STAR Protocols

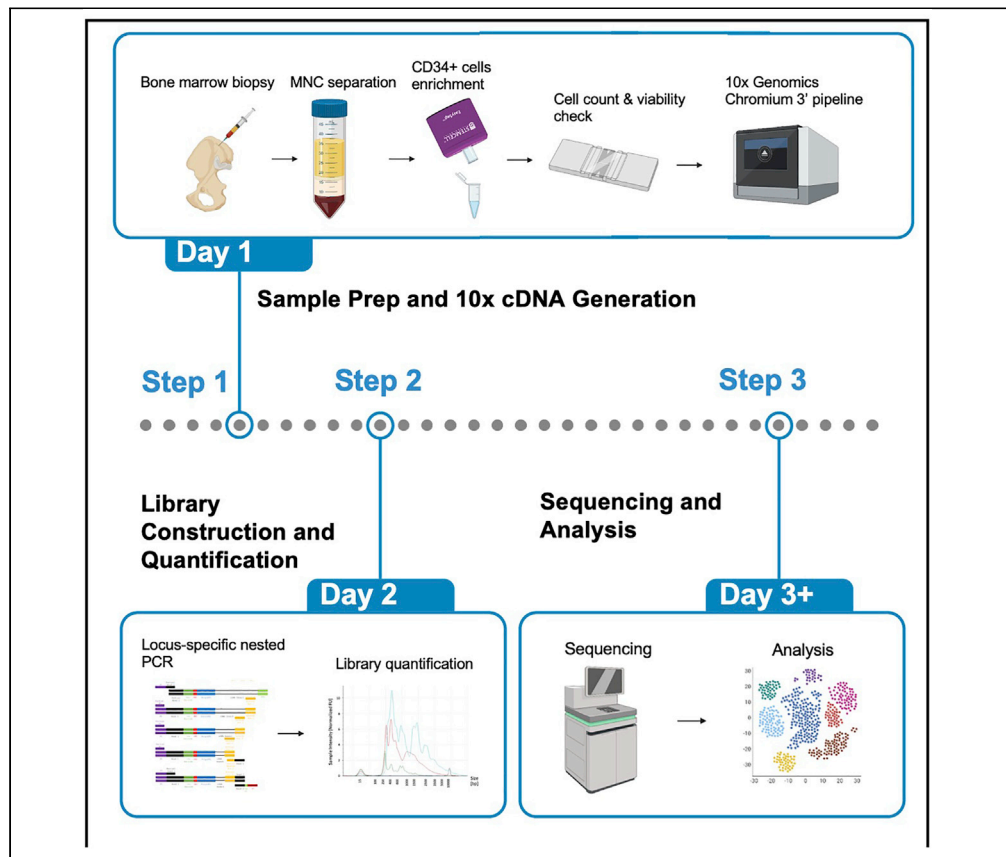**Protocol**

Integrating readout of somatic mutations in individual cells with single-cell transcriptional profiling

Shichen Liu,
Maximilian Nguyen,
Sahand Hormoz

shichen_liu@dfci.harvard.edu (S.L.)
sahand_hormoz@hms.harvard.edu (S.H.)

Highlights

Isolation of CD34+ cells from human bone marrow aspirates

Enrichment of target somatic mutations from single-cell cDNA

Protocol enables single-cell RNA sequencing alongside single-cell amplicon sequencing

In many biological applications, the readout of somatic mutations in individual cells is essential. For example, it can be used to mark individual cancer cells or identify progenies of a stem cell. Here, we present a protocol to perform single-cell RNA-seq and single-cell amplicon-seq using 10X Chromium technology. Our protocol demonstrates how to (1) isolate CD34+ progenitor cells from human bone marrow aspirate, (2) prepare single-cell amplicon libraries, and (3) analyze the libraries to assign somatic mutations to individual cells.

Protocol

Integrating readout of somatic mutations in individual cells with single-cell transcriptional profiling

Shichen Liu,[1,2,4,*] Maximilian Nguyen,[1,2] and Sahand Hormoz[1,2,3,5,*]

[1]Department of Data Science, Dana-Farber Cancer Institute, Boston, MA 02215, USA

[2]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

[3]Broad Institute of MIT and Harvard, Cambridge, MA, USA

[4]Technical contact

[5]Lead contact

*Correspondence: shichen_liu@dfci.harvard.edu (S.L.), sahand_hormoz@hms.harvard.edu (S.H.)

https://doi.org/10.1016/j.xpro.2021.100673

SUMMARY

In many biological applications, the readout of somatic mutations in individual cells is essential. For example, it can be used to mark individual cancer cells or identify progenies of a stem cell. Here, we present a protocol to perform single-cell RNA-seq and single-cell amplicon-seq using 10X Chromium technology. Our protocol demonstrates how to (1) isolate CD34+ progenitor cells from human bone marrow aspirate, (2) prepare single-cell amplicon libraries, and (3) analyze the libraries to assign somatic mutations to individual cells.

For complete details on the use and execution of this protocol, please refer to Van Egeren et al. (2021).

BEFORE YOU BEGIN

Design of locus-specific amplicon primers

⏱ Timing: 15 min per locus

1. Identify somatic mutations of interest and the mRNA in which they occur (Figure 1A).
2. Design locus-specific primers 1, 2, 3 to be roughly 300 bp, 150 bp and 50 bp upstream (5′) of the somatic mutation (Figures 1B, 1C, and 1D)

   Note: 1: 10× Genomics Chromium 3′ pipeline captures the polyadenylated (polyA) transcripts. The directionality of the three locus-specific primers reverses and become reverse primers that anneal to the 3′ end of cDNA during Polymerase chain reactions.

   ⚠ CRITICAL: All three primers should have a length of around 25 bp and have a melting temperature of around 65°C to reduce experimental complexity.

3. Design locus-specific primer 4 by adding Illumina Read 2 sequence to the 5′ end of the first 18–22 nucleotides of locus-specific primer 3 (Figures 1B and 1C, and Table 1).

