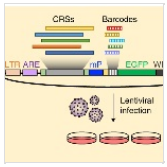



Oct 28, 2020

# Massive Parallel Reporter Assay (MPRA)

Gracie Gordon[1], Nadav Ahituv[2]

[1]UCSF; [2]University of California, San Francisco

1 Works for me This protocol is published without a DOI.

AhituvLab

Ofer Yizhar-Barnea
UCSF

ABSTRACT

Massively parallel reporter assays (MPRAs) can simultaneously measure the function of thousands of candidate regulatory sequences (CRSs) in a quantitative manner. In this method, CRSs are cloned upstream of a minimal promoter and reporter gene, alongside a unique barcode, and introduced into cells. If the CRS is a functional regulatory element, it will lead to the transcription of the barcode sequence, which is measured via RNA sequencing and normalized for cellular integration via DNA sequencing of the barcode. This technology has been used to test thousands of sequences and their variants for regulatory activity, to decipher the regulatory code and its evolution, and to develop genetic switches. Lentivirus-based MPRA (lentiMPRA) produces 'in-genome' readouts and enables the use of this technique in hard-to-transfect cells. Here, we provide a detailed protocol for lentiMPRA, along with a user-friendly Nextflow-based computational pipeline—MPRAflow—for quantifying CRS activity from different MPRA designs. The lentiMPRA protocol takes ~2 months, which includes sequencing turnaround time and data processing with MPRAflow.

EXTERNAL LINK

https://www.nature.com/articles/s41596-020-0333-5

THIS PROTOCOL ACCOMPANIES THE FOLLOWING PUBLICATION

Gordon MG, Inoue F, Martin B, Schubach M, Agarwal V, Whalen S, Feng S, Zhao J, Ashuach T, Ziffra R, Kreimer A. lentiMPRA and MPRAflow for high-throughput functional characterization of gene regulatory elements. Nature Protocols. 2020 Aug;15(8):2387-412.

ATTACHMENTS

s41596-020-0333-5.pdf

EXTERNAL LINK

https://www.nature.com/articles/s41596-020-0333-5

PROTOCOL CITATION

Gracie Gordon, Nadav Ahituv 2020. Massive Parallel Reporter Assay (MPRA). **protocols.io** https://protocols.io/view/massive-parallel-reporter-assay-mpra-bmbgk2jw

MANUSCRIPT CITATION please remember to cite the following publication along with this protocol

Gordon MG, Inoue F, Martin B, Schubach M, Agarwal V, Whalen S, Feng S, Zhao J, Ashuach T, Ziffra R, Kreimer A. lentiMPRA and MPRAflow for high-throughput functional characterization of gene regulatory elements. Nature Protocols. 2020 Aug;15(8):2387-412.

EXTERNAL LINK

https://www.nature.com/articles/s41596-020-0333-5



CREATED

Sep 12, 2020

LAST MODIFIED

Oct 28, 2020

PROTOCOL INTEGER ID

42056

ATTACHMENTS

s41596-020-0333-5.pdf

MATERIALS TEXT

MATERIALS

NEB 10-beta Electrocompetent E.coli - 6x0.1 ml **New England Biolabs Catalog #C3020K**

NEBNext High-Fidelity 2X PCR Master Mix - 250 rxns **New England Biolabs Catalog #M0541L**

I-SceI - 500 units **New England Biolabs Catalog #R0694S**

AgeI-HF - 300 units **New England Biolabs Catalog #R3552S**

NEBuilder HiFi DNA Assembly Master Mix - 250 rxns **New England Biolabs Catalog #E2621X**

QIAprep Spin Miniprep Kit **Qiagen Catalog #27104**

TURBO DNA-free™ Kit **Thermo Scientific Catalog #AM1907**

Qiaquick gel extraction kit **Qiagen Catalog #28704**

Wizard(R) SV Genomic DNA Purification System, 250 preps **Promega Catalog #A2361**

DMEM **Thermo Fisher Scientific Catalog #41966**

pMD2.G **addgene Catalog #12259**

psPAX2 **addgene Catalog #12260**

SbfI-HF **New England Biolabs Catalog #RS3642S**

Superscript II Reverse Transcriptase **Thermo Fisher Scientific Catalog #18064071**

HEK293 **ATCC Catalog #CRL-1573**

HighPrep™ PCR **Contributed by users Catalog #AC-60050**

2X SsoFast EvaGreen Supermix with Low ROX **BIO-RAD Catalog #1725211**

Trypsin-EDTA (0.05%), phenol red **Thermo Fisher Catalog #25300120**

SYBR™ Safe DNA Gel Stain **Thermo Fisher Catalog #S33102**

Penicillin-Streptomycin **Sigma Aldrich Catalog #P4333**

FBS **Contributed by users**

SurePrint 244K Oligonucleotide Libraries **Agilent Technologies Catalog #G7223A**

Qiagen Plasmid Plus Midi Kit **Qiagen Catalog #12945**

pLS-SV40-mP-EGFP **addgene Catalog #137724**

pLS-SceI **addgene Catalog #137725**

Lenti-X Concentrator **Takara Catalog #631232**

AllPrep DNA/RNA Mini Kit **Qiagen Catalog #80204**

DISCLAIMER:



ABSTRACT

Massively parallel reporter assays (MPRAs) can simultaneously measure the function of thousands of candidate regulatory sequences (CRSs) in a quantitative manner. In this method, CRSs are cloned upstream of a minimal promoter and reporter gene, alongside a unique barcode, and introduced into cells. If the CRS is a functional regulatory element, it will lead to the transcription of the barcode sequence, which is measured via RNA sequencing and normalized for cellular integration via DNA sequencing of the barcode. This technology has been used to test thousands of sequences and their variants for regulatory activity, to decipher the regulatory code and its evolution, and to develop genetic switches. Lentivirus-based MPRA (lentiMPRA) produces 'in-genome' readouts and enables the use of this technique in hard-to-transfect cells. Here, we provide a detailed protocol for lentiMPRA, along with a user-friendly Nextflow-based computational pipeline—MPRAflow—for quantifying CRS activity from different MPRA designs. The lentiMPRA protocol takes ~2 months, which includes sequencing turnaround time and data processing with MPRAflow.

