reads for output library.

Supplemental Table S1: Proposed guidelines for assay scaling based on the library size, fragment length, and estimated read depth are shown. For read depth, if the dynamic range of the library (the ratio of read count between the most active fragment in the output library to the least active fragment) can be estimated, then the total number of reads in both input and output libraries is given by the product of the dynamic range, the minimum number of reads per fragment for it to be considered active, and the total number of fragments in the complete library. However, to uncover the true dynamic range in each output library, input and output libraries should be sequenced using the same number of reads. Additionally, after library construction and validation, studies should report complete library details including the length of the target region, number of unique fragments obtained, and the final sequence depth or fold coverage for each unique fragment after sequencing to enable assay reproducibility.

Supplemental Table S2: List of oligo sequences used for in-house STARR-seq assays.

Oligo name	Sequence (5' - 3')	Function
Adapter I	/5Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTC ()	A' -- Read 1 sequence -- 'A'
Adapter II	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	Reverse complement of Read 2
LM_PCR forward	TAGAGCATGCACCGGACACTCTTTCCCTACACGACGCTCTTCCGATC*T	Cloning overhang in bold; *: phosphorothioate bond
LM_PCR reverse	GGCCGAATTCGTCGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	Cloning overhang in bold
Blocking oligo I	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTCCGGT	For hybridization and capture
Blocking oligo II	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-Amm	For hybridization and capture; Amm: modification
1st strand primer_ORI	CTCATCAATGTATCTTATCATGTCTG	Reverse transcription of STARR-seq reporter specific transcripts
2nd strand primer_ORI	GTCGTGAGGCACTGGGCA*G	Reverse transcription of STARR-seq reporter specific transcripts

i7_UMI_P7 primer	CAAGCAGAAGACGGCATACGAGATNNNNNNNNNN[i7_RC]GTGACTGGAGTTCAGACGTGT*G	N: UMIs; i7: Unique Dual Index barcode (in bold) (NEB #E6440S manual); RC: reverse complement
jPCR forward	TCGTGAGGCACTGGGCAG*G*T*G*T*C	Primer extends over spice junction (in bold)
jPCR reverse	CAAGCAGAAGACGGCATACG*A	Binds to P7 sequence
i5 primer	AATGATACGGCGACCACCGAGATCTACAC[i5]ACACTCTTTCCCTACACGACGCTCTTCCGATCT	i5: Unique Dual i5 Index barcode (NEB #E6440S manual)
P7 primer	CAAGCAGAAGACGGCATACGAGA*T	Binds to P7 sequence

Supplemental Table S2: Sequence information for all oligos used for in-house STARR-seq assays is shown. Sequence oligos were ordered from Integrated DNA Technologies, Inc for use. Dilutions required for each oligo is provided in STARR-seq protocol.

Supplemental Table S3 (A-C): Rubric used for scoring based on details provided for each feature

Steps	Rubric for each score based on details provided				
	0	1	2	3	4
Library size and DNA source	No detail	Only mentions genome wide/focused. Doesn't connect with goal of experiment	Library target justified and answers research question	Provides details on building target library including source of DNA and enrichment (if focused)	Explains scientific logic and provides complete detail on choosing library size, type, DNA source, enrichment.
Length selection	No detail	Fragment length	Fragment length and method of size selection	Justification of length and selection	Validation of selection
Insert Preparation	No detail	Mentions steps involved	Adapter ligation kit and LM_PCR reaction details including primers	Adapter ligation protocol parameters and number of reactions	Complete information on adapter sequences, explanation for choice
Library cloning	No detail	Insert source and vector name	Kit/reagent used	Total number of reactions, pooling information and purification steps	Insert and vector ratio with concentration and amount and optimization steps
Transformation	No detail	Transformation methodology	Competent cell name	Amount of competent cell and ligated product per reaction and number of reactions	Validation of transformation, complexity assessment, QC and optimization
Plasmid Library QC	No detail	Validation of any 1 step or QC checkpoint	Validation of any 2 steps or report both QC checkpoints	Validation of any 3 steps or provide intermediate reports for both QC checkpoints	Validation of all steps and report complete details for both checkpoints

Supplemental Table S3A: Rubric used for scoring each feature of each study for assessing plasmid library information is shown (for **Figure 4 A, B**).

Transfection	No detail	Cell type	Transfection methodology	Transfection protocol	Transfection efficiency, validation and optimization
RNA Isolation	No detail	Different sub-stages of RNA (total RNA, mRNA prior to RT	RNA extraction kits used for each sub-stage execution	Number of cells used, amount of final RNA, and description of protocol of kits used	Optimization and validation of RNA extracted at each sub-stage, purity and RNA integrity
Reverse Transcription	No detail	Reaction kit, primer and parameters	Amount of starting material and number of reactions.	Explanation for use of kit parameters	Validation and purification of cDNA obtained and optimizations.
Library screening QC	No detail	Explanation and validation of any 1 step	Explanation and validation of any 2 steps or report QC checkpoint	Explanation and validation of any 3 steps or report intermediate data for QC checkpoint	Explanation and validation of all steps and provide complete details for checkpoint

Supplemental Table S3B: Rubric used for scoring each feature of each study for assessing library screening information is shown (for **Figure 4 C, D**).

Sequencing Library preparation	No detail	Parameters of sequencing ready PCR reactions	Kit detail and protocol details for sequencing libraries (input and output)	Number of reactions, replicate information	Index information, lane and pooling information, final validation and amount sent for sequencing
Read Information	No detail	Mention sequencing platform and GEO	Read length, type of sequencing	Adapter type, total number of reads	Complete adapter sequence and reads per replicate
Read QC	No detail	Read QC steps reported in manuscript or supplementary document	Read QC tools used along with their parameters, raw data processing commands and post processing commands (after peak call) if applicable reported in manuscript or supplementary document	Final read depth or coverage per library reported in manuscript or supplementary document	Read loss due to mapping quality, N bases and bad/off-target reads reported in manuscript or supplementary document and enough information to calculate read depth or library coverage by a reader
Data QC	No detail	Correlation of replicates within input and/or output libraries reported in manuscript or supplementary document	Correlation across replicates within input and output libraries and details about transfection efficiency reported in manuscript or supplementary document	Correlation of enhancers reads between replicates and/or correlation of fold changes between replicates reported in manuscript or supplementary document	Number of reproducible peaks called across replicates and/or by sub-sampling filtered read file, reported in manuscript or supplementary document
Data transparency	No detail	Raw data and final enhancer activity data uploaded to public repository e.g. GEO	Filtered/Processed reads (after read QC) uploaded to public repo OR source code provided in public repo to reproduce filtered read file	Intermediate QC data uploaded to public repo OR source code provided in public repo to reproduce intermediate QC files	Complete analysis pipeline starting from raw data processing to analysis post peak calling uploaded to public repo. The source code should be sufficient to reproduce final results and QC data at all intermediate steps

Supplemental Table S3C: Rubric used for scoring each feature of each study for assessing library sequencing information (for **Figure 4 E, F**).

Supplemental Protocol

STARR-seq Protocol (Girirajan lab)

This mammalian (human) STARR-seq protocol is based on the protocol reported by Muerdter and colleagues (Muerdter et al. 2018) and Neumayr and colleagues (Neumayr et al. 2019) from Dr. Alexander Stark's lab, as well as the SeqCap EZ HyperCap Workflow by Roche (Roche Sequencing Solutions, Inc, CA 94588, USA) along with various modifications and adaptations. The major steps of the protocol are provided below, followed by detailed description of the steps.

