Jul 30, 2025

MPRA RNA-seq and DNA-seq library preparation

DOI
https://dx.doi.org/10.17504/protocols.io.e6nvw4pjzlmk/v1

Jessica McAfee[1], Hyejung Won[1]
[1]UNC Chapel Hill

IGVF MPRA

Jessica McAfee
UNC Chapel Hill

DOI: https://dx.doi.org/10.17504/protocols.io.e6nvw4pjzlmk/v1

Protocol Citation: Jessica McAfee, Hyejung Won 2025. MPRA RNA-seq and DNA-seq library preparation. **protocols.io** https://dx.doi.org/10.17504/protocols.io.e6nvw4pjzlmk/v1

Manuscript citation:
Sool Lee*, Jessica C. McAfee*, Jiseok Lee*, Alejandro Gomez, Austin T. Ledford, Declan Clarke, Hyunggyu Min, Mark B. Gerstein, Alan P. Boyle, Patrick F. Sullivan, Adriana S. Beltran, Hyejung Won. Massively parallel reporter assay investigates shared genetic variants of eight psychiatric disorders. *Cell.* (2024) https://doi.org/10.1016/j.cell.2024.12.022 *Co-first authors

Nana Matoba*, Jessica C. McAfee*, Oleh Krupa, Jess Bell, Brandon D. Le, Jordan M. Valone, Gregory E. Crawford, Hyejung Won, Jason L. Stein. Massively parallel assessment of gene regulatory activity at human cortical structure associated variants. *Biorxiv.* (2025) https://doi.org/10.1101/2025.02.08.635393 *Co-first authors

Protocol status: Working
We use this protocol and it's working

Created: June 30, 2025

Last Modified: July 30, 2025

Protocol Integer ID: 221314

Keywords: MPRA, library Prep, PJD MPRA, qPCR, Massively Parallel Reporter Assay, DNA-seq, RNA-seq, generating mpra rna, mpra rna, transfectable mpra backbone, seq library preparation this protocol, donor-egp2ap-rc, seq library preparation, seq library, seq, dna

Funders Acknowledgements:
NIMH
Grant ID: DP2MH122403
IGVF Consortium
Grant ID: UM1HG012003, U01HG011952
PsychENCODE Consortium
Grant ID: R01MH122509
NICHD
Grant ID: T32HD040127
NIGMS
Grant ID: 5T32GM067553, 5T32GM135128
University Cancer Research Fund, Comprehensive Cancer Center Core support grant
Grant ID: P30CA016086
UNC Center for Mental Health and Susceptibility grant
Grant ID: [P30ES010126

Abstract

This protocol is for generating MPRA RNA-seq and DNA-seq libraries for sequencing. This protocol uses a transfectable MPRA backbone: Donor_eGP2AP_RC (Addgene: ID#133784). For more detail, see methods section of: https://www.cell.com/cell/fulltext/S0092-8674(24)01435-1

Guidelines

All temperatures are in Celsius.

Materials

	Critical Commercial Assays	Brand	Cat. Number
	DNA Clean and Concentrator Kit-25	Zymo Research	Cat#D4033
	Zymo DNA Clean and Concentrator Kit-5	Zymo Research	Cat#D4033
	Zymo Gel DNA Recovery Kit	Zymo Research	Cat#D4008

	Reagents	Brand	Cat. Number
	AMPure XP Beads	Beckman Coulter	Cat#A63881
	NEBNext 2X Q5 Hifi HS Master Mix	New England Biolabs (NEB)	Cat#M0453S
	SparQ beads	VWR	Cat#76302-834
	SuperScript IV Reverse Transcriptase	Invitrogen	Cat#18090050

Before start

Extract RNA and DNA from your cells that have been transfected with your MPRA library.

RNA and DNA extraction

1 RNA and DNA extraction:
DNA and RNA were extracted from the cells by Quick-DNA/RNA Miniprep Plus Kit (Zymo Research, cat. no. D7003) using 600 uL of shield buffer and 600 uL of lysis buffer per well.
RNA and DNA eluted with water at ~55°C

Reverse transcription of RNA

2 Run reverse transcription (RT) to create cDNA from RNA

2.1 Run RT rxn per sample: Use up to 5ug (5ug=5000ng; weight max) or 11uL (volume max) of RNA.

2.2 For each RNA sample, make the following mastermix:

	A	1X (uL)
	2uM Lib_Han_RT	1
	10mM dNTP	1
	Total RNA 5ug	to 11
	Water	to 13
	Total	13

Heat for 65°C for 5 min, ice for 1 min

Add the following mixture to finish RT.

