Title

Protocol for mapping double-stranded DNA break sites across the genome with translocation capture sequencing

Journal

STAR Protocols, 4(2)

ISSN

2666-1667

Authors

Delaney, Joe R

La Spada, Albert R

Publication Date

2023-06-01

DOI

10.1016/j.xpro.2023.102205

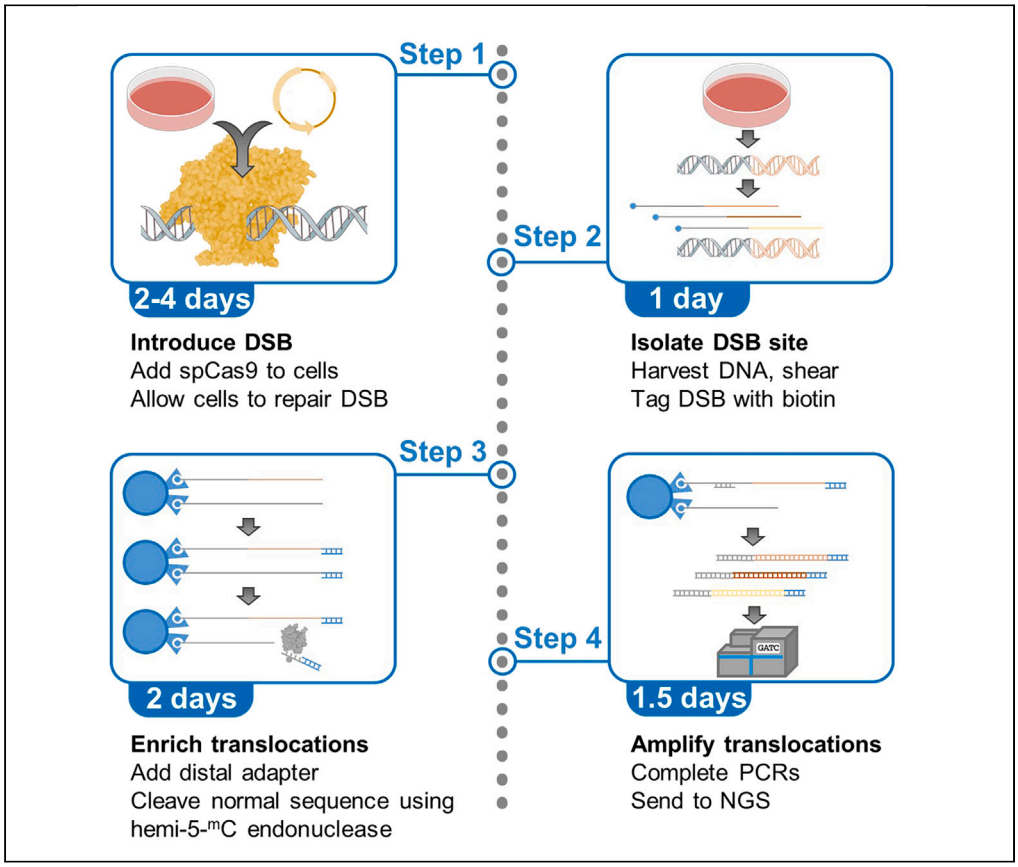
Peer reviewed

STAR Protocols

CellPress
OPEN ACCESS

Protocol

Protocol for mapping double-stranded DNA break sites across the genome with translocation capture sequencing



Translocation sequencing can be used to assess mechanisms of DNA repair and identify genome-wide double-strand breaks (DSBs) accessible to DNA repair machinery. Here, we present a protocol for mapping double-strand DNA break sites across the genome with translocation capture sequencing. Bait DSBs are introduced using a Cas9 nuclease and repaired by the host cell, connecting bait DSBs to other DSBs. Repair sites are detected by isolating bait site DNA, cleaving normal sequence to enrich off-site repair, and next-generation sequencing.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

Joe R. Delaney,
Albert R. La Spada

delaneyj@musc.edu
(J.R.D.)
alaspada@uci.edu
(A.R.L.S.)

Highlights

Protocol is used to map double-strand DNA break sites across the genome

Bait DNA break can be any nuclease-targeted site

Translocated DNA is enriched by cleaving locally repaired sequence

Flexibility of cell and *in vivo* systems is high

Delaney & La Spada, STAR Protocols 4, 102205
June 16, 2023
https://doi.org/10.1016/j.xpro.2023.102205

STAR Protocols

CellPress
OPEN ACCESS

Protocol

Protocol for mapping double-stranded DNA break sites across the genome with translocation capture sequencing

Joe R. Delaney[1,5,6,7,*] and Albert R. La Spada[2,3,4,5,*]

[1]Department of Biochemistry and Molecular Biology, Medical University of South Carolina, Charleston, SC 29425, USA

[2]Departments of Pathology & Laboratory Medicine, Neurology, and Biological Chemistry, School of Medicine, University of California, Irvine, Irvine, CA 92697, USA

[3]Department of Neurobiology & Behavior, School of Biosciences, University of California, Irvine, Irvine, CA 92697, USA

[4]UCI Center for Neurotherapeutics, University of California, Irvine, Irvine, CA 92697, USA

[5]These authors contributed equally

[6]Technical contact

[7]Lead contact

*Correspondence: delaneyj@musc.edu (J.R.D.), alaspada@uci.edu (A.R.L.S.)
https://doi.org/10.1016/j.xpro.2023.102205

SUMMARY

Translocation sequencing can be used to assess mechanisms of DNA repair and identify genome-wide double-strand breaks (DSBs) accessible to DNA repair machinery. Here, we present a protocol for mapping double-strand DNA break sites across the genome with translocation capture sequencing. Bait DSBs are introduced using a Cas9 nuclease and repaired by the host cell, connecting bait DSBs to other DSBs. Repair sites are detected by isolating bait site DNA, cleaving normal sequence to enrich off-site repair, and next-generation sequencing.
For complete details on the use and execution of this protocol, please refer to Switonski et al. (2021).[1]

BEFORE YOU BEGIN

Primer and sgRNA design

Timing: 4 h

Critical to the success of final base-pair resolution translocation mapping is the ability to map the translocation from the same read as the bait sequence. To do so, the single-guide RNA (sgRNA) must be designed in a location adjacent to where a hemi-methyl-sensitive restriction enzyme has a recognition site. With LpnPI (recognition site: C^{m}CDG), this compatibility is common. The overall schematic, including directionality and Watson or Crick orientation, is diagrammed in Figure 1. Additional quality-control primers are recommended prior to executing Genetic Engineering-Free TransLocation Capture Sequencing (GEF-TLC-seq) on experimental samples.

1. Design Nested Primer ~30–60 bp 5′ from Cas9 cleavage site.

   Note: Unique-molecular identifiers (UMIs) and/or in-line sample barcodes may be optionally included between the I5-sequencing portion of the Nested Primer and the genomic DNA sequence.