(1) STARR-seq plasmid library preparation

(A) Insert preparation: End repair and dA tailing
(B) Insert preparation: Adapter ligation
(C) Insert preparation: LM_PCR
(D) Hybridization and Capture
(E) Vector preparation: Vector culturing
(F) Vector preparation: Vector linearization
(G) Library amplification: Library cloning
(H) Library amplification: Transformation
(I) Library amplification: Library storage and extraction

(2) STARR-seq screening

(A) Culturing of cells
(B) Transfection
(C) Total RNA isolation
(D) mRNA isolation
(E) TURBO DNase treatment
(F) RNAClean XP treatment
(G) cDNA library preparation: 1st strand synthesis
(H) RNase A treatment and AMPure XP clean-up
(I) UMI addition: 2nd strand synthesis
(J) UMI addition: UMI_PCR
(K) Junction PCR
(L) Sequencing Ready PCR: Output library preparation
(M) Sequencing Ready PCR: Input library preparation

(1A) Insert preparation: End repair and dA tailing

Before starting:

- The first section of this protocol follows the SeqCap EZ HyperCap workflow by Roche. This protocol replaces the NEBNext Illumina library preparation protocol reported by Neumayr and colleagues (Neumayr et al. 2019) to incorporate hybridization and capture of the target library using SeqCap EZ Prime Choice XL probes (Roche catalog # 08247510001).
- The required reagents for this protocol include KAPA Hyper Prep reaction kit (Roche catalog #KK8500) that consists of End repair and A-tailing buffer, End repair and A-tailing enzyme, ligation buffer, DNA ligase and KAPA HiFi HotStart ReadyMix (2×). This kit also contains primer mixes for standard Illumina adapters however, use custom adapter sequences provided in supplementary table S2. Order additional polymerase for subsequent PCR steps KAPA HiFi HotStart ReadyMix (2×) (Roche catalog #07958935001)
- Prepare AMPure XP beads (Beckman Coulter catalog #A63881) for sample clean-up in 1.5ml Eppendorf tube (VWR catalog #87003-294) to equilibrate to room temperature for at least 30 mins.
- Human whole genome DNA (Promega catalog # G3041) is fragmented through sonication at core facility and then selected at specified length using blue pippin and fragment length distribution is verified on a bioanalyzer and provided in 5 tubes at ~ 33 ng/µl. Library is sheared and selected to ~ 500 bp.
- Take equal amounts (in ng) from each tube and pool to a total of 260ng DNA starting material for end repair and dA tailing.
- Starting DNA: 260 ng made up to 50 µl with ultra-water in 0.2 ml tubes (VWR catalog #20170-012) in 2 replicates.

End repair and dA tailing protocol:

1. Thaw reagents on ice. Assemble following reaction on ice. Mix thoroughly by pipetting and light tapping. Spin down and keep on ice.

	Reagent	Volume
1	KAPA End Repair & A-Tailing Buffer	7µl
2	KAPA End Repair & A-Tailing Enzyme mix	3µl
3	DNA + water	50µl
	Total	60µl

2. Incubate reactions on thermocycler (program: end_repair)

	Temperature	Time
1	20°C	30 mins
2	65°C	30 mins
3	4°C	hold

- Continue to adapter ligation without stopping.

(1B) Insert preparation: Adapter Ligation

1. Anneal adapter oligos: Resuspend adapter I and II sequences to 100 µM stock (All adapter sequences provided in supplementary table S2). Order fresh oligo sequences from IDT (Integrated DNA Technologies, Inc, IA-52241) for new library. Use NEBuffer2 (New England Biolabs catalog # B7002S).
2. Assemble following reaction in a 0.2 ml tube.

	Reagent/Oligo	Volume
1	Adapter I	10 µl
2	Adapter II	10 µl
3	NEBuffer 2	4 µl
4	Ultra-pure water	16 µl
	Total	40 µl (25µM)

3. Incubate reaction at 95°C for 5 mins on thermocycler. Keep on bench for 2 hours to cool slowly to room temperature.
4. Add 2 µl ultra-pure water + 3µl adapter mix for final conc at 15 µM for KAPA protocol. (5 µl adapter per reaction in the next step)
5. Thaw out adapter ligation reagents on ice and assemble following reaction in 0.2 ml tubes.

	Reagent/Oligo	Volume
1	Ultra-water	5 µl
2	KAPA Ligation Buffer	30 µl
3	KAPA DNA Ligase	10 µl
4	End Repair/dATail product (Step A)	60 µl
5	Adapters (15 µM)	5 µl
	Total	110 µl

6. Mix samples thoroughly by pipetting and spin down. Incubate in thermocycler (program: Adapter lig)

	Temperature	Time
1	20°C	15 mins

7. Clean-up reaction using 0.8× AMPureXP bead clean-up (protocol below).

AMPure XP Clean-up notes

- These are magnetic beads that bind to specific lengths of DNA depending on the ratio of bead volume and sample volume.
- Beads must be equilibrated to room temperature prior to use for at least 30 mins.
- Use only freshly prepared 80% ethanol (diluted in ultra-pure water) for wash steps.
- Ethanol wipe bench and magnetic rack prior to use.
- All further AMPure XP clean-up steps will follow the same protocol with different volume ratios and elution reagent type and volume.

AMPure XP Clean-up protocol:

1. Pipette appropriate amount of beads into sample tube. For 0.8× clean-up, pipette 88 µl beads into each 110 µl adapter ligation reaction and transfer mix to 1.5 ml Eppendorf tube.
2. Mix beads and sample thoroughly by vortexing for 10 secs and pipetting vigorously. Spin down the tube once finished.
3. Incubate mix for 5-10 mins at room temperature. Set-up magnetic rack during incubation and prepare 5 ml 80% ethanol.
4. Place tubes on magnetic rack and incubate for 5 mins (till a clear solution is observed). Carefully remove supernatant liquid without disturbing the beads. (Keep tube on the rack while discarding supernatant)
5. Add 200 µl 80% ethanol and incubate for 30 secs to 1 min and then remove promptly. Repeat this step. After 2 washes, leave tube on rack to dry.
6. If residue ethanol is scattered on the inner tube surface, spin down tube for 1-2 secs and then use P20 pipette to remove trace ethanol. Dry tube for 2-3 mins or till no liquid visible.
7. Do not over-dry. The beads should have a moist liquid coating and not crack up.
8. Remove tube from rack and add 40 µl 10mM Tris-HCl at pH 7.5 for elution. Mix thoroughly with beads and vortex lightly and spin down.
9. Incubate for 5 mins. Place tube back on magnetic rack. Wait till clear solution obtained and pipette solution into a fresh 1.5 ml Eppendorf tube.
10. Measure sample purity and concentration. Typical concentration observed is ~ 40 ng/µl but purity ratios will be high (>3 for both 260/280 and 260/230).
11. Use 3 µl adapter ligated DNA for subsequent LM_PCR step.

(1C) Insert preparation: Ligation Mediated PCR (LM_PCR)

- This step adds overhang arms to the adapter ligated fragments to facilitate library cloning.
- Use custom primers (LM_PCR forward and LM_PCR reverse) designed for human STARR-seq ORI vector (Muerdter et al. 2018) as provided in supplementary table S2.
- Order fresh primers from IDT for new library. Reconstitute oligos to 100 µM.

LM_PCR protocol:

1. Dilute primers and make primer pool according to the following:

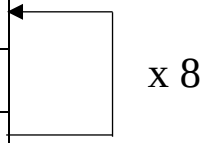
	Reagent/Oligo	Volume
1	LM_PCR forward	2 µl
2	LM_PCR reverse	2 µl
3	Ultra-pure water	16 µl
4	Total	20 µl

2. Thaw out LM_PCR reagents on ice and assemble following reaction on ice in 0.2 ml PCR strip tubes.

	Reagent/Oligo	Volume
1	Adapter ligated DNA (Previous step)	3µl
2	Primer Mix	5µl
3	Ultra-water	17µl
4	KAPA HiFi HotStart Ready Mix	25µl
	Total	50µl

3. Gently mix and spin down and incubate in Thermocycler (Protocol: LM_PCR)

	Temperature	Time	Step
1	98°C	45 secs	Initial denaturation
2	98°C	15 secs	Denaturation
3	65°C	30 secs	Annealing
4	72°C	30 secs	Extension
5	72°C	60 secs	Final extension
6	4°C	Hold	



4. Clean-up reactions using 1.8× AMPure XP bead clean-up. (Use 90 µl beads for each 50 µl reaction. Elute samples in 30 µl of ultra-pure water.

5. Measure concentration and purity. Typical yield is ~ 10-20 ng/µl.

6. **Checkpoint 1:** Assess samples on bioanalyzer. Send 3 µl of each sample to assess purity and fragment length distribution. Fragments should show a steady increase in length based on the size of the adapter sequences and serves as the first validation of library preparation. This also shows presence of adapter dimers or primer dimers as well as potential PCR biases.

