university of groningen

University Medical Center Groningen

University of Groningen

Clonal analysis of patient-derived samples using cellular barcodes

Jacobs, Sabrina; Bystrykh, Leonid V.; Belderbos, Mirjam E.

Published in:
Methods in Molecular Biology

DOI:
10.1007/978-1-0716-0810-4_20

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Jacobs, S., Bystrykh, L. V., & Belderbos, M. E. (2021). Clonal analysis of patient-derived samples using cellular barcodes. In *Methods in Molecular Biology* (pp. 317-344). (Methods in Molecular Biology; Vol. 2185). Humana Press. https://doi.org/10.1007/978-1-0716-0810-4_20

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 05-06-2022

Chapter 20

Clonal Analysis of Patient-Derived Samples Using Cellular Barcodes

Sabrina Jacobs, Leonid V. Bystrykh, and Mirjam E. Belderbos

Abstract

Cellular barcoding is a relatively simple method that allows quantitative assessment of the clonal dynamics of normal, nonmalignant hematopoietic stem cells and of leukemia. Cellular barcodes are (semi-)random synthetic DNA sequences of a fixed length, which are used to uniquely mark and track cells over time. A successful barcoding experiment consists of several essential steps, including library production, transfection, transduction, barcode retrieval, and barcode data analysis. Key challenges are to obtain sufficient number of barcoded cells to conduct experiments and reliable barcode data analysis. This is especially relevant for experiments using primary leukemia cells (which are of limited availability and difficult to transduce), when studying low levels of chimerism, or when the barcoded cell population is sorted in different smaller subpopulations (e.g., lineage contribution of normal hematopoietic stem cells in murine xenografts). In these settings, retrieving accurate barcode data from low input material using standard PCR amplification techniques might be challenging and more sophisticated approaches are required. In this chapter we describe the procedures to transfect and transduce patient-derived leukemia cells, to retrieve barcoded data from both high and low input material, and to filter barcode data from sequencing noise prior to quantitative clonal analysis.

Key words Barcode, Clone, Sequencing, Clonal analysis, Leukemia

1 Introduction

The progression, chemotherapeutic resistance, and relapse of leukemia are thought to develop through a process of clonal selection and evolution [1]. Hematopoietic stem or progenitor cells can acquire genomic aberrations, which might alter essential cell functions, and develop into a wide variety of genetically and phenotypically distinct clones with different growth properties and chemotherapeutic sensitivity [2, 3]. The relapsing clone is often already present as a minor clone at diagnosis, with additional mutations upon relapse, suggestive of clonal evolution [4–7]. Most studies on the clonal evolution of leukemia rely on sequencing of naturally occurring genomic aberrations in bulk diagnostic, remission, and relapsed patient-derived samples, and use complex

César Cobaleda and Isidro Sánchez-García (eds.), *Leukemia Stem Cells: Methods and Protocols*, Methods in Molecular Biology, vol. 2185, https://doi.org/10.1007/978-1-0716-0810-4_20, © Springer Science+Business Media, LLC, part of Springer Nature 2021

mathematical models to retrospectively reconstruct the evolutionary trajectories of the retrieved clones [8, 9]. However, due to ongoing clonal evolution, it is challenging to define minor clones in leukemia using this approach [10]. In addition, this approach is barely applicable to normal, nonmalignant hematopoietic stem cells (HSCs), in which genomic aberrations hardly occur.

Viral transduction, i.e., the integration of genetic material into the host cell genome, allows to study clonal dynamics of both leukemia cells and HSCs in a prospective manner [11]. One of the first developed techniques of this type used viral integration sites as unique clonal marks. Although this technique allows to track an unlimited number of clones, it requires fragmentation of genomic DNA (gDNA) and as a consequence it has a biased PCR amplification toward shorter DNA fragments, which hampers accurate quantitative clonal analysis [12]. The integration of (semi-) random synthetic DNA sequences of fixed length (i.e., barcodes) into these vectors resulted in more accurate quantification of clones [13–15]. Cellular barcoding has been shown to be of value in the study of both normal, nonmalignant HSCs and leukemia cells [14, 16–21]. However, as long as barcode libraries are generated in a probabilistic manner (e.g., random barcode design and mixed library pool of unknown composition and size), cellular barcoding is prone to problems of correct identification of library size and content [11]. In addition, the identification of barcodes is obscured by accompanying PCR and sequencing errors. As a result, the actual number and identity of barcodes in a library can differ substantially from what is approximated. Therefore, barcode libraries should be thoughtfully designed and validated, with accurate discrimination between "true" barcodes and noise [11, 22]. In the future, synthetic, high-throughput barcode library production using robotics may provide a better strategy for barcode library production, as it allows for the production of large numbers of individual barcodes which can be pooled into libraries of certain size, composition, and complexity (i.e., total number of barcodes).