	A	1X (uL)
	5X SSIV buffer	4
	100mM DTT	1
	RNase OUT	1
	SSIV Transcriptase	1
	Total	7

55°C for 1 hour, 80°C for 10 min.

Create qPCR standards

3 Generate qPCR standards from cDNA to determine transfection efficiency.

3.1 Run the following PCR for one of the samples.

	A	1X (uL)
	cDNA	1
	Water	19

	A	1X (uL)
	10uM Forward Primer: Lib_Hand	2.5
	10uM Reverse Primer: Lib_Seq_Luc_R	2.5
	PCR MM 2X (NEBNext Q5 Hot Start HiFi)	25
	Total	50

	Step	Temp	Time
	Initial denaturation	98	30 sec
	Denaturation	98	10 sec
	Annealing	63	15 sec
	Extension	72	25 sec
	Cycles	Go to 2	24 times
	Final extension	72	5 min
	Hold	4	

Extract the band at 205 bp using gel extraction (1.8% agarose gel, Zymo Gel DNA Recovery Kit, Cat#D4008)
Measure the concentration using Qubit.
Perform serial dilution, such that you have 1ng, 100pg, 10pg, 1pg, 100fg, and 10fg/uL stocks. These will serve as standards.
*Once standards are made, this step does not need to be repeated for each batch

qPCR to estimate transfection efficiency

4 Use qPCR to estimate the concentration of the MPRA RNA barcodes.

4.1 Master mix for cDNA and standards:

	Reagents	1x (uL)
	cDNA library	1
	NFW	7
	10uM Lib_Hand	1
	10uM Lib_Seq_Luc_R	1
	SYBR Green 2x MM	10
	Total	20

	Temp	Time	Cycle
	50	2 min	

	Temp	Time	Cycle
	95	2 min	
	95	1 sec	X45 Cycles
	60	30 sec	

Master mix for DNA samples:

	Master mix for DNA Samples	1X (uL)
	DNA library	1
	water	7
	10uM Lib_Hand_RT	1
	10uM Lib_Seq_Luc_R	1
	SYBR Green 2x MM	10
	Total	20

	Temp	Time	Cycle
	50	2 min	
	95	2 min	
	95	1 sec	X45 Cycles
	60	30 sec	

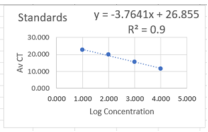
Run the qPCR with all of these together on the same plate.

4.2 **qPCR Data Analysis:**

Example of Data Analysis:

Sample Na	CT	Ct Mean	x=(y-26.855)/(-3.7641)	10^x fg
e1v1	14.304	16.559	2.735287341	543.6099
e1v1	18.815	16.559		
E1 V2	13.895	13.899	3.442027631	2767.118
E1 V2	13.903	13.899		
E1V3	11.415	12.535	3.804432233	6374.296
E1V3	13.654	12.535		
E1C1	12.345	12.892	3.70945903	5122.229
E1C1	13.440	12.892		
E1C2	13.727	13.807	3.466469827	2927.317
E1C2	13.887	13.807		

Sample Name	CT	log []	mean CT
10pg	11.705	4.000	11.540
10pg	11.376		
1pg	14.867	3.000	15.697
1pg	16.527		
100fg	19.972	2.000	19.834
100fg	19.695		
10fg	22.682	1.000	22.709
10fg	23.196		
10fg	22.248		



4.3 Make a plot of average CT (y axis) and log concentration (x axis) of your standard values. For the x axis log concentration: 100pg would be 5, 10pg would be 4, 1pg would be 3, 100fg would be 2, 10fg would be 1. Get the line equation and R squared value.

4.4 For each of your samples, get the average CT value between the two replicates.

4.5 Solve your standards' line equation for X (log concentration). Plug in the average sample CT value in the Y (average CT value), solve for X.
Retrieve the concentration (in fg) from X as 10^X. Generally, we are targeting for >700fg/replicate.

If your sample has a high transfection efficiency, continue to the next step. If the sample has low transfection efficiency, go back to optimizing transfection efficiency.

PCR 1

5 Amplify your RNA and DNA library.

5.1 Repeat the RT (step 2) for the rest of your RNA samples. Convert all of your RNA to cDNA.

5.2 **PCR 1 RNA (cDNA)**:
For RNA samples, use 5uL of your cDNA per 1 rxn. Generally, you will do 18 rxns per sample. Note that this uses **Lib_Hand** primer.