## Library Amplification 3h

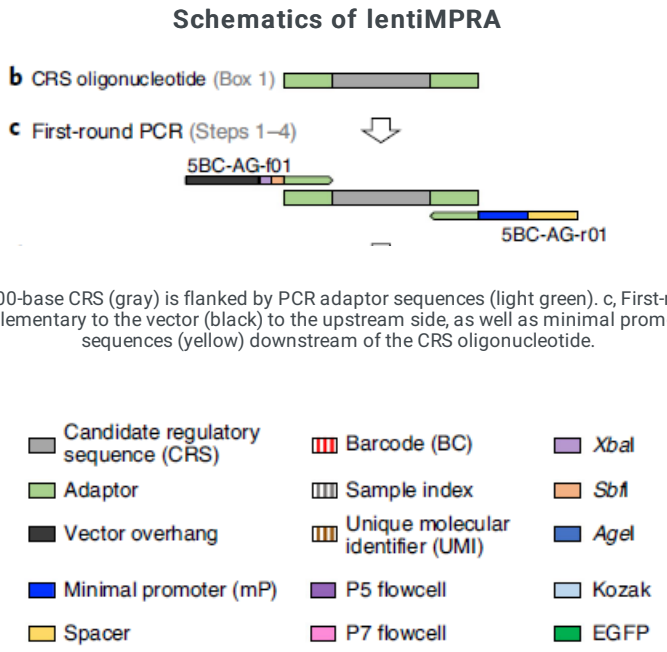
1 Dissolve the Agilent oligonucleotide (10 pmol) (Fig. 1b and Box 1) in 100 µL TE buffer to obtain a 100 nM solution.

2 Set up the first-round PCR reaction. This reaction adds a vector overhang sequence upstream and minimal promoter and adaptor sequences downstream of the CRSs

| Reagent | Volume (µL) | Final conc |
|---|---|---|
| Agilent oligonucleotide (100 nM) | 2 | 1nM |
| NEBNext High-Fidelity 2× PCR Master Mix | 100 | 1x |
| 5BC-AG-f01 (100 µM) | 1 | 0.5uM |
| 5BC-AG-r01 (100 µM) | 1 | 0.5uM |
| Ultrapure distilled H2O | 96 | |
| Total volume | 200 | |

Step 2 includes a Step case.
**Schematics of lentiMPRA**

step case




b, CRS oligonucleotide. A 200-base CRS (gray) is flanked by PCR adaptor sequences (light green). c, First-round PCR. PCR primers add sequences that are complementary to the vector (black) to the upstream side, as well as minimal promoter (mP, blue) and spacer sequences (yellow) downstream of the CRS oligonucleotide.

3 Split the premixture into five PCR tubes (40 µL per tube)

**CRITICAL STEP** Splitting the PCR reaction into multiple tubes is important to reduce the risk of PCR 'jackpotting' (errors that occur during the early PCR cycles and get amplified exponentially) or amplification bias accidentally occurring during PCR

4 Run the PCR reaction as follows:

| Cycle No. | Denature | Anneal | Extend |
|---|---|---|---|
| 1 | 98oC, 2min | | |
| 2-6 (5cycles) | 98oC, 15sec | 60oC, 20sec | 720C, 30sec |
| 7 | | | 72oC, 5min |

5 Combine the PCR products in a DNA LoBind tube.

6 Bring the HighPrep PCR reagent to RT for at least 30 min before use. Shake thoroughly to fully resuspend the magnetic beads. Add 1 volume (200 µL) of HighPrep PCR reagent and mix

thoroughly by pipetting up and down 6–8 times

7 Incubate the mixture for 5 min at RT.

8 Place the tube on the magnet for 2–3 min until the solution clears and beads pull to the side of the wall

9 Carefully remove the supernatant by pipetting without disturbing the beads. A small amount (10–20 µL) of supernatant can be left in the tube

10 With the tube on the magnet, add 500 µL of 80% (vol/vol) ethanol and incubate for 30 s.

11 Remove the ethanol by pipetting while the tube is still on the magnet.

12 Repeat the 80% (vol/vol) ethanol washing (Steps 10 and 11)

13 Flash-spin (200–1,000g, 22–25 °C, 3 s) the tube, immediately place it back on the magnet and remove the supernatant

14 Dry the bead pellet for 2–3 min. Do not overdry the beads.

15 Add 50 µL of Buffer EB to the beads and mix by pipetting and vortexing.

16 Place the tube on the magnet for 1–2 min until the solution is clear.

17 Transfer 45–50 µL of the eluate to a DNA LoBind tube.

18 Measure the DNA concentration using a NanoDrop spectrophotometer. The expected concentration is 5–20 ng/µL.

19 

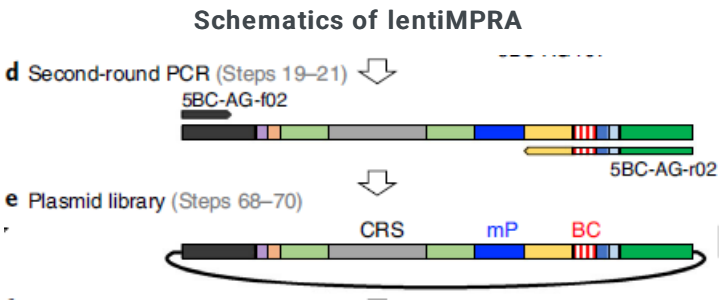
Set up the second-round PCR reaction. This reaction adds a 15-bp barcode and vector overhang sequence downstream of the first-round PCR fragment

| Reagent | Volume (µL) | Final conc |
|---|---|---|
| First-round PCR product | Variable (100ng) | 1nM |
| NEBNext High-Fidelity 2× PCR Master Mix | 200 | 1x |
| 5BC-AG-f01 (100 µM) | 2 | 0.5uM |
| 5BC-AG-r01 (100 µM) | 2 | 0.5uM |
| Ultrapure distilled H2O | Up to 400uL | |
| Total volume | 400 | |

Step 19 includes a Step case.