The continuously expanding genome editing toolbox [23] allows for novel possibilities to optimize the barcoding technique and to address some of its disadvantages. For instance, currently, the barcode composition of a given sample can only be determined retrospectively by PCR and sequencing, and sorting of individual barcode clones is not (yet) possible. Furthermore, cellular barcoding requires in vitro culture of target cells, which is an extra bottleneck that may result in the loss of clonogenic cells. In addition, the potential risks of lentiviral transduction and integration of (non-functional) barcode DNA into the host cell hamper the use of barcoding in humans.

A well-established, alternative method for low-resolution in vivo lineage tracking is Cre-LoxP switching color cassettes, which uses combinations of fluorescent proteins as clonal markers

[24, 25]. This method has the advantage that it can be directly applied in transgenic models using (tissue-specific) inducible Cre recombinase, and allows for sorting of individual clones [25]. However, viral transduction is still required in patient-derived material, and analysis of quantitative clonal dynamics is limited due to recombination bias and limited number of distinguishable color combinations [24–26].

A second, largely similar method replaced these color cassettes by DNA sequences [27]. Although individual clones can no longer be sorted, it increases the number of traceable clones. However, as by definition the library composition remains unknown and frequently redundant, it is still difficult to discriminate between noise and "true" barcodes. In addition, different clones can be potentially marked with identical barcode combinations [26].

A third approach is based on CRISPR-Cas9 technology, which uniquely barcodes individual cells by introducing indels ("scars") in targeted regions of the genome or in synthetic DNA sequences (e.g., GFP-repeat) via single-guided RNA [28–30]. "Scars" are first amplified by PCR followed by deep sequencing, and unique scars are then defined by the introduced insertions and deletions (indels; substitutions are a result of PCR or sequencing errors). Applying this method to study the development of zebrafish shows that—although limited—thousands of scars can be generated, with only a few clusters of scars (i.e., sharing specific ancestral indels) that contribute to the development of specific organs [29, 30]. The number of scars can be increased using a system that introduces scars into the guide RNA [31]. However, bias in the introduction of specific scars and reconstruction of the clonal signature faces a problem of low complexity and redundancy of scars. Altogether, although these recently suggested methods definitely look promising, considerable improvements are still needed to increase their efficiency and practicality. In the future, the abovementioned methods may allow modification of the barcoding method, to overcome some of its disadvantages, and to perform even more robust, quantitative clonal tracking.

Considerations for barcode design, barcode library preparation, and barcode data analysis have been discussed extensively elsewhere [11, 32]. Therefore, their essential features are only briefly mentioned here. In this chapter, we will focus on the technical procedures to introduce barcodes into target cells, specifically into patient-derived leukemia cells, and to retrieve barcode data, which can be especially challenging from low input material. Figure 1 represents a flowchart of the steps described in this chapter.

