

## Video Article

# Characterizing Mutational Load and Clonal Composition of Human Blood

Axel Rosendahl Huber<sup>1</sup>, Freek Manders<sup>1</sup>, Rurika Oka<sup>1</sup>, Ruben van Boxtel<sup>1</sup><sup>1</sup>Princess Máxima Center for Pediatric OncologyCorrespondence to: Ruben van Boxtel at [R.vanBoxtel@prinsesmaximacentrum.nl](mailto:R.vanBoxtel@prinsesmaximacentrum.nl)URL: <https://www.jove.com/video/59846>DOI: [doi:10.3791/59846](https://doi.org/10.3791/59846)

Keywords: Genetics, Issue 149, hematopoiesis, clonality, somatic mutations, development, lineage tracing, mutational signatures, leukemia

Date Published: 7/11/2019

Citation: Huber, A.R., Manders, F., Oka, R., van Boxtel, R. Characterizing Mutational Load and Clonal Composition of Human Blood. *J. Vis. Exp.* (149), e59846, doi:10.3791/59846 (2019).

## 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.

## Video Link

The video component of this article can be found at <https://www.jove.com/video/59846/>

## 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<sup>1</sup>. Gradual mutation accumulation in HSPCs<sup>1</sup> can result in age-related clonal hematopoiesis (ARCH)<sup>2,3</sup>, 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<sup>2,3</sup>. However, recent studies have shown an incidence of 95% of ARCH in elderly individuals<sup>4</sup>, 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<sup>5</sup>. For instance, exposure to UV light, alkylating agents, and defects in DNA repair pathways have each been associated with a different mutational signature<sup>6,7</sup>. In addition, due to 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<sup>8</sup>. 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<sup>9</sup>. Previous methods have obtained sufficient DNA from a single cell for WGS using whole-genome amplification (WGA)<sup>10</sup>. However, the main disadvantage of WGA is its relatively error-prone and unbalanced amplification of the genome, which can result in allele dropouts<sup>11</sup>. 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<sup>12</sup>. 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. Preparing fresh bone marrow, cord blood, or peripheral blood

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.
2. **OPTIONAL:** The recommended number of cells required to sort a full 384 well plate of HSPCs is  $1-2 \times 10^7$ . If more cells are isolated during density gradient centrifugation, store the surplus of cells in liquid nitrogen.
3. Resuspend cells in 500  $\mu\text{L}$  of IMDM + 10% FBS per  $1 \times 10^7$  cells, and add drop-by-drop an equal volume of IMDM + 30% FBS + 20% DMSO to achieve a suspension of  $1 \times 10^7$  cells in 1 mL of IMDM + 20% FBS + 10% DMSO.
4. Immediately transfer the mononuclear cells to 1 mL cryogenic vials and freeze cells at  $-80^\circ\text{C}$  in a controlled-rate cell freezing container overnight. Transfer the cells the next day to a liquid nitrogen storage upon further processing.

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

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  $37^\circ\text{C}$  water bath.
  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  $37^\circ\text{C}$  water bath.
  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.
  4. Add an additional pre-warmed 15 mL of IMDM + 10% FBS dropwise to the sample while gently swirling the tube.
  5. Pellet the cells by centrifugation for 5 min at  $350 \times g$ .
  6. Remove all but  $\pm 3$  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.
  7. Take 10  $\mu\text{L}$  of the cell suspension for cell counting. Dilute these 10  $\mu\text{L}$  by adding 20  $\mu\text{L}$  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%.
3. If working with bone marrow or umbilical cord blood cells, take up  $5 \times 10^6$  mononuclear cells for MSC culture (step 2.1). If working with peripheral blood, take up  $2-5 \times 10^6$  cells for T-cell isolation (step 2.2)
  4. Pellet remaining cells 5 min at  $350 \times g$  and resuspend in 3 mL of FACS buffer (0.05% BSA + 1 mM EDTA in PBS).
  5. Transfer  $1 \times 10^6$  cells to a microtube filled with 200  $\mu\text{L}$  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.