7. Samples can be stored at -20°C.

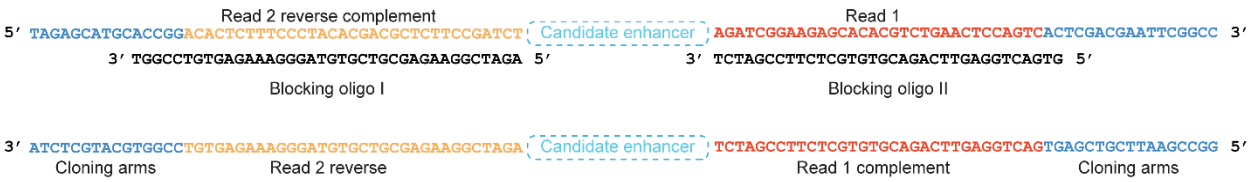
8. For whole genome libraries, repeat several replicates of LM_PCR as directed in protocol reported by Neumayr and colleagues (Neumayr et al 2019).

9. For focused libraries using hybridization and capture of target sites, perform at least 8 replicates. Assess each replicate for quality and length distribution.

10. Pool equal amounts from each replicate for hybridization and capture.

(1D) Hybridization and Capture

- This step allows for selectively capturing target library using custom designed hybridization and capture probes. The required kits needed for this include
  (1) SeqCap EZ Prime Choice XL probes (Roche catalog # 08247510001)
  (2) SeqCap EZ Hybridization and Wash kit (Roche catalog # 05634261001)
  (3) SeqCap EZ Pure Capture Bead Kit (Roche catalog # 06977952001)
  (4) Blocking Oligos custom designed (IDT), sequences provided in supplementary table S2
  (5) COT Human DNA (Sigma Aldrich catalog #11581074001)
- Upon receipt of probes, immediately aliquot into 4.5 µl aliquots in 0.2 ml tubes and store at -20°C. Each aliquot will be sufficient for 1 capture reaction.
- To prepare blocking oligos, reconstitute oligos to 400 µM. Use 2.5 µl (1000 pmol) of each for final reaction. Here is a schematic outline on the role of blocking oligos during hybridization and capture.



- Prepare pooled LM_PCR product from multiple replicates and measure final concentration and purity. Starting material should be ~ 1 µg of DNA.
- Prior to starting, thaw capture reaction aliquot on ice.
- Bring AMPure XP beads to room temperature for at least 30 mins for clean-up.

Hybridization protocol:

1. Prepare following reaction mix in a 1.5 ml Eppendorf tube.

	Component	Volume
1	COT Human DNA	5 µl
2	Blocking Oligos	2.5 + 2.5 µl
3	LM_PCR Pool	1 µg

2. Calculate total volume of the mixture and add 2× AMPure XP beads to mix and vortex for 10 seconds and incubate at room temperature for 10 mins.
3. Place tube onto magnetic rack and discard supernatant when it clears.
4. Wash beads with 190 µl 80% ethanol for 30 secs and remove and dry beads for 5 mins.

5. For each reaction, prepare a hybridization buffer mix:
   7.5 µl Hybridization Buffer + 3 µl Hybridization Component A = 10.5 µl
6. Add hybridization buffer mix to beads (per reaction) and vortex and incubate up to 2 mins and place on rack.
7. Pipette complete 10.5 µl solution from tube into fresh 200 µl tube containing 4.5 µl of SeqCap EZ Prime Choice XL probes and mix thoroughly.
8. Incubate on thermocycler (program: Hybridization)

	Temperature	Time
1	95°C	5 mins
2	47°C	20 hours

Wash and recover of captured library protocol:

- Follow SeqCap EZ HyperCap workflow (User's guide v2.3 pages 29 – 34) according to manufacturer's instructions completely for capturing target library from hybridized beads.
- Dilute all buffers according to the protocol.
- Prior to starting, allow capture beads and AMPure XP clean-up beads to equilibrate to room temperature for at least 30 mins.
- Proceed directly to post capture LM_PCR reaction. All reaction parameters and primers are same as previous LM_PCR.
- Use sample and beads as starting material for LM_PCR (~ 20 µl).
- Perform 1.8× AMPure XP clean-up of LM_PCR product. Elute in 50 µl ultra-pure water. Repeat AMPure XP clean-up step and elute in 40 µl. This significantly improves purity ratio of final product.
- Product now ready for library cloning.

(1E) Vector preparation: Vector culturing

Use human STARR-seq vector ADDgene #99296 as reported by Muerdter and colleagues (Muerdter et al. 2018). The hSTARR-seq_ORI vector was a gift from Alexander Stark (Addgene plasmid # 99296 ; http://n2t.net/addgene:99296 ; RRID:Addgene_99296)

1. Use all standard bacteria culture protocols for preparation of LB broth, LB-agar plates and antibiotic stocks. Prepare fresh broth and plates for use.
2. Prepare LB plates along with both ampicillin (10mg/ml stock) and chloramphenicol (25 mg/ml in ethanol stock) according to standard lab protocol. (final ampicillin concentration is 100 µg/ml and chloramphenicol is 25 µg/ml in plate) (Vector carries both resistance markers)
3. Streak LB plate using a sterile pipette tip with inoculum from vector stab and incubated overnight at 37°C to grow isolated colonies.
4. Following day, pick multiple colonies using sterile pipette tips and inoculate 4 ml LB broth with ampicillin and chloramphenicol in 14 ml bacteria culture tube (VWR catalog #60819-524) (1 culture tube per colony) and shake at 300 rpm at 37°C overnight in shaker-incubator.
5. Use 1 ml of culture broth to make glycerol stock for each colony (Use standard procedure for glycerol stock preparation). Spin down rest of the cultures and extract vector DNA using ZymoPURE plasmid Miniprep kit (Zymo catalog #D4209).
6. Measure purity and concentration and send 2 samples for sanger sequencing at the genomics core facility using multiple primers designed for ORI vector sequence verification to verify sequence. (Sequences available on request)
7. Following sequence verification regrow sample with highest concentration and purity from the glycerol stock. In the morning, pick glycerol stock with a sterile pipette and inoculate 2 ml of LB broth with ampicillin and chloramphenicol and incubate at 300rpm at 37C for 8 hours as a starter culture. In the evening, use 1 ml of starter culture to incubate 150 ml of LB broth with ampicillin and chloramphenicol and incubate overnight at 300 rpm at 37°C.
8. Make fresh glycerol stocks using 1ml of culture. Spin down rest of the culture and extract vector DNA using ZymoPURE II plasmid Midiprep kit (Zymo catalog #D4200) and verify concentration and purity.

(1F) Vector preparation: Vector linearization

- Reagents required include enzymes SalI-HF (NEB catalog #R3138S) and AgeI-HF (NEB catalog #R3552S) along with supplied Cutsmart buffer and 6× purple loading dye.
- Perform gel extraction using Zymoclean Gel DNA Recovery Kit (Zymo catalog #D4001). Further PCR purification or sample concentration can be carried out using DNA clean and concentrator kit (Zymo #D4003).
- Bring AMPure XP beads to room temperature for at least 30 mins.
- Prepare 1% agarose gel.

Vector linearization protocol:

1. Setup following reaction for each digest. Set-up 8 such reactions in 0.2ml PCR strip tubes and incubate in thermocycler at 37°C for 2 hours. Approximately 1/3rd of the product will be recovered from the gel so scale reactions accordingly for larger libraries.

	Digest mix	Volume
1	Vector	~ 1µg
2	SalI	1µl
3	AgeI	1µl
4	Cutsmart	5µl
5	Water	Up to 50µl
	Total	50µl

2. Run on 1% agarose gel for 1 hour at 100V till bands separate and cut out heavier band (~2kb) and weigh each slice. Add 3 volumes of ADB buffer and melt gel at 55°C for 30 min. Column purify each gel slice separately and pool 4 samples together and measure concentration and purity. (Purity may be low).

