Journal of Visualized Experiments

# Characterizing mutational load and clonal composition of human blood

--Manuscript Draft--

| | |
|---|---|
| **Article Type:** | Invited Methods Article - JoVE Produced Video |
| **Manuscript Number:** | JoVE59846R1 |
| **Full Title:** | Characterizing mutational load and clonal composition of human blood |
| **Keywords:** | hematopoiesis; Clonality; Somatic mutations; development; Lineage tracing; Mutational signatures; Leukemia |
| **Corresponding Author:** | Ruben van Boxtel<br><br>NETHERLANDS |
| **Corresponding Author's Institution:** | |
| **Corresponding Author E-Mail:** | R.vanBoxtel@prinsesmaximacentrum.nl |
| **Order of Authors:** | Ruben van Boxtel |
| | Axel Rosendahl Huber |
| | Freek Manders |
| | Rurika Oka |
| **Additional Information:** | |
| **Question** | **Response** |
| Please indicate whether this article will be Standard Access or Open Access. | Open Access (US$4,200) |
| Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations. | Utrecht, Utrecht, The Netherlands |

**TITLE:**

# Characterizing Mutational Load and Clonal Composition of Human Blood

**AUTHORS AND AFFILIATIONS:**

Axel Rosendahl Huber[1], Freek Manders[1], Rurika Oka[1], Ruben van Boxtel[1]

[1]Princess Máxima Center for Pediatric Oncology, Utrecht, the Netherlands

**Corresponding Author:**

Ruben van Boxtel (R.vanboxtel@prinsesmaximacentrum.nl)

**Email Addresses of Co-Authors:**

Axel Rosendahl Huber (A.K.M.RosendahlHuber@prinsesmaximacentrum.nl)
Freek Manders (F.M.Manders@prinsesmaximacentrum.nl)
Rurika Oka (R.Oka@prinsesmaximacentrum.nl)

**KEYWORDS:**

hematopoiesis, clonality, somatic mutations, development, lineage tracing, mutational signatures, leukemia

**SUMMARY:**

Somatic mutation patterns in cells reflect previous mutagenic exposure and can reveal developmental lineage relationships. Presented here is a methodology to catalogue and analyze somatic mutations in individual hematopoietic stem and progenitor cells.

**ABSTRACT:**

Hematopoietic stem and progenitor cells (HSPCs) gradually accumulate DNA mutations during a lifespan, which can contribute to age-associated diseases such as leukemia. Characterizing mutation accumulation can improve understanding of the etiology of age-associated diseases. Presented here is a method to catalogue somatic mutations in individual HSPCs, which is based on whole-genome sequencing (WGS) of clonal primary cell cultures. Mutations that are present in the original cell are shared by all cells in the clonal culture, whereas mutations acquired in vitro after cell sorting are present in a subset of cells. Therefore, this method allows for accurate detection of somatic mutations present in the genomes of individual HSPCs, which accumulate during life. These catalogues of somatic mutations can provide valuable insights into mutational processes active in the hematopoietic tissue and how these processes contribute to leukemogenesis. In addition, by assessing somatic mutations that are shared between multiple HSPCs of the same individual, clonal lineage relationships and population dynamics of blood populations can be determined. As this approach relies on in vitro expansion of single cells, the method is limited to hematopoietic cells with sufficient replicative potential.

## INTRODUCTION:

Exposure of hematopoietic stem and progenitor cells (HSPCs) to endogenous or extrinsic mutagenic sources contributes to the gradual accumulation of mutations in the DNA during a

lifespan[1]. Gradual mutation accumulation in HSPCs[1] can result in age-related clonal hematopoiesis (ARCH)[2,3], which is a non-symptomatic condition driven by HSPCs carrying leukemia-driver mutations. Initially, it was thought that individuals with ARCH have an increased risk for leukemia[2,3]. However, recent studies have shown an incidence of 95% of ARCH in elderly individuals[4], making the association with malignancies less clear and raising the question of why some individuals with ARCH eventually do or do not develop malignancies. Nonetheless, somatic mutations in HSPCs can pose serious health risks, as myelodysplastic disorders and leukemia are characterized by the presence of specific cancer driver mutations.

To identify the mutational processes and study blood clonality, mutation accumulation in individual HSPCs needs to be characterized. Mutational processes leave characteristic patterns in the genome, so-called mutational signatures, which can be identified and quantified in genome-wide collections of mutations[5]. For instance, exposure to UV light, alkylating agents, and defects in DNA repair pathways have each been associated with a different mutational signature[6,7]. In addition, due the stochastic nature of mutation accumulations, most (if not all) of the acquired mutations are unique between cells. If mutations are shared between multiple cells of the same individual, it indicates that these cells share a common ancestor[8]. Therefore, by assessing shared mutations, lineage relationships can be determined between cells and a developmental lineage tree can be constructed branch by branch. However, cataloguing rare somatic mutations in physiologically normal cells is technically challenging due to the polyclonal nature of healthy tissues.

Presented here is a method to accurately identify and determine somatic mutations in the genomes of individual HSPCs. This involves the isolation and clonal expansion of HSPCs in vitro. These clonal cultures reflect the genetic makeup of the original cell (i.e., mutations in the original cell will be shared by all other cells in the culture). This approach allows us to obtain sufficient DNA for whole genome sequencing (WGS). We have previously shown that mutations accumulated in vitro during clonal culture will be shared by a subset of cells. This enables the filtering of all in vitro mutations, as these will be present in a smaller fraction of reads compared to in vivo acquired mutations[9]. Previous methods have obtained sufficient DNA from a single cell for WGS using whole-genome amplification (WGA)[10]. However, the main disadvantage of WGA is its relatively error-prone and unbalanced amplification of the genome, which can result in allele dropouts[11]. Nonetheless, as this approach relies on in vitro expansion of single cells, it is limited to blood cells with sufficient replicative potential, which is not the case for WGA-dependent methods. Earlier efforts sequencing clonal cultures have relied on using feeder layers to ensure clonal amplification of single HSPCs[12]. However, DNA from the feeder layers can potentially contaminate the DNA of the clonal cultures, confounding the subsequent mutation calling and filtering. The method presented here solely relies on specified medium to clonally expand single HSPCs, and therefore avoids the issue of DNA contamination. Until now, we have successfully applied this method on human bone marrow, cord blood, viably frozen bone marrow, and peripheral blood.

**PROTOCOL:**

Samples must be obtained in accordance with appropriate ethics protocols, and donors must give informed consent prior to the procedure.

**1. Preparation of sample material**

NOTE: When working with freshly obtained material, start with step 1.1. When working with frozen material, start with step 1.2.

1.1. Preparing fresh bone marrow, cord blood, or peripheral blood

1.1.1. Isolate the mononuclear fraction from the sample using density gradient separation by following the manufacturers' instructions (see **Table of Materials**), and count the mononuclear cells using a hemocytometer. After isolation of the mononuclear cells, continue with step 1.3.

1.1.2. OPTIONAL: The recommended number of cells required to sort a full 384 well plate of HSPCs is 1–2 x $10^7$. If more cells are isolated during density gradient centrifugation, store the surplus of cells in liquid nitrogen.

1.1.3. Resuspend cells in 500 µL of IMDM + 10% FBS per 1 x $10^7$ cells, and add drop-by-drop an equal volume of IMDM + 30% FBS + 20% DMSO to achieve a suspension of 1 x $10^7$ cells in 1 mL of IMDM + 20% FBS + 10% DMSO.

1.1.4. Immediately transfer the mononuclear cells to 1 mL cryogenic vials and freeze cells at -80 °C in a controlled-rate cell freezing container overnight. Transfer the cells the next day to a liquid nitrogen storage upon further processing.

1.2. Preparing frozen mononuclear cells from bone marrow, cord blood, or peripheral blood

1.2.1. Prepare 50 mL of cell thawing medium containing 45 mL of Iscove's Modified Eagle's Medium (IMDM) and 5 mL of fetal bovine serum (FBS), and warm in 37 °C water bath.

1.2.2. Take the vial containing the sample from liquid nitrogen storage, transfer the sample to dry ice, and thaw as quickly as possible in a 37 °C water bath.

1.2.3. When the sample is almost thawed, wipe the vial with 70% ethanol and transfer its contents to a 50 mL conical tube. Rinse the vial with 1 mL of pre-warmed IMDM + 10% FBS to collect the remaining cells, and add this solution dropwise (5 s per drop) to the thawed sample while gently swirling the tube.

1.2.4. Add an additional pre-warmed 15 mL of IMDM + 10% FBS dropwise to the sample while gently swirling the tube.

1.2.5. Pellet the cells by centrifugation for 5 min at 350 x *g*.

1.2.6. Remove all but ±3 mL of the supernatant. Resuspend the cells in the remaining supernatant and dilute by adding 20 mL of IMDM + 10% FBS drop-by-drop while gently shaking the tube.

1.2.7. Take 10 µL of the cell suspension for cell counting. Dilute these 10 µL by adding 20 µL of 0.4% trypan blue solution and count the cells using a hemocytometer. The cell number can decrease upon thawing, with up to 50% cell loss after thawing. Cell viability should range between 70% and 90%.

1.3. If working with bone marrow or umbilical cord blood cells, take up 5 x $10^6$ mononuclear cells for MSC culture (step 2.1). If working with peripheral blood, take up 2–5 x $10^6$ cells for T-cell isolation (step 2.2)

1.4. Pellet remaining cells 5 min at 350 x *g* and resuspend in 3 mL of FACS buffer (0.05% BSA + 1 mM EDTA in PBS).

1.5. Transfer 1 x $10^5$ cells to a microtube filled with 200 µL of FACS buffer, which will serve as a negative control for flow cytometry (step 3.8), and keep on ice.

## 2. Cell culture

NOTE: To obtain catalogues of somatically acquired mutations, donor-specific germline variation needs to be filtered out. When starting with bone marrow biopsies or umbilical cord blood, mesenchymal stromal cells (MSCs) can be used as matched control to filter for germline variation. In this case, follow section 2.1. When using (mobilized) peripheral blood follow step 2.2 to isolate and use T-cells as matched control sample to filter for germline variation (**Figure 1**). The bulk T-cell population will share the same lineage relationship as HSPCs.