Preparation of bone marrow aspirates from patients and healthy donors

⏱ Timing: ∼ 1 h

4. Use a sterile syringe and tube with no heparin coating to aspirate the sample.
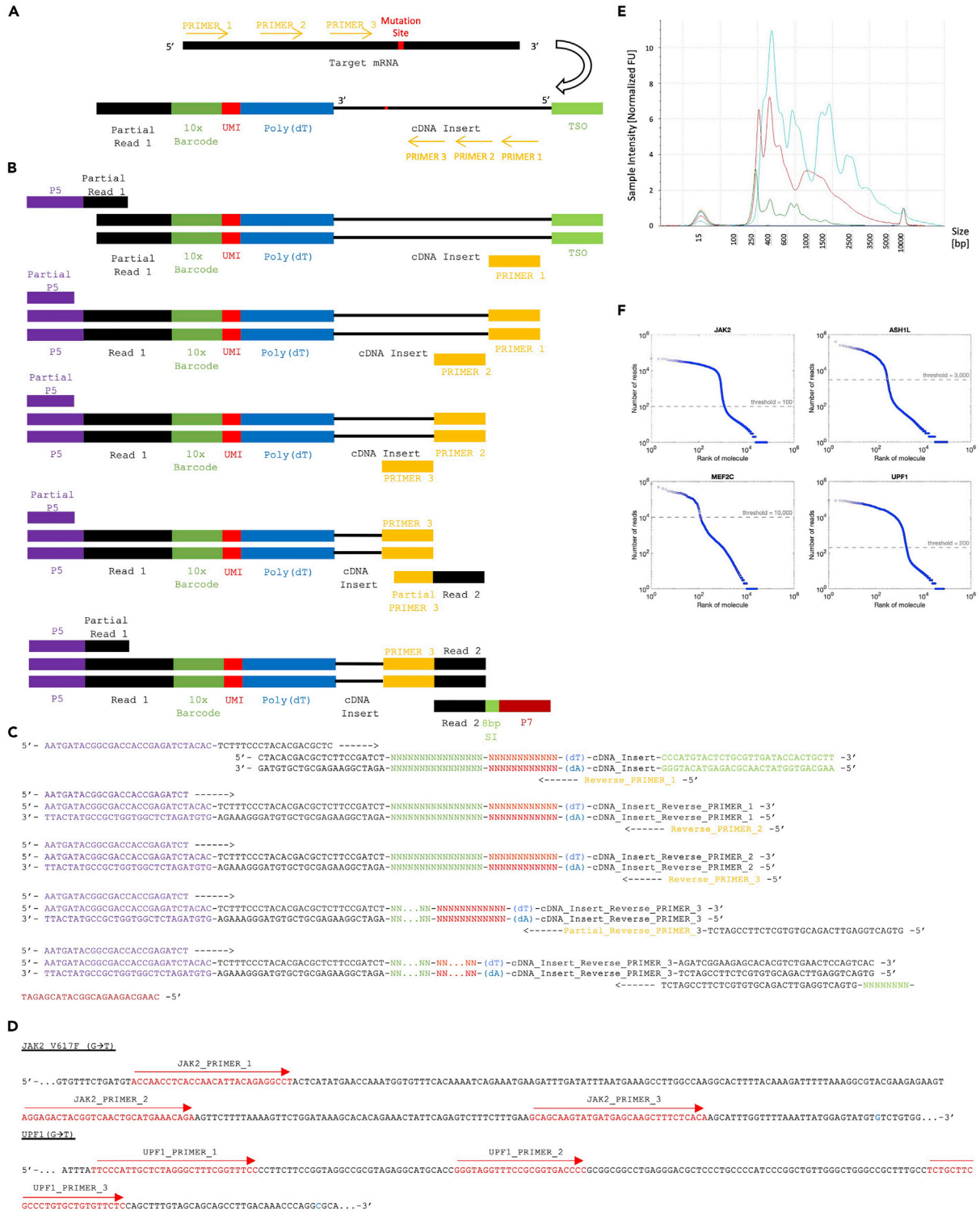5. Collect bone marrow aspirate (BMA) and place it in EDTA-coated tubes.

Figure 1. Primer designs, directions, locations, sequences of target mutation; accurate identification of the mutated cells from the amplicon libraries

(A) Illustration of primer direction against target mRNA and the change of primer directionality during amplification of sc-cDNA.

(B) Schematic diagrams of the nested PCR with locus-specific primers 1–4 (denoted as primers 1–4) from steps 1 to 5, respectively.

(C) Oligonucleotide sequences and localization of common primers and adaptors.

(D) Example primer positions and sequences of targeted mutation. Red indicates primer sequences and blue indicates mutation nucleotides.

(Figure reprinted with permission from Van Egeren et al., 2021).

Table 1. Primers and sequences for mutation-specific single-cell amplicon libraries (5′→3′)