#### 1. MSC culture

1. Prepare 50 mL of MSC medium containing 45 mL of DMEM/F12 medium, 10% FBS, 500  $\mu\text{L}$  of trypsin or trypsin alternative, and 500  $\mu\text{L}$  of penicillin/streptomycin solution.
2. Plate approximately  $5 \times 10^5$  mononuclear cells in 1.5 mL of MSC medium per well. Place the cells in a humidified incubator at  $37^\circ\text{C}$  with 5%  $\text{CO}_2$ .
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%.
4. If the MSCs are confluent, wash cells with 1 mL of PBS and harvest the MSCs by adding 200  $\mu\text{L}$  of trypsin or trypsin alternative per well. Incubate cells for 5 min at  $37^\circ\text{C}$ . Add 800  $\mu\text{L}$  of MSC medium, and pipet the cells up and down to loosen cells from the well plate.
5. Transfer MSCs to microcentrifuge tube and pellet the cells by centrifugation for 5 min at  $350 \times g$ . Remove the supernatant and continue with DNA isolation or store the pellet at  $-20^\circ\text{C}$  for later DNA isolation (section 4).

#### 2. T-cell isolation

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

1. Resuspend the cell pellet in 100  $\mu\text{L}$  of anti-CD3 staining solution (1:100 dilution of anti-CD antibody in FACS buffer).
2. Wash the cells by adding 1 mL of FACS buffer. Pellet the cells by centrifugation for 5 min at  $350 \times g$  and resuspend in 300  $\mu\text{L}$  of FACS buffer.
3. Isolate at least  $5 \times 10^5$  CD3+ cells using a FACS-sorter in a 5 mL polystyrene tube pre-filled with 1 mL of FBS.

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

1. Spin down at 1–2 x 10<sup>7</sup> 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<sup>7</sup> cells, increase the antibody mix and FACS buffer volumes accordingly.
2. Prepare 50 µL of 2x HSC staining mix according to the recipe seen in **Table 1**.

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	25.5

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

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.
4. Wash the cells by adding 1 mL of FACS buffer and pellet by centrifuging for 5 min at 350 x g.
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).
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**).
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.
8. **Sorting single HSPCs**
  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<sup>-</sup> fraction. Draw gates for CD34<sup>+</sup> cells and further characterize this subset by setting a specific gate for CD38<sup>-</sup> CD45RA<sup>-</sup> cells.
  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.
9. **Culturing singly-sorted HSCs**
  1. Directly transfer the 384 well plate to a humidified 37 °C incubator with 5% CO<sub>2</sub>.  
**NOTE:** To prevent evaporation during culture, wrap the 384 well culture plate (with lid) in transparent polyethylene wrap.
  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

1. After 4 weeks of culturing, determine which wells have a confluency of 30% or higher.
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.
3. Pre-wet a pipette tip with 1% BSA in PBS to minimize the number of cells sticking to the pipette tip.
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.
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 the bottom of well. Inspect the wells using a standard inverted light microscope to ensure whether all cells have been collected.
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.
7. Spin down the cell suspension for 5 min at 350 x g. A small pellet should be visible.
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

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:**
  1. Add 2  $\mu\text{L}$  of RNase A after addition of buffer AL during section 2. Incubate for 2 min before adding proteinase K.
  2. Incubate for 30 min at 56 °C instead of 10 min.
  3. Elute the DNA by loading the column with 50  $\mu\text{L}$  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.
2. Determine the DNA concentration using a DNA measuring 2  $\mu\text{L}$  per clone. The DNA yield typically varies between 0.5–3 ng/ $\mu\text{L}$ .

