

Aug 20, 2025

MPRA Protocol

DOI

<https://dx.doi.org/10.17504/protocols.io.q26g7nye3lwz/v1>

DBisht@mdanderson.org¹, Nicholas Pease², kmdeepabisht²

¹MD Anderson Cancer Center; ²University of Pittsburgh



Nicholas Pease

University of Pittsburgh

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.q26g7nye3lwz/v1>

Protocol Citation: DBisht@mdanderson.org , Nicholas Pease, kmdeepabisht 2025. MPRA Protocol. **protocols.io**
<https://dx.doi.org/10.17504/protocols.io.q26g7nye3lwz/v1>

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: August 12, 2025

Last Modified: August 20, 2025

Protocol Integer ID: 224555

Keywords: mpra protocol protocol for mpra vector assembly, mpra vector assembly, mpra protocol protocol, rna extraction, seq library construction, nature genetics

Abstract

Protocol for MPRA vector assembly, transfection, RNA extraction and Tag-seq library construction (adapted from Tewhey et al., *Cell* 2016 and Mouri et al., *Nature Genetics* 2022)

Guidelines

Tag-seq Library Construction

To minimize amplification bias during the creation of cDNA tag sequencing libraries, samples were amplified by qRT-PCR to estimate relative concentrations of GFP cDNA. The qRT-PCR was performed using 1 μ L of cDNA sample in a 10 μ L PCR reaction containing 5 μ L Q5 NEBNext Ultra II master mix, 1.7 μ L Sybr green I diluted 1:10,000 (Life Technologies, S-7567) and 0.5 μ M of TruSeq_Universal_Adapter and MPRA_Illumina_GFP_F primers (Table X). Samples were amplified with the following conditions: 98°C for 20 seconds, 40 cycles (98°C for 10 sec, 62°C for 15 sec, 72°C for 30 sec), 72°C for 2 min followed by melt curve analysis. For the plasmid library control, serial dilution of the plasmid library was prepared from 1000 pg to 1 fg using 10-fold dilutions. A standard curve was plotted with the plasmid library dilutions and based on the threshold cycles for the cDNA samples, the plasmid dilution with the same cycle threshold value was used for the adaptor ligation. To add Illumina sequencing adapters, cDNA samples and 5 plasmid library controls were diluted to normalized it the cDNA replicate with the lowest concentration or highest threshold cycle and 10 μ L of normalized sample was amplified using the reaction conditions from the qRT-PCR scaled to 50 μ L, without adding Sybr green to the reaction and by using only n-1 (where n= the cycle number obtained from qRT-PCR) amplification cycles. Amplified cDNA was 2x SPRI purified and eluted in 30 μ L of EB.

A second PCR was performed to add the individual index primers to each sample. The PCR was performed with 20 μ L of purified PCR 1 elute in a 50 μ L Q5 NEBNext Ultra II reaction with 0.5 μ M of TruSeq_Universal_Adapter primer and Illumina_Multiplex primer containing a unique 8 bp index for sample demultiplexing post-sequencing. Samples were amplified at 98°C for 20 seconds, 6 cycles (98°C for 10 sec, 62°C for 15 sec, 72°C for 30 sec), 72°C for 2 minutes. Indexed libraries were 2x SPRI purified and quantified on Qubit Flex (Thermo Fisher Scientific, Q33326) using dsDNA HS Assay Kit (Thermo Fisher Scientific, Q33230) and pooled according to molar estimates from TapeStation (Agilent Technologies, 4200 TapeStation) quantifications. Samples were sequenced using 1×20bp reads on the Illumina NextSeq 2000 platform targeting 80M reads per cDNA library replicate and 60M reads per plasmid library replicate.

Materials

Materials and Reagents Mentioned:

- Sfil digested pGL4:23:AlucΔxbal vector
- Barcoded oligos
- Gibson assembly reagents (NEB, E2611)
- SPRI purification reagents
- Electroporation cuvettes and device (2kV, 200 ohm, 25 µF)
- Electrocompetent 10-beta *E. coli* (NEB, C3020K)
- SOC media
- LB media
- Carbenicillin (Sigma-Aldrich, C1389)
- Qiagen plasmid isolation kit (Qiagen, 12963)
- Illumina sequencing reagents/platform (NovaSeq)
- AsiSI (NEB, R0630)
- Q5 NEBNext Hot-Start (NEB M0494)
- Primers (primer 200 and primer 201)
- DpnI
- RecBCD (NEB, M0345)
- BSA
- ATP
- NEB Buffer 4
- Qiagen Gigaprep kit (Qiagen, 12991)
- RPMI medium (Corning, 10-040-CV)
- FBS (Gibco A52568-01)
- GlutaMAX (Gibco 35050-061)
- Pen-Strep (Corning, 30-002-CI)
- Neon transfection system (Thermo Fisher Scientific, MPK5000)
- Neon transfection kit (MPK10096B)
- Qiagen Maxi RNeasy (Qiagen, 75162)
- DTT (G-Biosciences, 786227)
- Tissue homogenizer, TH (OMNI international)