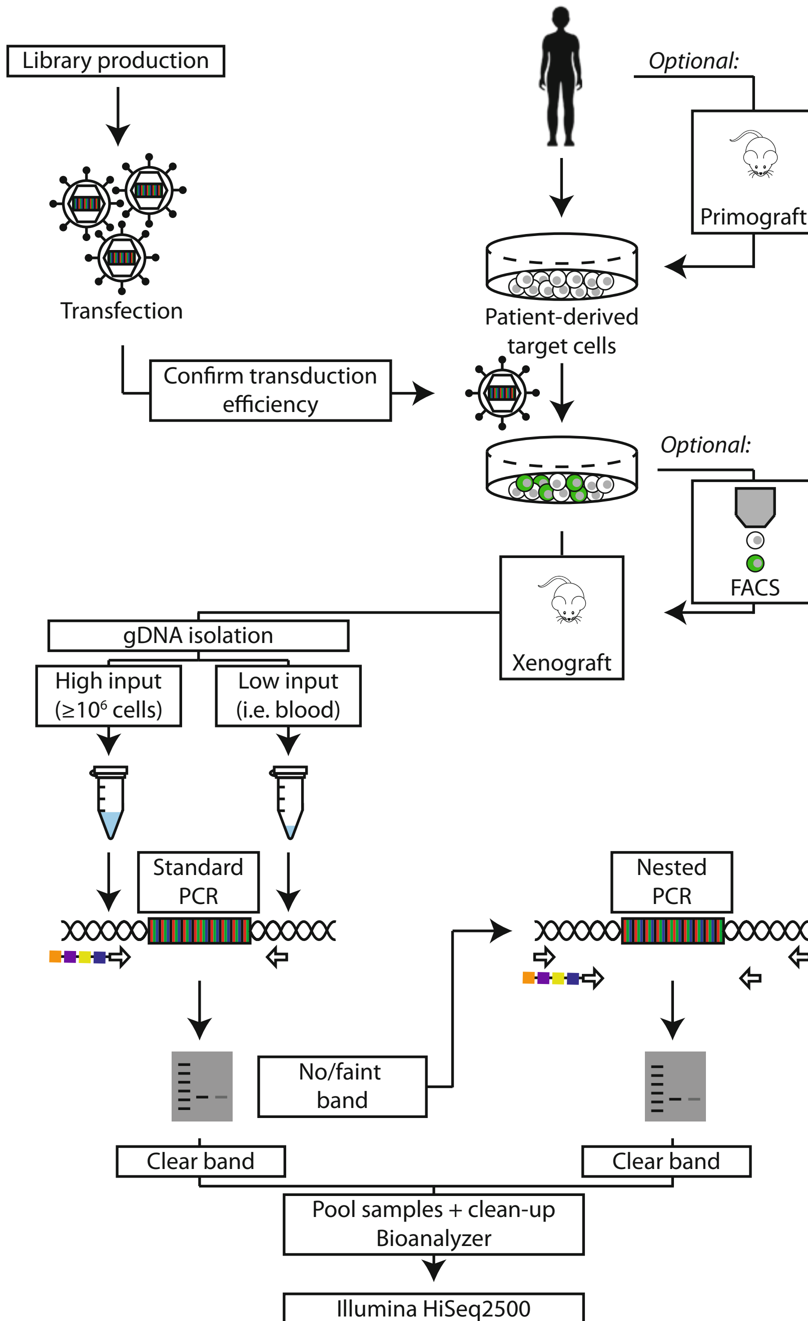

Fig. 1 Flowchart of a cellular barcoding experiment. After producing and validating the barcode library, the first step is to produce virus. Prior to transducing patient-derived leukemia cells, it is recommended to first confirm the transduction efficiency of the virus using a cell line. When this is confirmed, patient-derived leukemia cells are thawed and directly transduced. Since patient-derived leukemia cells are limited in number and difficult to transduce, they are often first transplanted and expanded in mice ("primograft"). When these mice develop leukemia, increased numbers of patient-derived cells can be harvested and transduced. The resulting barcoded cells—sorted or unsorted—can be used for in vitro or in vivo clone-tracking experiments. The

2 Materials

2.1 Transfection

1. 70% Ethanol in water.
2. HEK293FT cell line.
3. HEK293FT culture medium: DMEM supplemented with 10% heat-inactivated FCS and 1% penicillin and streptomycin.
4. Serum-free expansion medium (SFEM: Iscove's MDM, bovine serum albumin, recombinant human insulin, iron-saturated human transferrin, 2-mercaptoethanol, best purchased from a specialized supplier), supplemented with 10% heat-inactivated FCS and 1% penicillin and streptomycin.
5. 0.4% Trypan blue solution.
6. 0.1% Gelatin: Dissolve 0.5 g of gelatin type A in 500 mL Milli-Q water and autoclave. Cool down before use.
7. Dulbecco's PBS.
8. 0.05% Trypsin-EDTA.
9. Opti-MEM® I reduced serum medium.
10. Packaging plasmid: pCMV Δ8.91.
11. Envelope plasmid: VSV-G.
12. Vector construct (pEGZ2 B322 barcode library).
13. FuGENE® HD transfection reagent.
14. Disposables: T75 culture flasks, 15 mL collection tubes, 50 mL collection tubes, siliconized Eppendorf tubes, 20 mL syringes, 0.45 μm Millex HV low-protein-binding filters and cryovials.
15. Instruments: Hemocytometer, centrifuge, autoclave, vortex, and ML2-level cell culture facility.

2.2 Transduction

1. 70% Ethanol in water.
2. SupB15 cell line (ATCC®).
3. Patient-derived progenitor B-cell acute lymphoblastic leukemia cells (B-ALL).
4. SupB15 culture medium: RPMI 1640 medium supplemented with 10% heat-inactivated FCS and 1% penicillin and streptomycin.