INTERNAL_FORWARD	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC
SHORT_INT_FOR	AATGATACGGCGACCACCGAGATCT
JAK2_PRIMER1	ACCAACCTCACCAACATTACAGAGGCCT
JAK2_PRIMER2	AGGAGACTACGGTCAACTGCATGAAACAGA
JAK2_PRIMER3	GCAGCAAGTATGATGAGCAAGCTTTCTCACA
JAK2_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCAGCAAGTATGATGAGCAA
ASH1L_PRIMER1	GCATCTCACTCCTATCTGAAAAGTTGACAAGC
ASH1L_PRIMER2	TGGCCACAAAGAAAAACCTAGACCATGTCA
ASH1L_PRIMER3	GGAAATGTCCCCTTCAGGCTGTCGTATCAA
ASH1L_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGAAATGTCCCCTTCAGGCT
GOLIM4_PRIMER1	TCACCCTATGAGGAACAGTTGGAACAGCAG
GOLIM4_PRIMER2	GGGCACTTACTACGGCAGCAGGAACAG
GOLIM4_PRIMER3	TGCTATGGATAATGATATCGTTCAGGGAGCAG
GOLIM4_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGCTATGGATAATGATATCG
ABHD2_PRIMER1	GGGTGACACAGCAAGACCCTTCTCAAAATT
ABHD2_PRIMER2	AAATGGACAGAGCCTCTTACTTTGGGGCA
ABHD2_PRIMER3	TGTGAGACACAAATACTGCCCACTTCATTCA
ABHD2_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGTGAGACACAAATACTGCC
FHIT_PRIMER1	GCCTACTTAATCCTTTTCCTACTTCGTGGGGG
FHIT_PRIMER2	GGGATCACAAAAGTGAAGATTGGATGCCGT
FHIT_PRIMER3	CCAGTTGTGTTTTCTCATTTCCCTTAGAGCCA
FHIT_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGTTGTGTTTTCTCATTT
FRYL_PRIMER1	GGATTTTGTTCATTCCCTCTCTTGACTACTGGG
FRYL_PRIMER2	TGGTCCCATACTGTCTTAGTCCCTATTCCCC
FRYL_PRIMER3	GAGTCAACAGTCTGTTCATATCACTTCCTTTCTCA
FRYL_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGAGTCAACAGTCTGTTCATA
MAML3_PRIMER1	GCCAATTGCTTTTTCAAGTACACCCACTTTTACT
MAML3_PRIMER2	ACCCTGGTGTGAAAACAAAGTAAAACCGAGT
MAML3_PRIMER3	TGCCTAGATAACCATCTTTCTCCCCACCC
MAML3_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGCCTAGATAACCATCTTTC
PPIL3_PRIMER1	TRAPPC11_PRIMER1
PPIL3_PRIMER2	CAAGAGTTTGAAACCAGCCTGGGCAACAT
PPIL3_PRIMER3	TGTGCCTGTAGTCCCAGCTACTCAAGAGG
PPIL3_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGTGCCTGTAGTCCCAGCTA
RSF1_PRIMER1	GACACCTCCTTATCTCCCTTCACCTGGGT
RSF1_PRIMER2	GACACTCAACTCTCCCTCCCTCTGTTGG
RSF1_PRIMER3	GGGCTGCTTTAACCCCTAAAACTCCTTCC
RSF1_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGCTGCTTTAACCCCTAAA
TRAPPC11_PRIMER1	CCTGCATCACCTCTTCCAATACCGCTTTC
TRAPPC11_PRIMER2	CTCCCCGCATCTCTTCCTTGCTGAAGA
TRAPPC11_PRIMER3	ACCTGGATTTTGGAGATTACATGGTGCTGT
TRAPPC11_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACCTGGATTTTGGAGATTAC
HSPA9_PRIMER1	ACCTGACAAGAGTCTTAAGCAACCAAAGCA
HSPA9_PRIMER2	GTGGGTCATGCCTGTAATCCCAACACTTG
HSPA9_PRIMER3	GTGTGGGAGTTGAAGATCACCCTAGGCAA
HSPA9_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTGTGGGAGTTGAAGATCAC
MEF2_PRIMER1	TGTCTCAGTCACATTTCTCCAGGTTTCCGT
MEF2_PRIMER2	TGAGATATCGATGTCATTTTCAATGCAGAGGCA
MEF2_PRIMER3	ACTGCAGTTCACTATTGGCATAACAAGTAACCA
MEF2_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACTGCAGTTCACTATTGGCA
JAK2_EXON12_PRIMER1	ACCAGATGGAAACTGTTCGCTCAGACAAT
JAK2_EXON12_PRIMER2	TGTCCCCCAAAGCCAAAAGATAAATCAAACCT
JAK2_EXON12_PRIMER3	ACCAACCTCACCAACATTACAGAGGCCTAC
JAK2_EXON12_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACCAACCTCACCAACATTAC
LMO7_PRIMER1	AAAGGCATGGTTCACTGGAGGCCAC

(Continued on next page)

Table 1. Continued

INTERNAL_FORWARD	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC
LMO7_PRIMER2	ACTGAAGTACAGTCTTATCATATGAGCAGAATGACG
LMO7_PRIMER3	TCCATTTGTTTAAGACTGTTAAACACAAGCACATTGC
LMO7_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCCATTTGTTTAAGACTGTT
CES5A_PRIMER1	GGGTTAGGCATTGTAGTGGAGATAGGCATGGAA
CES5A_PRIMER2	AGGCTGGAGTGCAGTGGCATGATCTT
CES5A_PRIMER3	GGGCCTGTGCCACTACACTCAGCTAATTT
CES5A_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGCCTGTGCCACTACACTC
CDC42SE2_PRIMER1	TGAGGATATGTGCAAGTGATGGTGCTGGAGTT
CDC42SE2_PRIMER2	GATGGTGCTGGAGTTGCCACAGTGAA
CDC42SE2_PRIMER3	TCAAGGGAAGTGGTTGAGGAAACGGAGT
CDC42SE2_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCAAGGGAAGTGGTTGAGGA
NRROS_PRIMER1	GAATCCATCTGTCTCCTTTCCTCAGCTTTGCCT
NRROS_PRIMER2	AGTCCCGGAGCTGGTGGCAAAGA
NRROS_PRIMER3	TCTCACGGGCCCAGCCTTACTCA
NRROS_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCTCACGGGCCCAGCCTTAC
NSFP1_PRIMER1	GGCGTTCATTTTGTGACAGTTCA
NSFP1_PRIMER2	AGCATTCACTGAGAAAAACAATAATGA
NSFP1_PRIMER3	TGGGGAAGATGGTAGGGAGTTTG
NSFP1_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGGGGAAGATGGTAGGGAGT
PIWIL2_PRIMER1	GGAGTTTCACTCTTGTTGCCCAGGCT
PIWIL2_PRIMER2	GTGCAATCTCAGCTCACCGCAACCT
PIWIL2_PRIMER3	TCTCCTGCCTCAGCCTCCCAAGT
PIWIL2_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCTCCTGCCTCAGCCTCCCA
ZNF22_PRIMER1	AACCTACAATTTACACACCTCCCTGCCTTCA
ZNF22_PRIMER2	GCCAAGTGTCAGACTCTAATGAGCCCTCA
ZNF22_PRIMER3	GCTCAGGTTTTAATTTCTATTGAATGCTA
ZNF22_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCTCAGGTTTTAATTTCTAT
UPF1_PRIMER1	TTCCCATTGCTCTAGGGCTTTCGGTTTCC
UPF1_PRIMER2	GGGTAGGTTTCCGCGGTGACCCC
UPF1_PRIMER3	TCTGCTTCGCCCTGTGCTGTGTTCTC
UPF1_PRIMER4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCTGCTTCGCCCTGTGCTGT

Note: Cryopreservation of BMA is not recommended for the protocol because the impact of cryopreservation on transcription profile of BMA is not known.

Preparation before experiment

⏱ Timing: ~ 1 h

6. See key resources table for preparation of needed materials.
7. Set centrifuge to room temperature (20°C–22°C).
8. Prepare 2% FBS/PBS and acclimate to room temperature (20°C–22°C).