3. Perform secondary clean-up with 1× AMPure XP beads for each pooled sample and confirm concentration. (Should be above 35ng/µl)

(1G) Library amplification: Library cloning

- For cloning, use the NEBuilder HiFi DNA Assembly Master Mix (NEB catalog #E2621L) with the 2-3 fragment assembly protocol for efficient cloning. Alternatives include In-fusion HD cloning and Gibson Assembly.
- Try multiple ligation ratios to optimize library cloning. Repeat cloning using multiple replicates to maximize library complexity.
- Use DNA clean and concentrator kit or AMPure XP beads for reaction clean-up prior to transformation.

Library cloning protocol:

1. Calculation of [insert: vector ratio] for ligation

Considerations:
a) Vector and insert mass and associated molarity
b) Maximum molar capacity per reaction (30 fmol – 200 fmol per reaction for ligating 2-3 fragments)
c) Final volume of reaction
d) Number technical replicates to be performed

For calculations:

- Insert length: Use average insert fragment length as observed on bioanalyzer in step C.
- Vector length: 2543 bp

Ratio calculations can be done using the NEBioCalculator (https://nebiocalculator.neb.com/#!/ligation) to calculate amount of DNA for insert and vector required for final reaction assembly. Example parameter calculation for cloning:

Starting material:

DNA	conc ng/ul	260/280	260/230
Insert_1	29.8	1.81	2.32
Insert_2	25.7	1.81	2.22
Vector	38.2	1.87	2.33

Cloning parameters:

Ratio	Vector (ng)	Vector (fmol)	Insert (ng)	Insert (fmol)	total fmol
1:2	100	63.64	51.91	127.3	190.94
1:2	50	31.82	25.95	63.62	95.44
1:2	70	44.54	36.34	89.1	133.64
1:7	39.29	25	71.38	175	200

2. Set up cloning reactions according to the following volumes. Perform at least 4 reactions for each condition.

	Sample	Amount (volume)
1	Insert	X (depends on ratio)
2	Vector	Y (depends on ratio)
3	NEBuilder HiFi Assembly mix	Up to 10ul
4	Water	10
5	total	20 ul

3. Incubate reactions at 50°C for 1 hour. Pool 4 reactions of same conditions and purify using 1× AMPure XP beads and measure concentration and purity.

Ligation reactions	conc ng/ul	260/280	260/230	Ratio	# Reactions
HiFi Assembly_1	24.1	1.92	1.5	1:7	4
HiFi Assembly_2	36	2	1.87	1:2	4
HiFi Assembly_3	33.4	1.87	1.48	1:2	4
HiFi Assembly_4	31.8	1.88	2.22	1:2	4
HiFi Assembly_5	42.5	1.91	2.21	1:2	4
HiFi Assembly_6	41.6	1.87	2.21	1:2	4
HiFi Assembly_7	42	1.89	2.27	1:2	4

(1H) Library amplification: Library transformation

- Use either NEB5 alpha (NEB #C2989, currently discontinued) and NEB10 beta (NEB #C3020K) electrocompetent cells. Following protocol has been tested with only NEB electrocompetent cells.
- Perform test transformation for each cloning ratio in 25 µl competent cells. Vary between 2 – 5 µl of cloned product per transformation reaction.
- For control, use supercoiled plasmid of choice, either plasmid provided with NEB electrocompetent cells (pUC19 Vector at 50 pg/µl or plasmid PX459 (Ran et al. 2013) (ADDgene catalog #62988). Vector pSpCas9(BB)-2A-Puro (PX459) V2.0 was a gift from Feng Zhang (Addgene plasmid # 62988 ; http://n2t.net/addgene:62988 ; RRID:Addgene_62988) For PX459, dilute stock to 1:10 and use 2 µl per transformation.
- Prior to starting: Electroporation is a very time sensitive step and must be carried as quickly and efficiently as possible. Make sure everything is kept ready, labeled and within reach. After cells have thawed out, work as quickly as possible to proceed. If there are a large number of reactions, perform and thaw in batches of 4 or 8 reactions.
- Prepare adequate number of sterile LB-agar ampicillin plates according standard laboratory protocol for estimating cloning efficiency and CFU/µg. Prepare >20 plates from 500 ml LB-agar. Pre-warm to 37°C during transformation.
- Prepare adequate amount of sterile LB broth with ampicillin for inoculating and amplifying transformed library. Prepare 500 ml to 1 L LB broth depending on number of transformations performed. Use >600 ml for 100 µl (4 reactions) to 200 (8 reactions) µl of competent cells.
- Pre-chill electroporation cuvettes 200 µl and 20 µl tips overnight (or at least 1 hour) at 4°C.
- Tightly pack ice into an ice bucket and spray with ethanol to make a ‘chilled ice sludge’. Move cuvettes into ice bucket.
- Pre-warm SOC media provided with electrocompetent cells to 37°C for 1 hour and place back on bench.
- Take out required number of sterile 1.5ml Eppendorf tubes (1 for each transformation reaction) and place on ice to chill.
- Thaw out electrocompetent cells on the ice. Gently flick tube to check thawing.
- Label each tube with name of transformation reaction.
- Label 14 ml round bottomed bacterial culture tubes with similar labels as tubes.
- Make sure to work inside a sterile environment (under a lamp or inside a biosafety cabinet). Wipe everything down with ethanol prior to starting transformation.
- Ensure electroporation machine is within reach and set to correct parameters depending on competent cells. Use 1700V for NEB5 alpha or NEB10 beta cells.

Transformation procedure:

1. After cells have completely thawed out, pipette out 25 µl of cells onto the pre-chilled and labeled Eppendorf tube (DO NOT pipette more than once or necessary). Use pre-chilled tips while pipetting.

2. Add appropriate amount of cloned product (2 – 5 µl) to cell. Gently mix by flicking tube from side by holding the tube around the cap a few times. Use pre-chilled tips. (High number of unique colonies observed for 5 µl of product in 25 µl of cells)
3. Transfer mix of cells and cloned product into electroporation cuvette without any bubbles. Wipe cuvette from the side and electroporate cells with 2 pulses. Record the time constants for each transformation. (Typical range is 4.4 – 4.6 milliseconds).
4. Immediately add 975 ml of pre-warmed SOC media to cuvette and pipette up and down to mix.
5. Transfer the transformant mixture into pre-labeled 14 ml bacterial culture tubes and incubate in a shaker-incubator at 300 rpm at 37°C for 1 hour.
6. Repeat this for each transformation reaction one by one.
7. During 1 hour incubation, prepare LB broth and plates for overnight culture. Label plates with name of cloning ratio being tested and serial dilution number.
8. Label flasks with cloning ratio being amplified. Keep separate flasks for different ratios. If ratios show similar efficiency, products can be pooled during extraction.
9. Prepare 100% ethanol and sterile spreader for plating. Prepare 1.5 ml Eppendorf tubes for serial dilution. Add 900 µl of LB broth (no antibiotics) to each tube.
10. After 1 hour incubation, take 14 ml tubes off the shaker-incubator. Transformant sample should be cloudy.
11. Use 1 reaction per ratio to test efficiency in dilution series. Take 100 µl of transformant and add to 900 µl of LB broth in the 1.5ml Eppendorf tube for the 1st dilution (labeled as [-2]) and mix thoroughly.
12. Change tips and take 100 µl of 1st dilution and add to another 900 µl LB broth in the next tube for the 2nd dilution (labeled as [-3]. Repeat this up to 5th or 6th dilutions (labeled [-6] and [-7] respectively). Repeat this for each cloning ratio and control plasmid.
13. From each tube, take 100 µl of diluted transformant and spread evenly onto the pre-warmed plates and incubate overnight at 37°C.
14. Add rest of the transformants to the LB broth in conical flasks with ampicillin. Make sure volume of broth is ~ ¼ of maximum volume of flask. Use 500 ml flasks for 150 ml cultures. Use 1 L flasks for 250 ml cultures. Pool replicates of same cloning ratios.
15. Incubate cultures overnight on shaker-incubator at 300 rpm and 37°C.

(1I) Library amplification: Library storage and extraction

1. After incubation between 12-16 hours check for colonies on LB plates. Depending on the cloning and transformation efficiency, there should be at least >1 colony on the [-6] plate for all successfully transformed libraries. DH10B should show increased number of colonies.
2. This can be used to compare efficiency of each cloning ratio and show overall estimates of library complexity. A single colony on [-6] is equivalent to at least 1 million unique colonies. Higher number of colonies indicates greater library complexity.
3. Control plasmid should show significantly more colonies than cloned products and can be used to calculate the Colony Forming Units (CFU)/ μg of plasmid transformed using the formular provided by NEB. This helps estimate the transformation efficiency of the competent cell being used.
4. After validation of transformation, check the optical density of the cultures (~ 2.6).