### 2.1. **MSC culture**

2.1.1. Prepare 50 mL of MSC medium containing 45 mL of DMEM/F12 medium, 10% FBS, 500 µL of trypsin or trypsin alternative, and 500 µL of penicillin/streptomycin solution.

2.1.2. Plate approximately 5 x $10^5$ mononuclear cells in 1.5 mL of MSC medium per well. Place the cells in a humidified incubator at 37 °C with 5% $CO_2$.

2.1.3. Replace the medium after 24 h, and subsequently replace medium every 3 days to ensure that all hematopoietic cells are washed off. Continue to culture until the confluency is 100%.

2.1.4. If the MSCs are confluent, wash cells with 1 mL of PBS and harvest the MSCs by adding 200 µL of trypsin or trypsin alternative per well. Incubate cells for 5 min at 37 °C. Add 800 µL of MSC medium, and pipet the cells up and down to loosen cells from the well plate.

2.1.5. Transfer MSCs to microcentrifuge tube and pellet the cells by centrifugation for 5 min at 350 x *g*. Remove the supernatant and continue with DNA isolation or store the pellet at -20 °C for later DNA isolation (section 4).

2.2. **T-cell isolation**

NOTE: If using (mobilized) peripheral blood, T-cells can be isolated and used as germline control.

2.2.1. Resuspend the cell pellet in 100 µL of anti-CD3 staining solution (1:100 dilution of anti-CD antibody in FACS buffer).

2.2.2. Wash the cells by adding 1 mL of FACS buffer. Pellet the cells by centrifugation for 5 min at 350 x *g* and resuspend in 300 µL of FACS buffer.

2.2.3. Isolate at least 5 x $10^5$ CD3+ cells using a FACS-sorter in a 5 mL polystyrene tube pre-filled with 1 mL of FBS.

2.2.4. Pellet the sorted cells using centrifugation for 5 min at 350 x *g*, remove the supernatant, and continue directly with DNA isolation (section 4) or store the pellet at -20 °C for later DNA isolation.

## 3. HSPC isolation, sorting, and culture

3.1. Spin down at 1–2 x $10^7$ mononuclear cells for 5 min at 350 x *g* and resuspend in 50 µL of FACS buffer (see step 2.2.1). Transfer the cells to a microcentrifuge tube.

NOTE: When sorting with >2 x $10^7$ cells, increase the antibody mix and FACS buffer volumes accordingly.

3.2. Prepare a 50 µL of 2x HSC staining mix according to the recipe seen in **Table 1**.

[Place Table 1 here]

3.3. Mix 50 µL of cell solution with the prepared HSC staining mix and incubate the cells for 15 min at room temperature (RT) or for 1 h on ice for the antibodies to bind.

3.4. Wash the cells by adding 1 mL of FACS buffer and pellet by centrifuging for 5 min at 350 x *g*.

3.5. Resuspend the cells in 300 µL of FACS buffer and filter the cell suspension through a 35 µm cell strainer-capped 5 mL polystyrene tube to remove cell clumps before fluorescence-activated cell sorting (FACS).

3.6. Prepare 25 mL of HSPC culture medium, consisting of 1x SFEM medium supplemented with 100 ng/mL SCF, 100 ng/mL Flt3, 50 ng/mL TPO, 10 ng/mL IL-3, 20 ng/mL IL- 6, and 100 ng/mL antibiotic formulation (see **Table of Materials**).

3.7. Fill a 384 well cell culture plate with 75 µL of HSPC culture medium in each well.

NOTE: To prevent evaporation of the medium in the outer wells, fill the outer wells with 75 µL of sterile water or PBS, and do not use these wells for cell sorting.

3.8. Sorting single HSPCs

3.8.1. Set gates for the HSPC sorting based on an unstained control (step 1.9) and 10,000 cells from the stained sample. A representative result for setting gates is depicted in **Figure 1**. Gate single cells by drawing a gate around the linear FSC-height vs. FSC-area fraction. Use unstained control fraction to draw gate for lineage$^-$ fraction. Draw gates for CD34+ cells and further characterize this subset by setting a specific gate for CD38- CD45RA- cells.

3.8.2. Load the 384 well plate on the FACS machine and sort single cells.

NOTE: If applicable to the FACS-machine, toggle on the option to keep index sorting data to enable re-tracing of the sorted cells.

3.9. Culturing singly-sorted HSCs

3.9.1. Directly transfer the 384 well plate to a humidified 37 °C incubator with 5% $CO_2$.

NOTE: To prevent evaporation during culture, wrap the 384 well culture plate (with lid) in transparent polyethylene wrap.

3.9.2. Keep the 384 well plate in the incubator for 3–4 weeks until visible clones appear. Representative images of clonal culture are depicted in **Figure 2**. Based on the condition of the input material 5%–30% of sorted cells will clonally expand.

**4. Harvesting HSPC clones**

4.1. After 4 weeks of culturing, determine which wells have a confluency of 30% or higher.

4.2. Pre-fill (for each clonal outgrowth) 1.5 mL microtubes with 1 mL of 1% BSA in PBS and label the tube according to the corresponding well.

4.3. Pre-wet a pipette tip with 1% BSA in PBS to minimize the number of cells sticking to the pipette tip.

4.4. Pipet up/down the medium in the well fiercely (at least 5 times) with a 200 µL pipette (set at 75 µL) and scrape the bottom of the well to loosen cells in the well, and collect the cell suspension in the labeled microtube corresponding to the well.

4.5. Take up 75 µL of fresh 1% BSA in PBS and repeat pipetting in the well to ensure maximum uptake of cells.

NOTE: Clonally cultured cells can stick to bottom of well. Inspect the wells using a standard inverted light microscope to ensure whether all cells have been collected.

4.6. If all wells with >30% confluency have been harvested, place the 384 well plate back in incubator. Clonal cultures can proliferate for up to 5 weeks.

4.7. Spin down the cell suspension for 5 min at 350 x *g*. A small pellet should be visible.

4.8. Carefully remove all but about 5 µL of supernatant. Cell pellets can be frozen at -20 °C and stored for multiple months before DNA isolation.

**5. DNA isolation**

5.1. Isolate HSPC and MSC/T-cell DNA using a micro-scale DNA isolation kit according to the manufacturer's instruction with the following adjustments:

5.1.1. Add 2 µL of RNase A after addition of buffer AL during section 2. Incubate for 2 min before adding proteinase K.

5.1.2. Incubate for 30 min at 56 °C instead of 10 min.

5.1.3. Elute the DNA by loading the column with 50 µL of TE buffer with low EDTA (10 mM Tris, 0.1 mM EDTA). For optimal elution, reload the eluate again on the column and spin again.

5.2. Determine the DNA concentration using a DNA measuring 2 µL per clone. The DNA yield typically varies between 0.5–3 ng/µL.

**6. Sequencing**

6.1. Perform DNA sequencing as described by Jager et al.[13]

**7. Mapping and somatic mutation calling**

7.1. Map the output of sequencing (FASTQ files) to the reference genome and call mutations as described in Jager et al.[13]

7.2. Inspect the data for aberrant karyotypic changes in sequenced clone and bulk data using a copy number analysis tool, such as Control-FreeC[14]. Until now, we have not reported any HSPCs with karyotypic aberrances.

7.3. Generate a blacklist, which consists of a panel of unmatched normal samples for filtering purposes from an own set of samples, as previously described[13], or use the following uploaded blacklist: <https://data.mendeley.com/datasets/9y4yhwt5rp/draft?a=6ce9b571-187d-4cf9-b9af-7082c9f3652e>.

7.4. Filter single nucleotide variations using SNVFI <https://github.com/ToolsVanBox/SNVFI>.

7.4.1. Preset SNVFI.config file, such that all paths to helper functions are correct.

7.4.2. Run SNVFI with .ini file configured according to the settings seen in **Supplemental File 1** (SNVFI.ini). To exclude in vitro induced mutations we filter for a VAF ≥0.3[9].

7.5. Check the variant allele fraction (VAF) output of SNVFI (**Figure 4**). Check whether the peak of the density plot is near 0.5, indicating the sample is clonal.

7.6. OPTIONAL: To determine the part of the genome which is covered during filtering, determine the callable regions along germline and control using CallableLoci from GATK (Genome Analyziz ToolKit):

```
java -jar GenomeAnalysisTK.jar \
        -T CallableLoci \
        -R reference.fasta \
        -I myreads.bam \
        -summary table.txt \
        -o callable_status.bed
```

7.7. OPTIONAL: Retrieve the Callable regions from the CallableLoci output and perform pairwise intersections between the samples and the bulk using the python script CallableLoci_processor.py present at <https://data.mendeley.com/datasets/9y4yhwt5rp/draft#folder-13e1f141-12eb-4b05-85c4-d0b198a4ba9b>. The resulting bed files can be used to further filter the output of SNVFI and to inspect the mutational profile in section 9:

```
CallableLoci_processor.py dir_in dir_out sample_name bulk_name –samples sample1 sample2 sample3
```

**8. Indel calling**

8.1. Select all insertions and deletions (Indels) in the raw_variants.vcf file using GATK SelectVariants:

```
java -Xmx12G \
        -jar GenomeAnalysisTK.jar \
```

**-T SelectVariants **
**-R reference_genome.fasta **
**-V raw_variants.vcf **
**-o raw_INDELs.vcf **
**-selectType INDEL**

8.2. Filter raw_INDELS.vcf list using INDELFI <https://github.com/ToolsVanBox/INDELFI>:
**perl INDELFI.pl -i input.vcf (from step 8.1) **
**-s column test sample **
**-c column control sample**

**9. Mutational profile inspection**

9.1. Use the resulting .vcf files from SNVFI output from step 7.6 (or from step 7.9 with optional callable loci analysis) to analyze the genomic mutational profile, mutation types, and signature analysis using the R package MutationalPatterns[15]: <http://bioconductor.org/packages/release/bioc/html/MutationalPatterns.html>. For representative output that can be produced with the resulting .vcf file such as a 96-trinucleotide mutational spectrum, see **Figure 5**.