	PCR 1 cDNA Master Mix	1X (uL)	18X (uL)
	cDNA	5	90
	10uM Forward: **Lib_Hand**	2.5	45
	10uM Reverse: **Lib_Seq_Luc_R**	2.5	45
	PCR MM 2X (NEBNext Q5 HS HiFi)	25	450
	NFW	15	270
	Total	50	900

	Temp	Time	Cycles
	98	30 sec	1
	98	10 sec	X 9 cycles
	63	15 sec	
	72	25 sec	
	72	5 min	1
	4	forever	1

PCR clean up: Split each sample to 2 of the -25 zymo clean and concentrator columns, elute each in 25uL water→ 50uL total per sample

5.3 PCR 1 for DNA:
This one uses primer Lib_Hand_RT.
You will do 4X per one sample.

	PCR1 DNA Master Mix	1X (uL)	4X (uL)
	DNA	0.5	2
	10uM Forward: **Lib_Hand_RT**	2.5	10
	10uM Reverse: **Lib_Seq_Luc_R**	2.5	10
	PCR MM 2X (NEBNext Q5 HS HiFi)	25	100

	PCR1 DNA Master Mix	1X (uL)	4X (uL)
	NFW	19.5	78
	Total	50	200

	Temp	Time	Cycles
	98	30 sec	1
	98	10 sec	X8 cycles
	63	15 sec	
	72	25 sec	
	72	5 min	1
	4	forever	

PCR clean up: 1 sample use 1 -25 zymo clean and concentrator, elute each in 25uL water

qPCR 2

6 The purpose of qPCR 2 is to estimate the number of PCR cycles for PCR 2. PCR 2 is used to add sequencing adapters and sample indexes.

qPCR 2 is optional (you can apply this step only when the samples are overamplified with a certain cycle).

Do two reactions for each sample. Pick an index primer at random for qPCR2 and use it for all the samples.

6.1 qPCR 2 Master mix:

	A	1X (uL)
	cDNA or DNA from PCR 1	1
	water	7
	10uM Primer: P5_Seq_Luc_F	1
	10uM R Primer: P7_IND_#_Han	1
	SYBR Green 2X MM	10
	Total	20

	Temp	Time	Cycles
	50	2 min	1
	95	2 min	1

	Temp	Time	Cycles
	95	1 sec	X45 Cycles
	60	30 sec	

Use ¼ of the multicomponent plot (1/4 of the fluorescence from the top and bottom of your lines), follow down this ¼ line to the number of PCR cycles to use.

PCR 2

7 PCR 2 adds Illumina sequencing adapters and sample indexes.

Assign indexes to each sample (each RNA (cDNA) and DNA library should get a unique index).

Use all of the product you get from the cleanup steps of PCR 1.

DNA samples should have 1.5 rxns per sample. RNA should have 3 rxns per sample.

	A	B	RNA	DNA
	PCR 2 Master Mixes	**1X (uL)**	**3X (uL)**	**1.5X (uL)**
	Library	16.6	49.8	24.9
	NFW	3.4	10.2	5.1
	10uM F Primer: P5_Seq_Luc_F	2.5	7.5	3.75
	10uM R Primer: P7_IND_#_Han	2.5	7.5	3.75
	PCR MM 2X HIFI HS Q5 (NEB)	25	75	37.5
	Total	50	150	75

	Temp	Time	Cycles
	98	30 sec	1
	98	10 sec	X PCR cycles determined by qPCR 2
	69	15 sec	
	72	25 sec	
	72	5 min	1
	4	forever	

Expected band size: 273bp

The PCR product is cleaned up and size-selected using AMPure XP Beads (Beckman Coulter, cat. no. A63881) with 0.7X and 0.9X ratios to select for a DNA fragment at 273 bp.

Sequence your samples

8 The RNA and DNA library samples are pooled together in a 2:1 concentration ratio. Use NovaSeq SP with custom cycles of 35 × 8 × 0 and 20% PhiX and custom sequencing primers.