## 6. Sequencing

1. Perform DNA sequencing as described by Jager et al.<sup>13</sup>

## 7. Mapping and Somatic Mutation Calling

1. Map the output of sequencing (FASTQ files) to the reference genome and call mutations as described in Jager et al.<sup>13</sup>
2. Inspect the data for aberrant karyotypic changes in sequenced clone and bulk data using a copy number analysis tool, such as Control-FreeC<sup>14</sup>. Until now, we have not reported any HSPCs with karyotypic aberrances.
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<sup>13</sup>, or use the following uploaded blacklist: <<https://data.mendeley.com/datasets/9y4yhwt5rp/3>>.
4. Filter single nucleotide variations using SNVFI <<https://github.com/ToolsVanBox/SNVFI>>.
  1. Preset SNVFI.config file, such that all paths to helper functions are correct.
  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  $\geq 0.3$ <sup>9</sup>.
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.
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 Analysis Toolkit):
 

```
java -jar GenomeAnalysisTK.jar \
  -T CallableLoci \
  -R reference.fasta \
  -I myreads.bam \
  -summary table.txt \
  -o callable_status.bed
```
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://github.com/ToolsVanBox/CallableLoci\\_processor](https://github.com/ToolsVanBox/CallableLoci_processor). 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

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
```
2. Filter raw\_INDELS.vcf list using INDELF1 <<https://github.com/ToolsVanBox/INDELF1>>:
 

```
perl INDELF1.pl -i input.vcf (from step 8.1) \
  -s column test sample \
  -c column control sample
```

## 9. Mutational Profile Inspection

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<sup>15</sup>: <<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

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.
2. To identify mutations that are present in a subset of the clones and subclonally present in the bulk, perform the following steps.
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://github.com/ToolsVanBox/filterSomatic>. Before running this script, edit the filterSomatic.ini file (see **Supplemental File 2**) to set the paths and adjust the other parameters.
4. Run filterSomatic.py (**python3 filterSomatic.py -i filterSomatic.ini**).
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://github.com/ToolsVanBox/Identify\\_lowVAF\\_bulk\\_muts](https://github.com/ToolsVanBox/Identify_lowVAF_bulk_muts). 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
```
6. Determine all mutations shared between clones which are not present in the bulk sample by overlapping all mutation positions (concatenate column 1 and 2 of SNVFI output).
7. Exclude false positives obtained during steps 10.5 and 10.6 by manual inspection using IGV<sup>16</sup>. 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.
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.
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<sup>-</sup>, CD34<sup>+</sup>, CD38<sup>-</sup>, CD45RA<sup>-</sup>. Additionally, CD90 and CD49f can be used to distinguish between progenitor cells or self-renewing stem cells<sup>17</sup> (**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<sup>14</sup> 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<sup>15</sup>. 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**).