Library Storage

1. Pipette 5 ml of culture into a sterile round-bottomed 15 ml tube and spin down culture for 5 mins.
2. Pipette out LB broth and resuspend pellet in 750 ml of fresh LB broth. Mix with 750 ml of sterile glycerol and transfer to 2 ml screw-capped cryotube. Label tubes and transfer to -80°C for long term storage.
3. Make up to 5 glycerol stocks for each successfully transformed ratio.

Library extraction

1. Pipette remaining cultures into screw capped 500 ml centrifuge tubes and spin down. Aspirate media and use pellet to extract DNA using ZymoPURE II Plasmid Maxiprep kit (Zymo catalog #D4203) according to manufacturer's protocol.
2. Use up to 150 ml of culture per maxiprep reaction. Use centrifuge version of the protocol.
3. Perform optional endotoxin treatment on the extracted library using column provided in kit.
4. Verify concentration and purity of final library ready for transfection or input library preparation.

(2) STARR-seq Screening

(2A) Culturing of cells

- Follow standard cell culture techniques for all experiments. Conduct all cell culture inside biosafety cabinet. Use only cell culture grade labware including sterile serological pipettes (VWR catalog #89130-896 for 5 ml, catalog #89130-898 for 10 ml, catalog #89130-900 for 25 ml and catalog #414004-265 for 2 ml aspiration pipettes) and standard cell culture plates and culture dishes.
- Use HEK293T cells or cell line of choice for STARR-seq. Do appropriate literature review to see evidence of IFN response from cell line upon transfection. HEK293T cells have not shown to be highly susceptible.
- Make appropriate media for HEK293T cells using following recipes.

	Media	percentage	Vol (500 ml)	Catalog number
1	DMEM (High glucose)	87	435	Sigma Aldrich #D6429-500ML
2	PenStrep	1	5	Sigma Aldrich #P4333
3	FBS	10	50	Sigma Aldrich #F2442
4	Non-essential amino acids	1	5	Sigma Aldrich #M7145-100ML
5	HEPES	1	5	Sigma Aldrich #H0887-100ML

- For 500 ml solutions, take out 65 ml of DMEM from the bottle and prepare media inside DMEM bottle.
- Store FBS in 40-50 ml stocks in 50 ml tubes (VWR catalog #89039-656) and store at -20°C. Thaw out FBS overnight at 4°C.
- Store PenStrep at -20°C in 5-10 ml stocks in 15 ml tubes (VWR catalog #89039-666) and thaw out at room temperature.
- Scale recipe as needed.
- Thaw cells in to 6-welled plates (VWR catalog #10861-696) or a 6 cm dish (VWR catalog #25382-100) using standard lab protocol.
- Passage cells every 2-3 days or when ~80% confluent and track cell morphology. Pass cell through at least 3 passages prior to transfection.
- For transfection of STARR-seq library, use at least 30 to 50 million cells per replicate (1 x 15 cm dish (VWR catalog #430599). For efficiency, passage cells from 2 or 3, 80% confluent 10 cm dishes (VWR catalog #25382-166) per 15 cm dish.

(2B) Transfection

Use Lipofectamine 3000 reagent (Thermo Fisher catalog #L3000008) for HEK293T cells with OptiMEM (Thermo Fisher catalog #31985070). Transfection optimized and shown to be >70% efficient and up to 90%.

1. Ensure ~70-80% confluency and healthy morphology of cells prior to starting transfection.
2. Pre-warm optiMEM media to 37°C in water bath. Label 1.5 ml sterile Eppendorf tubes to be used for transfection. Keep everything inside biosafety cabinet.
3. Aspirate DMEM culture media from the dishes to be transfected, wash cells with 10 ml sterile 1× PBS (Sigma Aldrich catalog #806552-500ML), aspirate and add 30 ml of OptiMEM media to cells, 30 mins prior to transfection and place cells back into the incubator. (While aspirating, be careful not to detach cells from the surface. Add media dropwise or from the side of dish to prevent detachment.)
4. Scale lipofectamine 3000 protocol according to number of cells or size of culture dish/plate. For 15 cm dish, use following reagent volumes. Remaining protocol steps are unchanged.
5. Perform up to 3 transfections per cell line. To avoid batch effects, transfect replicates on separate days.

	Tube A	Vol
1	Lipofectamine 3000	57 µl
2	OptiMEM	1443 µl
3	total	1.5 ml

	Tube B	Vol
1	STARR-seq plasmid library	38 µg
2	P3000 enhancer reagent	76 µl
3	OptiMEM	1375 µl
4	Total	1.5 ml

6. Prepare tubes A and B and add mix from tube B into tube A dropwise, gently mix by tapping (DO NOT vortex) and incubate for 15 mins.
7. After incubation, gently pipette combined mix onto the cells and incubate for 24 hours.

(2C) Total RNA Isolation

- Any RNA isolation kit may be used and should be scaled according to number of cells and culture volume. Use Trizol plus RNA purification kit (Thermo Fisher Catalog #12183555) for HEK293T cells. This kit combines trizol and chloroform extraction with PureLink RNA Mini Kit.
- RNA isolation should be carried out inside fume hood. Wipe everything down with RNaseZap RNase Decontamination Solution (Thermo Fisher Catalog # AM9782) and use separate tips and pipettes for RNA. Eppendorf tubes should be clean, RNAse free and labeled for each replicate.
- Move tabletop centrifuge to cold room if temperature-controlled centrifuge not available at least 1 hour prior to starting and wipe down with RNAseZap.
- Harvest cells in trizol in the biosafety cabinet and then move to fume hood. Trizol and chloroform are biohazard chemicals and all associated waste should be discarded separately and disposed only in satellite waste area inside fume hood.

Total RNA Isolation Protocol:

1. For 15 cm dish, use up to 6 ml of trizol.
2. To collect cells, aspirate out all media from cells and wash the cells with 10 ml of PBS.
3. Add 6 ml trizol to cells directly and collect ~1 ml of cells in trizol per 1.5 ml Eppendorf tube (6 tubes per replicate) and incubate for 5 mins at room temperature.
4. Add 0.2 ml of chloroform to each 1 ml of cells and trizol mix vigorously by shaking the tube for 15 secs. Make sure tube cap is firmly closed and not leaking. Incubate at room temperature for 3 mins.
5. Centrifuge tubes at 12,000 g for 15 mins at 4°C. Gently place tubes back inside fume hood. Sample will form 2 phases inside tube. Lower phase will be red phase comprising of phenol-chloroform, an interphase and upper aqueous phase will comprise of RNA.
6. Carefully pipette 450 – 500 µl of aqueous phase into fresh Eppendorf tubes without disturbing other phases. Pipette 150 µl at a time for ease.
7. Add equal volume (450 – 500 µl) of 70% ethanol to each tube, vortex and spin down.
8. Add up to 700 µl of mix into spin column and centrifuge at 12,000 g for 30 secs at room temperature and discard flow-through. Repeat process until all sample has been processed using the same column for all tubes per replicate. (Use the same column for all 6 tubes per replicate).
9. After all sample has been processed per replicate add 700 µl wash buffer I and centrifuge at 12,000 g for 30 secs and discard flow-through.
10. Add 500 µl of wash buffer II and repeat centrifugation and discard flow-through. Repeat step.
11. Perform a final dry centrifuge spin at 12,000 g for 1 min to remove trace ethanol from the column.
12. Remove collection tube and place column into fresh tube for RNA collection. Add 100 µl of DEPC water to column and incubate sample for 1 min at room temperature.

13. Centrifuge tube at 12,000 g for 2 mins at room temperature and measure concentration and purity.
14. Add 100 µl DEPC water for 2nd elution and 3rd elution by repeating process and measure concentration and purity of each elution. If values are consistent, pool eluates and measure final concentration and purity. Purity ratio for 260/280 and 260/230 should be above 2.0.
15. Send samples for RIN analysis to genomics core facility for quality assessment.
16. RNA can be stored at -80°C or used for mRNA isolation.
17. Dilute total RNA to 750 ng/µl for mRNA isolation. Typical yield is ~1200 µl of dilute total RNA (~900 µg of total RNA).