Sequencing Primers:
Index primer: TCGGCAGTTGGGAAGAGCATAGTCGTAGAGCACGC
R1 primer: CCAAGAAGGGCGGCAAGATCGCCGTGTAATAATTCTAGA

9 Primers:

	Primer Name	Sequence
	Lib_Hand_RT	ATGCTCTTCCCAACTGCCGACGGGGAGTGTACTAGT
	Lib_Hand	TGCTCTTCCCAACTGCCGA
	Lib_seq_Luc_R	TACAACCGCCAAGAAGCTGC
	P5_seq_Luc_F	AATGATACGGCGACCACCGAGATCTACACTACAACCGCCAAGAAG CTGC
	P7_Ind_#_Han	CAAGCAGAAGACGGCATACGAGATNNNNNNNNGCGTGCTCTACG ACTATGCTCTTCCCAACTGCCGA
	Index primer	TCGGCAGTTGGGAAGAGCATAGTCGTAGAGCACGC
	R1 primer	CCAAGAAGGGCGGCAAGATCGCCGTGTAATAATTCTAGA

P7_ind_#_Han Index primers:

	Index Primer Name	Index Sequence in Primer 5'-3'	Index Sequence after Sequencing 5'-3' (for demultiplexing)	Nextera Name	Full Primer Sequence
	P7_Ind_1_Han	TCGCCTTA	TAAGGCGA	[H/N]701	CAAGCAGAAGACGG CATACGAGATTCGC CTTAGCGTGCTCTA CGACTATGCTCTTC CCAACTGCCGA
	P7_Ind_2_Han	CTAGTACG	CGTACTAG	[H/N]702	CAAGCAGAAGACGG CATACGAGATCTAGT ACGGCGTGCTCTAC GACTATGCTCTTCC CAACTGCCGA
	P7_Ind_3_Han	TTCTGCCT	AGGCAGAA	[H/N]703	CAAGCAGAAGACGG CATACGAGATTTCTG CCTGCGTGCTCTAC GACTATGCTCTTCC CAACTGCCGA
	P7_Ind_4_Han	GCTCAGGA	TCCTGAGC	[H/N]704	CAAGCAGAAGACGG CATACGAGATGCTC AGGAGCGTGCTCTA CGACTATGCTCTTC CCAACTGCCGA
	P7_Ind_5_Han	AGGAGTCC	GGACTCCT	[H/N]705	CAAGCAGAAGACGG CATACGAGATAGGA GTCCGCGTGCTCTA CGACTATGCTCTTC CCAACTGCCGA
	P7_Ind_6_Han	CATGCCTA	TAGGCATG	[H/N]706	CAAGCAGAAGACGG CATACGAGATCATGC CTAGCGTGCTCTAC GACTATGCTCTTCC CAACTGCCGA
	P7_Ind_7_Han	GTAGAGAG	CTCTCTAC	[H/N]707	CAAGCAGAAGACGG CATACGAGATGTAGA GAGGCGTGCTCTAC GACTATGCTCTTCC CAACTGCCGA
	P7_Ind_8_Han	CCTCTCTG	CAGAGAGG	[H/N]708	CAAGCAGAAGACGG CATACGAGATCCTCT CTGGCGTGCTCTAC GACTATGCTCTTCC CAACTGCCGA