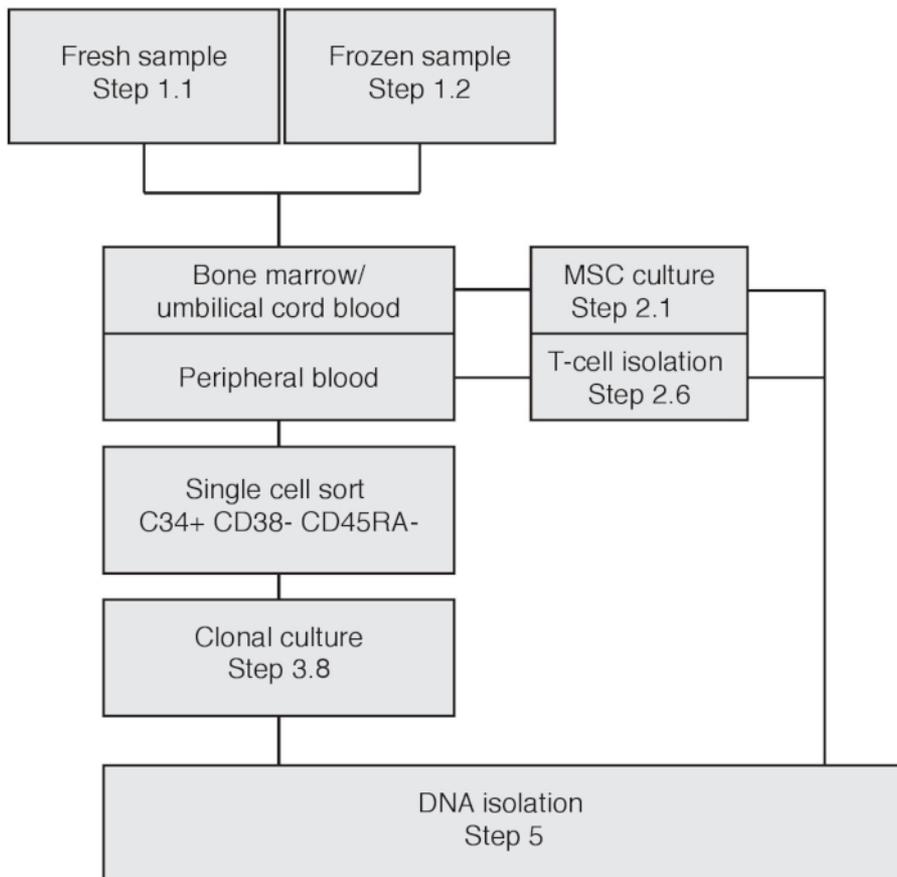


Figure 1: Flowchart depicting experimental procedure based on input material. Please click here to view a larger version of this figure.

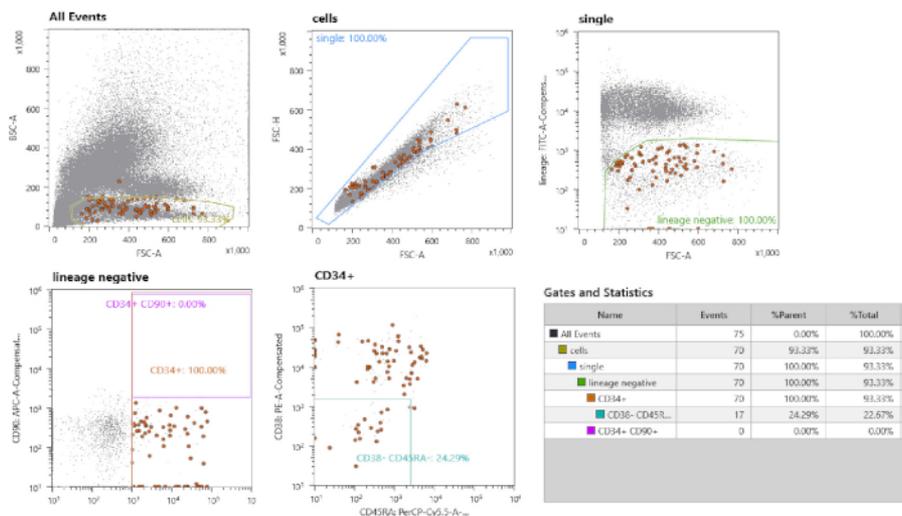
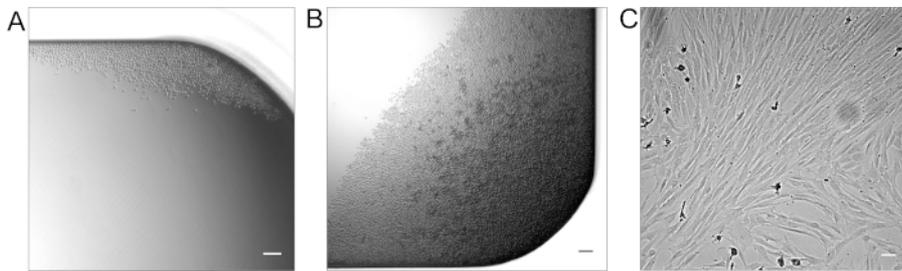
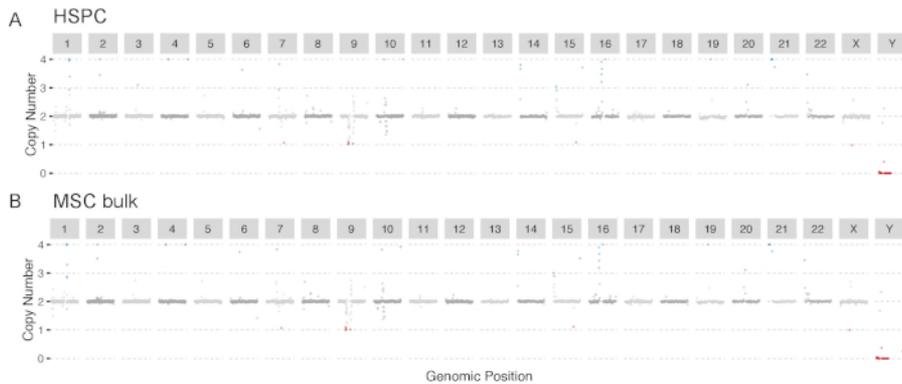