**10. Construction of a developmental lineage tree using base substitutions**

10.1. To construct a developmental lineage tree, detect shared mutations between clones. Mutations present in the first branches of the lineage tree can also be subclonally present in the bulk sample (MSCs/T-cells). Later branching lineages will be defined by mutations shared by HSPC clones only.

10.2. To identify mutations that are present in a subset of the clones and subclonally present in the bulk, perform the following steps.

10.3. In order to filter for somatic mutations shared between clones, run the filterSomatic.py script in a Unix-based terminal. The script can be found at <https://data.mendeley.com/datasets/9y4yhwt5rp/draft#folder-13e1f141-12eb-4b05-85c4-d0b198a4ba9b>. Before running this script, edit the filterSomatic.ini file (see **Supplemental File 2**) to set the paths and adjust the other parameters.

10.4. Run filterSomatic.py (**python3 filterSomatic.py -i filterSomatic.ini**).

10.5. Filter for mutations that are subclonally present in the bulk using the Determine_lowVAF_bulk.R script in a Unix-based terminal. The script can be found at <https://data.mendeley.com/datasets/9y4yhwt5rp/draft#folder-13e1f141-12eb-4b05-85c4-d0b198a4ba9b>. This will generate separate .vcf files for shared and unique SNVs:
**Rscript Determine_lowVAF_bulk.R**
**--vcf Path/To/Filter_somatic_output.vcf**

```
        --bulk bulk_name
        --sample_name sample-name
        --gender [M|F]
        --out_dir out_dir
```

10.6. Determine all mutations shared between clones but are not present in the bulk sample by overlapping all mutation positions (concatenate column 1 and 2 of SNVFI output).

10.7. Exclude false positives obtained during steps 10.5 and 10.6 by manual inspection using IGV[16]. Mutations are considered false when not present, when the mutation is present in the germline or when present in poorly mapped regions, see **Figure 7**.

NOTE: We highly recommend to re-sequence all shared loci independently using targeted or sanger sequencing.

10.8. Use the shared mutations obtained during steps 10.1 and 10.2 to build a binary table of mutations versus sequenced clones, with 0 indicating that the mutation is not present and 1 indicating presence of the mutation.

10.9. Output the mutation binary table as in a heatmap together with a dendrogram indicating lineage relationships between cells using R. The heatmap indicates mutations status for each cell. See the output of this function (**Figure 6**).

```
Clones <- read.table("Path/To/BinaryTable")

my_palette <- colorRampPalette(c("#cccccc", "#333333"))(n = 2)
col_breaks <- c(0,0.5,1)

heatmap.2(clones, distfun=function(x) dist(x,method = 'binary'),
    hclustfun=function(x) hclust(x,method = average),
    dendrogram = "column", Rowv = F,
    col=my_palette, breaks=col_breaks,
    trace="none", density.info="none")
```

**REPRESENTATIVE RESULTS:**

**Experimental procedure**

The experimental workflow is depicted in **Figure 1**. Based on the type of input material, different steps must be followed. In **Figure 2** a flow cytometric output of a cord blood cell sort is depicted. First, all monocytic cells are selected by loosely drawing a gate around this population. Then, singlets are isolated by selecting for cells with a linear FSC-H/FSC-A ratio, as a lower FSC-H/FSC-A ratio includes doublets or cell clumps. The unstained control sample is used to define cell sorting gates for lineage$^-$, CD34$^+$, CD38$^-$, CD45RA$^-$. Additionally, CD90 and CD49f can be used to distinguish between progenitor cells or self-renewing stem cells[17] (**Figure 2**). Index sorting

enables the re-tracing of individual cells, and the sorted cells are depicted as brown dots. During cell culture, individual clones can expand at a different pace, with some clones expanding within 3 weeks, while other clones are only fully expanded until the fifth week of culture. See **Figure 3A,B** for representative colony outgrowth. A representative picture is shown of a nearly confluent MSC bulk culture at 11 days after plating (**Figure 3C**).

**Checking quality after sequencing and mutation analysis**

Shown is an example output of the copy number analysis generated by Control-FreeC[14] to check for copy number alterations (**Figure 4**). Karyotypic information can indicate which chromosomes to exclude during a SNVFI run (step 7.6). The VAF plot created by SNVFI (**Figure 5**) is a histogram of variant allele frequencies in the sample. A peak in the density plot at 0.5 indicates the sample is clonal. To get more insight in the underlying biological causes behind mutations, these can be analyzed using the R package MutationalPatterns[15]. Depicted here is a typical analysis producing a 96-trinucleotide plot (**Figure 6**). In addition to quantification of different mutation types, signature extraction can be also performed with this tool.

**Constructing a developmental lineage tree**

Mutations shared amongst clones or present in a clone (and at low VAF) in the germline control are validated using IGV. Mutations are considered true when present in the sample and not at high VAF levels in the germline (**Figure 7A**). Mutations are considered false when not present in IGV, which can happen in poorly mapped regions (**Figure 7B**). In other cases, events detected by SNVFI are missed germline mutations (**Figure 7C**). Independent re-sequencing of mutations by targeted re-sequencing is highly recommended for these mutations in selected clones. After detection of shared somatic mutations between clones, a binary matrix is generated (step 10.8). A heatmap is constructed containing cells with and without the shared mutations A-M. Above this heatmap the developmental lineage tree is indicated (**Figure 8**).

**FIGURE AND TABLE LEGENDS:**

**Figure 1: Flowchart depicting experimental procedure based on input material.**

**Figure 2: Cell sorting strategy.** First, gating is performed on small mononuclear cells. Second, single cells are gated by selection of the linear fraction. Lineage negative cells are gated. All CD34+ CD38- CD45- cells are single cell-sorted. The fraction of cells in brown should be noted, which are the sorted cells highlighted by the option “index sorting”.

**Figure 3: Representative cell culture results.** Representative HSPC clones in a 384 well plate at (**A**) 2 weeks after plating and (**B**) 4 weeks after plating. (**C**) MSC culture after 2 weeks of medium replacement. Scale bar = 100 µm.

**Figure 4: Karyotypes.** (**A**) Clonal HSPC culture and (**B**) MSC bulk sample. The karyotypes were determined by read-depth analysis. Both graphs indicate a karyotypically normal sample.

**Figure 5: Histogram of variant allele frequencies.** Histogram of variant allele frequencies of the

variants in a clone before the last filtering step of SNVFI (VAF >0.3). A peak at VAF = 0.5 indicates that the sample is clonal. The subclonal mutations with low VAF are excluded during last filtering step of SNVFI (VAF >0.3).

**Figure 6: Representative mutational spectrum analysis of somatic mutations in a HSPC sample.** Depicted is the relative contribution of each trinucleotide change (of which the middle base is mutated) to the total spectrum.

**Figure 7: Manual inspection of mutations using IGV[16]. (A)** Mutations are considered true when present in the clone and not in the bulk sample. (**B)** Mutations are considered as false positives when present in a poorly mapped region. (**C)** Mutations are considered as false positives when present in a germline control. The vertical line indicates the position of a called mutation.

**Figure 8: Construction of a developmental lineage tree.** Depicted is a dendrogram indicating developmental lineages splitting off during development. The heatmap under the dendrogram indicates the presence of mutations in different clones.

**Table 1: HSC sorting mix.** Shown is a table indicating the dilutions of antibodies used to sort the HSCs.

**DISCUSSION:**

Presented here is a method to detect mutations that accumulated during life in individual HSPCs and to construct an early developmental lineage tree using these mutation data.

Several critical requirements must be met in order to successfully perform these assays. First, the viability of the sample must be ensured. Quick handling of the sample is key to ensure the efficiency of the procedure. Second, loss of growth factor potency will negatively affect the clonal expansion of HSPCs. To ensure high growth factor potency, it is important to avoid freeze-thaw cycles and prepare single-use aliquots. Third, after performing WGS, mutation calling and filtering, it is crucial to validate the clonality of the clonal culture. To confirm the clonality of the culture, the VAF of the mutations should cluster around of 0.5 in a karyotypically normal sample (**Figure 3**). In cells with a low mutational load, such as cord blood HSPCs, it is more difficult to determine clonality due to the low mutation numbers.

Our approach relies on in vitro expansion of single cells to allow for WGS. Therefore, our approach is restricted to cells that have the replicative potential to clonally expand, such as HSPCs. In our hands, about 5%–30% of all single-sorted cells are able to expand adequately. Reduced outgrowth rates can potentially result in a selection bias. As discussed previously, methods using WGA can overcome this selection bias as this technique is does not rely on the expansion of cells. However, WGA has its own shortcomings, and clonal amplification remains the only method to accurately determine the number of mutations in the whole genome without allelic dropouts and equal coverage along the genome, especially in samples with low true somatic mutation numbers.

The data generated using this approach can be used to determine phylogenies of the

hematopoietic system, as the mutations detected in single cells can be used to dissect cell lineages, as depicted in **Figure 6**. Typically, one or two mutations can define each branch in a healthy donor[1]. Since lineages branch early after conception, mutations defining these first branches will also be present with a low VAF in the matched normal sample that was used for filtering the germline variants[1,18,19]. In this case, the use of non-hematopoietic cells, such as MSCs, are preferred as they are expected to separate very early during development from the hematopoietic system. As T-cells are of hematopoietic origin, the use of these cells as a matched normal sample to filter germline variants could therefore confound the construction of the earliest branching of the developmental lineage tree. Subclonal presence of branch-specific mutations in certain mature blood populations, which can be measured by targeted deep sequencing, will indicate that the progeny of that branch can give rise to that mature cell type. In addition, our approach allows for assessing the mutational consequences of mutagenic exposure in vivo and ultimately how this may contribute to leukemia development.

**ACKNOWLEDGMENTS:**

This study was supported by a a VIDI grant of the Netherlands Organization for Scientific Research (NWO) (no. 016.Vidi.171.023) to R. v. B.

**DISCLOSURES:**

The authors have nothing to disclose.