**Schematics of lentiMPRA**

step case




d, Second-round PCR. Reverse primer adds the barcodes (red-striped section) and GFP complementary sequences (green).
e, Plasmid construct

20 Split the premixture into 10 PCR tubes (40 µL per tube).

21 Run the PCR reaction as follows:

| Cycle No. | Denature | Anneal | Extend |
|---|---|---|---|
| 1 | 98oC, 2min | | |
| 2-13 (12 cycles) | 98oC, 15sec | 60oC, 20sec | 720C, 30sec |
| 14 | | | 72oC, 5min |

22 Combine the PCR products into a DNA LoBind tube.

23 Add 200 µL of 6× gel loading dye (final conc. 2×) and mix the solution by vortexing.

24 Run the sample on two 1.5% (wt/vol) TAE–agarose gels (30 mL of 5 × 6-cm mini gels with 3-cm-width wells) and visualize the DNA using SYBR Safe DNA gel stain.

25 Cut the DNA band (419 bp) using a blue-light Safe Imager.

**CRITICAL STEP** We highly recommend using a blue-light Safe Imager, because the UV transilluminator markedly decreases the recombination efficiency.

26 Purify the DNA from the gel slice, using QIAquick Gel Extraction Kit according to the manufacturer's protocol.

27 Elute the DNA in 50 µL of Buffer EB per column. If multiple columns are used, combine the eluate.

28 Purify the DNA using 1.2 volumes of HighPrep PCR reagent and following Steps 6–17.

29 Measure the DNA concentration using a NanoDrop spectrophotometer. The expected concentration is ~25 ng/µL.

**PAUSE POINT** Purified DNA can be stored at −20 °C for months.

## Vector Linearization

30

Set up the vector digestion reaction as follows:

| Reagent | Volume (uL) | Final Conc. |
|---|---|---|
| pLS-SceI | Variable (10ug) | |
| CutSmart (10x) | 20 | 1x |
| AgeI-HF (20U/uL) | 5 | 0.5U/uL |
| SbfI-HF (20U/uL) | 5 | 0.5U/uL |
| Ultrapure distilled H2O | make up to 200uL | |
| Total volume | 200 | |

31 Incubate the reaction at 37 °C for 3 h to overnight.

32 To complete the plasmid digestion, add 5 µL of AgeI-HF (20 U/µL) and 5 µL of SbfI-HF (20 U/µL) to the reaction.

33 33 Incubate the reaction at 37 °C for 3 h to overnight.

34 Vortex for 30 s and incubate at 80 °C for 20 min.

35 Purify the DNA using 0.65 volume (136.5 µL) of HighPrep PCR reagent and following Steps 6–17.

36 Measure the DNA concentration using a NanoDrop spectrophotometer.

37 To check the DNA size and quality, run 100–200 ng of the linearized vector and purified insert DNA (from Step 29) on a 1% (wt/vol) gel along with a 1-kb DNA ladder. Make sure that specific single bands (7.8-kb linearized vector and 419-bp insert DNA) appear, but that no other bands appear on the gel.

Recombination and electroporation 3d

38 Set up the recombination reaction as follows:

| Reagent | Volume (uL) | Final Conc. |
|---|---|---|
| Linearized pLS-SceI (from step 36) | Variable (1 µg) | |
| Purified insert DNA (from step 29) | Variable (250 ng) | |
| NEBuilder HiFi DNA Assembly Master Mix | 100 | 1X |
| Ultrapure distilled H2O | Make up to 200 µL | |
| Total volume | 200 | |

39 Incubate the reaction at 37 °C for 1 h.

40 Place the tube on ice

41 Purify the DNA using 0.65 volume (136.5 µL) of HighPrep PCR reagent

42 Set up the digestion reaction to get rid of undigested vectors as follows:

| Reagent | Volume (uL) | Final Conc. |
|---|---|---|
| Recombination product | 44 | |
| CutSmart buffer (10×) | 5 | 1X |
| I-SceI (20 U/µL) | 1 (20U) | 0.4 U/uL |
| Total volume | 50 | |

43 Incubate the reaction at 37 °C for 1 h.

44 Purify the DNA using 1.8 volume (90 µL) of HighPrep PCR reagent and following Steps 6–14.

45 To elute the DNA, add 20 µL of Buffer EB to the beads and mix by pipetting and vortexing

46 Place the tube on the magnet for 1–2 min and transfer 18 µL of the eluate to a DNA LoBind tube.

47 ⏸

Measure the DNA concentration using a NanoDrop spectrophotometer.
Make sure the DNA the concentration of the recombination product (the eluate) is >25 ng/µL, so as not to need to add >4 µL to 100 µL of competent cells in Step 50.

**PAUSE POINT** The recombinant product can be stored at −20 °C for at least a month.

48 Prewarm 4–5 mL of Stable Outgrowth Medium (from the NEB 10-beta electrocompetent cells) at 37 °C for at least 30 min.

49 ⚠

Thaw NEB 10-beta electrocompetent cells on ice. We usually use 100 µL (one tube) of the competent cells for low-complexity libraries (0.5–2 million total barcodes) and 400 µL (four tubes) for high-complexity libraries (8–12 million total barcodes).

**CRITICAL STEP** Competent cells and cuvettes should be kept on ice during the following procedure.

50 Add 100 ng of the recombination product per 100 µL competent cells. The volume of DNA should be <4 µL per 100 µL competent cells.

51 Gently transfer 50 µL of the competent cells to a 1-mm-gap cuvette without creating bubbles. Two cuvettes are prepared for 100 µL of competent cells.

52 Gently tap the cuvettes on the counter to move the cells to the bottom.

53 Place the cuvettes in a Gemini X2 electroporator and shock the cells with the following settings: voltage, 2.0 kV; resistance, 200 ohms; capacitance, 25 µF, number of pulses, 1; gap width, 1 mm.

54 Immediately add 450 µL of prewarmed Stable Outgrowth Medium to the cuvettes, thoroughly mix by pipetting up and down, and transfer to a 14-mL conical tube.

55 Repeat the electroporation for all cuvettes, combining the electroporated bacteria in a single tube (total 1 mL culture per 100 µL competent cells).

56 Add fresh Stable Outgrowth Medium and scale up to 4 mL in total. If 400 µL competent cells were used, there is no need to add more.

57 Dilute 2 µL of the bacteria in 400 µL of fresh LB medium in a 1.5-mL tube and plate the entire tube of diluted bacteria (402 µL) in a prewarmed 15-cm LB agar plate along with 20 µL of 100 mg/mL carbenicillin. This plate will be used for colony counting and plasmid mini prep.