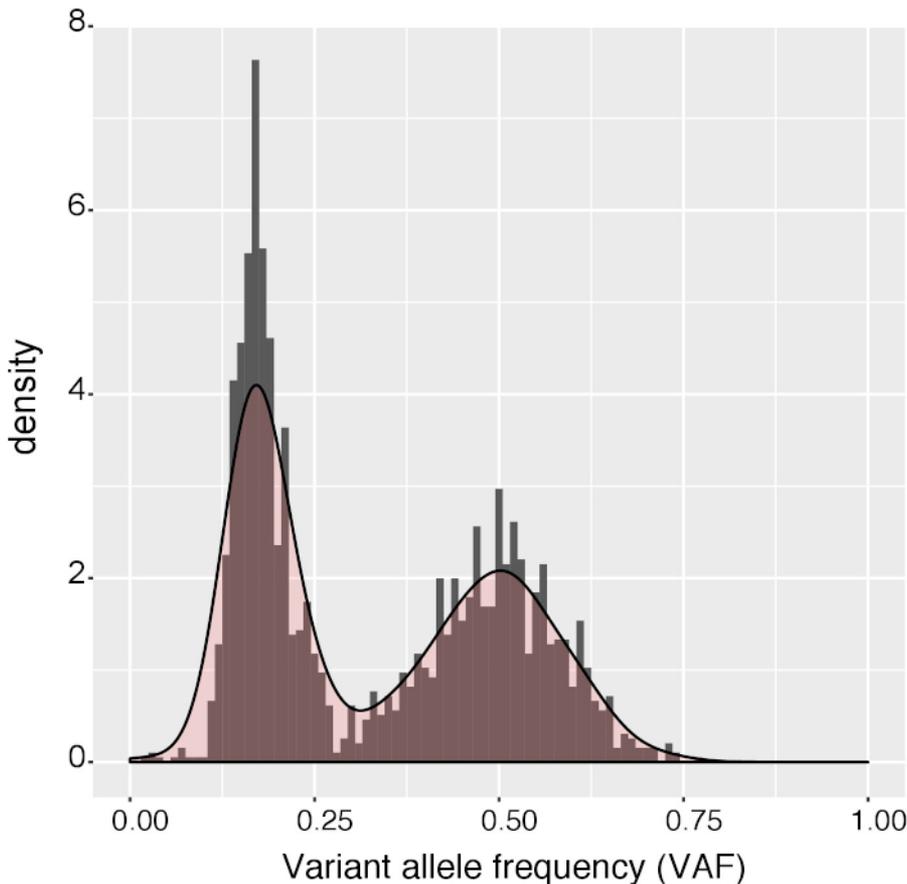
Figure 2: Cell sorting strategy. First, gating is performed on small mononuclear cells. Second, single cells are gated by selection of the linear fraction. Lineage negative cells are gated. All CD34<sup>+</sup> CD38<sup>-</sup> CD45<sup>-</sup> cells are single cell-sorted. The fraction of cells in brown should be noted, which are the sorted cells highlighted by the option “index sorting”. Please click here to view a larger version of this figure.