(2D) mRNA Isolation

- Use dynabeads mRNA Purification Kit (Thermo Fisher catalog #61006) for mRNA isolation. Each kit contains reagents for 10 mRNA isolations.
- However, binding buffer and wash buffer B (Thermo fisher catalog #11900D) and elution buffer (Thermo Fisher catalog #A33566) will not be enough if beads are to be reused and can be ordered or prepared additionally. Binding buffer is not available separately and needs to be prepared or ordered with kit.
- Additional Lysis/Binding buffer (Thermo Fisher catalog #A33562) also needs to be ordered for reusing beads). Instructions are provided on manufacturer website as well as in STARR-seq protocol by Neumayr and colleagues (Neumayr et al 2019).
- Dilute total RNA sample to 750 ng/µl prior to starting mRNA isolation. (Beads can process up to 75 µg of RNA at a time and can be regenerated for reuse. Process 100 µl of diluted total RNA at a time.
- Use 200 µl of dynabeads per 100 µl of total RNA. Regenerate beads up to 6 times. Therefore, use ~ 600 µl of total RNA for 1 x 200 µl dynabead reaction. Reuse beads within same replicate to avoid cross contamination.
- Work inside RNA hood and wipe everything with RNAseZap.
- Set heat block A to 65°C and heat block B to 80°C.
- Set tabletop shaker incubator to 25°C (room temperature) and 60 rpm.
- Thaw out total RNA on ice before proceeding.

mRNA Isolation Protocol:

1. Separate the diluted total RNA into 100 µl batches in individual 1.5 ml Eppendorf tubes and keep on ice.
2. Transfer tube containing the sample being processed to 65°C for 2 mins (to remove secondary structures) and then immediately replace on ice.
3. To prepare beads, pipette 200 µl of the beads to a RNAse free 1.5 ml Eppendorf tube and place on magnetic rack for 30 secs. Carefully discard the supernatant without disturbing the beads adhering to the magnet.
4. Remove tubes from the rack and add 100 µl of binding buffer to the tube, lightly mix and place tubes back on rack. Remove supernatant. Add another 100 µl of binding buffer (or same volume of total RNA being processed) and mix. Beads are not ready for mRNA isolation.
5. Add 100 µl of total RNA to 100 µl of beads and mix by pipetting. Incubate on shaker incubator at room temperature and low rpm (~60 rpm) for 5 mins.
6. Transfer tubes to magnetic rack and remove supernatant. Add 200 µl of wash buffer B, mix and place tubes on magnetic rack. Remove supernatant and repeat wash step.
7. Remove the tubes from the rack and add 6-10 µl of elution buffer per 100 µl total RNA and incubate samples at 80°C for 2 mins and immediately place tubes back on rack.
8. Pipette supernatant (mRNA) into a fresh tube and measure concentration and purity.

9. To reuse the beads, remove the beads from the rack and add 300 µl of Lysis/Binding buffer to wash beads. Place back on the rack, remove the supernatant and add 100 µl binding buffer to proceed with the next round of mRNA isolation for the same replicate. (Note, this is only for reuse of beads and not regeneration which has a separate protocol provided on manufacturer's website).
10. Repeat until all total RNA has been processed to mRNA. Pool multiple batches for each replicate and measure final concentration and purity. Typical yield is between 45 – 70 µl of mRNA at ~ 400 ng/µl or 18 – 28 µg of mRNA and depends on starting amount of total RNA (around 1 – 5% of total RNA).
11. Store mRNA at -80°C or proceed with TURBO DNase treatment, RNA clean-up and reverse transcription.

(2E) TURBO DNase treatment

- This is required to remove any residual DNA from sample prior to reverse transcription. Use TURBO DNA-free Kit (Thermo Fisher catalog #AM1907).
- Starting material is around ~400 ng/µl or ~400 µg/ml and thus requires the rigorous protocol for TURBO DNase treatment and thus use 2 µl or 4 units of TURBO DNase enzyme per reaction.
- Process entire sample together per replicate. Scale protocol according to volume of mRNA.
- Set heat block to 37°C prior to starting.

TURBO DNase treatment protocol:

1. Add 0.1 volume of 10× TURBO DNase buffer to sample.
2. Add 2 µl (4 units) of TURBO DNase enzyme and mix gently (DO NOT vortex).
3. Incubate at 37°C for 60 mins.
4. Add 0.2 volumes of DNase inactivation reagent and mix gently (DO NOT vortex) and incubate at room temperature for 5 mins. Lightly flick tube to redistribute reagent every 1 min.
5. Centrifuge tubes at 10,000 g for 1.5 mins at room temperature and transfer supernatant to fresh 1.5 ml Eppendorf tube for the next step.

(2F) RNAClean XP treatment

- Perform clean-up of each TURBO DNase reaction using RNAClean XP beads (Beckman Coulter catalog #A63987)
- Warm beads to room temperature for at least 30 mins.
- Prepare fresh 80% ethanol for sample wash.
- Set tabletop shaker incubator to 37°C and 60 rpm (low rpm).
- Process 1 replicate per clean-up reaction. Scale volume of beads according to volume of mRNA.

RNAClean XP protocol:

1. Add 1.8 volume of bead per 1 volume of mRNA sample and mix by thoroughly pipetting >20 times and incubate at room temperature for 15 mins.
2. Transfer mix to magnetic rack and incubate for 10 mins or till when all beads have adhered to magnet.
3. Remove supernatant (should be clear liquid) and add 500 µl of 80% ethanol for 1st wash. Incubate for up to 2 mins, remove supernatant and add 500 µl of 80% ethanol for 2nd wash. Incubate for 2 mins and remove without disturbing beads.
4. Dry beads completely at room temperature for 5 mins. Check for any residual ethanol inside tube and remove.
5. Remove tubes from the rack, and add 20 µl DEPC water to beads for elution, mix by pipetting and vortexing and incubate in tabletop shaker incubator at 37°C at low rpm for 3 mins and place tubes back on magnetic rack.
6. Incubate for 1 min on the rack and pipette pure mRNA into a fresh 1.5 ml Eppendorf tube for further processing.
7. Assess final concentration and purity of mRNA. This is the final checkpoint for samples prior to cDNA library preparation.
8. Typical yield is ~12 µg of mRNA per replicate (600 – 900 ng/µl of mRNA in 19 µl of water (1 µl used for purity and concentration assessment)

(2G) cDNA library preparation: 1st Strand Synthesis

- For 1st strand synthesis, use Superscript III kit
- Divide total sample amount by 5 and round off to nearest multiple of 5 to determine number of reactions required per sample. Use 5 reactions using 2.4 μg of mRNA per reaction for ~12 μg samples.
- Adjust all samples to 12 μg in 20 μl (600 ng/μl) of DEPC water for uniformity across replicates. Use 4 μl (2.4 μg mRNA) per RT reaction.
- Though protocol suggests using 500 ng of mRNA at a time for 1st strand synthesis, we are performing reverse transcription only for fraction of mRNA that was self-transcribed due to enhancer activity and so taking larger amounts of mRNA is permissible.
- Dilute STARR-seq reporter specific primer to 2 μM. Dilute dNTP solution to 10 μM.

1st Strand Synthesis Protocol:

1. Assemble following reaction as reaction 1:

	Reagent/sample	Vol (μl)
1	mRNA	4
2	dNTP	1
3	Primer	1
4	Water	7
5	Total	13

2. Incubate at 65°C for 5 mins followed by 4°C for 1 min on a thermocycler (program RT_1)
3. Assemble following reaction as reaction II:

	Reagent	Vol (μl)
1	1st strand buffer	4
2	DTT	1
3	RNaseOUT	1
4	Superscript III	1
5	total	7

4. Add reaction II to reaction I for each RT reaction (final volume 20 µl)
5. Incubate according to following conditions on thermocycler (program RT_II):

1	50°C	60 mins
2	75°C	15 mins
3	4°C	hold

(2H) cDNA clean-up: RNase A treatment and AMPure XP clean-up

- Treat all reverse transcribed samples with RNase A to remove residual RNA.
- Pool 5 RT reactions (total = 100 µl per replicate) (5 reactions carried out per replicate)
- Move to DNA bench to avoid pipetting RNase A in RNA isolation bench
- Prepare 80% ethanol for AMPure XP bead clean-up wash steps
- Bring beads to room temperature for at least 30 mins prior to starting.