2. Design Bait Biotinylated Oligonucleotide (oligo) ~50–100 bp 5′ from 5′ end of Nested Primer.

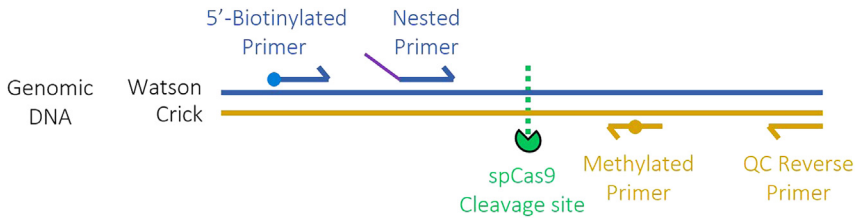


Figure 1. GEF-TLC-seq primer design schematic

3. Design Methylated Oligo for LpnPI cleavage ~20–60 bp 3′ of Cas9 cleavage site, with methyl group flanked by ≥ 12 bp on each side. If your bait and nested primer are designed from the Watson strand, this primer should be designed from the Crick strand (reverse complement orientation).

   Note: Design quality control primer further 3′ of the methylated oligo, to test PCR conditions.

4. The complete, finished NGS library structure is diagrammed below in Figures 2 and 3 for reference.

   Note: The LpnPI site will not often appear in the library if LpnPI cleavage is efficient. An additional schematic is also found in Figure S5 of the original publication.

Prepare stock solutions

Timing: 1 h

5. 5 M NaCl (73.05 g NaCl to 250 mL in nuclease free water).
6. 1 M Tris-HCl pH 7.4 (60.57 g Tris base to 500 mL in nuclease free water, pH to 7.4 with HCl).
7. 10% SDS w/v (10 g SDS to 100 mL in nuclease free water).
8. 0.5 M EDTA pH 8.0 (18.61 g–100 mL in nuclease free water, pH with NaOH).
9. Hexamine cobalt (III) chloride, 20 mM (0.53 g–100 mL in nuclease free water).
10. PEG8000, 50% w/v (5 g–10 mL in nuclease free water).

All stock solutions may be stored at 22°C or 4°C.

Prepare buffers and working solutions

Timing: 1 h

11. Prepare annealing buffer, 2× bead buffer, and 1× bead buffer (see materials and equipment).
12. Filter sterilize each buffer. Use a 30 mL Leur-lock syringe to aspirate each buffer, attach a 0.45 μm syringe filter, and eject sterilized buffer into fresh 50 mL conical tube. Do not re-use filters or syringes between buffers.

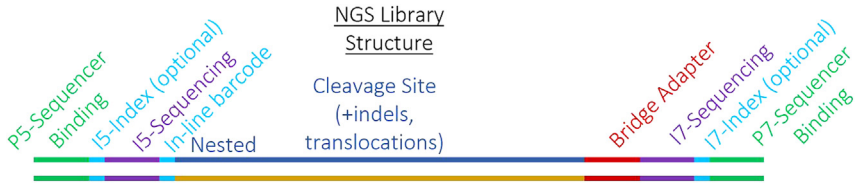


Figure 2. GEF-TLC-seq sequencing library schematic

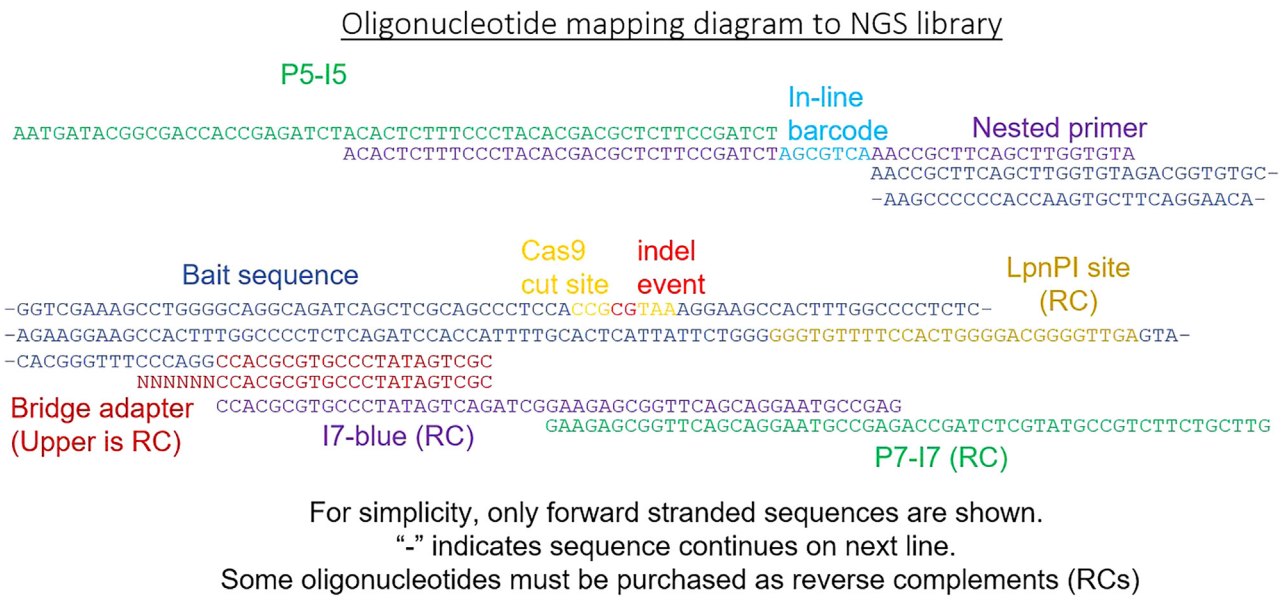


Figure 3. GEF-TLC-seq sequencing library nucleotide details

13. All buffers may be stored at 22°C or 4°C.

Quality control: DNA shearing

Timing: 4 h

14. Isolate DNA from cells to be used in the experiment.
15. Test DNA sonicator or enzymatic shearing strategy until a 250 bp–750 bp smear is achieved, peaking at 500 bp.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Cutsmart buffer (10×)	NEB	Cat#B6004S
DMEM, high glucose	ThermoFisher	Cat#11965092
EDTA	Millipore Sigma	Cat#E5134
EndoFectin Max transfection reagent	GeneCopoeia	Cat#EF014
Fetal bovine serum (FBS), characterized	HyClone	Cat#SH30071.03HI
Hexamine cobalt III chloride	Millipore Sigma	Cat#481521
Individually-capped 0.2 mL PCR strips	ThermoFisher	Cat#A30588
LpnPI enzyme	NEB	Cat#R0663S
Nuclease-free water	Fisher Scientific	Cat#10-977-023
OptiMEM	Fisher Scientific	Cat#31985070
Phosphate buffered saline (PBS) pH 7.4	VWR	Cat#45000-446
Phusion DNA polymerase	ThermoFisher	Cat#F530L
Polyethylene glycol 8000 (PEG-8000)	Promega	Cat#V3011
Proteinase K	NEB	Cat#P8107S