58 ⚠

Incubate the cells at 37 °C for 1 h with agitation (200 r.p.m.). Prewarm ten 15-cm LB agar plates at 37 °C.

**CRITICAL STEP** We recommend using 15-cm plates rather than larger plates because these enable fine-tuning of the colony numbers collected.

59 Plate undiluted bacteria onto the other nine prewarmed 15-cm LB agar plates (400 µL/plate), along with 100 µL/plate of 100 mg/mL carbenicillin. A higher amount of carbenicillin than usual is added because the dense culture conditions increase the risk of non-transformed bacteria growth.

60 Incubate the plates at 37 °C overnight.

61 ⏸

To check the plasmid sequence of individual colonies, pick 16 colonies from the diluted-bacteria plate, purify the plasmids using a QIAprep Spin Miniprep Kit, and send them for Sanger sequencing using n40.dn.F and EGFP.up.R primers (Supplementary Table 3). Confirm that the sequence structure corresponds to the design (Fig. 1e, Extended Data Fig. 1c).

**PAUSE POINT** The plates can be stored at 4 °C for a month.
? TROUBLESHOOTING

## Colony counting and plasmid library prep

62 Count the number of colonies on the diluted-bacteria plate. If there are too many colonies, count colonies in a quarter of the area and multiply by four to estimate the total number of colonies on the plate.

63 Estimate the total number of colonies per undiluted-bacteria plate by multiplying the colony count in the diluted-bacteria plate by 200 (Supplementary Table 1).

64 ⚠

Determine the number of undiluted-bacteria plates to be used for the following plasmid preps. The total number of colonies needed can be determined by multiplying the number of designed CRSs by the desired number of barcodes per CRS

**CRITICAL STEP** The ideal number of barcodes per CRS is between 50 and 200. Fewer barcodes per CRS may reduce reproducibility. More barcodes per CRS requires more cells, more virus and deeper sequencing reads, which increase costs. In addition, associating >200 barcodes per CRS does

not increase reproducibility.

65 Add 5–6 mL of LB medium to each bacterial plate and gently scrape the colonies, using a cell lifter without disturbing solid agarose.

66 Collect the bacterial suspension and combine into a few 50-mL tubes.

67 Add 5–6 mL of fresh LB medium again to the plates and collect as much leftover bacteria as possible into the tubes.

68 Purify the plasmids using a Qiagen Plasmid Plus Midi Kit, following the 'standard protocol' in the manufacturer's protocol. The number of columns to be used depends on the amount of bacteria. We usually use four columns of Qiagen Plasmid Plus Midi Kit per undilutedbacteria plate.

69 Measure the plasmid concentration using a NanoDrop spectrophotometer. The expected concentration is 0.5–2 µg/µL.

70

To check the DNA size and quality, run 100–200 ng of the plasmid on a 1% (wt/vol) agarose gel along with a 1-kb DNA ladder.

**PAUSE POINT** The purified plasmid library can be stored at −20 °C for years.

Sequencing for CRS–barcode association 2d

71 Set up the PCR reaction. This reaction adds a P5 flowcell sequence and the sample index sequence upstream and a P7 flowcell sequence downstream of the CRS–barcode fragment .

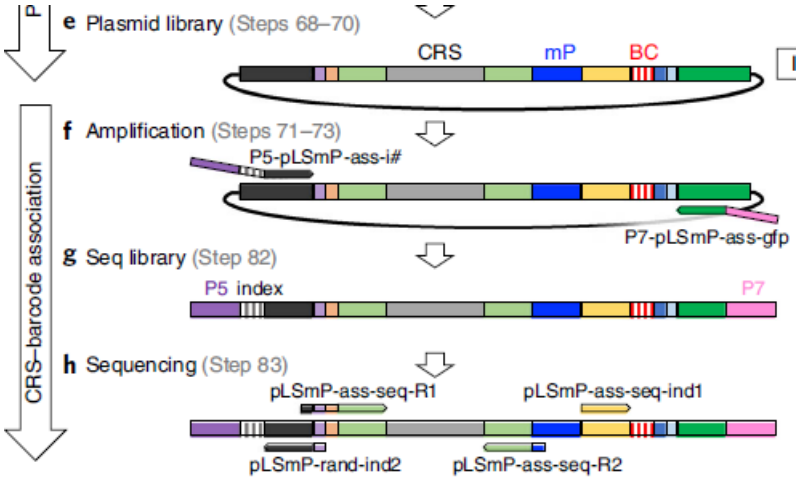
**NOTICE: Use different sample index sequences for pLSmP-ass-i# if multiple libraries are generated and multiplexed**

| Reagent | Volume (uL) | Final Conc. |
|---|---|---|
| Plasmid library | Variable (40ng) | |
| NEBNext High-Fidelity 2× PCR Master Mix | 100 | 1x |
| pLSmP-ass-i# (100 µM) | 1 | 0.5uM |
| pLSmP-ass-gfp (100 µM) | 1 | 0.5uM |
| Ultrapure distilled H2O | Make up to 200uL | |
| Total volume | 200 | |

Step 71 includes a Step case.
**Schematics of lentiMPRA**

step case

**Schematics of lentiMPRA**




e, Plasmid construct. f, Amplification for CRS–barcode association. Primers add P5 (purple) and sample index (gray-striped section) upstream and P7 (pink) downstream. g, Sequencing library structure. h, Sequencing reaction. Paired-end reads specify the CRS sequence, with index read 1 providing the barcode and index read 2 reading the sample index for multiplexing. i, Integrated DNA and expressed RNA in infected cells.

72 Split the premixture into five PCR tubes (40 µL per tube).

73 ⚠

Run the PCR reaction as follows:

| Cycle No. | Denature | Anneal | Extend |
|---|---|---|---|
| 1 | 98oC, 1min | | |
| 2-16 (15 cycles) | 98oC, 15sec | 60oC, 20sec | 720C, 3min |
| 17 | | | 72oC, 5min |

**CRITICAL STEP** Incomplete DNA elongation may create chimeric DNA annealing in the next cycle and can cause CRS–barcode swapping. A longer extension time (3 min) can help to reduce this risk.