**REFERENCES:**

1. Osorio, F. G. et al. Somatic Mutations Reveal Lineage Relationships and Age-Related Mutagenesis in Human Hematopoiesis. *Cell Reports.* **25,** 2308–2316.e4 (2018).
2. Genovese, G. et al. Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *New England Journal of Medicine.* **371,** 2477–2487 (2014).
3. Jaiswal, S. et al. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine.* **371,** 2488–2498 (2014).
4. Young, A. L., Challen, G. A., Birmann, B. M., Druley, T. E. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nature Communications.* **7,** 1–7 (2016).
5. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J., Stratton, M. R. Deciphering Signatures of Mutational Processes Operative in Human Cancer. *Cell Reports.* **3,** 246–259 (2013).
6. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature.* **500,** 415–421 (2013).
7. Alexandrov, L. et al. The Repertoire of Mutational Signatures in Human Cancer. *bioRxiv*. doi:10.1101/322859 (2018).
8. Behjati, S. et al. Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature.* **513,** 422–425 (2014).
9. Blokzijl, F. et al. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature.* **538,** 260–264 (2016).
10. Gawad, C., Koh, W., Quake, S. R. Single-cell genome sequencing: Current state of the science. *Nature Reviews Genetics.* **17,** 175–188 (2016).

11. Dong, X. et al. Accurate identification of single-nucleotide variants in whole-genome-amplified single cells. *Nature Methods.* **14,** 491–493 (2017).
12. Welch, J. S. et al. The origin and evolution of mutations in acute myeloid leukemia. *Cell.* **150,** (2012).
13. Jager, M. et al. Measuring mutation accumulation in single human adult stem cells by whole-genome sequencing of organoid cultures. *Nature Protocols.* **13,** 59–78 (2018).
14. Boeva, V. et al. Control-FREEC: A tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics.* **28,** 423–425 (2012).
15. Blokzijl, F., Janssen, R., van Boxtel, R., Cuppen, E. MutationalPatterns: Comprehensive genome-wide analysis of mutational processes. *Genome Medicine.* **10,** 1–11 (2018).
16. Thorvaldsdóttir, H., Robinson, J. T., Mesirov, J. P. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics.* **14,** 178–192 (2013).
17. Notta, F. et al. Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science.* doi:10.1126/science.1201219 (2011).
18. Lee-Six, H. et al. Population dynamics of normal human blood inferred from somatic mutations. *Nature.* **561,** 473–478 (2018).
19. Behjati, S. et al. Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature.* doi:10.1038/nature13448 (2014).

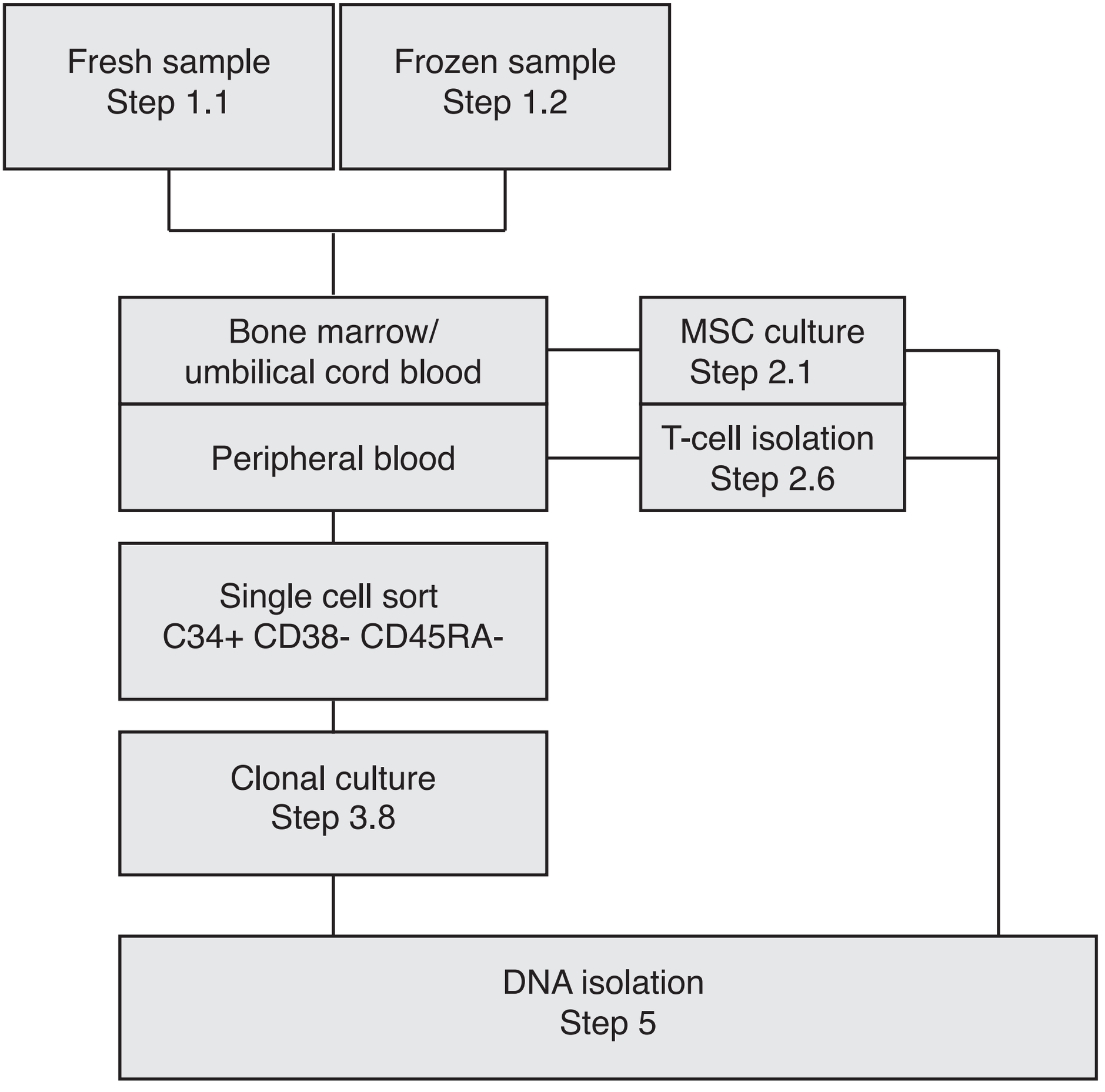
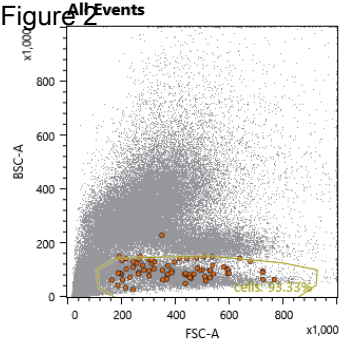
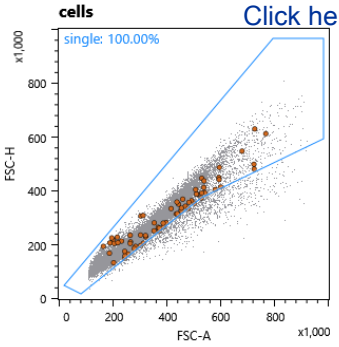
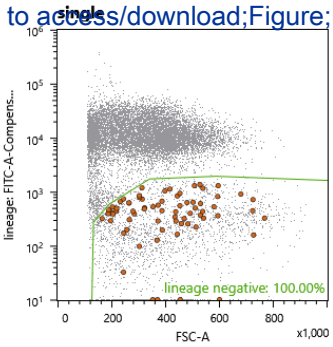

Fresh sample
Step 1.1
Frozen sample
Step 1.2
Bone marrow/
umbilical cord blood
Peripheral blood
MSC culture
Step 2.1
T-cell isolation
Step 2.6
Single cell sort
C34+ CD38- CD45RA-
Clonal culture
Step 3.8
DNA isolation
Step 5

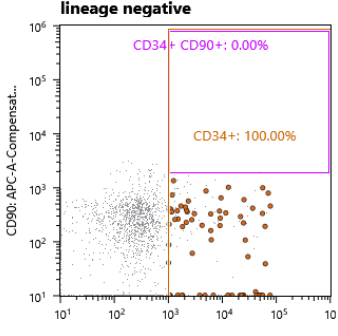
Figure 2

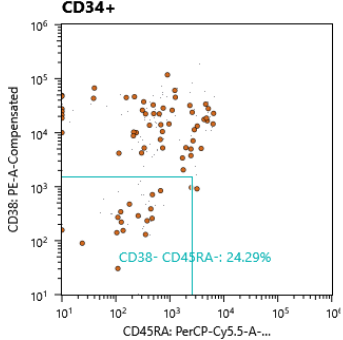
Click here to access/download;Figure;Figure 2.ai