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Ribonuclease A (RNaseA) from bovine pancreas	Merck Millipore	Cat#R6513
Streptavidin Dynabeads, C1	ThermoFisher	Cat#11205D
Syringe filters, 0.45 µm	Millipore Sigma	Cat#SLHPR33RS
Syringes, leur lock tip, 30 mL	VWR	Cat#BD302832
T4 DNA ligase kit	NEB	Cat#M0202S
Tris base (Trisma®)	Millipore Sigma	Cat#T1503
Trypsin-EDTA	Millipore Sigma	Cat#T3924-100ML
Critical commercial assays		
Blood and Tissue DNA Kit	QIAGEN	Cat#69506
Invitrogen PureLink PCR cleanup kit	Fisher Scientific	Cat#K182001
Deposited data		
Raw and processed NGS data	This paper	GEO: GSE166119
Experimental models: Cell lines		
Mouse Neuro2a cell line	ATCC	Cat#CCL-131; RRID: CVCL_0470
Oligonucleotides		
5-Cytosine methylated primer site for example target site 5′TCAACCCCGTCCC/iMe-dC/AGT GGAAAACACCC 3′	IDT	N/A
Biotinylated primer 5′ /5Biosg/GGTTTCCCCTCC GCTCTTTA 3′	IDT	N/A
Bridge Adapter (upper) 5′ GCGACTATAGG GCACGCGTGGNNNNNN/3AmMO/ 3′	IDT	N/A
Bridge Adapter (lower) 5′ /5Phos/CCACGCGT GCCCTATAGTCGC 3′	IDT	N/A
I7-Blue 5′ CTCGGCATTCCTGCTGAACCGCTCTT CCGATCTGACTATAGGGCACGCGTGG 3′	IDT	N/A
P5-I5 5′ AATGATACGGCGACCACCGAGATCTA CACTCTTTCCCTACACGACGCTCTTCCGATCT 3′	IDT	N/A
P7-I7 5′ CAAGCAGAAGACGGCATACGA GATC GGTCTCGGCATTCCTGCTGAACCGCTCTTC 3′	IDT	N/A
I5-Nested 5′ ACACTCTTTCCCTACACGACGCTCTT CCGATCTAGCGTCAAACCGCTTCAGCTTGGTGTA 3′	IDT	N/A
Recombinant DNA		
Plasmid expressing sgRNA to example target site on Chr14 (5′ CTCGCAGCCCTCCACCGTAA 3′)	This paper and Switonski et al. [1]	N/A
Plasmid expressing non-targeting sgRNA [genome-dependent! example here is for mice or human cells] (e.g., 5′ GTATTACTGATATTGGTGGG 3′)	This paper and Switonski et al. [1]	N/A
Plasmid to clone sgRNAs into for new targets	Addgene	Cat#48138
Software and algorithms		
Galaxy	Afgan et al., 2018[2]	https://usegalaxy.org/ RRID: SCR_006281
Other		
Bioruptor Pico	Diagenode	Cat#B01060010
Cell counter	BIORAD	Cat#TC20
Dry heat blocks	Southern Labware	Cat#BSH1001
High-speed microcentrifuge	Beckman Coulter	Cat#Microfuge 16
Illumina sequencer	Illumina	Cat#MiSeq
Low-speed PCR strip centrifuge	Southern Labware	Cat#C1012
Magnetic bead capture stand	Fisher Scientific	Cat#12-321-D
Nanodrop	ThermoFisher	Cat#ND2000
Qubit Fluorometer	ThermoFisher	Cat#Q33238
TapeStation	Agilent	Cat#4150
Thermal cycler for PCR	BIORAD	Cat#C1000

MATERIALS AND EQUIPMENT

Annealing Buffer		
Reagent	Final concentration	Amount
5 M NaCl	25 mM	250 μL
1 M Tris-HCl (pH7.4)	10 mM	500 μL
500 mM EDTA (pH8.0)	0.5 mM	50 μL
Nuclease free water	N/A	49.2 mL
Total	**1× Annealing Buffer**	**50 mL**

2× Bead Buffer		
Reagent	Final concentration	Amount
5 M NaCl	2 M	20 mL
1 M Tris-HCl (pH7.4)	10 mM	500 μL
500 mM EDTA (pH8.0)	1 mM	100 μL
Nuclease free water	N/A	29.4 mL
Total	**2× Bead Buffer**	**50 mL**

1× Bead Buffer		
Reagent	Final concentration	Amount
5 M NaCl	1 M	10 mL
1 M Tris-HCl (pH7.4)	5 mM	250 μL
500 mM EDTA (pH8.0)	0.5 mM	50 μL
Nuclease free water	N/A	39.7 mL
Total	**1× Bead Buffer**	**50 mL**

STEP-BY-STEP METHOD DETAILS

Introduce DSB via CRISPR-spCas9 and isolate genomic DNA

Timing: 2–4 days

Introduce plasmids containing spCas9 and the sgRNA targeting the desired bait site to induce DSBs. Allow the cells to repair the DSB for 24–72 h. This allows for one slow cell division while limiting subsequent apoptosis or necrosis from cytotoxic translocations.

1. Per sample, plate two replicates of 5 million cells into two 10cm tissue-culture dishes, each containing 3 mL media without antibiotics.

   Note: In this example, mouse Neuro2a cells are used, cultured in DMEM with 10% fetal bovine serum at 37°C with 5% CO_2.

   a. Allow to adhere, beginning transfection 6–24 h post-plating.

   Note: This step is flexible and should be adjusted to culture conditions pertinent to the experimental situation specific to user cell lines and DNA repair environment.

2. Per sample, transfect CRISPR-spCas9 expressing plasmid containing one sgRNA which should target either:
   a. Non-targeting control (the example here does not target the mouse or human genome, `GTATTACTGATATTGGTGGG`. Other genomes may require a different guide).

b. Targeted bait site (the example here targets the mouse genome, CTCGCAGCCCTCCACCG TAA).
   i. In two separate 1.5 mL microcentrifuge tubes per sample, aliquot 1 mL optiMEM.
   ii. In one tube, add 30 μL EndoFectin Max transfection reagent.
   iii. In the other tube, add 3 μg plasmid.
   iv. Incubate at 22°C, 5 min.
   v. Combine diluted transfection agent with plasmid solution. Mix gently by pipetting up and down slowly.
   vi. Incubate at 22°C, 10 min.
   vii. Add the 2 mL of optiMEM+plasmid+transfection reagent drip-wise onto each plate of cells.

3. Once cells are adhered to the plate, transfect for 6 h–24 h in a tissue culture incubator (6 h recommended for most cells).
4. Aspirate transfection media and replace with normal media (10 mL per sample). Incubate in tissue culture incubator for 24–72 h (just over 1 cell division, 24 h for most cells [including transfection time: 30 h]).
5. Aspirate media. Rinse once with PBS (gently so as to not disturb cells). Aspirate PBS.
6. Add 2 mL Trypsin-EDTA to cells. Incubate in tissue culture incubator until cells are no longer adherent (5 min for most cells).
7. Combine Trypsin-EDTA with 1 mL normal media and split resuspended cells into two 1.5 mL microcentrifuge tubes (1.5 mL each). Spin cells 750 g 2 min.
8. Aspirate media from cell pellet. Wash cells once with 1 mL PBS per tube, repeat spin.
9. Add 200 μL PBS and resuspend cells using P200 pipette. Add 20 μL Proteinase K (NEB, or Qiagen kit, 600 mAU/mL). Incubate 56°C 30 min to lyse cells. Heat inactivate 95°C 5 min.
10. Add 5 μL RNaseA (8 mg/mL) and incubate 37°C for 30 min.
11. Purify DNA using Qiagen Blood and Tissue Kit. Elution step should be: 100 μL elution buffer, twice from same column with two sequential spins. Pool eluted DNA from each pair of replicate samples, here.
12. Use NanoDrop to quantify DNA. A260/280 should be > 1.8 if kit isolated DNA properly.
13. Save purified DNA in −20°C until ready to continue protocol.