74 Combine the PCR products in a DNA LoBind tube.

75 Add 100 µL of 6× gel loading dye (final conc. 2×) and mix the solution by vortexing.

76 Run the sample on a 1.5% (wt/vol) agarose gel
(30 mL of 5 cm × 6-cm mini gel with 3-cm widthwell).

77 Cut the DNA band (470 bp) using a blue-light Safe Imager.

78 Purify the DNA from the gel slice using the QIAquick Gel Extraction Kit according to the manufacturer's protocol.

79

79 Elute the DNA in 50 µL Buffer EB per column. If multiple columns are used, combine the eluate.

80 Purify the DNA using 1.8 volumes of HighPrep PCR reagent, following Steps 6–17.

81 Measure the DNA concentration using Qubit dsDNA HS Assay Kit according to the manufacturer's protocol.

82

To check the DNA size and quality, run 50–100 ng of the DNA on a 1.5% (wt/vol) agarose gel along with a 100-bp DNA ladder.

**PAUSE POINT** Purified DNA can be stored at −20 °C for months.

83 Send the purified DNA and custom primers for sequencing.

The sequencing should be done using paired-end reads covering the full CRS with some overlap (here 146 bp each), with 15 cycles for index read 1 and 10 cycles for index read 2. Index read 1 provides the barcode sequence, and index read 2 provides the sample index

Let the sequencing facility know that index read 2 should be used for demultiplexing and that short reads should not be masked, bcl2fastq parameters:
*--minimum-trimmed-read-length 0 --mask-short-adapter-reads 0.*

A minimum 10× coverage of sequencing reads based on the total number of barcodes is required. For example, we use an Illumina MiSeq v.2 run (15 million reads) for a 0.5-million-barcode library or an Illumina NextSeq mid-output run (120 million reads) for an 8- to 12-million-barcode library

| READ | CYCLES | PRIMER | OUTPUT |
|---|---|---|---|
| Read 1 | 146 | pLSmP-ass-seq-R1 | CRS (upstream, forward) |
| Read 2 | 146 | pLSmP-ass-seq-R2 | CRS (downstream, reverse) |
| Index read 1 | 15 | pLSmP-ass-seq-ind1 | Barcode (forward) |
| Index read 2 | 10 | pLSmP-rand-ind2 | Sample index |

Step 83 includes a Step case.
**NGS primer schematics**

Lentivirus packaging

step case

## NGS primer schematics


84 Culture 293T cells in DMEM (with 10% heat-inactivated FBS).

85 Seed 10–12 million 293T cells per T225 flask. The number of flasks depends on the library complexity and the infectability of the cells. For example, we use one flask for the 0.5-millionbarcode library or six T225 flasks for the 8- to 12-million-barcode library when carrying out lentiMPRA in HepG2 cells.

86 Incubate the cells for 2 d. 2d

87 Prepare premixtures A and B as follows:
- Premix A: 800 µL/flask OPTI-MEM and 60 µL/flask EndoFectin.
- Premix B: 800 µL/flask OPTI-MEM, 10 µg/flask plasmid library, 6.5 µg/flask psPAX2, and 3.5 µg/flask pMD2.G.

88 Add **premix A** to **premix B** by pipetting and mix the mixture by inverting the tube.

89 Incubate the tube at RT for 15 min.

90 Add 1.6 mL/flask of the mixture to 293T cells (from Step 86).

91 Incubate the cells for 8–14 h. ! CAUTION During the following procedure (Steps 92–103), the liquid and plastic waste should be discarded into 10% (vol/vol) bleach, because they are contaminated with lentivirus. Culture plates and tubes for storage should be clearly labeled as lentiviral contaminants.

92 Replace the media with 30 mL/flask DMEM (with 5% heat-inactivated FBS) supplemented with 1× ViralBoost reagent (60 µL reagent per 30 mL medium).

93 Incubate the cells for 36–48 h. 2d

94 Check GFP expression using a fluorescence microscope. The majority of cells are expected to express strong GFP, because viral RNA, including the GFP gene, is transcribed via the 5′ long terminal repeat (LTR).

95 Filter the supernatant through a 0.45-µm PES filter system. Use multiple filters (one filter for up to three T225 flasks), so as not to get clogged.

96 Transfer the flow-through into multiple 50-mL tubes (30 mL per tube).

97 Add 1/3 volume (10 mL per tube) of Lenti-X concentrator reagent, close the lid tightly and mix gently by inverting the tubes.

98

Seal the lid with Parafilm and place the tube in a refrigerator at 4 °C for at least 4 h.

**PAUSE POINT** The tubes can be stored at 4 °C up to 1 week.

99 Centrifuge the tubes at 1,500g for 45 min at 4 °C.

100 Discard the supernatant into 10% (vol/vol) bleach by gentle decanting.

101 Discard the remainder of the supernatant into 10% bleach by pipetting without disturbing the pellet.

102 Gently resuspend the pellet in cold DPBS. We usually use 600 µL DPBS per T225 flask.

103

Store the lentivirus at 4 °C. **CRITICAL STEP** We do not recommend freezing the virus, especially in the case of high complexity libraries, because freeze–thaw cycles substantially decrease the viral titer. We did not see amarked loss of the viral titer when stored at 4 °C for up to 3 weeks. The following infection experiments should be done within 3 weeks.

**PAUSE POINT** The virus can be stored at 4 °C up to 3 weeks

Lentivirus titration

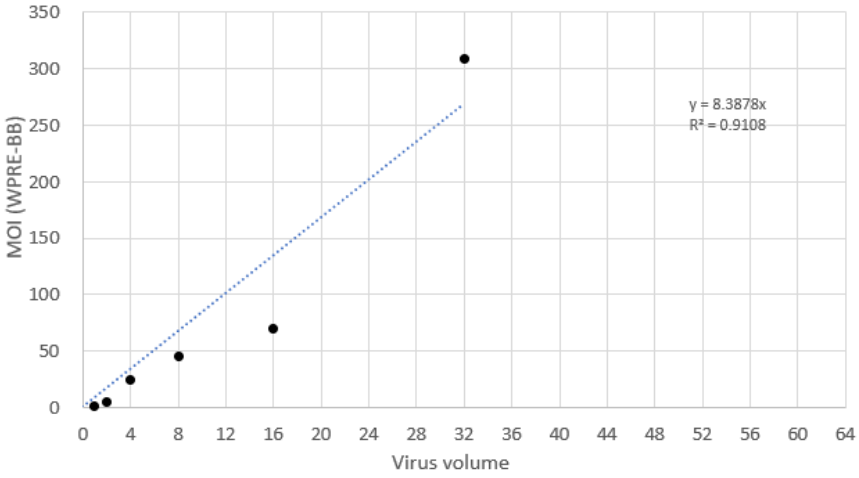
104 Infect the lentivirus library (0, 1, 2, 4, 8, 16, 32, 64 µL) into the cells to be used, extract genomic DNA from the cells, and perform qPCR as described in Box 2 (steps 1–14).