Cell Lines:

- GM12878s (Coriell Institute)
- K562s (CIMR)

Additional Materials and Reagents (from these pages):

- RNase free DNase (Qiagen 79254)
- SUPERase-In (Thermo Fisher Scientific, AM2696)
- Turbo DNase (Thermo Fisher Scientific, AM2238)
- 10% SDS (Life Technologies, 15553-035)
- 0.5M EDTA (Thermo Fisher Scientific, AM9260G)



- 20X SSC (Life Technologies, 15557-044)
- Formamide (Sigma Aldrich, 75-12-7)
- Biotin-labeled GFP probes (*GFPBiotinCapture1-3*)
- Streptavidin beads (Thermo Fisher Scientific, 65002)
- NaOH (Sigma-Aldrich, S2770)
- NaCl (Thermo Fisher Scientific, AM9760G)
- HulaMixer (Thermo Fisher Scientific, V.3A01)
- DynaMag magnet (Thermo Fisher Scientific, 12321D)
- DEPC treated water (Ambion, AM9906)
- RNAClean XP SPRI beads (Beckman, A63987)
- SuperScript III
- AMPure SPRI (Beckman, A63881)
- Q5 NEBNext Ultra II master mix
- Sybr green I (Life Technologies, S-7567)
- TruSeq_Universal_Adapter and MPRA_Illumina_GFP_F primers
- Illumina_Multiplex primer
- Qubit Flex (Thermo Fisher Scientific, Q33326)
- dsDNA HS Assay Kit (Thermo Fisher Scientific, Q33230)
- TapeStation (Agilent Technologies, 4200 TapeStation)

Newly Mentioned/Clarified on These Pages:

- Illumina NextSeq 2000 platform

MPRA vector assembly

- 1 To create the mpra Δ orf library, barcoded oligos were inserted into SfiI digested pGL4:23: Δ luc Δ xbal by Gibson assembly (NEB, E2611) using 1.1 μ g of oligos and 1 μ g of digested vector in a 50 μ L reaction incubated for 60 min at 50°C. The reaction is then purified by 1.2x SPRI and eluted in 20 μ L of elution buffer.
- 2 A test transformation was performed to determine the library coverage using 1 μ L of ligated vector into 50 μ L of electrocompetent 10-beta E. coli (NEB, C3020K) by electroporation (2kV, 200 ohm, 25 μ F).
- 3 For the scaled-up transformation, 1 μ L of the library was transformed into 50 μ L of electrocompetent cells. The electroporated cells were recovered in 950 μ L SOC media and then immediately split into ten 1 mL aliquots of SOC. These cultures were recovered for an hour at 37°C and then each culture individually expanded in 20 mL LB supplemented with 100 μ g/mL of carbenicillin (Sigma-Aldrich, C1389) in shaker at 37°C for 6.5 hours. After completing the incubation, all the cultures were pooled and proceeded for plasmid isolation (Qiagen, 12963).
- 4 To validate library complexity and connect barcodes to CE sequences, Illumina libraries were prepared for the mpra: Δ orf library as described in Mouri et al. Nature Genetics 2022 and sequenced using 2 $\times$ 150 PE reads on the Illumina NovaSeq platform with 15% PhiX spike-in.
- 5 To create the final mpra:gfp library, 10 μ g of mpra: Δ orf library plasmid was linearized with 100 units of AsiSI (NEB, R0630) in a final 400 μ L reaction and incubated overnight at 37°C. The linearized product was then column purified (Qiagen, 28104) and eluted in 30 μ L.
- 6 The GFP amplicon was amplified from pMPRAv3:minP-GFP using Q5 NEBNext Hot-Start (NEB M0494), 0.5 μ M primer 200 and 0.5 μ M primer 201 (table...X), following cycle conditions, 98°C for 30 s, 20 cycles (98°C for 10 s, 60°C for 15 s, 72°C for 45 s), 72°C for 5 min. The amplified product was incubated with DpnI for 30 min at 37°C, followed by 0.5x reverse SPRI and 1.5x forward SPRI purification and eluted in 40 μ L EB.
- 7 A second PCR was then performed using the 1:100 diluted purified PCR1 GFP amplicon as PCR1 and then column purified. This amplicon containing a minimal promoter, GFP open reading frame and a partial 3' UTR was then inserted by Gibson assembly using 1.6 μ g of AsiSI linearized mpra Δ orf plasmid and 5.28 μ g of the GFP amplicon in a 400 μ L reaction for 90 minutes at 50°C followed by a 1.5x SPRI purification.
- 8 The total recovered volume was re-digested to remove remaining uncut vectors by incubation with 50 Units of AsiSI, 5 Units of RecBCD (NEB, M0345), 10 μ g BSA, 1 mM ATP, 1x NEB Buffer 4 in a 100 μ L reaction incubated overnight at 37°C followed by 1.5x SPRI purification and elution with 40 μ L of EB.