Pause point: DNA is stable frozen.

Shear DNA and linearly-amplify Cas9-cut bait site

Timing: 1 day

These steps utilize a primer upstream of the spCas9 cut site to isolate and amplify the bait region ≥70- fold. The linear amplification utilizes a biotinylated primer, which will be used to bind to beads.

14. Use Biorupter Pico or other DNA shearing instrument to sonicate DNA samples, shearing to ~500 bp according to manufacturer protocol [alternatively, use enzymatic shearing methods].

Note: This shearing length is assuming 2 × 250–300 bp long NGS reads. If using 2× 150 bp NGS, shearing to ~250–300 bp average may enable some R1-R2 read overlap, but this change from 500 bp is not necessary.

15. Run 1 μL sheared DNA on a 1% agarose gel.
   a. Validate that the sheared DNA recapitulates quality control step, for each sample.
   b. Redo shearing with any samples which fail this quality control step, altering the shearing conditions appropriately.

Note: Due to RNase, depending on quality of DNA purification, the most intense band may appear very small - this is degraded RNA. Degraded RNA will not affect the assay.

16. Set up 8 × 50 μL bait-amplification reactions for each sample:

Reagent	Rxn Vol	Master mix
5× Phusion buffer	10 μL	(x9 = 90 μL)
dNTPs (2.5 mM each)	1.5 μL	(x9 = 13.5 μL)
Biotinylated-primer (1 μM)	0.5 μL	(x9 = 4.5 μL)
Phusion polymerase (2 U/μL)	0.5 μL	(x9 = 4.5 μL)
Sheared DNA	25 μL	(x9 = 225 μL)
Nuclease-free water	12.5 μL	(x9 = 112.5 μL)

Note: Error-proofing polymerase (such as Phusion) is recommended, but not required for accurate mapping of translocations. Cycling conditions should be adjusted to be appropriate for the polymerase used.

Note: This is not a polymerase-CHAIN-reaction protocol; only one primer is used, resulting in linear amplification.

17. Run thermal cycler with the following cycling conditions:

Steps	Temperature	Time	Cycles
Initial Denaturation	98°C	2 min	1
Denaturation	95°	15 s	70 cycles
Annealing	58°C	15 s	
Extension	72°C	1 min	
Final extension	72°C	2 min	1
Hold	4°C	Forever	

Optional: Add polymerase again to each tube and repeat. Error-proofing polymerases decline in activity over time. This fresh addition allows for further amplification in a non-chain reaction condition, as is the case with the linear-amplification done here.

Pause point: Save PCR strips in −20°C. Note that error-proofing polymerases often have resection activity and these tubes should not be stored at 4°C or 22°C.

18. Use Invitrogen PureLink PCR cleanup kit (Binding Buffer 2) here to further remove residual RNA from original RNase use and any excess biotin primer. Elute 2 × 100 μL in E1 buffer.

Pause point: Save purified DNA in −20°C or 4°C.

Isolate bait DNA and ligate adapters on beads

Timing: 1–2 days

Streptavidin beads capture biotinylated DNA containing translocated, normal, or Non-Homologous End-Joining (NHEJ) sequences at the bait site.

⚠ CRITICAL: All steps until elution from beads should be completed daily at defined pause points. Do not allow bead-containing reactions sit idle for days or weeks, even in cold storage. Do not freeze beads.

19. Pool bait-amplified DNA from PCR strips in a 1.5 mL microcentrifuge tube. Add 50 µL 5 M NaCl. Add 2.5 µL 0.5 M EDTA pH8.
20. Transfer 40 µL (per sample) Dynabeads C1 streptavidin beads (well vortexed!) into new 1.5 mL microcentrifuge tube. Add 600 µL 1× Bead Buffer. Mix by pipetting 10×.
21. Capture beads on magnetic stand 1 min. Discard supernatant.
22. Resuspend beads off magnetic stand in 600 µL 1× Bead Buffer. Mix well.
23. Capture beads on magnet stand 1 min. Discard supernatant.
24. Resuspend beads with pooled bait-amplified DNA. Rotate at 22°C a minimum of 2 h (4 h or 16 h is best).

Pause point: 16 h bead-DNA capture.

25. Thaw bridge adapter **on ice** (a slow thaw is important to ensure primers remain bound).
26. Capture DNA-beads on magnet stand. Wash with 600 µL 1× Bead Buffer, 3 washes.
27. Resuspend beads in 1 mL nuclease-free water. Capture beads and discard supernatant.
28. Resuspend beads in 45 µL nuclease-free water.
29. Set up 100 µL ligation reactions in PCR strips (1 per sample):

Reagent	Rxn Vol
DNA-beads complex	all (45 µL)
10× T4 ligation buffer	10 µL
Hexamine cobalt (III) chloride (20 mM)	5 µL
Bridge Adapter (50 µM)	5 µL
T4 DNA ligase (3 U/µL)	5 µL
50% PEG8000	30 µL

30. Mix well. Incubate reactions at 25°C 1 h. 22°C 1 h.
31. Resuspend mixture well. Incubate another 22°C 1 h.
32. Incubate reactions at 16°C 1–16 h.

Pause point: 16 h T4 DNA ligase reaction at 16°C.

Digestion of normal sequence to enrich for translocations

Timing: 3–5 h

Digestion of normal sequence is the key step to enable a reasonable amount of sequencing to find hundreds to thousands of translocated sequences within your DNA samples. Otherwise, almost all sequencing reads will contain short insertion-deletions (indels) from NHEJ or completely normal sequence. Digestion of normal sequence is accomplished using a methylated oligo which binds to non-translocated DNA and a hemi-methyl sensitive restriction enzyme (the example here is LpnPI). This process removes the adapter from the above step, preventing PCR of cleaved sequence.

33. Add 100 µL 2× Bead Buffer in each PCR tube made in step 13. Combine in new 1.5 mL microcentrifuge tube.
34. Capture beads on magnetic stand, remove supernatant. Wash with 200 µL 1× Bead Buffer twice.