105 Plot the MOI for each condition and draw a linear approximation with the virus volume on the x axis and the MOI on the y axis (Supplementary Table 2).

https://static-content.springer.com/esm/art%3A10.1038%2Fs41596-020-0333-5/MediaObjects/41596_2020_333_MOESM3_ESM.xlsx
Step 105 includes a Step case.
**Lenti titeration graph example**

step case

**Lenti titeration graph example**




106 On the basis of its slope and the number of cells seeded, calculate the virus titer (in transducing units per microliter), using Supplementary Table 1.

https://static-content.springer.com/esm/art%3A10.1038%2Fs41596-020-0333-5/MediaObjects/41596_2020_333_MOESM2_ESM.xlsx

## Lentivirus infection and DNA/RNA extraction 1w

107

Determine the number of integrations per barcode (i.e., total number of a particular barcode existing in an entire cell population)
Use the spreadsheet provided in Supplementary Table 1.
https://static-content.springer.com/esm/art%3A10.1038%2Fs41596-020-0333-5/MediaObjects/41596_2020_333_MOESM2_ESM.xlsx

We recommend a range between 50 and 500 integrations per barcode. Fewer numbers increase the risk of barcode loss during the downstream procedure. Higher numbers are better but increase the cost.

108 Seed an appropriate number of cells in 10-cm or 15-cm dishes.

The number of cells required is determined as total barcode integrations (total number of any barcodes existing in the entire cell population) divided by the MOI of the cells (Supplementary Table 1).

Three independent biological replicates should be performed, and each replicate sample should be treated separately during the following procedures

109 Incubate the cells overnight.

110 Refresh the culture media (culture conditions depend on the cell type used) and add Polybrene at the appropriate concentration (Box 2).

111 111 Add an appropriate amount of the lentivirus library. The volume of virus required is given as total barcode integrations divided by virus titer (Supplementary Table 1).

112 Refresh the culture media (culture conditions depend on the cell type used) with no Polybrene the

following day.

113 After 2 d (3 d of culture in total), check GFP fluorescence to confirm proper lentiviral integration and expression. GFP expression depends on library design. If the majority of the CRSs in the library are active enhancers, the cells should have strong GFP expression.

114 Remove the culture media and wash the cells with DPBS three times. Remove the DPBS completely.

115 Add RLT Plus lysis buffer (from the AllPrep DNA/RNA Mini Kit) supplemented with 2-mercaptoethanol (10 µL of 2-mercaptoethanol per 1 mL of RLT Plus).
We usually use 1,200 µL or 2,400 µL of lysis buffer per 10-cm dish or 15-cm dish, respectively.

116

Scrape the cells using a cell lifter and homogenize the cell lysis, using a 3-mL syringe and 20-gauge needle.

**PAUSE POINT** The cell lysate can be frozen and stored at −80 °C for months.

117 Transfer the lysate to DNA columns. Use two or four columns per 10-cm dish or 15-cm dish, respectively

118 Extract genomic DNA and total RNA simultaneously using the AllPrep DNA/RNA Mini Kit according to the manufacturer's protocol. For RNA samples, perform DNase treatment between two of the 350-µL RW1 washes using Qiagen's RNase-free DNase Set, according to the manufacturer's protocol

119 Elute DNA in 30 µL/column of Buffer EB and combine the eluates of each replicate in a single tube. Keep each of the three replicates separate.

120 Elute RNA in 30 µL/column of RNase-free H2O and combine the eluates of each replicate in a single tube. Keep each of the three replicates separate

121

Measure the concentrations of the DNA and RNA samples using a NanoDrop spectrophotometer. At least 12 µg DNA and 60 µg RNA per replicate are required.

**PAUSE POINT** DNA can be stored at −20 °C for years. RNA can be stored at −80 °C for months

Reverse transcription 4h

122 Treat RNA samples with DNase, using the TURBO DNA-free Kit and following the manufacturer's protocol for 'Rigorous DNase treatment'. Because the RNA sample has already been treated with DNase during the AllPrep procedure (Step 93), TURBO DNase treatment can be done using a high-concentration condition (without any dilution).

123

Measure the RNA concentration using the Qubit RNA HS Assay Kit. At least 60 µg or 240 µg total

RNA per replicate is required for a low- (0.5–2 million total barcodes) or high- (8–12 million total barcodes) complexity library, respectively.

**PAUSE POINT** DNase-treated RNA can be stored at −80 °C for months

124 For RNA samples, perform a reverse-transcription reaction in 8-strip PCR tubes. This reaction adds a 16-bp UMI and a P7 flowcell sequence downstream of the barcode.

For a low-complexity library (0.5–2 million total barcodes), use the amounts given in the table below. For a high-complexity library (8–12 million total barcodes), we recommend multiplying all amounts in the following table by 4

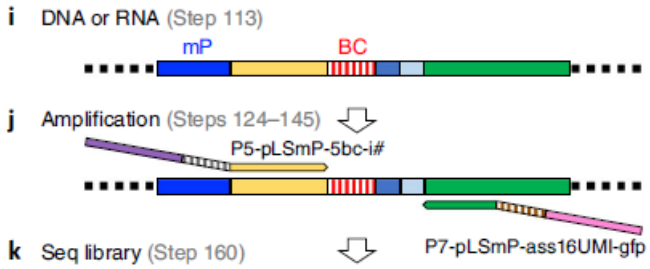
| Reagent | Volume (µL) | Final conc. |
|---|---|---|
| RNA | Variable (60 µg total RNA) | |
| P7-pLSmP-ass16UMI-gfp (100 µM) | 0.25 | 0.25 µM |
| dNTP mix (10 mM, from SuperScript II Reverse Transcriptase) | 5 | 0.5 mM |
| UltraPure distilled H2O | Make up to 65 µL | |
| Total volume | 65uL | |

Step 124 includes a Step case.