Note: 2% FBS/PBS should be stored at 4°C for up to 6 months before and after usage.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
EasySep™ Human CD34 Positive Selection Kit II	STEMCELL Technologies	Cat# 17856

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Whole bone marrow samples	Massachusetts General Hospital; Dana-Farber Cancer Institute	N/A
Chemicals, peptides, and recombinant proteins		
Q5 High-Fidelity 2X Master Mix	New England Biolab	M0492S
UltraPure Distilled Water	Invitrogen	10977
1X RBC Lysis Buffer	eBioscience	Cat# 00-4333-57
Fetal Bovine Serum	VWR	Cat# 89510-186
SPRIselect	Beckman Coulter	Cat# B23318
Lymphoprep	STEMCELL Technologies	Cat# 07801
EasySep Buffer	STEMCELL Technologies	Cat# 20144
1× Dulbecco's phosphate-buffered saline	Thermo Fisher Scientific	Cat# 14040133
eBioscience 1X RBC Lysis Buffer	Thermo Fisher Scientific	Cat# 00-4333-57
Critical commercial assays		
Chromium Single Cell 3′ GEM, Library & Gell Bead Kit v3, 16 rxns	10× Genomics	Cat# 1000075
Chromium Chip B Single Cell Kit, 48 rxns	10× Genomics	Cat# 120262
High Sensitivity D5000 ScreenTape	Agilent	Cat# 5067-5592
High Sensitivity D5000 Reagents	Agilent	Cat# 5067-5593
Qubit dsDNA HS Assay Kit	Invitrogen	Cat # 32854
Chromium i7 Multiplex Kit, 96 rxns	10× Genomics	Cat# 120262
Oligonucleotides		
Primers	Hormoz Lab	Table 1
Software and algorithms		
Amplicon analysis script	Hormoz Lab	N/A
Other		
SepMate™-15/50 tubes	STEMCELL Technologies	Cat# 85450
Plastic whole blood tube with spray-coated K2EDTA	Becton, Dickinson and Company	Cat# 366643
Alternatives		
Any DNA ScreenTapes and Reagents (for step 38)	Agilent	N/A
Any Bioanalyzer DNA Chips and Reagents (for step 38)	Agilent	Cat# 366643
Red Blood Cell Lysis Buffer	Sigma-Aldrich	Cat# 11814389001

STEP-BY-STEP METHOD DETAILS
Density gradient centrifugation

⏱ Timing: 60–100 min

This section describes how to isolate mononuclear cells from BMA. Lymphoprep/Ficoll with a density of 1.077g/mL is used in this protocol (Figure 2).

1. Add 22 mL Lymphoprep/Ficoll to the 50 mL SepMate Tube by carefully pipetting it through the central hole of the SepMate insert without creating bubbles.
2. Dilute BMA with an equal volume of 2% FBS/DPBS and mix gently with wide-bore pipette.

   Note: The maximum volume of diluted BMA is 8 mL for SepMate-15 and 34 mL for SepMate-50.

3. Overlay diluted BMA to Lymphoprep/Ficoll in the SepMate tube by slowly pipetting it down the side of the tube slightly above liquid level.

   ⚠ CRITICAL: Mixing diluted BMA and density gradient may result in incomplete separation or loss of cells and recovery will decrease.
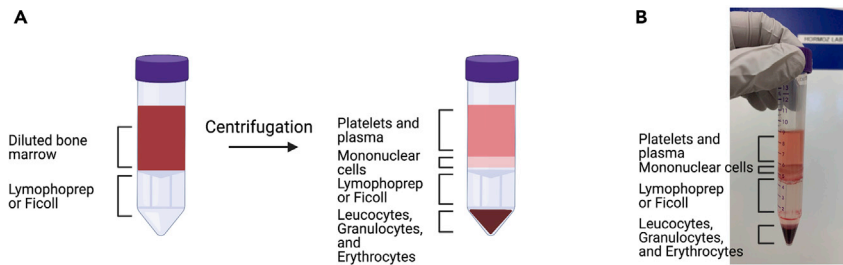
Figure 2. Lymophoprep/Ficoll separation demonstration and expected layering after centrifugation
(A) Illustration of layering of Lymophoprep/Ficoll and diluted bone marrow before and after centrifugation in a SepMate tube.
(B) Picture of separated layers of bone marrow after centrifugation.

4. Centrifuge tube at 1200 × *g* for 20 min at room temperature (20°C–22°C), with brake off.

  ⚠ CRITICAL: It is important that the brake is off otherwise the buffy layer of mono-nuclear cells (MNCs) will be disrupted and recovery will decrease.

  ⚠ CRITICAL: Polycythemia vera samples with high red blood cell (RBC) counts may require additional RBC lysis. The enriched cell layer (layer above the SepMate barrier) should be poured off into a new tube before lysing the RBCs using eBioscience 1X RBC Lysis Buffer. The manufacturer's protocol can be found here: https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2F00-4333.pdf&title=VGVjaG5pY2FsIERhdGEgU2hlZXQ6IDFYIFJCQyBMeXNpcyBCdWZmZXI=

5. Remove the top layer of plasma/platelet and move the buffer layer containing MNCs to a new 50 mL tube.
6. Top up the MNCs to 45 mL with 2% FBS/PBS and mix well with a wide-bore pipette.
7. Centrifuge MNCs at 300 × *g* for 12 min at room temperature (20°C–22°C), with low brake.
8. Remove and discard supernatant.
9. Top up the MNCs again until 45 mL with 2% FBS/PBS and mix well with a wide-bore pipette.
10. Centrifuge MNCs at 120 × *g* for 12 min at room temperature (20°C–22°C), with no brake

  ⚠ CRITICAL: If excess platelets remain (common when samples are from patients with essential thrombocythemia), repeat steps 9 and 10 once. Excess platelets can be identified by visual inspection of the supernatant after step 10. If the supernatant appears cloudy, we identify the sample as having excess platelets.

11. Discard supernatant and resuspend MNC in 1 mL of EasySep buffer on ice, mix with a wide-bore pipette.
12. Measure cell concentration with a hemocytometer and an automatic cell counter.

Magnetic-beads cell enrichment

⏱ Timing: 45 min

The subsequent steps are performed to obtain a single-cell suspension of CD34+ cells.

13. Add MNCs (at concentration of >10^8 cells/mL) to 5 mL polystyrene round-bottom tube.
14. Add 100 μL EasySep Human CD34 Positive Selection Cocktail to the sample.

△ CRITICAL: If sample volume is greater than 1 mL, add the selection cocktail at a ratio of 100 µL per 1 mL of sample.

15. Incubate the sample at room temperature (20°C–22°C) for 10 min.
16. Vortex EasySep Dextran RapidSpheres for 30 s immediately before use.
17. Add 75 µL of EasySep Dextran RapidSpheres to the sample (Van Egeren et al., 2021).

△ CRITICAL: If sample volume is greater than 1 mL, add the RapidSpheres at a ratio of 75 µL per 1mL of sample.

18. Mix the sample and incubate at room temperature (20°C–22°C) for 5 min.
19. Top up the sample to 2.5 mL with EasySep Buffer.
20. Place the tube containing the sample in an EasySep magnet and incubate at room temperature (20°C–22°C) for 3 min.
21. Discard supernatant by inverting the magnet with the tube inside.
22. Rinse the tube with EasySep buffer.