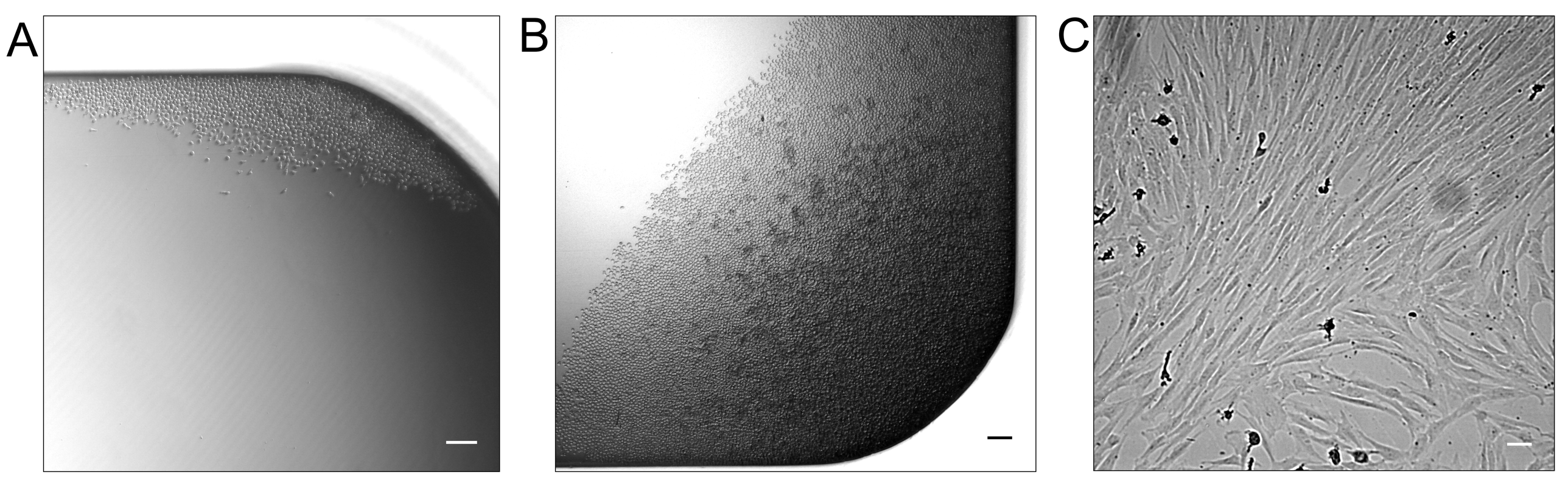



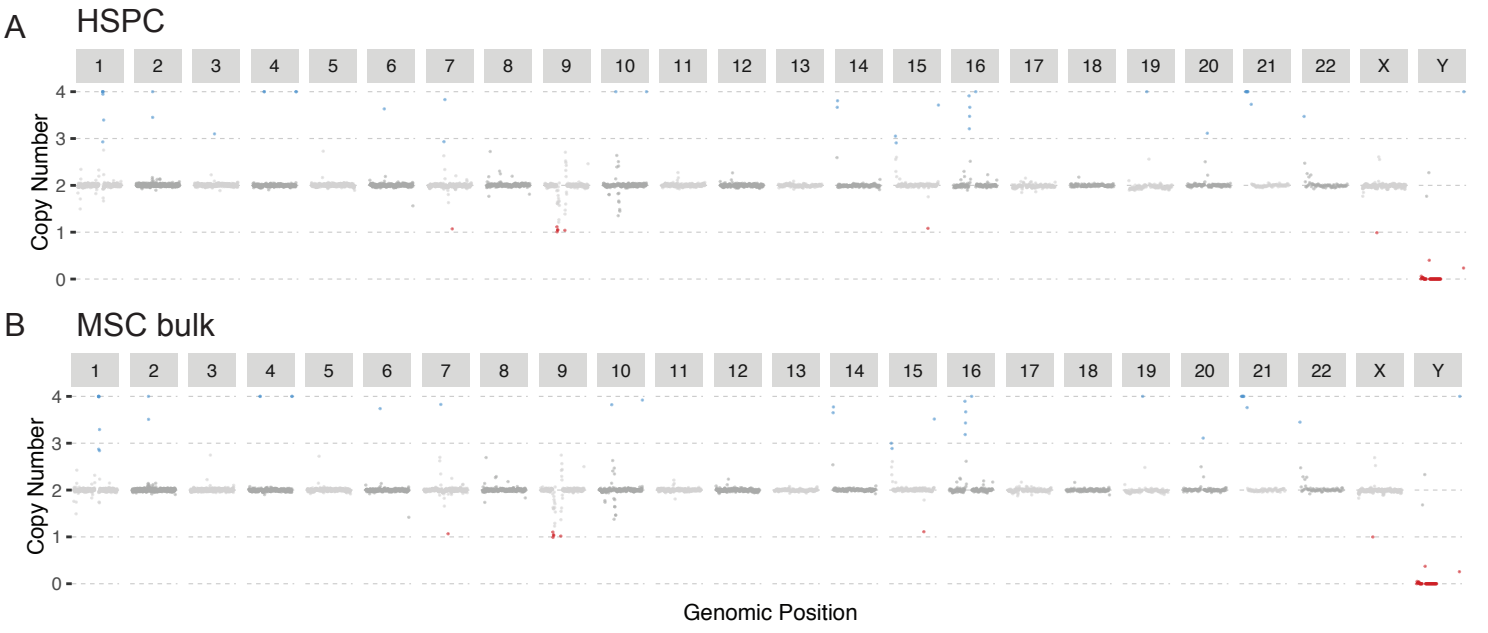



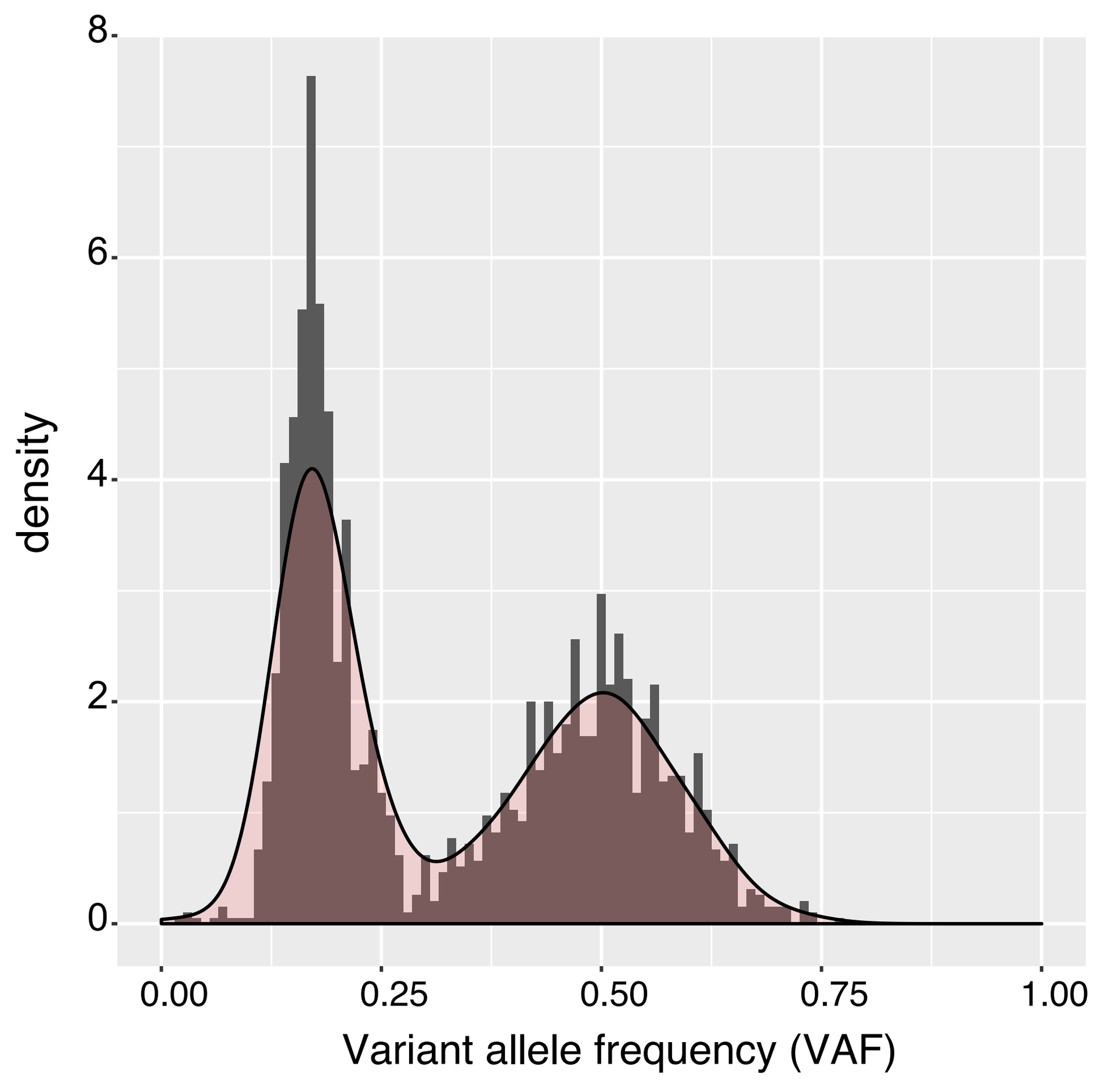



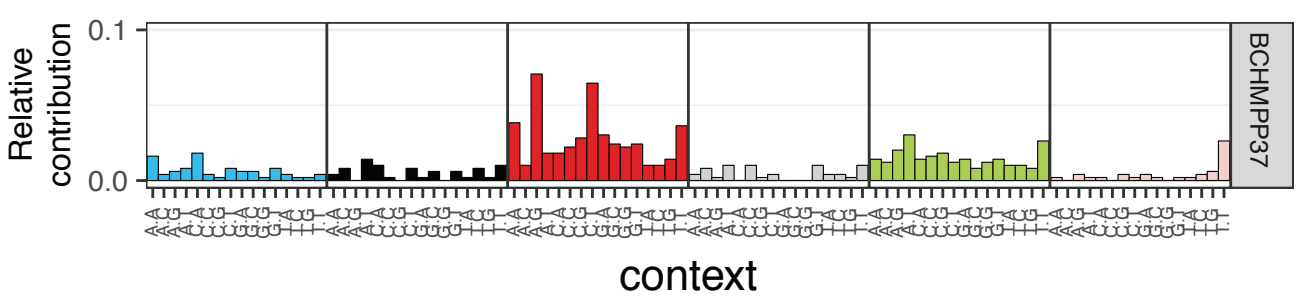

Gates and Statistics

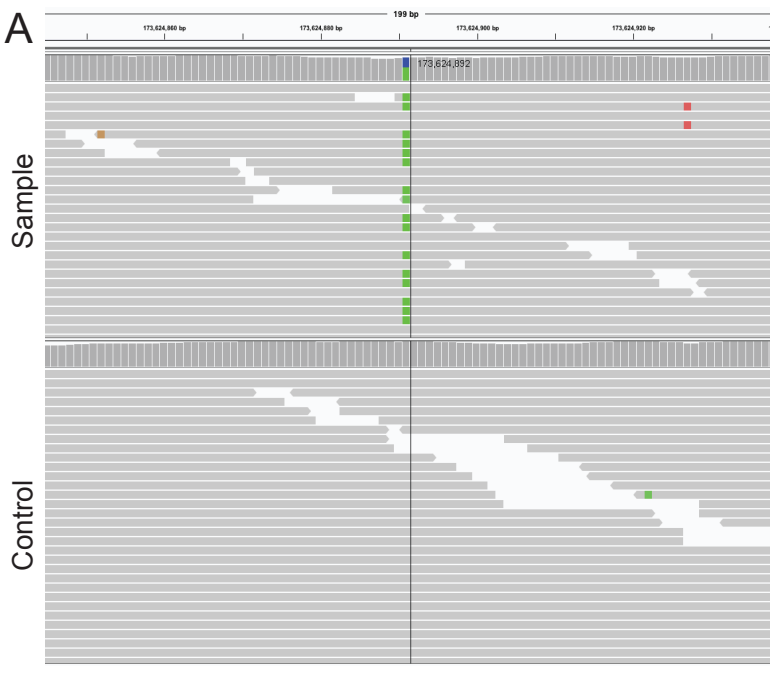
| Name | Events | %Parent | %Total |
|---|---|---|---|
| All Events | 75 | 0.00% | 100.00% |
| cells | 70 | 93.33% | 93.33% |
| single | 70 | 100.00% | 93.33% |
| lineage negative | 70 | 100.00% | 93.33% |
| CD34+ | 70 | 100.00% | 93.33% |
| CD38- CD45R... | 17 | 24.29% | 22.67% |
| CD34+ CD90+ | 0 | 0.00% | 0.00% |

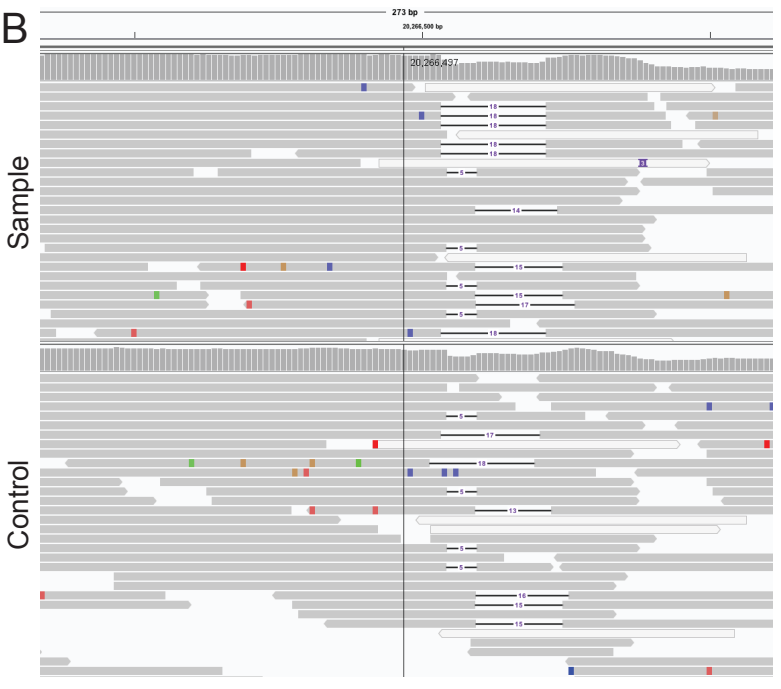

A
B
C


A