RNase A treatment and AMPure XP bead clean-up Protocol:

1. Add 1 µl of RNase A (10mg/ml) per 5 RT reactions.
2. Incubate at 37°C for 1 hour in thermocycler (program: RNase A treatment)
3. Perform 1.4× AMPure XP bead clean-up protocol
4. Elute cDNA in 43 µl DEPC water for next step.
5. Measuring cDNA concentration is advised but is not accurate and hence it is also okay to proceed to either 2nd strand synthesis and UMI_PCR or directly to jPCR steps.

(2I) UMI addition: 2nd strand synthesis

- Unique Molecular Identifiers (UMIs) are essential to label and filter out PCR duplicates from the libraries that occur during library preparation.
- For addition of UMIs, follow UMI-STARR-seq protocol reported by Neumayr and colleagues (Neumayr et al. 2019) including a 2nd strand synthesis reaction and UMI_PCR reaction prior to jPCR.
- Start 2nd strand synthesis using 2nd strand primer as shown by Neumayr and colleagues (Neumayr et al. 2019). Dilute primer to 10 μM in DEPC water for use.
- Use KAPA 2× HiFi Hot Start Ready mix polymerase (Roche catalog #07958927001)

2nd strand synthesis protocol:

1. Process entire replicate cDNA (~43 μl) per reaction. Set-up following reaction for each replicate:

	Sample/reagent	Vol (μl)
1	cDNA	42.5
2	KAPA HiFi 2× Ready mix	50
3	2nd strand primer	7.5
4	Total	100

2. Incubate reactions in thermocycler (program: 2nd strand synth)

	Temperature	Time	Step
1	98°C	60 secs	Denaturation
2	65°C	30 secs	Annealing
3	72°C	30 secs	Extension

3. Reaction purification using 1.4× AMPure XP bead clean-up.
4. Elute library in 43 μl of DEPC water for UMI_PCR step.

(2J) UMI addition: UMI_PCR

- To add UMI, use custom designed i7-UMI-P7 primer to add i7 index sequence and UMI sequence simultaneously to cDNA. This helps retain unique dual indexing to mitigate index hopping as well as allow for PCR duplicate filtration as opposed to Neumayr and colleagues protocol where the i7 is replaced by the UMI.
- Dilute primers to 10 μM for use. Keep record of which index is being added to which sample and label all tubes.
- Process complete cDNA for each reaction.

UMI_PCR protocol:

1. Set-up following reaction for each replicate:

	Sample/reagent	Vol (μl)
1	cDNA	42.5
2	KAPA HiFi 2× Ready mix	50
3	i7-UMI-P7 primer	7.5
4	Total	100

2. Incubate reaction in thermocycler (program: 2nd strand synth) Same conditions as 2nd strand synthesis.

	Temperature	Time	Step
1	98°C	60 secs	Denaturation
2	65°C	30 secs	Annealing
3	72°C	30 secs	Extension

3. Clean-up reactions using 1.4× AMPure XP bead clean-up and elute in 50 μl of DEPC water per replicate. These samples are now ready for jPCR step.

(2K) Junction PCR (jPCR)

- Perform same number jPCR reactions as RT reactions. Use 10 µl of cDNA library per jPCR reaction.
- Use modified jPCR primers as provided by Neumayr and colleagues in their protocol (Neumayr et al. 2019).
- jPCR allows amplification of only self-transcribed fragments and filters out plasmid transcripts

jPCR protocol:

1. Set-up following reactions for each replicate. (5 reactions per replicate)

	Sample/reagent	Vol (µl)
1	cDNA library	10
2	DEPC water	10
3	jPCR forward primer	2.5
4	jPCR reverse primer	2.5
5	KAPA 2× HiFi Ready Mix	25
	Total	50

2. Incubate reactions in thermocycler (program: jPCR)

	Temperature	Time	Step
1	+98°C	45 secs	Initial denaturation
2	+98°C	15 secs	Denaturation
3	+65°C	30 secs	Annealing
4	+72°C	30 secs	Extension
5	+72°C	60 secs	Final extension
6	+4°C	Hold	

X 16

3. Clean-up jPCR reactions by pooling all 5 reactions per replicate (250 µl per replicate) and perform a 0.8× AMPure XP clean-up using 200 µl of beads per replicate.
4. Elute in 50 µl DEPC water. Measure concentration and purity of final jPCR products prior to performing final sequencing ready PCR step.
5. Typical yield is around 50 – 130 ng/µl of 50 µl product. Yields may vary based on STARR-seq activity and starting amount for samples.

(2L) Sequencing ready PCR: Output library preparation

- This step is carried out to add i5 and P5 index adapters to the library before sending samples for sequencing.
- Use standard i5 primer as forward primer and P7 primer as shown in Neumayr et al 2019 and supplementary table S2.
- Dilute all primers to 10 μM prior to use.
- Bring AMPure XP beads to room temperature for at least 30 mins for PCR clean-up
- Prepare 1% agarose gel for test PCR run to determine number of cycles for PCR. Start with low cycle number. 5 cycles are sufficient for jPCR products obtained in the previous step.
- If performing test PCR, test on all samples since they may have different starting amounts.

Sequencing Ready PCR protocol:

1. Setup the following reaction for each replicate:

	Sample/reagent	Vol (μl)
1	jPCR product	10
2	DEPC water	10
3	i5 primer	2.5
4	P7 primer	2.5
5	KAPA 2× HiFi Ready Mix	25
	Total	50

2. Incubate reactions on thermocycler (program: SeqReady PCR)

	Temperature	Time	Step	
1	+98°C	45 secs	Initial denaturation	
2	+98°C	15 secs	Denaturation	
3	+65°C	30 secs	Annealing	X 5
4	+72°C	30 secs	Extension	
5	+72°C	60 secs	Final extension	
6	+4°C	Hold		

3. Run 10 µl sample with 2 µl of 6× loading dye (for each replicate) on 1% agarose gel for 30 mins at 130V and observe band. Check for characteristic smear for each band at right size as shown by Neumayr and colleagues (Neumayr et al. 2019). If band is concentrated, repeat PCR with lower cycles (if cycle >5) or try lower starting material (5 µl of jPCR).
4. Repeat PCR for correct starting amount and number of cycles as determined.
5. Clean up each reaction with 1× AMPure XP beads and elute in 20 µl. Typical yield can vary and may need to be diluted prior to sequencing.
6. Samples once sent to core facility will be reassessed for quality, checked for distribution on tapestation, further diluted and pooled into an equimolar pool for sequencing.

(2M) Sequencing ready PCR: Input library preparation

- To assess enhancer activity, plasmid library needs to be sequenced alongside output libraries for normalization.
- Library sequence architecture should be identical to enable pooling of library along with output libraries.
- Input library also goes through PCR steps and so also important to add UMIs to these libraries prior to sequencing similar to output libraries.
- Use custom i7-UMI-P7 primers designed for UMI addition. Use i5 primer to add i5 index and P5 adapter. Assign unique dual index pairs to each replicate of input library and use at least 3 replicates. Record indexes assigned. Dilute primers to 10 µM prior to use.
- Bring AMPureXP clean-up beads to room temperature for at least 30 mins for clean-up.
- Prepare 1% agarose gel for final library fragment length verification and purification.
- For starting material, use 1 µl of STARR-seq plasmid library (~750 ng/ µl) added with 39 µl of DEPC water.

UMI_PCR protocol for input library preparation:

1. Setup following reaction for each input replicate:

	Sample/reagent	Vol (µl)
1	STARR-seq plasmid library	40
2	KAPA HiFi 2× Ready mix	50
3	i7-UMI-P7 primer	5
4	i5 primer	5
5	Total	100

2. Incubate reaction in thermocycler (program: UMI_PCR)
3. Perform 1.4× AMPure XP clean-up of each replicate reaction. Elute in 20 µl DEPC water.