Fig. 1 (continued) first step towards assessing clonal complexity is the isolation of gDNA. Depending on the input material (high vs. low), different gDNA isolation kits can be used. Barcode sequences are amplified by standard PCR, which is confirmed on an agarose gel. Samples that show no band or a faint band can be repeated, or subjected to nested PCR. Samples that show a clear band can be cleaned up and pooled together in batches of 200–300 samples. Quality of the sample is confirmed on the BioAnalyzer, after which the sample is sent for deep sequencing

5. B-ALL culture medium: SFEM supplemented with 10% heat-inactivated FCS, 1% penicillin and streptomycin, 100 ng/mL human recombinant thrombopoietin (TPO), 10 ng/mL human recombinant IL-7, 20 ng/mL human recombinant Fms-related tyrosine kinase 3 ligand (FLT-3 L), and 50 ng/mL human recombinant stem cell factor (SCF).
6. 0.4% Trypan blue solution.
7. RetroNectin® (Takara) in PBS at a final concentration of 0.025 mg/mL. Stock of 2.5 mg is dissolved in a total volume of 100 mL PBS. Prepare aliquots and store at −20 °C.
8. Propidium iodide (PI).
9. SFEM supplemented with 1% penicillin and streptomycin.
10. SFEM supplemented with 10% heat-inactivated FCS and 1% penicillin and streptomycin.
11. Dulbecco's PBS.
12. Dulbecco's PBS supplemented with 0.2% or 2.0% bovine albumin fraction V (7.5% stock solution).
13. Dulbecco's PBS supplemented with 20%, 10%, and 5% heat-inactivated FCS.
14. Disposables: T75 culture flasks, 6-well culture plates, 12-well culture plates, 15 mL collection tubes, 50 mL collection tubes, FACS tubes, parafilm, and cell scrapers.
15. Instruments: Hemocytometer, centrifuge, flow cytometer with blue (488 nm, eGFP) and yellow (561–568 nm, PI) laser, and ML2-level cell culture facility.

2.3 DNA Isolation

1. 70% Ethanol in water.
2. Molecular BioProducts™ RNase away™ surface decontaminant.
3. Dulbecco's PBS.
4. Ethanol absolute (96–100%).
5. DNA isolation kit for high input material: DNeasy Blood and Tissue (Qiagen):
   (a) Add the appropriate amount of ethanol absolute (96–100%) to buffer AW1 and AW2 to obtain the working solution.
6. DNA isolation kit for low input material: QIAamp DNA micro kit (Qiagen):
   (a) Add the appropriate amount of ethanol absolute (96–100%) to buffer AW1 and AW2 to obtain the working solution.

7. Disposables: Sterile 1.5 mL Eppendorf collection tubes for elution.
8. Instruments: Heat block, microcentrifuge, and Nanodrop 2000 spectrophotometer.

2.4 PCR

1. Ethanol 70% in water.
2. RNase away™ surface decontaminant.
3. Oligonucleotide primers: Primers are diluted to a stock concentration of 100 μM in nuclease-free H_2O. Primers are stored at −20 °C:
   (a) Outer primer set flanking the barcode sequence and the priming region of the second set of primers.
   (b) Indexed forward primer and the universal reverse primer flanking the barcode sequence.
4. DreamTaq Green PCR Master Mix (2×), including nuclease-free water.
5. Disposables: Sterile 1.5 mL Eppendorf tubes and sterile PCR strips.
6. Instruments: UV3 HEPA PCR workstation and thermal cycler with heated lid.

2.5 Agarose Gel

1. TAE buffer (50×): Dissolve 242 g of Tris base in 700 mL of Milli-Q water. Add 57.1 mL of acetic acid and 100 mL of EDTA (0.5 μM). Add up to 1 L with Milli-Q water.
2. DNA Ladder Mix, ready to use (100–10,000 bp).
3. Instruments: Gel tray, well combs, electrophoresis tank including power supply, and UV transilluminator.

2.6 PCR Product Purification

1. QiaQuick PCR purification kit (Qiagen):
   (a) Add the appropriate amount of ethanol absolute (96–100%) to buffer PE to obtain the working solution.
2. Disposables: Sterile 1.5 mL Eppendorf collection tubes for elution.
3. Instruments: Microcentrifuge.

2.7 PCR Product Quality Control

1. Qubit® dsDNA HS assay kit (Thermo Fisher Scientific).
2. Agilent High Sensitivity DNA kit (Agilent Technologies).
3. Disposables: Qubit™ Assay Tubes (Thermo Fisher Scientific).
4. Instruments: Qubit® 2.0 Fluorometer, Agilent Chip Priming Station, IKA model MS3 vortex mixer, and Agilent 2100 Bioanalyzer System.