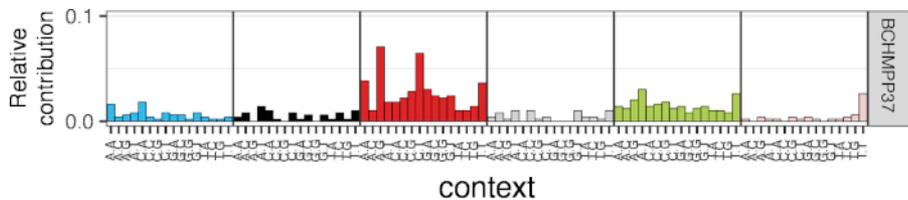
**Figure 3: Representative cell culture results.** Representative HSPC clones in a 384 well plate at (A) 2 weeks after plating and (B) 4 weeks after plating. (C) MSC culture after 2 weeks of medium replacement. Scale bar = 100  $\mu$ m. [Please click here to view a larger version of this figure.](#)



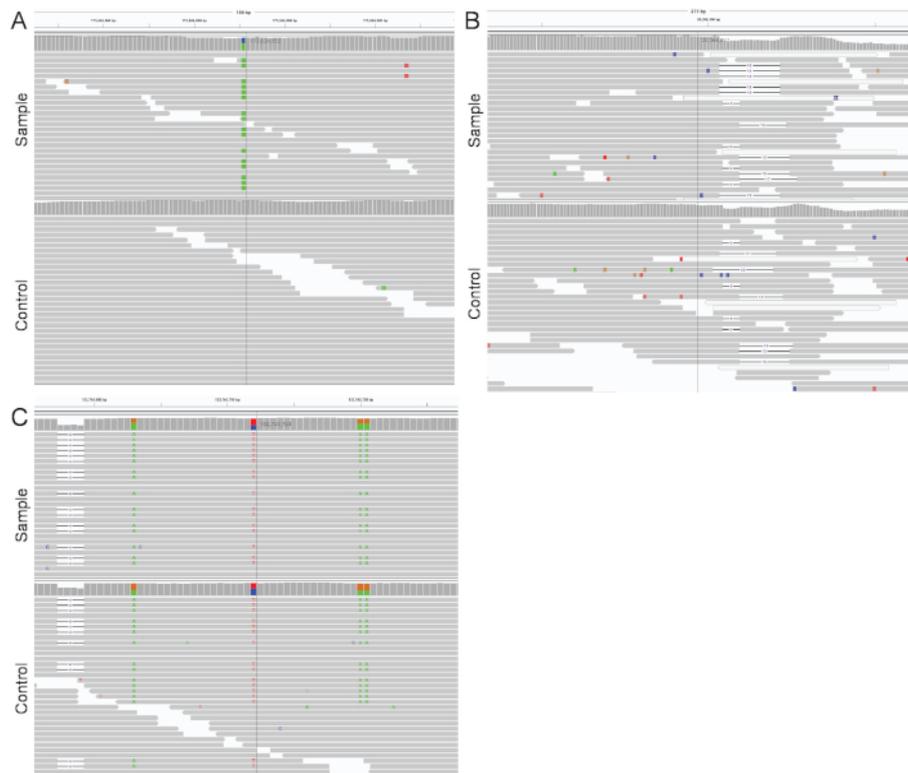
**Figure 4: Karyotypes.** (A) Clonal HSPC culture and (B) MSC bulk sample. The karyotypes were determined by read-depth analysis. Both graphs indicate a karyotypically normal sample. [Please click here to view a larger version of this figure.](#)