Note: Rinse the side of the tube to achieve maximum recovery.

23. Repeat steps 19 to 22 four more times.
24. Resuspend MNC in 1 mL of EasySep buffer on ice, mix with a wide-bore pipette.
25. Measure cell concentration and cell viability with a hemocytometer and an automatic cell counter. The target final concentration is 700–1,200 cells/µL. The typical percent CD34+ cells viability obtained by following this procedure ranges from 85%–95% based on trypan blue staining.

Note: Measuring cell concentration with both a manual hemocytomter and an automatic cell counter is recommended to minimize the counting error.

26. Keep isolated cells on ice and proceed immediately to the 10× Genomic Chromium Single Cell protocol.

Chromium single cell 3′ GEM cDNA and library construction

⊙ Timing: 45 min

27. Generate the full transcriptomic libraries according to 10× Genomic Chromium Single Cell v3 manufacturer's protocol. The manufacturer's protocol can be found at https://support.10xgenomics.com/single-cell-gene-expression/library-prep/doc/user-guide-chromium-single-cell-3-reagent-kits-user-guide-v3-chemistry.

⊔⊔ Pause point: The cDNA can be stored in a freezer at −20°C for 1–2 months until the following procedures.

Locus-specific single-cell amplicon libraries

⊙ Timing: 4–8 h

The five-step locus-specific PCR amplification first amplifies the somatic mutation of interest (PCR 1–3) in a nested fashion. Illumina sequencing adaptors are subsequently added to the amplified products by PCR (PCR 4–5). Libraries are then quantified on an Agilent Tapestation or Bioanalyzer. (Tapestation traces can be found in Figure 3)
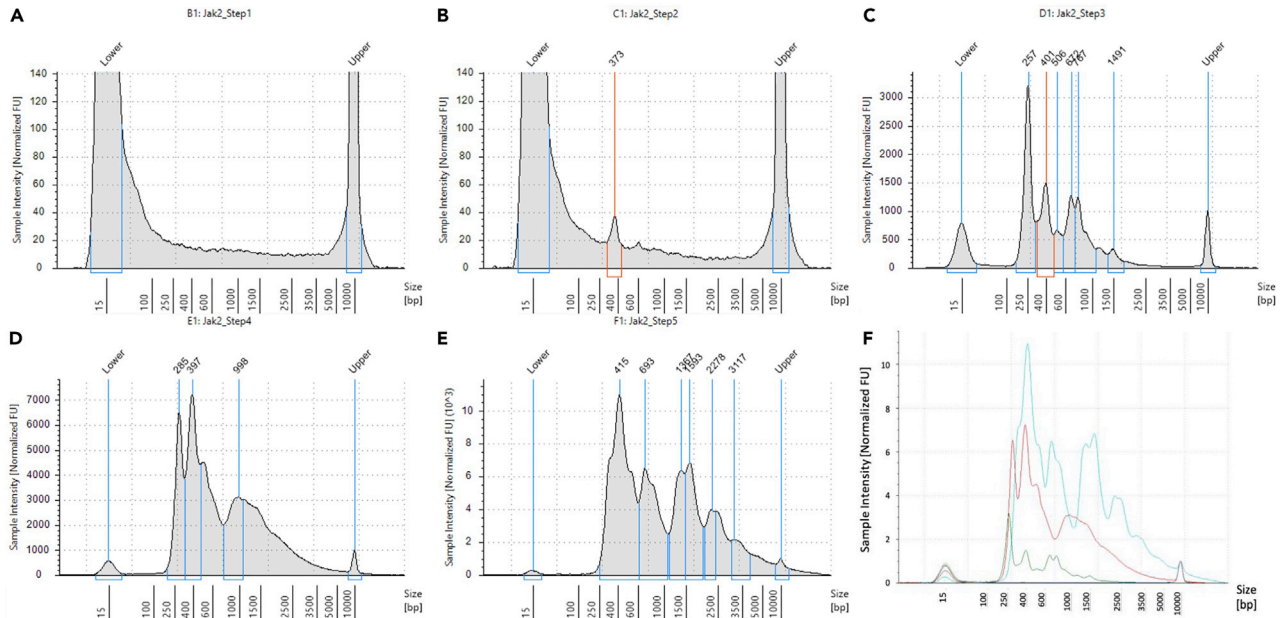
Figure 3. Step-by-step tapestation traces for locus-specific amplicon library construction
(A–E) Step by step tapestation traces from locus-specific PCR 1 to SI PCR.
(F) Collapsed view of tapestation of PCR steps 1–5.

28. Nested locus specific PCR 1.
    a. Prepare master mix in a 0.2 mL thin-wall PCR tube.

Reagent	Final concentration	Amount
Internal_forward (10 μM)	0.5 μM	1.25 μL
Locus-Specific Primer 1 (10 μM)	0.5 μM	1.25 μL
Single-cell cDNA	0.08 ng/μL	2 ng
2× Q5 master mix	1×	12.5 μL
ddH₂O	n/a	Variable
Total	n/a	25 μL

Note: Internal_forward (5′-AATGATACGGCGACCACCGAGATCTACAC-TCTTTCCCTA CACGACGCTC) completes Illumina Read 1 and adds partial Illumina P5 sequence to the 5′ end of the cDNA library.
    b. Mix by pipetting 10 times or gently vortexing, centrifuge briefly.
    c. Insert tubes into a thermocycler and incubator with the following protocol.

Locus-specific PCR 1

Steps	Temperature	Time	Cycles
Initial Denaturation	98°C	45 s	1
Denaturation	98°C	20 s	10 cycles
Annealing	67°C	30 s	
Extension	72°C	180 s	
Final extension	72°C	60 s	1
Hold	4°C	hold	

29. Remove primer-dimers using SPRIselect beads.

△ CRITICAL: SPRIselect beads should be acclimated to room temperature (20°C–22°C) before continuing.
   a. Add 25 μL nuclease-free water to each sample.
   b. Vortex to resuspend the SPRIselect reagent. Add 40 μL SPRIselect reagent (0.8×) to each sample and pipette mix (15×) or gently vortex mix.
   c. Incubate 5 min at room temperature (20°C–22°C).
   d. Place on the 10× magnetic separator (high orientation) until the solution clears.

   Note: The high orientation is used when the tube contains more than 50 μL of liquid and the low orientation is used when the tube contains less than 50 μL of liquid for easy liquid handling.
   e. Remove the supernatant.
   f. Add 200 μL 80% ethanol to the pellet. Wait 30 s.