	Index Primer Name	Index Sequence in Primer 5'-3'	Index Sequence after Sequencing 5'-3' (for demultiplexing)	Nextera Name	Full Primer Sequence
	P7_Ind_9_Han	AGCGTAGC	GCTACGCT	[H/N]709	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGCGTGCTCTACGACTATGCTCTTCCCAACTGCCGA
	P7_Ind_10_Han	CAGCCTCG	CGAGGCTG	[H/N]710	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGCGTGCTCTACGACTATGCTCTTCCCAACTGCCGA
	P7_Ind_11_Han	TGCCTCTT	AAGAGGCA	[H/N]711	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGCGTGCTCTACGACTATGCTCTTCCCAACTGCCGA
	P7_Ind_12_Han	TCCTCTAC	GTAGAGGA	[H/N]712	CAAGCAGAAGACGGCATACGAGATTCCTCTACGCGTGCTCTACGACTATGCTCTTCCCAACTGCCGA
	P7_Ind_13_Han	TCATGAGC	GCTCATGA	[H/N]714	CAAGCAGAAGACGGCATACGAGATTCATGAGCGCGTGCTCTACGACTATGCTCTTCCCAACTGCCGA
	P7_Ind_14_Han	CCTGAGAT	ATCTCAGG	[H/N]715	CAAGCAGAAGACGGCATACGAGATCCTGAGATGCGTGCTCTACGACTATGCTCTTCCCAACTGCCGA
	P7_Ind_15_Han	TAGCGAGT	ACTCGCTA	[H/N]716	CAAGCAGAAGACGGCATACGAGATTAGCGAGTGCGTGCTCTACGACTATGCTCTTCCCAACTGCCGA
	P7_Ind_16_Han	GTAGCTCC	GGAGCTAC	[H/N]718	CAAGCAGAAGACGGCATACGAGATGTAGCTCCGCGTGCTCTACGACTATGCTCTTCCCAACTGCCGA
	P7_Ind_17_Han	TACTACGC	GCGTAGTA	[H/N]719	CAAGCAGAAGACGGCATACGAGATTACTACGCGCGTGCTCTACGACTATGCTCTTCCCAACTGCCGA
	P7_Ind_18_Han	AGGCTCCG	CGGAGCCT	[H/N]720	CAAGCAGAAGACGGCATACGAGATAGGCTCCGGCGTGCTCTACGACTATGCTCTTCCCAACTGCCGA
	P7_Ind_19_Han	GCAGCGTA	TACGCTGC	[H/N]721	CAAGCAGAAGACGGCATACGAGATGCAGCGTAGCGTGCTCTACGACTATGCTCTTCCCAACTGCCGA
	P7_Ind_20_Han	CTGCGCAT	ATGCGCAG	[H/N]722	CAAGCAGAAGACGGCATACGAGATCTGCGCATGCGTGCTCTACGACTATGCTCTTCCCAACTGCCGA
	P7_Ind_21_Han	GAGCGCTA	TAGCGCTC	N723	CAAGCAGAAGACGGCATACGAGATGAGCGCTAGCGTGCTCTACGACTATGCTCTTCCCAACTGCCGA
	P7_Ind_22_Han	CGCTCAGT	ACTGAGCG	N724	CAAGCAGAAGACGGCATACGAGATCGCTCAGTGCGTGCTCTACGACTATGCTCTTCCCAACTGCCGA

	Index Primer Name	Index Sequence in Primer 5'-3'	Index Sequence after Sequencing 5'-3' (for demultiplexing)	Nextera Name	Full Primer Sequence
	P7_Ind_23_Han	GTCTTAGG	CCTAAGAC	N726	CAAGCAGAAGACGG CATACGAGATGTCTT AGGGCGTGCTCTAC GACTATGCTCTTCC CAACTGCCGA
	P7_Ind_24_Han	ACTGATCG	CGATCAGT	N727	CAAGCAGAAGACGG CATACGAGATACTGA TCGGCGTGCTCTAC GACTATGCTCTTCC CAACTGCCGA

Protocol references

Nana Matoba*, Jessica McAfee*, Oleh Krupa, Jess Bell, Brandon D. Le, Jordan M. Valone, Gregory E. Crawford, Hyejung Won, Jason L. Stein. Massively parallel assessment of gene regulatory activity at human cortical structure associated variants. *Biorxiv.* (2025) https://doi.org/10.1101/2025.02.08.635393 *Co-first authors

Sool Lee*, Jessica C. McAfee*, Jiseok Lee*, Alejandro Gomez, Austin T. Ledford, Declan Clarke, Hyunggyu Min, Mark B. Gerstein, Alan P. Boyle, Patrick F. Sullivan, Adriana S. Beltran, Hyejung Won. Massively parallel reporter assay investigates shared genetic variants of eight psychiatric disorders. *Cell.* (2024) https://doi.org/10.1016/j.cell.2024.12.022 *Co-first authors

Acknowledgements

We thank members of the Won lab, Dr. Michael Love, Ariana Marquez Gonzalez, and Rachel Sharp for helpful discussions and comments about this paper. This research was supported by the NIH New Innovator Award from the NIMH (DP2MH122403, H.W.), the IGVF Consortium (UM1HG012003, H.W. and U01HG011952, A.P.B.), the PsychENCODE Consortium (R01MH122509, H.W.), NICHD (T32HD040127, J.L.), NIGMS (5T32GM067553, S.L. and 5T32GM135128, J.C.M. and A.G.), and Genomics of ASD: Pathways to Genetic Therapies award from the Simons Foundation Autism Research Initiative (H.W.). We also acknowledge the technical support from the UNC High Throughput Sequencing Facility (University Cancer Research Fund, Comprehensive Cancer Center Core support grant [P30CA016086], and UNC Center for Mental Health and Susceptibility grant [P30ES010126]), as well as UNC Advanced Analytics Core (Center for Gastrointestinal Biology and Disease grant [P30DK034987]), and UNC Lenti-shRNA Core. Lastly, we would like to send our utmost gratitude toward our pets—Achilles, Daisy, Nahla, and Jiji—for their unconditional love and moral support.