HSPC
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y
Copy Number
4
3
2
1
0
B
MSC bulk
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y
Copy Number
4
3
2
1
0
Genomic Position


8
6
4
2
0
density
0.00
0.25
0.50
0.75
1.00
Variant allele frequency (VAF)


Relative contribution
0.1
0.0
BCHMPP37
AAA AAC AAG AAT ACA ACC ACG ACT AGA AGC AGG AGT ATA ATC ATG ATT
context


A
199 bp
173,624,960 bp
173,624,880 bp
173,624,900 bp
173,624,920 bp
173,624,892
Sample
Control


B
273 bp
20,266,500 bp
20,266,497
Sample
Control

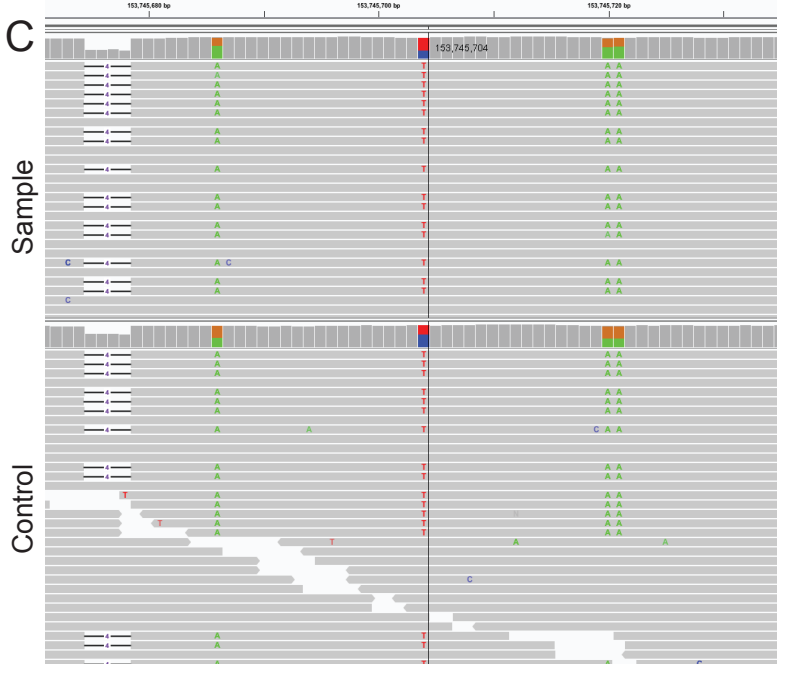
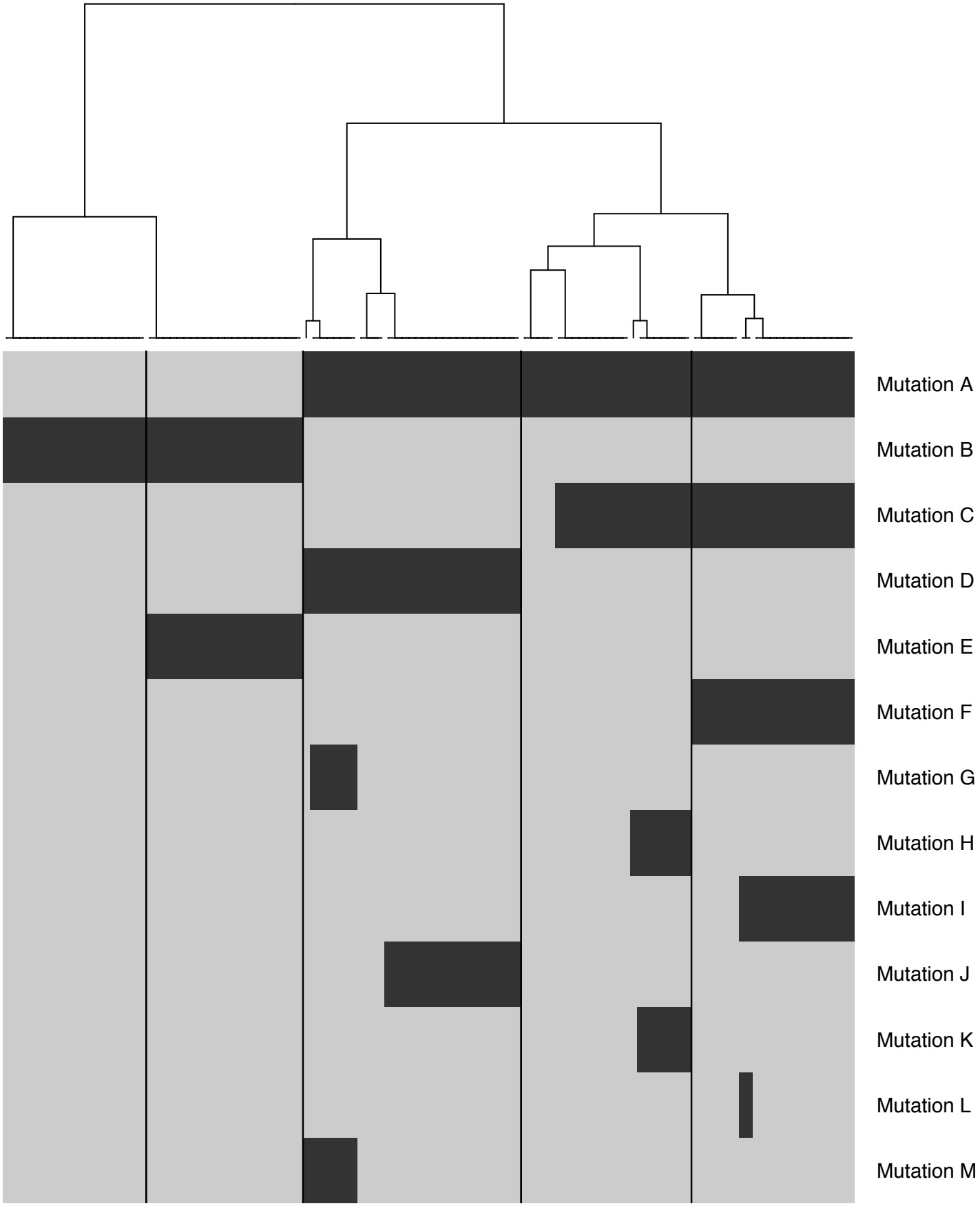