35. Wash once with 200 μL nuclease-free water. Remove water.
36. Set up normal-sequence digestion reaction:

Reagent	Rxn Vol
Nuclease-free water	25 μL
10× NEB Cutsmart Buffer	3 μL
Cytosine methylated oligo (100 μM)	0.5 μL

⚠ CRITICAL: When designing the methylated oligo, ensure the methylated cytosine is within LpnPI's recognition sequence: C^{m}CDG (D is A or G or T), and 20-60 bp from spCas9 target cleavage site (distal to biotinylated primer sequence). It is recommended to perform quality control of LpnPI digestion on a carefully controlled PCR of the bait site prior to using on translocated DNA.

Note: The enzyme LpnPI is not included in above reaction, as the following 95°C step will diminish its activity. 95°C also diminishes streptavidin bead efficiency, so denaturation is intentionally short.

37. Anneal methylated oligo on a thermal cycler:

Steps	Temperature	Time	Cycles
Brief denaturation	95°C	2 min	1
Initial cooling	70°C	30 s	1
Slow primer annealing	Ramp 0.1°C/s to 60°C		1
Slow primer annealing	Ramp 0.1°C/s to 50°C		1
Set to enzyme temperature	37°C	forever	

38. Add 0.5 μL LpnPI enzyme (5,000 U/mL) and 1 μL 30× enzyme activator per reaction.
39. Incubate 37°C for 2–4 h [can prepare nested PCR reactions during incubation].

Optional: Heat inactivate LpnPI at 65°C for 20 min. This step does not appear to affect assay outcome.

40. Add 30 μL 2× Bead Buffer.
41. Transfer reactions to 1.5 mL microcentrifuge tubes.
42. Wash beads with 600 μL 1× Bead Buffer three times, once with 1 mL nuclease-free water.
43. Resuspend beads and LpnPI-digested DNA in 225 μL nuclease-free water. **Proceed immediately to Nested PCR**, do not pause.

Nested PCR

Timing: 3 h

Streptavidin-biotin interactions are exceptionally strong, so instead of trying to disrupt this interaction chemically, a nested PCR is performed. To achieve this with variable translocated sequences, a reverse primer binds to bridge adapter sequence, resulting in chain-reaction amplification of translocated sequences without LpnPI cleavage. Once PCR is complete, the amplified nested sequence is released from the beads.

44. Set up 8 × 50 μL Nested PCR reactions per sample:

Reagent	Rxn Vol	Master mix
Nuclease-free water	6.5 μL	(x9 = 112.5 μL)
DNA-beads complex	~25 μL	(all of it)
5× Phusion buffer	10 μL	(x9 = 90 μL)
dNTPs (2.5 mM each)	4 μL	(x9 = 36 μL)
I5-nested (10 μM)	2 μL	(x9 = 18 μL)
I7-blue (10 μM)	2 μL	(x9 = 18 μL)
Phusion polymerase (2 U/μL)	0.5 μL	(x9 = 4.5 μL)

Note: Either the I5-nested and/or I7-blue primers may be designed with in-line sample barcodes or unique-molecule-identifiers (UMIs) for downstream bioinformatics. Sample barcodes are used, note which sample barcode was used with each sample. These in-line barcodes are not to be confused by the "index" sequences used with Illumina sequencers, which are often used by sequencing lab technicians to allow the sequencer to automatically de-multiplex samples when the sequencing run is complete.

45. Mix beads (do not spin tubes down). Run thermal cycler as follows:

Steps	Temperature	Time	Cycles
Initial Denaturation	95°C	30 s	1
Denaturation	95°	15 s	18 cycles
Annealing	60°C	30 s	
Extension	72°C	1 min	
Final extension	72°C	6 min	1
Hold	4°C	forever	

46. Pool 8 PCR products in new 1.5 mL microcentrifuge tube.
47. Centrifuge 15,000 g for 5 min at 22°C.
48. Use Invitrogen PureLink PCR cleanup kit (Binding Buffer 2) to purify supernatant. Elute once with 66 μL nuclease-free water.

Pause point: Save purified nested PCR DNA in −20°C or 4°C indefinitely for further steps.

Sequencer adapter PCR and library purification

Timing: 1 day

These steps add the adapters necessary for proper next-generation sequencing of the translocated-DNA libraries. Next-generation sequencers have an optimal DNA library base-pair range. The size selection step (either before or after pooling libraries) results in a sample ready to be sent to an Illumina next-generation sequencer.

49. Check concentration of each DNA sample with NanoDrop or Qubit (preferred).
50. Calculate PCR cycles:
    a. [DNA] ≥ 10 ng/μL: N = 11 cycles.
    b. 7 ≤ [DNA] < 10 ng/μL: N = 12–13 cycles.
    c. [DNA] < 7 ng/μL: N = 14–20 cycles.

51. Set up four 50 µL PCR reactions:

Reagent	Rxn Vol	Master mix
Nuclease-free water	16.5 µL	(x4.5 = 74.25 µL)
5× Phusion buffer	10 µL	(x4.5 = 45 µL)
dNTPs (2.5 mM each)	4 µL	(x4.5 = 18 µL)
P5-I5 primer (10 µM)	2 µL	(x4.5 = 9 µL)
P7-I7 primer (10 µM)	2 µL	(x4.5 = 9 µL)
Phusion polymerase (2 U/µL)	0.5 µL	(x4.5 = 2.25 µL)
DNA product	~15 µL	all of it

52. NGS-adapter PCR. Run thermal cycler as follows:

Steps	Temperature	Time	Cycles
Initial Denaturation	95°C	3 min	1
Denaturation	95°	15 s	N cycles
Annealing	62°C	30 s	
Extension	72°C	1 min	
Final extension	72°C	6 min	1
Hold	4°C	forever	

53. Perform size selection of final DNA libraries. This may be done before or after pooling samples. There are multiple adequate procedures to size-select, and each should be confirmed on a 1% agarose gel or TapeStation to result in DNA within the 350–1,000 bp range.
    a. Option 1: Simple column purification using specialized buffer to remove long adapter primer dimers.
        i. Use Invitrogen PureLink PCR cleanup kit (using Binding Buffer **High-Cutoff (B3)**).
        ii. Be sure that no high molecular weight DNA is observed (>2,000 bp or in the sample wells), or the sequencer may not cluster efficiently.
    b. Option 2: Run the entire DNA library on a 1% agarose gel.
        i. Physically cut out the 350–1,000 bp section using a razor blade.
        ii. Purify DNA from the agarose using a PureLink Quick Gel Extraction Kit.

    Note: if UV light is used, mutations will be introduced. Specialized dyes or running of a parallel size control gel may enable cutting out this range without using UV light.

    c. Option 3: Bead-based size selection. AMPure XP is an example of a commonly used bead-based DNA size selection kit.
54. Before sending libraries to an NGS facility, perform a simple Sanger sequencing check of the final libraries.
    a. Using P5/I5 Illumina adapters as PCR primers and sequencing primers, the Sanger sequence should appear fairly normal until the bait DSB site nears, where insertion-deletions and translocations begin to create random A's T's C's and G's at all following locations. An example of this quality control step is shown in Figure 4.

    Optional: Perform qPCR using either (a) NGS sequencing primers or (b) NGS sequencer-binding primers to ensure linear amplification of properly constructed libraries. Most NGS facilities do this for user-created custom libraries.