- 9 To generate final transfection ready MPRA library, 8 μL of mpra:gfp plasmid was electroporated (2kV, 200 ohm, 25 μF) into 200 μL of 10-beta cells. Electroporated bacteria was recovered in 12 mL SOC and split across 6 2 mL aliquots and incubated for 1 hour at 37°C then each 2 mL culture was expended to 500 mL of LB with 100 $\mu\text{g}/\text{mL}$ of carbenicillin and incubated for 16 hours at 37°C and followed by plasmid isolation using Qiagen Gigaprep kit (Qiagen, 12991).

Transfection

- 10 GM12878s (Coriell Institute for medical research) were cultured in RPMI medium (Corning, 10-040-CV) containing 15% FBS (Gibco A52568-01), 1% GlutaMAX (Gibco 35050-061) and 1% Pen-Strep (Corning, 30-002-CI). For each transfection, 1×10^7 cell were mixed with 10 μg of MPRA library in 100 μL RPMI. These cells were transfected with Neon transfection system (Thermo Fisher Scientific, MPK5000) and kit (MPK10096B) using 3 pulses of 1200 V for 20 ms. Total five replicates, grown on different days and maintained to ~ 1 million cells/ml, were transfected. For each replicate, a total of 1.5×10^8 million cells were pelleted at 200g and resuspended in 1.5 ml of RPMI medium containing 150 μg of the TF Motif library. After transfection, each replicate was recovered at a density of 5×10^5 cells/ml, in 300 ml of RPMI medium containing 15% FBS, 1% GlutaMAX and 1% Pen-Strep. After 24 hours the cells were pelleted at 200g and washed in PBS once. The cells were resuspended in 15ml RLT buffer provided with Qiagen Maxi RNeasy (Qiagen, 75162) and 300mM DTT (G-Biosciences, 786227). The cells were then homogenized with tissue homogenizer, TH (OMNI international) for 1 min at the maximum speed and stored at -80°C .
- 11 K562s (CIMR) were cultured in RPMI medium containing 10% FBS and 1% Pen-Strep. For each transfection, 1×10^7 cell were mixed with 5 μg of MPRA library in 100 μL RPMI were transfected with Neo Neon transfection system and kit using three pulses of 1450 V for 10 ms. Five total replicates, grown on different days and maintained at a density of ~ 1 million cells/ml, were transfected. For each replicate, a total of 1.5×10^8 million cells were pelleted at 200g and resuspended in 1.5 ml of buffer R provided in Neon transfection kit containing 150 μg of the TF Motif library. After transfection, each replicate was recovered with a density of 5×10^5 cells/ml, in 300 ml of RPMI medium containing 15% FBS and 1% Pen-Strep. After 24 hours the cells were pelleted at 200g and washed in PBS once. The cells were resuspended in 15ml RLT buffer and 300mM DTT. The cells were then homogenized for 1 min at the maximum speed and stored at -80°C .

RNA Extraction and cDNA Synthesis

- 12 Total RNA was isolated from cells using Qiagen Maxi RNeasy (Qiagen, 75162) following the manufacturer's protocol including the on-column DNase digestion by RNase free DNase (Qiagen 79254). The final purified RNA was treated with 5 μL SUPERase-In

(Thermo Fisher Scientific, AM2696). Another DNase treatment was performed on total RNA using 5 μ L of Turbo DNase (Thermo Fisher Scientific, AM2238) in 750 μ L of total volume for 1 hour at 37°C. The reaction was terminated by the addition of 7.5 μ L 10% SDS (Life Technologies, 15553-035) and 75 μ L of 0.5M EDTA (Thermo Fisher Scientific, AM9260G) followed by incubation at 70°C for 5 minute.