△ CRITICAL: 80% ethanol should be prepared fresh for best wash performance and yields.
   g. Remove the ethanol.
   h. Repeat steps f and g for a total of 2 washes.
   i. Centrifuge briefly and place on the magnet (low orientation).
   j. Remove any remaining ethanol and air dry for 2 min to avoid ethanol carryover.
   k. Add 26 μL nuclease-free water. Pipette mix (15×) or gently vortex mix.
   l. Incubate 2 min at room temperature (20°C–22°C).
   m. Place the tube strip on the magnet (low orientation) until the solution clears.
   n. Transfer 25 μL sample to a new tube strip.
   o. Store at 4°C for up to 72 h or at −20°C for up to 4 weeks, or proceed to the next step immediately.

   ⸬ **Pause point:** The PCR product can be stored in a freezer at −20°C for 1–2 months or in a fridge at 4°C for 24 h until the following procedures.

30. Nested locus specific PCR 2.
   a. Prepare master mix in a 0.2 mL thin-wall PCR tube.
   b. Mix by pipetting 10 times or gently vortexing, centrifuge briefly.
   c. Insert tubes into a thermocycler and incubator with the following protocol.

Reagent	Final concentration	Amount
Short_Int_For (10 μM)	0.5 μM	1.25 μL
Locus-Specific Primer 2 (10 μM)	0.5 μM	1.25 μL
PCR product from PCR 1	n/a	5 μL
2× Q5 master mix	1×	12.5 μL
ddH$_2$O	n/a	5 μL
Total	n/a	25 μL

   Note: Short_Int_For consists of partial Illumina P5 (5′-AATGATACGGCGACCACCGAGATCT).

Locus-specific PCR 2			
Steps	Temperature	Time	Cycles
Initial Denaturation	98°C	45 s	1
Denaturation	98°C	20 s	10 cycles
Annealing	67°C	30 s	
Extension	72°C	180 s	
Final extension	72°C	60 s	1
Hold	4°C	hold	

31. Repeat step 30 to remove primer-dimers.

⏸ **Pause point:** The PCR product can be stored in a freezer at −20°C for 1–2 months or in a fridge at 4°C for 24 h until the following procedures.

32. Nested locus specific PCR 3.
    a. Prepare master mix in a 0.2 mL thin-wall PCR tube.
    b. Mix by pipetting 10 times or gently vortexing, centrifuge briefly.
    c. Insert tubes into a thermocycler and incubator with the following protocol.

Reagent	Final concentration	Amount
Short_Int_For (10 μM)	0.5 μM	1.25 μL
Locus-Specific Primer 3 (10 μM)	0.5 μM	1.25 μL
PCR product from PCR 2	n/a	5 μL
2× Q5 master mix	1×	12.5 μL
ddH₂O	n/a	5 μL
Total	n/a	25 μL

Locus-specific PCR 3			
Steps	Temperature	Time	Cycles
Initial Denaturation	98°C	45 s	1
Denaturation	98°C	20 s	10 cycles
Annealing	67°C	30 s	
Extension	72°C	180 s	
Final extension	72°C	60 s	1
Hold	4°C	hold	

33. Repeat step 30 to remove primer-dimers.

⏸ **Pause point:** The PCR product can be stored in a freezer at −20°C for 1–2 months or in a fridge at 4°C for 24 h until the following procedures.

34. Nested locus specific PCR 4.
    a. Prepare master mix in a 0.2 mL thin-wall PCR tube.
    b. Mix by pipetting 10 times or gently vortexing, centrifuge briefly.
    c. Insert tubes into a thermocycler and incubator with the following protocol.

Reagent	Final concentration	Amount
Short_Int_For (10 μM)	0.5 μM	1.25 μL
Locus-Specific Primer 4 (10 μM)	0.5 μM	1.25 μL
PCR product from PCR 3	n/a	5 μL
2× Q5 master mix	1×	12.5 μL
ddH₂O	n/a	5 μL
Total	n/a	25 μL

Note: Locus-Specific Primer 4 contains Illumina Read 2 (GTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCT) and 18–22 nucleotides from the 5′ end of locus-specific primer 3.

Locus-specific PCR 4

Steps	Temperature	Time	Cycles
Initial Denaturation	98°C	45 s	1
Denaturation	98°C	20 s	10 cycles
Annealing	67°C	30 s	
Extension	72°C	180 s	
Final extension	72°C	60 s	1
Hold	4°C	hold	

35. Repeat step 30 to remove primer-dimers.
36. Sample Index PCR.
    a. Prepare master mix in a 0.2 mL thin-wall PCR tube.
    b. Mix by pipetting 10 times or gently vortexing, centrifuge briefly.
    c. Insert tubes into a thermocycler and incubator with the following protocol.

Reagent	Final concentration	Amount
SI Primer	n/a	0.5 μL
Chromium i7 Sample Index	n/a	2.5 μL
PCR product from PCR 4	n/a	2 μL
10× Amp Mix	1×	12.5 μL
ddH$_2$O	n/a	7.5 μL
Total	n/a	25 μL

Note: SI Primer, Chromium i7 Sample Index, and 10× Amp Mix are included in the 10× Chromium 3′ kit. Record the i7 Sample Index for each library for downstream analysis.

Sample index PCR

Steps	Temperature	Time	Cycles
Initial Denaturation	98°C	45 s	1
Denaturation	98°C	20 s	10 cycles
Annealing	54°C	30 s	
Extension	72°C	180 s	
Final extension	72°C	60 s	1
Hold	4°C	hold	

37. Repeat step 30 to remove primer-dimers.
38. Quantify the locus-specific library using Tapestation.

⏸ **Pause point:** The PCR product can be stored in a freezer at −20°C for 1–2 months or in a fridge at 4°C for 24 h until the following procedures.

Note: High Sensitivity D5000 ScreenTapes and reagents are recommended dur to the library size. However, other Agilent ScreenTapes or Agilent Bioanalyzer reagents can be used for quantifying the libraries. If a TapeStation or Bioanalyzer is not available, quantification can be done by running a 2%–3% agarose gel with appropriate ladder. Quantification with NanoDrop and Qubit is not sufficient due to the lack of the fragment size information.