55. Design your submission based on the assumption that 0.1%–1% of your reads are translocations.

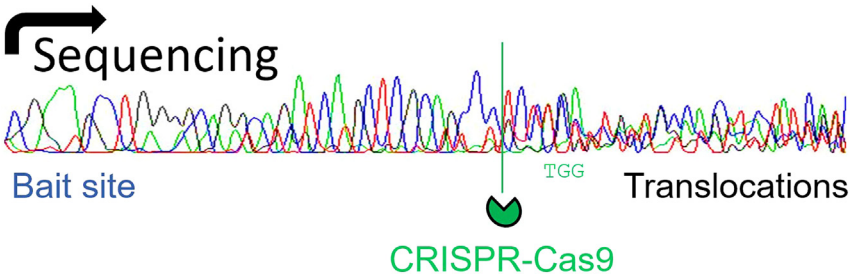


Figure 4. Sanger sequencing example of successful translocation capture

Note: Many of these will be on the same chromosome as the induced DSB, with a smaller fraction on other chromosomes. This percentage may be higher depending on your optimization of steps 33–36: LpnPI digestion.

56. When all libraries are ready, submit to an Illumina next-generation sequencing facility which routinely processes ≥150 bp reads.

Note: This protocol was initially developed for Illumina MiSeq, but modern high-throughput Illumina sequencers readily perform 150 bp or higher single- or paired-end sequencing. Depending on how the primers are designed, this allows for ~30–50 bp upstream base-pairs from the spCas9 cleavage site, and the remainder allows for accurate mapping of (a) NHEJ indels or (b) translocations.

Analyze the data

Timing: 1 h - 2 days, or up to weeks longer depending on analysis needs

Analysis of translocations depends entirely on the research question related to the designed DSB. However, most analyses will include (a) quality control of correctly captured DSB, (b) trimming of bait sequence, and (c) mapping of translations to the reference genome. These procedures are outlined below, using the public, free web-based bioinformatic suite UseGalaxy.org.

57. Upload "*.fastq.gz" sequencing file (or two per sample for paired-end sequencing) to your account, using the "Upload Data" button. If using large files, utilize FileZilla to send via FTP via UseGalaxy.org instructions.
    a. In the upload menu in UseGalaxy.org, set uploaded files as datatype "*.fastqsanger.gz" so downstream tools sense the sequencing data. This works better for more tools using Illumina NGS data, despite there being a "*.fastqillumina.gz" option.
    b. If your file fails to process in UseGalaxy.org for any steps, try sending your input file to "FASTQ Groomer" first. While this makes the file size large, duplicates the file, and adds processing time, the tool can fix unexpected file type errors.
58. Trim the bait sequence prior to the bait DSB, and retain only those reads which contained the bait DNA.
    a. Use the "cutadapt" tool, pasting ~40 bp of bait DNA near the DSB into the "5′ (Front) Adapter" input box, using the "Enter custom 5′ adapter sequence" within the "Enter custom sequence" sub-menu. For the example in this protocol, the sequence is: "CAAGCCCCC CACCAAGTGCTTCAGGAACAGGTCGAAAGCC". Under "Filter Options", check "Yes" for "Discard Untrimmed Reads".
    b. When this step is complete, only sequences which had the captured DSB site will be analyzed.

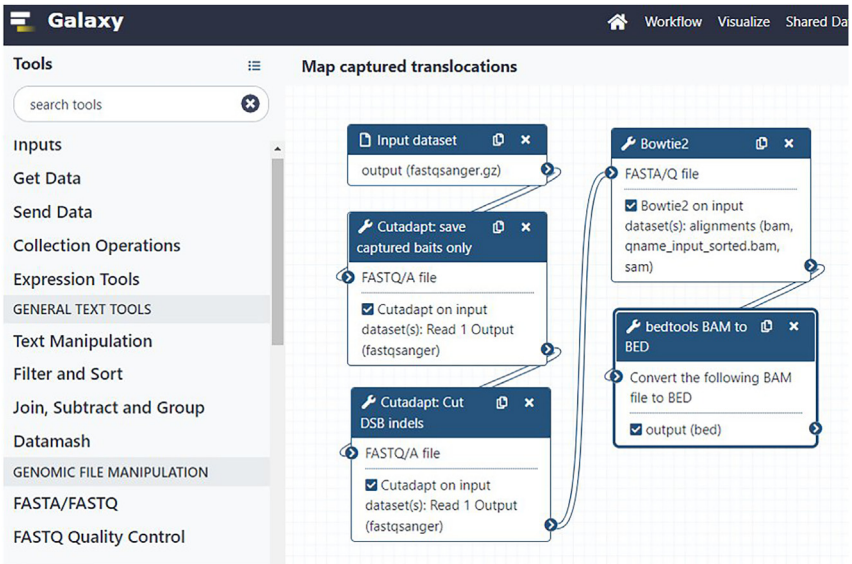


Figure 5. Example GEF-TLC-seq bioinformatic workflow in Galaxy

Optional: Perform sequencing-quality trimming using tools and settings appropriate for the sequencer used.

Optional: In our hands, A-tailing can be observed. In the first "cutadapt" trim step, this can be mitigated using a "3′ (End) Adapters" inclusion of a "AAAAAAA" sequence in the "Enter custom 3′ adapter sequence" option. This may also be considered with NextSeq instruments using a polyG sequence, to remove low signal sequences. However, these repetitive sequences do not appear to alter Bowtie2 mapping drastically using the settings below.

59. Trim ~10 bp past the DSB site to keep only translocated sequence. In this example, this is an additional 45 bp. Use "cutadapt" a second time, using the output from step 58. Set "Read 1 options", "Cut bases from reads before adapter trimming" to 45. In filter options, set "Minimum length (R1)" to 50, resulting in an output which only contains translocations with 50 bp of sequence or more.
60. Map your translocations to the genome. Use Bowtie2. In this example, the mouse genome (mm10 or mm39 at the time of writing) is the reference genome.

Note: While it is common to use RmDup to remove PCR duplicates after mapping, in this case, it will remove many normal sequences with set insertion-deletions around the bait site. RmDup may still be used to ensure only unique translocations are output for quantitative purposes.

61. If chromosome coordinates are desired to determine where the translocation map to, utilize the BAM to BED tool. This will output a tabular file of each mapped translocation for your sample. This "*.BED" file is a reasonable "final processed data output" for translocation analysis.

A screenshot of an example workflow in UseGalaxy.org is provided as Figure 5.

Or alternatively, with RmDup to include only unique translocations in Figure 6.