**Schematics of lentiMPRA**

step case

**Schematics of lentiMPRA**




i, Integrated DNA and expressed RNA in infected cells. j, Amplification for barcode counting. Primers add P5 and sample index upstream and P7 and UMI, brown stripe) downstream

125 Incubate the reaction at 65 °C for 5 min using a thermal cycler.

126 Place the tubes on ice.

127 Add 20 µL 5× First Strand buffer (from SuperScript II Reverse Transcriptase) and 10 µL 0.1 M DTT (from SuperScript II Reverse Transcriptase).

128 Incubate the tubes at 42 °C for 2 min using a thermal cycler.

129 Add 5 µL of Superscript II.

130

130 Incubate the tubes at 42 °C for 50 min, followed by incubation at 70 °C for 15 min using a thermal cycler.

**PAUSE POINT** cDNA can be stored at −20 °C for months.

Library prep and sequencing for RNA and DNA barcode counts

131 Dilute DNA samples (from Step 121) to a final concentration of 120 ng/µL. For RNA samples (from Step 96; we refer to RT products as RNA samples in the downstream steps to distinguish them from samples derived from genomic DNA), use all 100 µL of RT products for the following first-round PCR reaction.

132

Perform a first-round PCR reaction with all three replicates of both DNA and RNA samples.

This reaction adds the P5 flowcell sequence and sample index sequence upstream and a 16-bp UMI and P7 flowcell sequence downstream of the barcode .

Use different sample index sequences for each sample
https://static-content.springer.com/esm/art%3A10.1038%2Fs41596-020-0333-5/MediaObjects/41596_2020_333_MOESM4_ESM.xlsx
For a low-complexity library (0.5–2 million total barcodes), use the amounts given in the table below. For a high-complexity library (8–12 million total barcodes), we recommend multiplying all amounts in the following table by 4.

| Reagent | Volume (µL) | Final conc. |
|---|---|---|
| DNA or cDNA | 100 (12 µg DNA or entire RT product) | |
| NEBNext High-Fidelity 2× PCR Master Mix | 200 | 1x |
| P7-pLSmP-ass16UMI-gfp (100 µM) | 2 | 0.5 µM |
| P5-pLSmP-5bc-i# (100 µM) | 2 | 0.5 µM |
| UltraPure distilled H2O | 96 | |
| Total volume | 400 | |

for Low-Comlexity

Split the premixture into 8 PCR tubes (50 µL per tube).

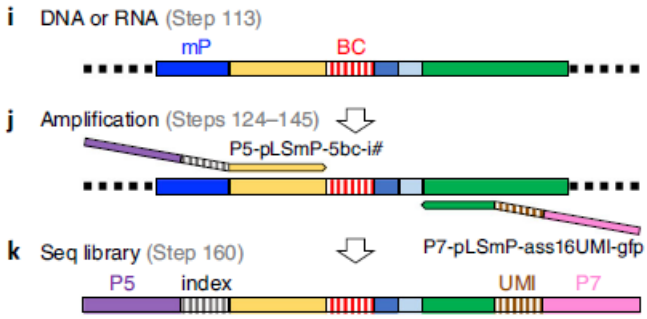
| Reagent | Volume (µL) | Final conc. |
|---|---|---|
| DNA or cDNA | 400 (48 µg DNA or entire RT product) | |
| NEBNext High-Fidelity 2× PCR Master Mix | 800 | 1x |
| P7-pLSmP-ass16UMI-gfp (100 µM) | 8 | 0.5 µM |
| P5-pLSmP-5bc-i# (100 µM) | 8 | 0.5 µM |
| UltraPure distilled H2O | 384 | |
| Total volume | 1600 | |

for High-Complexity

Split the premixture into 32 PCR tubes (50 µL per tube). 4x 8-strips
Step 132 includes a Step case.
**Schematics of lentiMPRA**

step case

**Schematics of lentiMPRA**




i, Integrated DNA and expressed RNA in infected cells. j, Amplification for barcode counting. Primers add P5 and sample index upstream and P7 and UMI, brown stripe) downstream. k, Sequencing library structure.

133

Run the PCR reaction as follows.

| Cycle | Denature | Anneal | Extend |
|---|---|---|---|
| 1 | 98oC, 1min | | |
| 2-4 (3 cycles) | 98oC, 10sec | 60oC, 30sec | 72oC, 1min |
| 5 | | | 72oC, 5min |

134 After the PCR reaction, combine all eight tubes for each sample in a DNA LoBind tube.

135 Purify the DNA, using 1.8 volumes (700 µL) of HighPrep PCR reagent and following Steps 6–14.

136 Add 60 µL of Buffer EB to the beads and elute the DNA by pipetting and vortexing.

137 Place the tube on the magnet for 1–2 min and transfer 55–58 µL of the eluate to a LoBind tube. Store the tubes on ice.

138

Set up the preliminary PCR reaction for each sample in a 96-well qPCR plate. This run finds the number of PCR cycles required for the following second-round PCR reaction.

| Reagent | Volume (µL) | Final conc. |
|---|---|---|
| First-round PCR product (from step 112) | 5 | |
| NEBNext High-Fidelity 2× PCR Master Mix | 10 | 1x |
| P7 (100uM) | 0.1 | 0.5 µM |
| P5 (100uM) | 0.1 | 0.5 µM |
| SYBR green (100X) | 0.1 | |

| | | |
|---|---|---|
| UltraPure distilled H2O | 4.7 | |
| Total volume | 20 | |

Run the qPCR reaction using a qPCR instrument as follows

| Cycle | Denature | Anneal | Extend |
|---|---|---|---|
| 1 | 98oC, 1min | | |
| 2-31 (30 cycles) | 98oC, 10sec | 60oC, 30sec | 72oC, 1min |

139 On the basis of the raw amplification curve of each sample, determine the number of cycles at which the amplification nearly plateaus for each sample.