Table 2. Expected outcomes for each step

Outcome	Result
Post Ficoll recovery	MNC with low RBC and platelet contamination
Post CD34+ magnetic beads enrichment	0.2%–0.8% of the post-Ficoll cell count, 85%–95% viability on the post-column hemocytometer count
10× Genomics Chromium 3′ library construction	Single-cell cDNA, and single-cell transcriptome library
Locus-specific nested PCR	Specific somatic mutation enriched cDNA library
Bioinformatic processing	csv file for wildtype/mutation calling visualization on UMAP

39. Sequence the locus-specific library using the following cycles:
    a. Read 1 28 cycles
    b. I7 Index 8 cycles
    c. I5 index 0 cycles
    d. Read 2 91 cycles

Note: The locus-specific library can be pooled with the transcriptome library for sequencing or be sequenced alone. We sequence the pooled transcriptome and locus-specific libraries on a NovaSeq SP Flowcell (800 million reads) and the locus specific libraries are usually allocated with 2% of the total reads (16 million reads). The libraries can also be sequenced on MiSeq Flowcells for testing and troubleshooting. Please take into account the multiple peaks in fragment size distribution when calculating library concentration.

EXPECTED OUTCOMES

See Table 2 for step-by-step expected outcomes.

QUANTIFICATION AND STATISTICAL ANALYSIS

⏱ Timing: Variable, depending on processing power

This section describes the usage and logic of the bioinformatic script for analysis of the single-cell amplicon library sequences. The script maps the somatic mutation of interest to 10× single-cell barcodes using MATLAB. The bioinformatic analysis scripts are available at https://gitlab.com/hormozlab/scamplicon-library-analysis.

1. The computational pipeline consists of 1. the bioinformatic analysis MATLAB script and 2. two fastq files (read 1 and read 2) of locus specific amplicon libraries
2. Move the MATLAB script to the folder containing fastq files of locus specific amplicon libraries.
3. Change 'GeneName' and 'filename' to match R1, and R2 fastq files of the library.
4. Change 'corrGene2' to th sequence 91 bp downstream of the locus-specific primer 3 including the primer sequence.
5. Change 'posMut' (denotes the position of the mutation from primer 3) and 'baseMut' (denotes the nucleotide of the mutation) to appropriate values. For JAK2V617F, posMut = 61 and baseMut = 'T'.
6. Optional: Change 'barcode_inputfile' to match with barcode tsv file from 10× Chromium computational pipeline.
7. Run the MATLAB script. We recommend running the script on a server with at least 100 GB of RAM for 12 h. The script will do the following:
    a. Parse Read 1 (containing single-cell barcodes and unique molecular identifier, UMI), and Read 2 (containing mRNA information) along with their QC score and index into arrays.
    b. Filter out reads with quality control (QC) score lower than 30 and calculate fraction of reads that passed QC threshold.
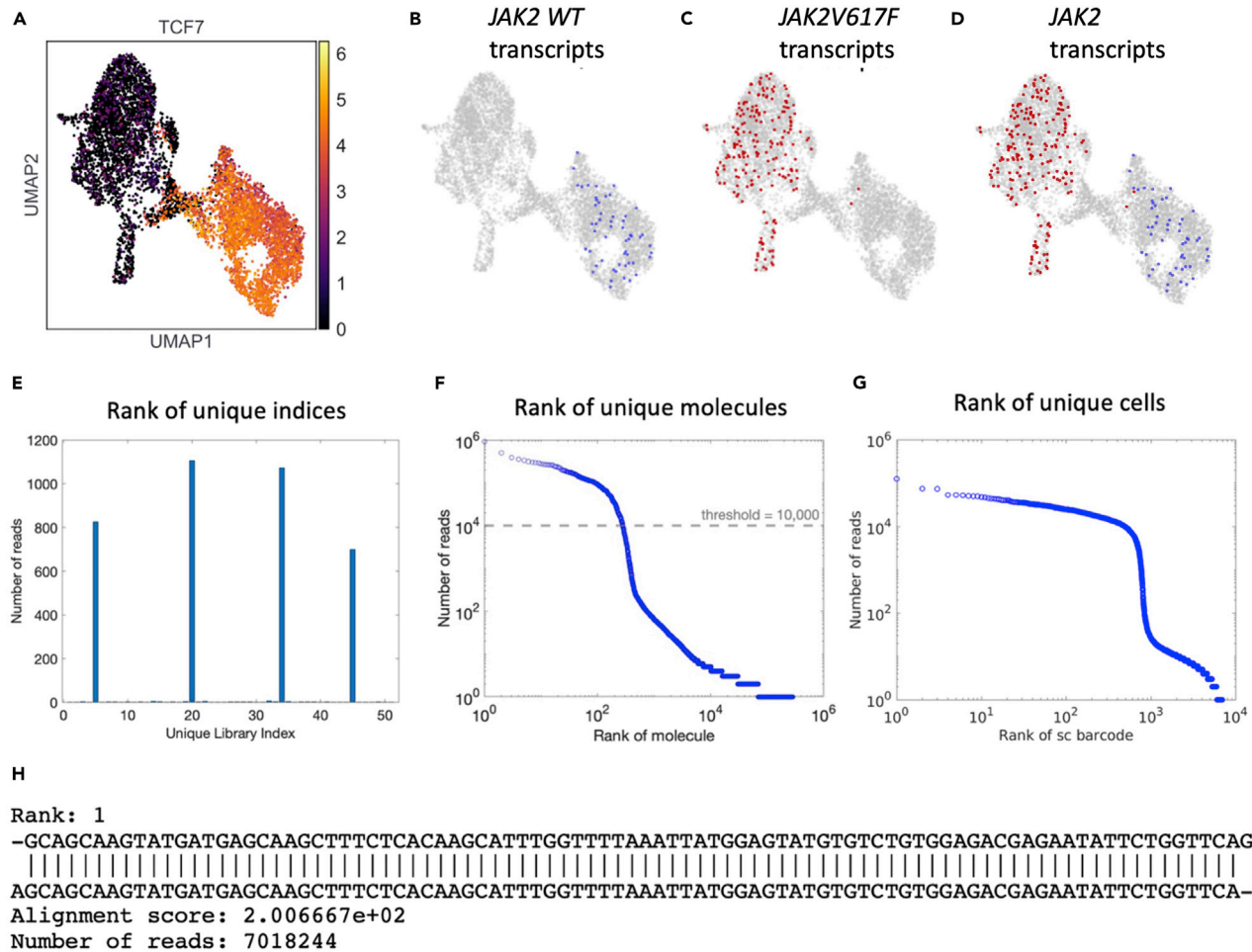
Figure 4. Accurate identification of the mutated cells from the amplicon libraries

(A–C) (A) In a control experiment MOLT4 (WT cells) were mixed with UKE-1 cells (homozygous JAK2-V617F mutation) and ran through the experimental and analysis pipeline. The two cell populations could be distinguished based on their transcriptional profiles: two distinct clusters were seen when transcriptomes of the cells were visualized using UMAP. Marker genes (TCF7 shown here) were used to identify the clusters as either MOLT4 or UKE-1 cells. Cells in which a mutated JAK2 transcript (B) or a WT JAK2 transcript (C) were detected in the amplicon libraries are shown as colored points. All other cells are shown in gray.