Sequencing Ready PCR protocol for input library preparation:

1. Setup following reaction for each input replicate: (use complete product)

	Sample/reagent	Vol (µl)
1	Input replicate DNA	20
2	i5 primer	2.5
3	P7 primer	2.5
4	KAPA 2× HiFi Ready Mix	25
	Total	50

2. Incubate reactions in thermocycler (program: SeqReady PCR)
3. Run entire sample (50 µl) with 10 µl 6× loading dye on a 1% agarose gel for 30 mins at 130V.
4. Visualize gel and verify fragment length. Slice out band of correct length and weigh gel slice.
5. Extract DNA form gel slice using Zymoclean Gel DNA recovery kit.
6. Clean-up extracted DNA using 1× AMPureXP bead clean-up and elute in 25 µl of DEPC water and measure concentration and purity. Typical yield is ~ 10 – 20 ng/µl.

- All input and output libraries are sent to sequencing core facility for sequencing.

References

Arnold. Cosmas D, Gerlach. Daniel, Stelzer. Christoph, Boryń. Łukasz M, Rath. Martina SA.

2013. Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science (80-)* 339.

Barakat TS, Halbritter F, Zhang M, Rendeiro AF, Perenthaler E, Bock C, Chambers I. 2018. Functional Dissection of the Enhancer Repertoire in Human Embryonic Stem Cells. *Cell Stem Cell* 23: 276-288.e8.

Bergman DT, Jones TR, Liu V, Ray J, Jagoda E, Siraj L, Kang HY, Nasser J, Kane M, Rios A, et al. 2022. *Compatibility rules of human enhancer and promoter sequences*. Springer US.

Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447: 799–816.

Chaudhri VK, Dienger-Stambaugh K, Wu Z, Shrestha M, Singh H. 2020. Charting the cis-regulome of activated B cells by coupling structural and functional genomics. *Nat Immunol* 21: 210–220. http://dx.doi.org/10.1038/s41590-019-0565-0.

Glaser L V., Steiger M, Fuchs A, Van Bömmel A, Einfeldt E, Chung HR, Vingron M, Meijsing SH. 2021. Assessing genome-wide dynamic changes in enhancer activity during early mESC differentiation by FAIRE-STARR-seq. *Nucleic Acids Res* 49: 12178–12195.

Hansen TJ, Hodges E. 2022. ATAC-STARR-seq reveals transcription factor–bound activators and silencers within chromatin-accessible regions of the human genome. *Genome Res* 32: 1529–1541.

Inoue F, Kircher M, Martin B, Cooper GM, Witten DM, McManus MT, Ahituv N, Shendure J. 2017. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res* 27: 38–52.

Johnson GD, Barrera A, McDowell IC, D'Ippolito AM, Majoros WH, Vockley CM, Wang X, Allen AS, Reddy TE. 2018. Human genome-wide measurement of drug-responsive regulatory activity. *Nat Commun* 9: 1–9. http://dx.doi.org/10.1038/s41467-018-07607-x.

Kalita CA, Brown CD, Freiman A, Isherwood J, Wen X, Pique-Regi R, Luca F. 2018. High-throughput characterization of genetic effects on DNA-protein binding and gene transcription. *Genome Res* 28: 1701–1708.

Kim YS, Johnson GD, Seo J, Barrera A, Cowart TN, Majoros WH, Ochoa A, Allen AS, Reddy TE. 2021. Correcting signal biases and detecting regulatory elements in STARR-seq data. *Genome Res* 31: 877–889.

Klein JC, Agarwal V, Inoue F, Keith A, Martin B, Kircher M, Ahituv N, Shendure J. 2020. A systematic evaluation of the design and context dependencies of massively parallel reporter assays. *Nat Methods* 17: 1083–1091. http://dx.doi.org/10.1038/s41592-020-0965-y.

Lambert JT, Su-Feher L, Cichewicz K, Warren TL, Zdilar I, Wang Y, Lim KJ, Haigh J, Morse SJ, Canales CP, et al. 2021. Parallel functional testing identifies enhancers active in early postnatal mouse brain. *Elife* 10: 1–27.

Lea AJ, Vockley CM, Johnston RA, Del Carpio CA, Barreiro LB, Reddy TE, Tung J. 2018.

Genome-wide quantification of the effects of DNA methylation on human gene regulation. *Elife* 7: 1–27.

Lee D, Shi M, Moran J, Wall M, Zhang J, Liu J, Fitzgerald D, Kyono Y, Ma L, White KP, et al. 2020. STARRPeaker: uniform processing and accurate identification of STARR-seq active regions. *Genome Biol* 21: 1–24.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 00: 1–3. http://arxiv.org/abs/1303.3997.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079.

Liu S, Liu Y, Zhang Q, Wu J, Liang J, Yu S, Wei GH, White KP, Wang X. 2017a. Systematic identification of regulatory variants associated with cancer risk. *Genome Biol* 18: 1–14.

Liu Y, Yu S, Dhiman VK, Brunetti T, Eckart H, White KP. 2017b. Functional assessment of human enhancer activities using whole-genome STARR-sequencing. *Genome Biol* 18: 1–13.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 1–21.

Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, Feizi S, Gnirke A, Callan CG, Kinney JB, et al. 2012. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol* 30: 271–277.

Muerdter F, Boryn ŁM, Woodfin AR, Neumayr C, Rath M, Zabidi MA, Pagani M, Haberle V, Kazmar T, Catarino RR, et al. 2018. Resolving systematic errors in widely used enhancer activity assays in human cells. *Nat Methods* 15: 141–149.

Neumayr C, Pagani M, Stark A, Arnold CD. 2019. STARR-seq and UMI-STARR-seq: Assessing Enhancer Activities for Genome-Wide-, High-, and Low-Complexity Candidate Libraries. *Curr Protoc Mol Biol* 128: e105.

Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, Lee C, Andrie JM, Lee SI, Cooper GM, et al. 2012. Massively parallel functional dissection of mammalian enhancers in vivo. *Nat Biotechnol* 30: 265–270.

Peng T, Zhai Y, Atlasi Y, Ter Huurne M, Marks H, Stunnenberg HG, Megchelenbrink W. 2020. STARR-seq identifies active, chromatin-masked, and dormant enhancers in pluripotent mouse embryonic stem cells. *Genome Biol* 21: 1–27.

Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*.

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8: 2281–2308.

Sahu B, Hartonen T, Pihlajamaa P, Wei B, Dave K, Zhu F, Kaasinen E, Lidschreiber K, Lidschreiber M, Daub CO, et al. 2022. Sequence determinants of human gene regulatory elements. *Nat Genet* 54: 283–294.

Schöne S, Bothe M, Einfeldt E, Borschiwer M, Benner P, Vingron M, Thomas-Chollier M, Meijsing SH. 2018. Synthetic STARR-seq reveals how DNA shape and sequence modulate transcriptional output and noise. *PLoS Genet* 14: 1–24.

Shlyueva D, Stelzer C, Gerlach D, Yáñez-Cuna JO, Rath M, Boryń ŁM, Arnold CD, Stark A. 2014. Hormone-Responsive Enhancer-Activity Maps Reveal Predictive Motifs, Indirect Repression, and Targeting of Closed Chromatin. *Mol Cell* 54: 180–192.

Vanhille L, Griffon A, Maqbool MA, Zacarias-Cabeza J, Dao LTM, Fernandez N, Ballester B, Andrau JC, Spicuglia S. 2015. High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat Commun* 6.

Verfaillie A, Svetlichnyy D, Imrichova H, Davie K, Fiers M, Atak ZK, Hulselmans G, Christiaens V, Aerts S. 2016. Multiplex enhancer-reporter assays uncover unsophisticated TP53 enhancer logic. *Genome Res* 26: 882–895.

Wang X, He L, Goggin SM, Saadat A, Wang L, Sinnott-Armstrong N, Claussnitzer M, Kellis M. 2018. High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human. *Nat Commun* 9. http://dx.doi.org/10.1038/s41467-018-07746-1.

Zabidi MA, Arnold CD, Schernhuber K, Pagani M, Rath M, Frank O, Stark A. 2015. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* 518: 556–559. http://dx.doi.org/10.1038/nature13994.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9.