- 13 All the DNase treated RNA was then proceeded for GFP mRNA pulldown. To this reaction 900 μ L of 20X SSC (Life Technologies, 15557-044), 1800 μ L of Formamide (Sigma aldrich, 75-12-7) and 2 μ L of 100 uM biotin-labeled GFP probes (GFP*BiotinCapture*1-3, table-x) mixture were added and the volume was made up to 3600 μ L. The reaction was then incubated at 65°C for 2.5 hours with intermittently inverting the tubes in every 30 min. The biotin probes were captured using 400 μ L of pre-washed Streptavidin beads (Thermo Fisher Scientific, 65002). The Streptavidin beads were washed twice with Buffer-A containing 0.1 M NaOH (Sigma-Aldrich, S2770) and 0.05 M NaCl (Thermo Fisher Scientific, AM9760G), and once with Buffer-B containing 0.1 M NaCl. The beads then finally eluted in 500 μ L of 20X SSC and added to the RNA-probe sample. The hybridized RNA-probe-beads mixture was mixed on HulaMixer (Thermo Fisher Scientific, V.3A01) at room temperature for 15 minutes. Beads were captured by DynaMag magnet (Thermo Fisher Scientific, 12321D) and washed once with 1x SSC and twice with 0.1x SSC.
- 14 Elution of RNA was performed by the addition of 25 μ L DEPC treated water (Ambion, AM9906) and heating of the mixture for 2 minutes at 80°C followed by immediate collection of eluent on a magnet. A second elution was performed by incubating the beads with an additional 25 μ L of water at 80°C. The eluted RNA was then processed for the third or the final DNase treatment with 1 μ L of Turbo DNase in a total of 56 μ L reaction. The reaction was incubated for 60 minutes at 37°C followed by inactivation with 1 μ L of 10% SDS. The final DNase treated GFP mRNA was eluted using RNAClean XP SPRI beads (Beckman, A63987) in 35 μ L DEPC treated water.
- 15 First-strand cDNA was synthesized using 30 μ L of the DNase-treated GFP mRNA with SuperScript III and a primer 19 (table x) using the manufacturer's recommended protocol. Single-stranded cDNA was purified using AMPure SPRI (Beckman, A63881) beads and eluted in 30 μ L EB.

Tag-seq Library Construction

- 16 To minimize amplification bias during the creation of cDNA tag sequencing libraries, samples were amplified by qRT-PCR to estimate relative concentrations of GFP cDNA. The qRT-PCR was performed using 1 μ L of cDNA sample in a 10 μ L PCR reaction containing 5 μ L Q5 NEBNext Ultra II master mix, 1.7 μ L Sybr green I diluted 1:10,000 (Life Technologies, S-7567) and 0.5 μ M of TruSeq_Universal_Adapter and MPRA_Illumina_GFP_F primers (Table X). Samples were amplified with the following conditions: 98°C for 20 seconds, 40 cycles (98°C for 10 sec, 62°C for 15 sec, 72°C for 30 sec), 72°C for 2 min followed by melt curve analysis. For the plasmid library control,

serial dilution of the plasmid library was prepared from 1000 pg to 1 fg using 10-fold dilutions. A standard curve was plotted with the plasmid library dilutions and based on the threshold cycles for the cDNA samples, the plasmid dilution with the same cycle threshold value was used for the adaptor ligation.

- 17 To add Illumina sequencing adapters, cDNA samples and 5 plasmid library controls were diluted to normalize to the cDNA replicate with the lowest concentration or highest threshold cycle and 10 μ L of normalized sample was amplified using the reaction conditions from the qRT-PCR scaled to 50 μ L, without adding Sybr green to the reaction and by using only $n-1$ (where n = the cycle number obtained from qRT-PCR) amplification cycles. Amplified cDNA was 2x SPRI purified and eluted in 30 μ L of EB.
- 18 A second PCR was performed to add the individual index primers to each sample. The PCR was performed with 20 μ L of purified PCR 1 elute in a 50 μ L Q5 NEBNext Ultra II reaction with 0.5 μ M of TruSeq_Universal_Adapter primer and Illumina_Multiplex primer containing a unique 8 bp index for sample demultiplexing post-sequencing. Samples were amplified at 98°C for 20 seconds, 6 cycles (98°C for 10 sec, 62°C for 15 sec, 72°C for 30 sec), 72°C for 2 minutes. Indexed libraries were 2x SPRI purified and quantified on Qubit Flex (Thermo Fisher Scientific, Q33326) using dsDNA HS Assay Kit (Thermo Fisher Scientific, Q33230) and pooled according to molar estimates from TapeStation (Agilent Technologies, 4200 TapeStation) quantifications.
- 19 Samples were sequenced using 1×20bp reads on the Illumina NextSeq 2000 platform targeting 80M reads per cDNA library replicate and 60M reads per plasmid library replicate.