(D) All cells in which either a WT or mutated JAK2 transcript was detected in the amplicon libraries. JAK2 transcript were detected in ~ 4% of cells (249 out of 6563 cells). The rate of erroneously detecting a mutated transcript in a MOLT4 cell or a wildtype transcript in a UKE-1 cell in less than 1%.

(E–G) Output plots from MATLAB analysis script. (E) Rank of unique indices. Index sequence can be found in MATLAB cell array 'uniqueindices'. (F) Number of reads vs rank of unqiue molecules and the threshold for calling the detected molecules as either wildtype or mutated. (G) Number of reads vs rank of unique cells. (H) Example of top 200 most common Read 2 and its align results. Related to Figure 1

Note: The expected output at this step are the sequencing reads with QC scores higher or equal to 30.

c. Count and plot the most frequent unique indices (Figure 4E).

d. Extract 11 bp around the mutated nucleotide with wildtype and mutated nucleotide substitutions for use in step m.

e. Extract 10× cell barcodes from Read 1 and either use: Option 1, raw barcode from the sequencing files or Option 2, Collapse the extracted raw barcodes to the list of barcodes generated by the 10× Chromium computational pipeline. If Option 2 is used, the script will do the following:

   i. Read and extract barcodes from 10× barcode tsv file.

ii. Compare raw barcode to 10× barcode and compute distance between the two strings.
iii. Collapse the raw barcode to the 10× barcode if the distance is equal to or smaller than 2 for barcode correction.

f. Calculate the occurrences of each cell by counting the single-cell barcodes.

g. Calculate the occurrences of each unique molecule by counting the unique-molecular-identifier (UMI).

h. Plot the ranked ordered number of reads detected for each unique molecular identifier (Figure 4F), and the ranked ordered number of reads detected in each cell (Figure 4G).

   Note: In a successfully amplified library with 5–20 million reads generated from ~10,000 cells, the number of Jak2 molecules with more than 1,000 reads should range from 200–500.

i. Calculate the occurrences of each unique Read 2 sequence and plot the occurrences of each unique Read 2 versus the corresponding number of reads.

j. Align top 200 most common Read 2 sequences to 'corrGene2' and print the alignment results and the number of reads to a txt file. (Figure 4H)

   Note: In a successfully amplified library, the top two reads should align perfectly (wild type) or have one mismatch (mutated) compared to the 'corrGene2'.

k. For each unique cell barcode, go through every Read 2 sequence for each unique UMI. Record and count the number of Read 2 sequences aligned to 'corrGene2' with an alignment score greater or equal to 150.

l. For each unique cell barcode, go through all UMIs detected in that cell. Calculate the distance between the selected UMI and the most common UMI (UMI with highest number of reads). If the difference between the two UMI sequences is smaller or equal to two, merge UMI with the most common UMI and remove the redundant UMI. Repeat with the second most common UMI and so on until no further mergers are possible for UMI correction.

m. For each unique cell barcode, align every Read 2 sequences with wildtype and mutated sequence for each unique UMI.

n. Call a UMI as either wildtype or mutated when the number of wildtype or mutated Read 2 sequences exceed 50% of the total number of reads for that UMI.

o. Store the UMI information and its wildtype/mutation call if the UMI contains more reads than a predetermined threshold.

   Note: The thresholds are typically picked by selecting the middle point of the "knee" in number of reads versus the rank (occurrence) of molecules plot (see Figure 4F).

p. Store mutation calling information in a csv file by assigning two columns to each cell barcode where the first column indicates the number of wildtype transcripts detected in that cell and the second columns the number of mutated transcripts.

8. Visualize wildtype and mutated population on UMAP.

LIMITATIONS

Here, we presented a protocol to apply amplicon sequencing to single-cell cDNA. A disadvantage of this protocol is the high number (50) of PCR cycles required to successfully amplify specific somatic mutations. For lowly expressed genes, even larger amplification needs to be performed which presents a significant challenge in terms of PCR recombination and crossover.

TROUBLESHOOTING

Problem 1

Inadequate separation of layers during Ficoll/Lymphoprep isolation (steps 4 and 5).

Potential solution

Make sure that the acceleration and brake settings of the centrifuge are set correctly. At high speed, the phases might mix and disturb the gradient. Make sure that the amount of blood per tube does not exceed the recommendations. The Ficoll/Lymphoprep solution should be at room temperature (20°C–22°C).

Problem 2

Too much clumping of the cells during the separation. (steps 12, 20, and 23)

Potential solution

Make sure you use the EasySep buffer supplied from NEB. The EasySep buffer contains EDTA which will reduce the clumping.

Problem 3

Low cell viability. (steps 11, 25, and 26)

Potential solution

Spin the cells at low RPM (120 rpm) for 2–3 min. The centrifugation is enough to pellet live cells, but most dead cells will still be in suspension. Make sure to always leave the cells on ice.

Problem 4

No/low peak below 700 bp after locus-specific amplicon library generation. (steps 25–39)

Potential solution

Perform a Tapestation measurement after each PCR. For lowly expressed genes, add 3–5 cycles to the first locus-specific PCR. If the Tapestation result is still not satisfactory, add 3–5 cycles to the subsequent cycles.

Problem 5

Overclustering during amplicon library sequencing. (step 40)

Potential solution

If the Tapestation trace of the library shows multiple peaks, it is difficult to accurately calculate the molar concentration of the library after qPCR or qubit quantification. We generally assume that at least half of the reads generated in the fastq files come from library strands smaller than 700 bp. It is also important to mix in > 10% of PhiX control to identify the cause of the overclustering and decrease the homogeneity of the amplicon libraries.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Sahand Hormoz (sahand_hormoz@hms.harvard.edu).

Materials availability

Oligonucleotides could be purchased from manufacturers.

Data and code availability

Amplicon analysis code is available at https://gitlab.com/hormozlab/scamplicon-library-analysis.

AUTHOR CONTRIBUTIONS

S.L., M.N., and S.H. developed and optimized the experimental and computational pipeline. S.L., and M.N. performed experiments. S.H. and S.L. analyzed and interpreted the data. S.L. and S.H. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Van Egeren, D., Escabi, J., Nguyen, M., Liu, S., Reilly, C., Patel, S., Kamaz, B., Kalyva, M., DeAngelo, D., Galinsky, I., et al. (2021). Reconstructing the lineage histories and differentiation trajectories of individual cancer cells in myeloproliferative neoplasms. Cell Stem Cell *28*, 514–523.e9.