3 Methods

3.1 Considerations for Barcode Library Production

Although the concept of barcode design and library production was discussed extensively in previous publications, here we would like to emphasize a few of it's essential features [11, 32]. First, any barcoding method relies on some combinatorial principles. For instance, any semi-random barcode is a stretch of DNA that consists of variable and fixed nucleotides. Accordingly, the size of the barcode library (i.e., its maximum complexity) is limited by the theoretical number of combinations, which is proportional to the number of variable nucleotide positions in the barcode. It is important to realize that the size of any experimental barcode library is smaller than its theoretical maximal size, as it is always a subset of the combinations. Accordingly, the subset defines the true library size and determines the distance between barcodes, which can be validated experimentally [14]. In addition, every barcode in the library should be of equal probability of occurrence. Most approaches for lineage tracking, like cellular barcoding, rely on random processes. We assume random synthesis and transduction of barcode DNA sequences, random introduction of scars, and random generation of color combinations in alternative lineage tracking approaches. Although randomness was confirmed for cellular barcoding, in reality we often face the fact of nonrandomness (e.g., Cre-LoxP recombination bias and nonrandom introduction of scars) [14, 16, 26, 30]. Such nonrandomness severely decreases the complexity of the library, and should be taken into account when reporting the final results.

3.2 Transfection

Note that all steps are performed in a ML2-level laboratory and that a GMO permit is required.

We previously reported protocols for making retro- and lentiviral barcode libraries based on semi-random barcode tags integrated into the viral vector backbone [11, 32]. Because of the pitfalls mentioned above, we advocate to use a library of known size, content, and complexity. To generate such a library, we subcloned individual barcode combinations as separate *E. coli* preps, and collected approximately 800 barcoded vectors in the freezer. Depending on the experimental aim, these barcodes can be pooled in equimolar ratios to libraries of the desired complexity.

After validating the barcode library, the library needs to be incorporated into viral particles, which are used to transduce patient-derived (leukemia) cells. The transduction efficiency of these viral particles depends on multiple factors. Obviously, the produced virus should have a suitable envelope protein and a sufficient viral titer. The use of healthy, low-passage HEK293FT cells improves the viral titer. The transduction efficiency can be further facilitated by improving cell-virus contact (e.g.,

RetroNectin®, polybrene, and/or spinfection) and by removing viral inhibitors (e.g., pre-coating of culture plates with virus or use of purification kits).

3.2.1 Day-7: Thaw HEK293FT Cells

1. Pre-warm culture medium.
2. Thaw cells rapidly, and resuspend in 10 mL of culture medium.
3. Centrifuge cells at 450 × *g* for 5 min at 5 °C and discard the supernatant.
4. Resuspend cells in culture medium and count the number of living cells using trypan blue.
5. Grow cells at a concentration of 0.25×10^6 cells/mL in 10 mL of culture medium in a T75 culture flask.
6. Incubate at 37 °C and 5% CO_2.
7. Passage cells 2–3 times a week.
8. To this aim, remove the culture medium from each T75 culture flask and gently rinse with 10 mL of PBS.
9. Remove PBS and add 2 mL of trypsin-EDTA (0.05%) to the bottom of the standing T75 culture flask. Gently swirl to cover the attached HEK293FT cells and directly remove trypsin.
10. Incubate the T75 culture flask at 37 °C and 5% CO_2 for 5 min to allow the cells to detach.
11. Resuspend the cells in 10 mL of culture medium and transfer to a 50 mL tube. Note that multiple T75 cell culture flasks can be combined.
12. Centrifuge cells at 450 × *g* for 5 min at 5 °C and discard supernatant.
13. Resuspend the cells in 10 mL of culture medium and count the number of living cells using trypan blue.
14. Grow cells at a concentration of 0.25×10^6 cells/mL in 10 mL of culture medium in a T75 culture flask.
15. Incubate at 37 °C and 5% CO_2.

3.2.2 Day 0: Plate HEK293FT Cells

1. Pre-coat the required number of T75 culture flasks with 10 mL of 0.1% gelatin for 2 h at 37 °C (*see* **Note 1**). Every T75 culture flask will yield approximately 5 mL of virus.
2. In the last ~60 min before starting transfection, collect HEK293FT cells.
3. To this aim, first remove the culture medium from each T75 culture flask and gently rinse with 10 mL of PBS.
4. Remove PBS and add 2 mL of trypsin-EDTA (0.05%) to the bottom of the standing T75 culture flask. Gently swirl to cover the attached HEK293FT cells and directly remove trypsin.

5. Incubate the T75 culture flask at 37 °C and 5% CO_2 for 5 min to allow the cells to detach.
6. Resuspend the cells in 10 mL of culture medium and transfer to a 50 mL tube. Multiple T75 culture flasks can be combined.
7. Centrifuge cells at 450 × *g* for 5 min at 5 °C and discard supernatant.
8. Resuspend the cells in 10 mL of culture medium and count the number of living cells using trypan blue.
9. Dilute the cells to a concentration of 0.15×10^6 cells/mL in culture medium.
10. Next, remove the gelatin from the T75 culture flasks.
11. Rinse the T75 culture flasks with 10 mL of PBS.
12. Grow the cells at a concentration of 0.15×10^6 cells/mL in 10 mL of culture medium in the gelatin-coated T75 culture flasks.
13. Incubate for 2 days at 37 °C and 5% CO_2 to allow the cells to attach.