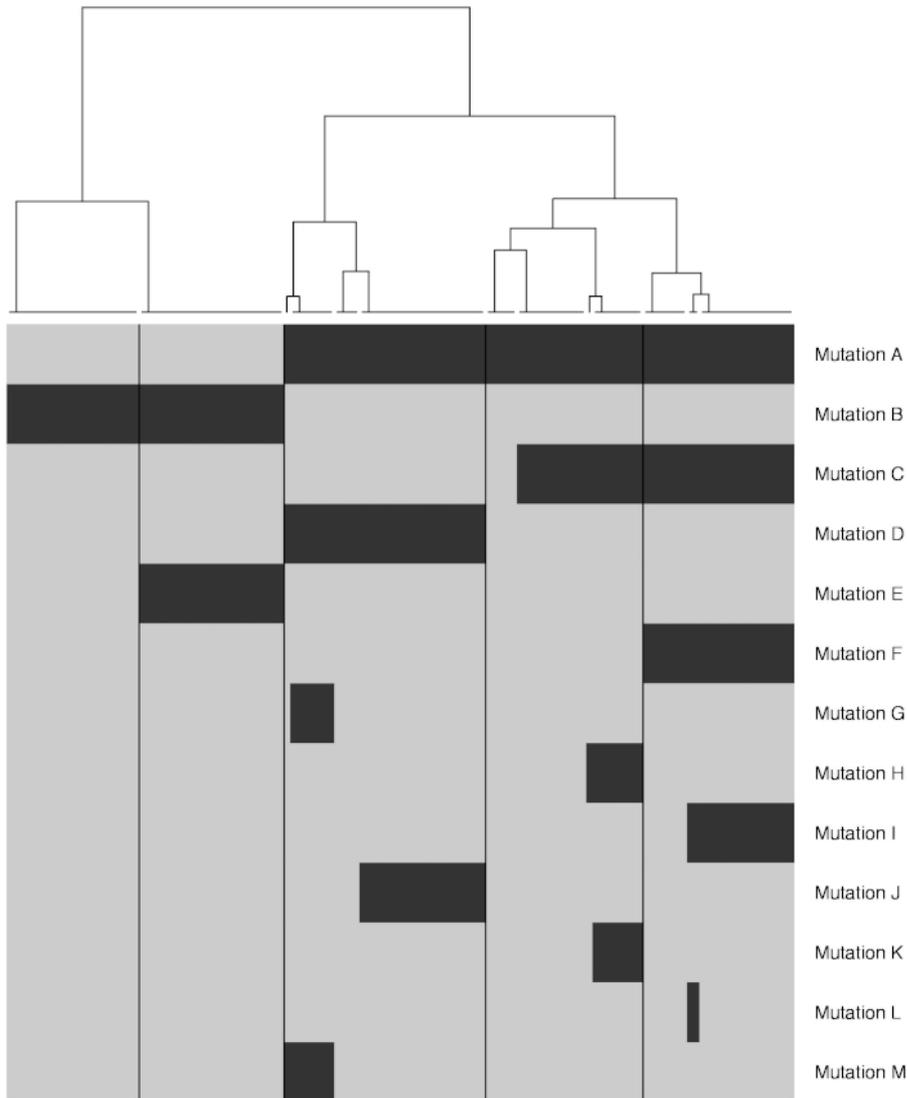
**Figure 5: Histogram of variant allele frequencies.** Histogram of variant allele frequencies of the variants in a clone before the last filtering step of SNVFI (VAF >0.3). A peak at VAF = 0.5 indicates that the sample is clonal. The subclonal mutations with low VAF are excluded during last filtering step of SNVFI (VAF >0.3). [Please click here to view a larger version of this figure.](#)



**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. [Please click here to view a larger version of this figure.](#)



**Figure 7: Manual inspection of mutations using IGV<sup>16</sup>.** (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. [Please click here to view a larger version of this figure.](#)



**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. [Please click here to view a larger version of this figure.](#)

## 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 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<sup>1</sup>. 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<sup>1,18,19</sup>. 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.

## Disclosures

The authors have nothing to disclose.

## 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.

## 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*. (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*. (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*. (2014).