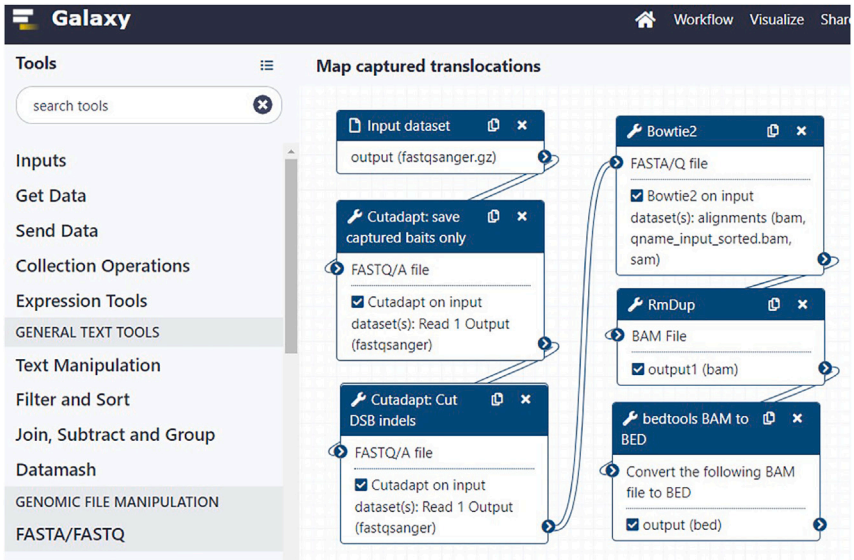


Figure 6. Example GEF-TLC-seq bioinformatic workflow in Galaxy with RmDup

EXPECTED OUTCOMES

If the entire wet-lab and *in silico* analysis worked properly, the end History result in UseGalaxy.org will appear similar to Figure 7.

If the BAM file is transferred to the UCSC Genome Browser, the translocations can be shown relative to where they map on the reference genome. The example above with the small control file mapped to Chromosome 18 appears as Figure 8. Each vertical line indicates a mapped, unique translocation.

When the mapped BAM reads are viewed at the DSB bait site on the UCSC Genome Browser, it appears similar to Figure 9. It is clear many different repairs near the break site are mapped.

QUANTIFICATION AND STATISTICAL ANALYSIS

Since most breaks are repaired locally through NHEJ and/or on the same chromosome via a variety of mechanisms, quantitative measures of translocations need careful consideration of what hypothesis is being tested. For instance, in the example here 68% of our mapped reads were located within 1 Mb of the bait site, 6% were on the same chromosome but distal to the bait site, and 26% were translocations to other chromosomes. If sequence specificity is the question, then it may be advisable to only use translocations which are not on the chromosome containing the bait site. If any regions are strongly homologous to the bait site, these may need to be masked in quantitative analyses as well.

LIMITATIONS

It should be noted that highly repetitive regions, like centromeres, telomeres, or other repetitive elements, may be sites of translocations, but are not mapped by using Bowtie2 on default settings. Culturing conditions likely influence DNA repair. Many cell lines, especially cancer cell lines, have mutations or epigenetic suppression of DNA repair enzymes, which will affect which translocations are observed. Proper controls may mitigate some of these limitations.

The quantity of unique translocations depends entirely on[1] the number of genomes initially present (approximately 2× the number of cells if using a diploid cell line, with a small percentage lost at each protocol step), (2) the efficiency of DSB generation, and (3) the efficiency of local break site repair, which most cells favor over DNA repair resulting in translocations.

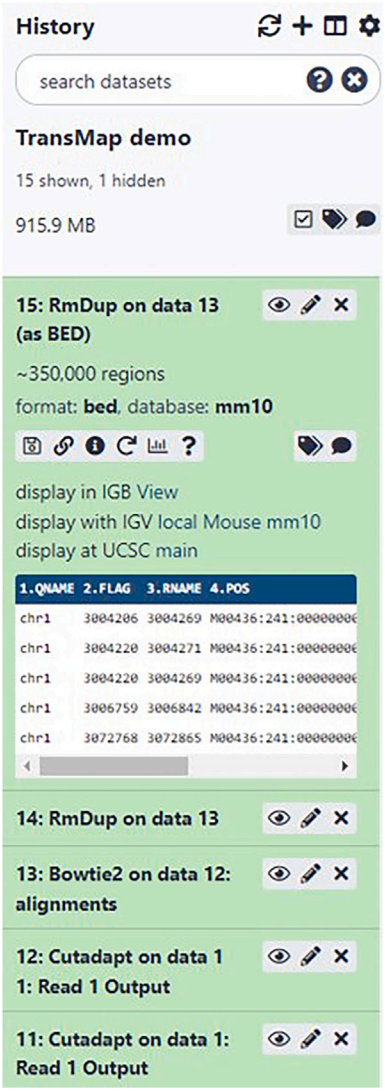


Figure 7. Example history view in Galaxy upon workflow completion

TROUBLESHOOTING

Problem 1

Sequencer incompatibility.

A major issue which may arise over time may be sequence-specific incompatibility with new sequencers, as technology continues to improve and change at rapid pace.

Potential solution

Before designing and ordering primers, it is advised to consult the sequencer manufacturer's descriptions for appropriate sequences for[1] binding to the sequencer and (2) sample identification in pooled libraries. Long-read (150 bp or more from one end) is essential for the quality-control and mapping bioinformatic analysis steps.

Problem 2

Poor DSB generation.

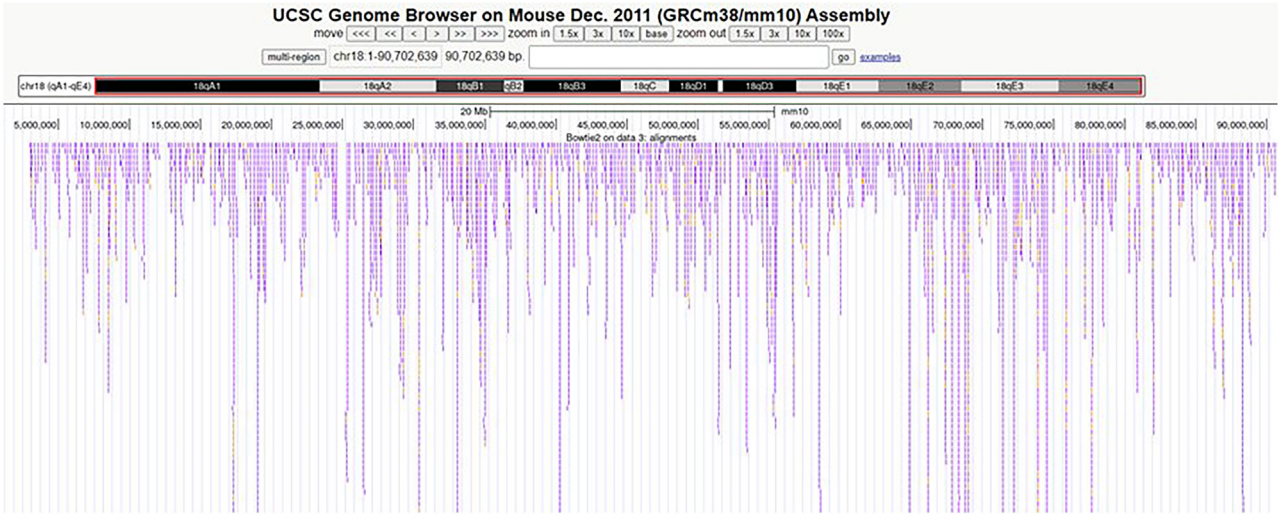


Figure 8. Example chromosome-wide translocation overview in UCSC Genome Browser

If few short indels or translocations are observed in your analysis, the bait DSB may never have been efficiently generated.