3.2.3 Day 2: Transfection

1. Prior to starting the transfection, let Opti-MEM™ and FuGENE® reach room temperature.
2. For *n* × T75 culture flasks, label *n* x siliconized Eppendorf tubes with "tube 1." Label one siliconized Eppendorf tube with "tube 2" (*see* **Note 2**).
3. Add 400 μL of Opti-MEM™ to each "tube 1."
4. Prepare "tube 2" for *n* × T75 culture flasks according to Table 1.
5. Gently vortex "tube 2" and transfer the corresponding volume for one T75 culture flask of "tube 2" to "tube 1."
6. Vortex FuGENE® and add 21 μL directly into the medium of each "tube 1".
7. Mix gently by ticking (not vortexing) and incubate at room temperature for 15 min to allow the transfection precipitates to be formed.

Table 1
Preparations of "tube 2." Required amounts to transfect one T75 culture flask

Opti-MEM™	100 μL
Packaging plasmid (pCMV Δ8.91)	3 μg
Envelope plasmid (VSV-G)	0.7 μg
Vector construct (pEGZ B322 barcode library)	3 μg

8. Dropwise and gently transfer the content of one "tube 1" to each T75 culture flask.
9. Gently swirl and incubate overnight at 37 °C and 5% CO_2.

3.2.4 Day 3: Medium Change

1. Carefully replace the medium in the T75 culture flask by 7.5 mL of serum-free medium (e.g., SFEM) for the target cells supplemented with 1% penicillin and streptomycin.
2. Incubate overnight at 37 °C and 5% CO_2.

3.2.5 Day 4: Harvest Virus

1. Carefully collect a maximum of 15 mL virus supernatant from every T75 culture flask into a 50 mL tube.
2. To this, collect 15 mL of virus in a 20 mL syringe and filter virus through a 0.45 micron Millex-HV low-protein-binding filter into a new, clean 50 mL tube.

   Optional: In case the HEK293FT cells detach, it is recommended to centrifuge at 450 × *g* for 5 min at room temperature to prevent clogging of the filter.
3. Aliquot filtered virus into cryovials and store at −80 °C.

3.3 Transduction of a Cell Line: Quality Control of Produced Virus

Note that all steps are performed in a ML2-level laboratory and that a GMO permit is required.

It is recommended to confirm the quality of the virus using a representative cell line (e.g., SupB15 B-acute lymphoblastic leukemia cell line in case of experiments with primary B-ALL cells) prior to transducing patient-derived target cells. Irrespective of the cell line used, the cells should be low in passage and recovered from thawing before transduction.

3.3.1 Day-7: Thaw SupB15 Cells

1. Pre-warm culture medium.
2. Thaw cells rapidly and collect in 10 mL of culture medium.
3. Centrifuge at 450 × *g* for 5 min at 5 °C and discard the supernatant.
4. Resuspend the cells in culture medium and count the number of living cells using trypan blue.
5. Grow cells at a concentration of 0.25×10^6 cells/mL in 15 mL of culture medium in a T75 culture flask.
6. Incubate overnight at 37 °C and 5% CO_2.

3.3.2 Day-6: Refresh Culture Medium

1. Collect the SupB15 cells in a 50 mL tube.
2. Rinse the T75 culture flask with 5 mL culture medium to collect the remaining cells and repeat if required.
3. Centrifuge at 450 × *g* for 5 min at 5 °C and discard the medium.

4. Resuspend the cells in culture medium and count the number of living cells using trypan blue.
5. Grow cells at a concentration of 0.25×10^6 cells/mL in 15 mL of culture medium in a T75 culture flask.
6. Passage cells 2–3 times a week at a concentration of 0.25×10^6 cells/mL.

3.3.3 Day 0: Pre-coat Wells with RetroNectin®

1. Add 1 mL of RetroNectin® per well to 6 wells of a 12-well culture plate. The other 6 wells will not be used.
2. Seal the plate with parafilm and incubate overnight at 4 °C.