C
153,745,680 bp
153,745,700 bp
153,745,720 bp
153,745,704
Sample
Control

Figure 8

Click here to access/download;Figure;Figure 8.ai



| Antibody | volume [µL] |
|---|---|
| BV421-CD34 | 5 |
| FITC-Lineage mix (CD3/14/19/20/56) | 5 |
| PE-CD38 | 2 |
| APC- CD90 | 0.5 |
| PerCP/Cy5.5 - CD45RA | 5 |
| PE/Cy7- CD49f | 1 |
| FITC -CD16 | 1 |
| FITC-CD11 | 5 |
| FACS Buffer | 24.5 |

| Name of Material/Equipment | Company | Catalog Number |
|---|---|---|
| 0.20 µm syringe filter | Corning | 431219 |
| 50 mL Syringe, Luer lock | BD | 613-3925 |
| Bovine Serum Albumin (BSA) | Sigma-Aldrich | A9647-50G |
| CD11c FITC | BioLegend | 301603 |
| CD16 FITC | BioLegend | 302005 |
| CD3 BV650 | Biolegend | 300467 |
| CD34 BV421 | BioLegend | 343609 |
| CD38 PE | BioLegend | 303505 |
| CD45RA PerCP/Cy5.5 | BioLegend | 304121 |
| CD49f PE/Cy7 | BioLegend | 313621 |
| CD90 APC | BioLegend | 328113 |
| Cell Strainer 5 mL tube | Corning | 352235 |
| CELLSTAR plate, 384w, 130 µL, F-bc | Greiner | 781182 |
| Cryogenic vial | Corning | 430487 |
| Dimethyl sulfoxide (DMSO) | Sigma-Aldrich | D2650 |
| DMEM/F12 | ThermoFisher | 61965059 |
| EDTA | Sigma-Aldrich | E4884-500G |
| Fetal Bovine Serum | ThermoFisher | 10500 |
| GlutaMAX | ThermoFisher | 25030081 |
| Human Flt3-Ligand, premium grade | Miltenyi Biotech | 130-096-479 |
| Human Recombinant IL-3 (E. coli-e> | Stem Cell Technologies | 78040.1 |
| Human Recombinant IL-6 (E. coli-e> | Stem Cell Technologies | 78050.1 |
| Human SCF, premium grade | Miltenyi Biotech | 130-096-695 |
| Human TPO, premium grade | Miltenyi Biotech | 130-095-752 |
| Integrative Genomics Viewer 2.4 | Broad Institute | |
| Iscove's Modified Eagle's Medium | ThermoFisher | 12440061 |
| Lineage (CD3/14/19/20/56) FITC | BioLegend | 348701 |
| Lymphoprep | Stem Cell Technologies | #07861 |
| PBS | | |

| | | |
|---|---|---|
| Penicillin-Streptomycin | ThermoFisher | 15140122 |
| Primocin | Invivogen | ant-pm-1 |
| QIAamp DNA Micro Kit | Qiagen | 56304 |
| Qubit 2.0 fluorometer | ThermoFisher | Q32866 |
| Qubit dsDNA HS Assay Kit | ThermoFisher | Q32854 |
| RNAse A | Qiagen | 19101 |
| SH800S Cell Sorter | Sony | SH800S |
| StemSpan SFEM, 500mL | Stem Cell Technologies | 9650 |
| TE BUFFER PH 8.0, LOW EDTA | G-Biosciences | 786-151 |
| TrypLE Express | ThermoFisher | 12605-10 |

| | Comments/Description |
|---|---|
| Clone 3.9 | |
| Clone 3G8 | |
| Clone UCHT1 | |
| 561 | |
| Clone HIT2 | |
| Clone HI100 | |
| Clone GoH3 | |
| Clone 5E10 | |
| Reconsititute in single-use aliquots (25 µL) at 100 µg/mL in 0.1% BSA in PBS | |
| Reconsititute in single-use aliquots (2.5 µL) at 100 µg/mL in 0.1% BSA in PBS | |
| Reconsititute in single-use aliquots (5 µL) at 100 µg/mL in 0.1% BSA in PBS | |
| Reconsititute in single-use aliquots (25 µL) at 100 µg/mL in 0.1% BSA in PBS | |
| Reconsititute in single-use aliquots (12.5 µL) at 100 µg/mL in 0.1% BSA in PBS | |
| https://software.broadinstitute.org/software/igv/download | |
| Clones: UCHT1, HCD14, HIB19, 2H7, HCD56 | |
| Used for Density gradient separation | |
| Made in at Institute's facility. Commerically available PBS can also be used | |

Antibiotic formulation

1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

# ARTICLE AND VIDEO LICENSE AGREEMENT

| | |
|---|---|
| Title of Article: | Characterizing mutational load and clonal composition of human blood |
| Author(s): | Axel Rosendahl Huber, Freek Manders, Rurika Oka, Ruben van Boxtel |

Item 1: The Author elects to have the Materials be made available (as described at http://www.jove.com/publish) via:

☐ Standard Access ☒ Open Access

Item 2: Please select one of the following items:

☒ The Author is **NOT** a United States government employee.

☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

## ARTICLE AND VIDEO LICENSE AGREEMENT

1. **Defined Terms.** As used in this Article and Video License Agreement, the following terms shall have the following meanings: "**Agreement**" means this Article and Video License Agreement; "**Article**" means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; "**Author**" means the author who is a signatory to this Agreement; "**Collective Work**" means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; "**CRC License**" means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode; "**Derivative Work**" means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; "**Institution**" means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; "**JoVE**" means MyJove Corporation, a Massachusetts corporation and the publisher of The Journal of Visualized Experiments; "**Materials**" means the Article and / or the Video; "**Parties**" means the Author and JoVE; "**Video**" means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. **Background.** The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. **Grant of Rights in Article.** In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4** and **7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and(c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

4. **Retention of Rights in Article.** Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. **Grant of Rights in Video – Standard Access.** This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. **Grant of Rights in Video – Open Access.** This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this **Section 6** is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. **Government Employees.** If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. **Protection of the Work.** The Author(s) authorize JoVE to take steps in the Author(s) name and on their behalf if JoVE believes some third party could be infringing or might infringe the copyright of either the Author's Article and/or Video.

9. **Likeness, Privacy, Personality.** The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

10. **Author Warranties.** The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

11. **JoVE Discretion.** If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

# ARTICLE AND VIDEO LICENSE AGREEMENT

discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

12. **Indemnification.** The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in JoVE or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

13. **Fees.** To cover the cost incurred for publication, JoVE must receive payment before production and publication the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

14. **Transfer, Governing Law.** This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to me one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement is required per submission.

**CORRESPONDING AUTHOR**

| | |
|---|---|
| Name: | Ruben van Boxtel |
| Department: | Research |
| Institution: | Princess Máxima Center for Pediatric Oncology |
| Title: | Dr. |
| Signature: |  |
| Date: | 14-2-2019 |

Please submit a **signed** and **dated** copy of this license by one of the following three methods:

1. Upload an electronic version on the JoVE submission site
2. Fax the document to +1.866.381.2236
3. Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02140

***Rebuttal***

***Editorial comments:***

*General:*

*1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues.*

*2. JoVE cannot publish manuscripts containing commercial language. This includes trademark symbols (™), registered symbols (®), and company names before an instrument or reagent. Please limit the use of commercial language from your manuscript and use generic terms instead. All commercial products should be sufficiently referenced in the Table of Materials and Reagents.*

*For example: GlutaMAX, TrypLE, Qiagen, Qubit, ThermoFisher etc.*

*Protocol:*

*1. Everything in the protocol (except for the introductory ethics statement) should be in a numbered step (in the imperative tense and of no more than 4 sentences), header, or 'Note'. Please move the introductory paragraphs of the protocol to the Introduction, Results, or Discussion (as appropriate) or break into steps. Please avoid bullet points.*

*2. There is a 10 page limit for the Protocol, but there is a 2.75 page limit for filmable content. If revisions cause the highlighted portion to be more than 2.75 pages, please highlight 2.75 pages or less of the Protocol (including headers and spacing) that identifies the essential steps of the protocol for the video, i.e., the steps that should be visualized to tell the most cohesive story of the Protocol.*

*3. For each protocol step/substep, please ensure you answer the "how" question, i.e., how is the step performed? Alternatively, add references to published material specifying how to perform the protocol action. If revisions cause a step to have more than 2-3 actions and 4 sentences per step, please split into separate steps or substeps.*

*Specific Protocol steps:*

*1. Please combine 1A/1B and 2A/2B into single sections; the numbering is currently somewhat confusing.*

We have adapted the protocol in a way a single sequence of steps can be followed. Depending on input the type and condition of input material different steps can be skipped. In addition, we have added a flowchart indicating the order of the experimental steps for which tissue.

*2. 10.3, 10.4: How exactly do you run these scripts (i.e., the exact command)?*

We thank the editors for pointing this part out. We added the execution command in step 10.4 which uses the the .ini file edited at step 10.3.

*3. 10.5: How exactly is this done?*
We have more thoroughly explained the procedure and have given a more detailed description of the commands used.

*Representative Results*
*1. Representative Results: Please include at least one paragraph of text to explain the Representative Results in the context of the technique you have described, e.g., how do these results show the technique, suggestions about how to analyze the outcome, etc. The paragraph text should refer to all of the figures (except those referenced elsewhere). Data from both successful and sub-optimal experiments can be included.*

*Figures:*
*1. Please remove the embedded Figures from the manuscript.*
We have removed the embedded figures from the manuscript.

*2. Please remove the embedded table from the manuscript. All tables should be uploaded separately to your Editorial Manager account in the form of an .xls or .xlsx file. Each table must be accompanied by a title and a description after the Representative Results of the manuscript text.*
*3. Figure 4: There should be a closing parenthesis after VAF.*

We have corrected this typographical error.

*References:*
*1. Please do not abbreviate journal titles.*
We have changed our reference style and do no longer abbreviate the journal titles.

*Table of Materials:*
*1. Please ensure the Table of Materials has information on all materials and equipment used, especially those mentioned in the Protocol.*

We have double-checked the table of materials so it contains all used material and equipment.