Potential solution

Lentiviral transduction, alternate sgRNAs, and different cell density may be altered to optimize formation of DSBs.

If possible, it is ideal to design a sgRNA to cut at a known restriction enzyme site. In this way, quality control can be performed prior to any sequencing by PCR followed by restriction enzyme digests of amplified product. If spCas9 is appropriately cutting the DNA at the bait site, the restriction site is often destroyed by NHEJ-mediated short insertion-deletions, resulting in less efficient restriction digests.

Alternatively, some NGS companies offer cheaper quality control sequencing of a small quantity of reads (such as Azenta/GENEWIZ, at 50,000 reads via the AMPLICON-EZ option), which can directly observe and give an estimate of frequency for the short insertion deletions after administration of spCas9 and the sgRNA.

Problem 3

Low translocation rate.

If few translocations are observed in an analysis despite clear efficiency of the cut site, it is possible that the cell biology is not optimal for translocations.

Potential solution

Specific repair factor efficiency is low, which depends on the cell type studied. Many cancer cell lines have different efficiencies for NHEJ, homology-directed repair, alternative end-joining, and other repair processes. Try another cell line or model with a better characterized repair capacity.

Problem 4

Low DNA yield.

If at any PCR step or purification step it seems there is too little DNA, it may be that too little starting material was used.

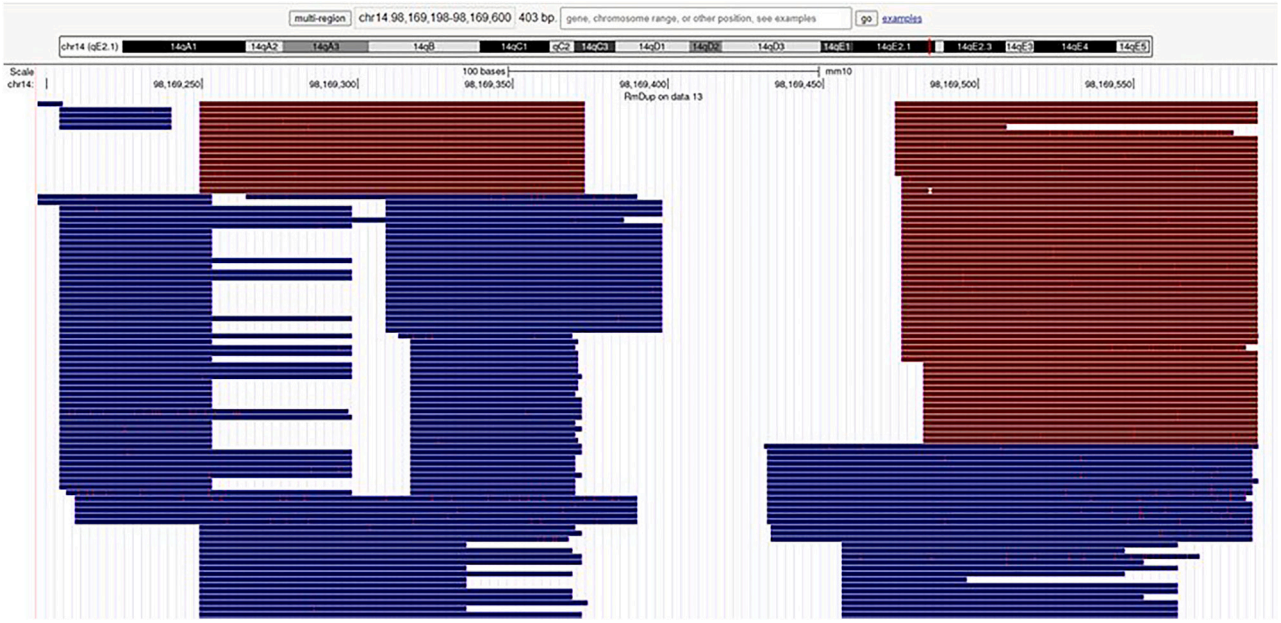


Figure 9. Example bait site zoomed-in view in UCSC Genome Browser

Potential solution

The number of cells in this protocol is a lower limit to discover translocations. Remember, each translocation came from a single original nucleus, so starting cell number is absolutely critical to discovering thousands of rare events. These events may not exist, allowing for amplification after LpnPI cleavage, if too few cells were used. It should be noted that many steps require the complete usage of previous step material. While it may be comforting to save half with every major step, you may need to double your material and reactions each time to allow for enough material for the next steps.

Problem 5

Bioinformatic errors.

Unfortunately, bioinformatics is just as complex as the bench science and requires a different skill set. Often even user-friendly web-based software like UseGalaxy can provide errors without clear reasons for why they occurred.

Potential solution

Start with reading the original paper and manual associated with the bioinformatic tool. Get an understanding for its intended usage, inputs, and outputs. Try to find example datasets specific to the tool to understand the format expected by the tool. After this basic understanding is achieved, if the problem persists, reach out to a bioinformatician at your institution to see is they can help communicate what is causing the error.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Joe R. Delaney (delaneyj@musc.edu).

Materials availability

Control plasmid reagents used in the example protocol here can be obtained by reasonable request to Joe R. Delaney. SgRNAs were cloned into the PX458 plasmid, available from Addgene (#48138).

Data and code availability

Control and experimental raw and processed datasets generated during this study are available at GEO: GSE166119 and SRA: SRR13622084. The "DACH1_posCtrl" sample refers best to the protocol written here.

ACKNOWLEDGMENTS

We are grateful to the UCSD Institute for Genomic Medicine for valuable advice in the planning and performance of the NGS. This work was supported by grants from the NIH (R01 EY014061 and R01 AG033082 to A.R.L.S. and R00 CA207729 and R03 CA256104 to J.R.D.) and includes data generated at the UC San Diego IGM Genomics Center supported by the Diabetes Research Center grant (P30 DK063491). This research was supported (in part) by a grant from the Rivkin Center for Ovarian Cancer (J.R.D.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the funding agencies.

AUTHOR CONTRIBUTIONS

A.R.L.S. and J.R.D. provided the conceptual framework for the study. J.R.D. and A.R.L.S. designed the experiments. J.R.D. performed the experiments. A.R.L.S. and J.R.D. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Switonski, P.M., Delaney, J.R., Bartelt, L.C., Niu, C., Ramos-Zapatero, M., Spann, N.J., Alaghatta, A., Chen, T., Griffin, E.N., Bapat, J., et al. (2021). Altered H3 histone acetylation impairs high-fidelity DNA repair to promote cerebellar degeneration in spinocerebellar ataxia type 7. Cell Rep. *37*, 110062. https://doi.org/10.1016/j.celrep.2021.110062.

2. Afgan E., Baker D., Batut B., van den Beek M., Bouvier D., Cech M., Chilton J., Clements D., Coraor N., Grüning B.A., et al. (2018 Jul 2). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. *46*, W537–W544. https://doi.org/10.1093/nar/gky379.