3.3.4 Day 1: Transduction

1. Thaw the required volume of virus (see below).
2. Remove RetroNectin® from the 12-well culture plate and add 1 mL of PBS supplemented with 2% bovine albumin to each of the 6 wells to prevent nonspecific binding (*see* **Note 3**).
3. Incubate for 30 min at room temperature.
4. Remove PBS supplemented with 2% bovine albumin and rinse wells with PBS.
5. Add 500 μL, 250 μL, 125 μL, 62.5 μL, 31.3 μL, or 15.6 μL of virus to the wells and add up to 500 μL with viral collection medium (i.e., SFEM supplemented with 0.1% penicillin and streptomycin).
6. To promote binding of the virus to the RetroNectin® coating, centrifuge at $1000 \times g$ for 45 min at room temperature (acceleration set at 1, brake set at 0).
7. Incubate for another 4 h at 37 °C and 5% CO_2.
8. In the last ~60 min, before continuing with the transduction protocol, collect SupB15 cells:
9. To this aim, first collect cells into 15 mL falcon tubes, and centrifuge at $450 \times g$ for 5 min at 5 °C.
10. Discard supernatant and resuspend cells in 10 mL culture medium. Reduce the volume if low cell numbers are expected.
11. Count the number of living cells using trypan blue and dilute the cells to a concentration of 2.5×10^6 cells/mL in culture medium.
12. Remove the viral supernatant from the 12-well culture plate.
13. Gently rinse each of the 6 wells with 1 mL PBS supplemented with 2% bovine albumin.
14. Replace the PBS supplemented with 2% bovine albumin in the 12-well culture plate by 800 μL of SupB15 cells (2×10^6 cells/well). Prevent the wells from drying out during this step.

15. Perform a spinfection by centrifuging the plate at 900 × *g* for 45 min at room temperature (acceleration set at 1, brake set at 0).
16. Incubate overnight at 37 °C and 5% CO_2.

3.3.5 Day 2: Remove Virus

1. Gently collect the transduced SupB15 cells from each of the 6 wells of the 12-well culture plate and transfer to individual 50 mL tubes.
2. Add 1 mL of culture medium to each of the 6 wells of the 12-well culture plate and gently detach cells from RetroNectin® by scraping. Transfer the cells to the corresponding 50 mL tube and check whether all SupB15 cells are collected. Repeat if required.
3. Centrifuge the cells at 450 × *g* for 5 min at room temperature and discard supernatant.
4. Resuspend the cells in 50 mL of PBS supplemented with 20% FCS, centrifuge at 450 × *g* for 5 min at room temperature, and discard supernatant. Repeat this for PBS supplemented with 10% FCS and 5% FCS.
5. Resuspend the cells in culture medium and count the number of living cells using trypan blue.
6. Grow the cells at a concentration of ~1 × 10^6 cells/mL in a suitable flask or culture plate.
7. Incubate overnight at 37 °C and 5% CO_2.

3.3.6 Day 3: Determine Transduction Efficiency

1. Resuspend the transduced cells (keep cells transduced with different viral volumes separate) and transfer 10 μL to Eppendorf tube to count the number of living cells using trypan blue.
2. Transfer 0.2–0.5 × 10^6 SupB15 cells to a FACS tube, centrifuge at 450 × *g* for 5 min at room temperature, and discard supernatant.
3. Wash the cells with 3 mL of PBS supplemented with 0.2% bovine albumin, centrifuge at 450 × *g* for 5 min at room temperature, and discard supernatant.
4. Resuspend the cells in 200 μL PBS supplemented with 0.2% bovine albumin and add 100 μL PI to discriminate between cells that are alive or dead.
5. Measure samples at the flow cytometer and determine the percentage of GFP^+ PI^- cells. Use GFP^- and GFP^+ cells to set the gates.

3.4 Transduction of Patient-Derived B-ALL Cells

When the quality of the produced virus is validated, and a sufficient transduction efficiency of at least 10% on a cell line is reached, the virus can be used to transduce patient-derived B-ALL cells. Here, it

is important to optimize the transduction efficiency, by titrating the amount of virus added to the target cells. On the one hand, the transduction efficiency should be sufficiently high to produce enough barcoded cells for experimental purposes. On the other hand, if the transduction efficiency is too high, this will increase the probability of having multiple barcode integrations in a single cell. This probability should be assessed both theoretically and experimentally. Theoretically, viral transduction follows a Poisson distribution, allowing to assess the probability of multiple integrations for any given transduction efficiency [33]. In our experiments, we generally aim for a transduction efficiency of ~10%, resulting in 0.5% chance of integrating multiple barcode vectors into one cell. However, in practice, the transduction efficiency of patient-derived cells may be substantially lower, as it is difficult to maintain the viability of patient-derived B-ALL cells in vitro. As a consequence, it is challenging to obtain a sufficient yield of barcoded patient-derived B-ALL cells. Therefore, the time of in vitro culture should be limited and it might be necessary to expand the number of patient-derived B-ALL cells via transplantation into sublethally irradiated (1.0 Gy) Nod/SCID/IL2R$\gamma^{-/-}$ mice, either before or after barcoding [16, 21].