140 Set up the second-round PCR reaction for each sample as follows.

Because it is expected that th numbers of cycles required for DNA and RNA samples will be different, we recommend running the PCRs for them separately. For a low-complexity library (0.5–2 million total barcodes), use the amounts given in the table below. For a high-complexity library (8–12 million total barcodes), we recommend multiplying all amounts in the following table by 4.

141

Set up the second-round PCR reaction for each sample as follows. Because it is expected that the numbers of cycles required for DNA and RNA samples will be different, we recommend running the PCRs for them separately. For a low-complexity library (0.5–2 million total barcodes), use the amounts given in the table below. For a high-complexity library (8–12 million total barcodes), we recommend multiplying all amounts in the following table by 4.

| **Reagent** | **Volume (µL)** | **Final conc.** |
|---|---|---|
| DNA or cDNA | 50 | |
| NEBNext High-Fidelity 2× PCR Master Mix | 100 | 1x |
| P7 (100 µM) | 1 | 0.5 µM |
| P5 (100 µM) | 1 | 0.5 µM |
| UltraPure distilled H2O | 48 | |
| Total volume | 200 | |

for Low-Comlexity

Split the premixture into 5 PCR tubes (40 µL per tube).

| Reagent | Volume (µL) | Final conc. |
|---|---|---|
| DNA or cDNA | 50 | |
| NEBNext High-Fidelity 2× PCR Master Mix | 400 | 1x |
| P7 (100 µM) | 4 | 0.5 µM |
| P5 (100 µM) | 4 | 0.5 µM |
| UltraPure distilled H2O | 192 | |
| Total volume | 800 | |

for High-Complexity

Split the premixture into 20 PCR tubes (40 µL per tube). 4x 8-strips

142 Run the qPCR reaction using a qPCR instrument as follows

According to cycles numbers established in step 113-114

| Cycle | Denature | Anneal | Extend |
|---|---|---|---|
| 1 | 98oC, 1min | | |
| 2-31 (X cycles) | 98oC, 10sec | 60oC, 30sec | 72oC, 1min |

143 Combine each sample into a DNA LoBind tube.

144 Purify the DNA using 1.8 volume (360 µL) of HighPrep PCR reagent following Steps 6–14.

145 Elute the DNA in 20 µL EB.

146 Measure the DNA concentration using a NanoDrop spectrophotometer.

147

Pool three replicates of DNA samples or RNA samples using an equal amount (1 µg) from each replicate, using Supplementary Table 4. Keep DNA and RNA samples separated until these are pooled at the later step (Step 160).

**PAUSE POINT** DNA can be stored at −20 °C for months.

148 Add equal volume of 6x gel loading dye (final conc. 3×) and mix the solution by vortexing.

149 Run the pooled sample on a 1.8% (wt/vol) agarose gel (30 mL of 5 cm ×6 cm mini gels with 1.3 cm-width well

150 Cut the DNA bands (162 bp) using a blue-light Safe Imager.

151 Purify the DNA from the gel slices using QIAquick Gel Extraction Kit according to the manufacturer's protocol.
Use one column for each of DNA or RNA samples.

152 Elute the DNA in 50 µL Buffer EB per column.

153 Purify the DNA using 1.8 volume (90 µL) of HighPrep PCR reagent

153.1 Bring the HighPrep PCR reagent to RT for at least 30 min before use. Shake thoroughly to fully resuspend the magnetic beads. Add 1 volume (200 µL) of HighPrep PCR reagent and mix thoroughly by pipetting up and down 6–8 times

153.2 Incubate the mixture for 5 min at RT

153.3 Place the tube on the magnet for 2–3 min until the solution clears and beads pull to the side of the wall

153.4 Carefully remove the supernatant by pipetting without disturbing the beads. A small amount (10–20 µL) of supernatant can be left in the tube.

153.5 With the tube on the magnet, add 500 µL of 80% (vol/vol) ethanol and incubate for 30 s.

153.6 Remove the ethanol by pipetting while the tube is still on the magnet.

153.7 Repeat the 80% (vol/vol) ethanol washing

153.8 Flash-spin (200–1,000g, 22–25 °C, 3 s) the tube, immediately place it back on the magnet and remove the supernatant

153.9 Dry the bead pellet for 2–3 min. Do not overdry the beads

154 Add 20 µL of Buffer EB to the beads and elute the DNA by pipetting and vortexing.
.

155 Place the tube on the magnet for 1-2 min and transfer 18 µL of the eluate to a LoBind tube.

156 Measure the DNA concentration using a Qubit dsDNA HS Assay Kit according to manufacturer's protocol.

To check the DNA size and quality, run 20-30 ng of the DNA on a 1.8% (wt/vol) gel along with

157 100-bp DNA ladder.

158

Pool the DNA and RNA samples in a single LoBind tube with 1:3 ratio to obtain 100 µL mixture at the final concentration of 10 nM (1 ng/µL), using Supplementary Table 4.

**PAUSE POINT** The pooled DNA can be stored at −20 °C for months.

159

Send the sequencing library and custom primers for sequencing

The sequencing should be done with paired-end 15 bp (barcode length), 16 cycles for index read 1 and 10 cycles for index read 2. Index read 1 provides the UMI sequence, and index read 2 provides the sample index.
let the sequencing facility know that the index read 2 should be used for demultiplexing and that short reads should not be masked, bcl2fastq parameters:

*--minimum-trimmed-read- length 0 --mask-short-adapter-reads 0*

An average of 10×(DNA) and 30×(RNA) coverage of the library (based on number of barcodes) via sequencing reads is required. For example, we use an Illumina NextSeq high-output run (400M reads over 3 replicates) for the 0.5M barcode library or three runs (1.2B reads over three replicates) for the 8–12 million barcode library (Supplementary Table 1). Sequencing cycles, primers, and expected output for each read is shown below

| READ | CYCLES | PRIMER | OUTPUT |
|---|---|---|---|
| Read 1 | 15 | pLSmP-ass-seq-ind1 | Barcode (forward) |
| Read 2 | 15 | pLSmP-bc-seq | Barcode (reverse) |
| Index read 1 | 16 | pLSmP-UMI-seq | UMI |
| Index read 2 | 10 | pLSmP-5bc-seq-R2 | Sample index |

Step 159 includes a Step case.
**Schematics of lentiMPRA**