***Reviewers' comments:***

*Reviewer #1:*

*Manuscript Summary:*
*The manuscript by Rosendahl Huber et al provides a well-written overview of a method that is likely to be of significant use to the research community. Overall the paper provides a clear protocol of the approach, which should be relatively*

*easily transferable to other labs. I have raised several issues that I think would be useful to address:*

*Minor Concerns:*

*1. The methods describe the procedure for human cells. Is there any reason to believe that the method for murine cells would be different? If not, please state this. If yes, it would be useful to get some insight into what additional challenges might be confronted (personally I believe the method should also work for murine cells, if the sorting strategy and the culture conditions are amended).*

We thank the reviewer for pointing this out and agree that this method is not uniquely applicable to human material. Currently, we have not tested our approach on murine HSCs; however, in the past, our group has performed a similar assay on murine intestinal stem cells, see Extended data figure 8 (Blokzijl et al., 2016).We agree with the reviewer that the wet-lab protocol required amending the sorting strategy and culture conditions. Adaptation of scripts is required to determine mutations as well. We adapted the discussion to state the possibility of appending the protocol to mouse material.

*2. Line 126: I would rewrite "When the sample IS almost thawed, wipe the vial with 70% ethanol and transfer ITS CONTENT to A 50 ml conical tube".*

We have corrected the typographical errors in the text.

*3. Line 143: may be it is useful to specify what a FACS buffer constitutes of?*

We thank the reviewer for pointing out this ordering mistake and will bring the explanation (present at line 174) to line 143.

*4. Line 148: It is little odd to only start culturing MSCs after step 1.9, where hematopoietic cells are essentially ready to be sorted. Obviously, MSC cultures should have started much earlier so that these cells are ready to go after step 1.10. I.e. some textual reshuffling is needed.*

We agree with the reviewer the sequence of steps in the protocol is not optimal. We have revised the ordering of the manuscript. As some steps have to be performed simultaneously, we have added a flow-chart for a more clear representation of the laboratory procedure.

*5. Line 160: How do you ensure that all hematopoietic cells are washed off? If they are not specifically depleted, these cultures will still contain a very significant number of monocytes/macrophages that are pretty sticky and will not be washed off. This may not be important for this specific application, but good to mention whether or not it is essential that ALL hematopoietic cells be washed off.*

We thank the reviewer for pointing this out. During culture, we indeed observe some colonies will differentiate and stick to the bottom of the well. Due to our washing technique we usually all, or almost all cells are taken up, due to fiercely pipetting and scraping the bottom of the well. We have not mentioned the scraping of the well, which is addressed in the revision.

*6. How many HSPCs can be stained with the antibody mix shown in Table 3.2? The text reads "spin down AT LEAST 10e7 cells, but that suggest that there is no maximum number, which is clearly not the case.*

We thank the reviewer for pointing out this inconsistency. We have successfully sorted up to $2*10^7$ cells with our technique, and it is now mentioned.

*7. Line 241: CD45- should be CD45RA-. Also, is CD90 used in the sorting protocol? The figures suggest it is, but in the text I do not read this.*

We thank the reviewer for pointing out this typographical error and will change CD45 in CD45RA. We agree with the reviewer we use CD90 and CD49f to distinguish multipotent progenitors from long-term HSCs. This is now added in the representative results section.

*8. How do you confirm after sorting that indeed a single cell was deposited in each well? Is there no visual inspection shortly after sort? What are the consequences if 2 cells end up in one well?*

We understand the concern of the reviewer. In our culture plates it is very hard to distinguish the individual sorted cells, and therefore we do not perform visual inspection after sorting. If multiple cells end up in one cell and both cells expand, the variant allele frequency (VAF) will not be 0.5 anymore for mutations present in all cells, but lower. This can be identified a peak-shift towards lower frequency in the VAF plots from the SNVFI (step 7.6 and 7.7).

*9. Why would you want to harvest the first cultures after 4 weeks, and culture the plate for an additional week to see whether new colonies emerge? Why not rather just harvest all colonies after 5 weeks?*

We thank the reviewer for addressing this procedure and we now explained the rationale behind this step in the protocol. The first colonies will emerge after 3-4 weeks and can survive for several days in the medium. However, we observed that fully grown colonies quickly deplete the medium and change its' pH, inducing cell death. To ensure optimal DNA recovery, wells are harvested when full during the $4^{th}$ and the $5^{th}$ week.

*Reviewer #2:*

*Manuscript Summary:*
*The manuscript by Huber and colleagues presents a detailed "how to" for the*

*determination of mutational burden in human hematopoietic stem cells, which will facilitate the determination for how mutational burden changes with age and disease. There are a couple of minor concerns that are listed below, but generally the presentation is clear and the methods straightforward.*

*Major Concerns:*
*none*

*Minor Concerns:*

*1. Regarding the code, the authors should alter the code presentation such that the recommended settings no longer require the user to manually input them. We usually build a file that is referenced to do this, so that the user can just hit a button and run the program.*

We thank the reviewer for pointing this out. We have provided or referenced to .ini files where both recommended settings and paths to input files and output folders are indicated. We have removed the recommended input settings from the protocol, as these can be found in the .ini files.

*2. Regarding the method, it would be nice to know the cell viability and DNA yield that they get from their method of banking mononuclear cells.*

In our revision, we have provided numbers on cell viability and recovery after thawing of mononuclear cells.

*3. They should better explain the logic for using MSCs or T-cells as germline controls. While using MSCs make sense, since this is a different lineage, how is the use of T-cells better than using a pool of HSC? Perhaps clonal hematopoiesis (such as in the elderly) could lead to the inability to discriminate a somatic mutation from a germline one, but is clonal hematopoiesis less likely to be observed in T-cells?*

We agree with the reviewers that T-cells are not an ideal sample due to their hematopoietic origin. Ideally to dissect a developmental lineage tree based on somatic mutations acquired during life, a bulk sample which does not share a developmental origin with the hematopoietic system is required, such as MSCs or a buccal swap sample. However, when only peripheral blood is available, T-cells can be considered a viable option, which is also being used in other studies focusing on studying the origin of leukemia within the blood lineage (Shlush et al, Nature 2017).

*4. In step 1.9, they use a volume of isolation rather than cell counts. The isolation is just for a flow control, but they are giving specifics on the antibody concentrations, and cell count differences will alter fluorescence intensities.*

*Therefore, it might be helpful to just specify the number of cells to use in the flow controls. Ok, so this is being pretty picky!*

We thank the reviewer for pointing out this step in the protocol. We specified the recommended cell number ($1*10^5$) required for the flow cytometry control sample.

*5. For step 10.6: If we understand this correctly, they are suggesting that those locations which show a number of different mutations would be considered a bit low confidence and should just be thrown out. If this is the case, we really don't like the idea that this would be a judgement call without specific criteria, and secondly manual inspection using IGF seems like an unsatisfactory way of eliminating these mutations as it might take days or weeks per sample. Perhaps this is a misunderstanding on our part.*

We thank the reviewer for pointing out the unclear passage in our manuscript. We agree with the reviewer we were not specific enough in our description. We have now rewritten this protocol step to be more specific in our validation criteria.
It is noteworthy that processing of sequenced reads to obtain genome alignments and then to call variants is a complicated process. A number of factors, including sequencing errors, actual variants diverging the read sequences from the references, could influence mapping quality. The variants are called from resulting mapping data assuming the mapping is done properly. This assumption holds true at the regions with unique sequences and in the genome, but not with the low-complexity regions. Even with the unique regions, the number of variants in the sequenced sample could limit the precision and accuracy. Although those regions we do not trust the variant calling from is obvious by eye as shown in the new Figure 7, we cannot distinguish such regions automatically with our current method. Adding to the technical difficulty, only a low number of mutations, typically less than 10, will be shared amongst clones, as most mutations are unique for each stem cell. Therefore, it does not take much time to evaluate all shared mutations. The manual check we perform here is also to validate our filtering and to ensure we only include high confidence mutation since a single mutation can determine the position of a cell in the lineage tree.
We agree with the reviewer this is a judgement call; however, we hope to convince the reviewer with the above explanation and the newly provided IGV examples that our manual check is reasonable and straightforward.

We have now written this protocol step as: “Mutations are considered false when not present, when the mutation is present in the germline or when present in poorly mapped regions”

*6. Table 3.4 is jumbled in the pdf we received.*

We thank the reviewer for pointing this out. We have supplemental table 3.4 in a separate excel file.

*Reviewer #3:*

*The manuscript described a pipeline about analyzing de novo somatic mutations in HSPCs by clone sequencing. I hope the following questions would be clarified before the manuscript published.*

*1. At the Step 3.1, the author pointed out 1x107 cells required at least. What are these cells? Are they PBMCs?*

We thank the reviewer for pointing this step out. We are now more specific and alter the sentence to: '$1x10^7$ mononuclear cells are required.

*2. At the Step 3.11, the single HSPC was cultured in the incubator at 37°C for 3-4 weeks. Is it necessary to refill medium? If yes, how often? Does it need to transfer cells from 384-well plate to 96-well plate for clone expansion? Still at this step, the author found 5%-30% of sorted cells would clonally expand. Are the success rates different for fresh cells or frozen ones? Is there any difference between the rates of bone marrow, cord blood or peripheral blood?*

While a 4-week culture period without changing medium seems long, the cells are perfectly able to proliferate for up to 4 weeks. We have observed enhanced growth when passaging colonies in 96-wells plates, and this might enhance the DNA load of a colony even further.

Regarding the question about success rates for clonal expansion: We have not specified for which samples we see a lower clonal outgrowth, as we see higher colony outgrowth in bone marrow, while we observe lower colony outgrowth in samples from peripheral blood. We clarified this further in the Representative results section.

*3. At both step 4.6 and 4.7, the author mentioned ">30% confluency" or a small visible pellet" as the marker for cell collection. However, this description is not very specific. Is it possible to use the number of cells instead?*

We agree with the reviewer that the size description for cell harvest and cell numbers are not highly specific. However, the primary objective of colony outgrowth is to obtain enough DNA to perform whole-genome sequencing and this increases with the number of cells. For cell counting cells have to be sacrificed, losing more cells in the process. When colonies have a confluency <30%, we typically observe too low DNA yields (<50ng) to perform whole genome sequencing.

Click here to access/download

**Supplemental Coding Files**

SNVFI.ini