3.4.1 Day 0: Pre-coat Wells with RetroNectin®

1. Add 2 mL of RetroNectin® to each well of a 6-well culture plate (*see* **Note 3**). Prepare three plates in total.
2. Seal plate with parafilm and incubate overnight at 4 °C.

3.4.2 Day 1: Thaw and Transduce Patient-Derived B-ALL Cells

1. Thaw the required volume of virus.
2. Remove RetroNectin® from the 6-well culture plate and add 2 mL of PBS supplemented with 2% bovine albumin to each well to prevent a specific binding.
3. Incubate for 30 min at room temperature.
4. Remove PBS supplemented with 2% bovine albumin and rinse each well with 2 mL PBS.
5. Add 1.5 mL of virus to each well of one 6-well culture plate, add 1.0 mL or 0.5 mL of virus to each well of the other 6-well cell culture plates, and add up to 1.5 mL with SFEM supplemented with 0.1% penicillin and streptomycin (*see* **Note 4**).
6. To promote binding of the virus to the RetroNectin® coating, centrifuge at 900 × *g* for 45 min at room temperature (acceleration set at 1, brake set at 0).
7. Incubate for another 4 h at 37 °C and 5% CO_2.
8. In the last ~60 min, before continuing with the transduction protocol, thaw patient-derived B-ALL cells.
9. To this aim, first pre-warm B-ALL culture medium.

10. Rapidly thaw patient-derived B-ALL cells and collect in 10 mL of B-ALL culture medium without added cytokines.
11. Centrifuge cells at 500 × *g* for 10 min at 5 °C and discard supernatant.
12. Resuspend cells in 10 mL B-ALL medium without added cytokines and count cells using trypan blue.
13. Centrifuge the cells at 500 × *g* for 10 min at 5 °C and discard supernatant.
14. Resuspend cells in B-ALL cell culture medium (concentration: 1.0×10^6–2.0×10^6 cells/mL).
15. Remove viral supernatant from the 6-well culture plates.
16. Gently rinse each well of the 6-well culture plate with 2 mL PBS supplemented with 2% bovine albumin.
17. Replace the PBS supplemented with 2% bovine albumin by 2 mL (=2.0–4.0×10^6 cells/well) of patient-derived B-ALL cells. Prevent the wells from drying out during this step.
18. Perform a spinfection by centrifuging the cells at 900 × *g* for 45 min at room temperature (acceleration set at 1, brake set at 0).
19. Incubate overnight at 37 °C and 5% CO_2.

3.4.3 Day 2: Remove Virus

Remove the virus as described in Subheading 3.3.5. As patient-derived B-ALL cells are difficult to maintain in vitro, the number of viable cells will decrease substantially during the transduction procedure. Cells can be cultured at a concentration between 0.5×10^6 and 1.0×10^6 cells/mL, in a suitable culture flask or dish.

3.4.4 Day 3: Determine Transduction Efficiency

Determine the transduction efficiency as described in Subheading 3.3.6. After confirming the transduction efficiency by flow cytometry, barcoded, live cells can be sorted for GFP^+ PI^-. Depending on the experimental question, sorted/unsorted barcoded cells can be transplanted into sublethally irradiated Nod/SCID/IL2R$\gamma^{-/-}$ mice.

3.5 Barcode Retrieval by Next-Generation Sequencing

The clonal dynamics of the transplanted barcoded cells in vivo can be assessed by barcode analysis on longitudinally acquired blood samples. However, several factors may limit the number of GFP^+-barcoded cells available for analysis. First, nonterminal blood collection from mice is limited to approximately 200 μL once every 3–4 weeks [34]. Furthermore, especially at early time points, the levels of human GFP^+ chimerism may be low. Finally, in certain experiments, one may want to sort different hematopoietic cell populations from either blood or bone marrow, which might result in limited cell numbers as well. Low cell numbers, and as a consequence low copy numbers of the barcode sequence, might hamper

successful barcode retrieval by high-throughput sequencing. Therefore, efficient strategies for DNA isolation and PCR amplification are needed, as described below.

3.5.1 Isolation of Genomic DNA

Successful barcode retrieval depends on the copy number of the barcode sequence and on the quality of the isolated gDNA. There are a wide variety of kits to isolate high-quality gDNA, which are constantly improving. The selected kit should be applicable to the number of cells from which gDNA can be isolated. In our experimental design, we use the Qiagen DNeasy Blood and Tissue for high input material ($\geq 1 \times 10^6$ cells). To prevent dilution of the barcode sequence, we use the Qiagen QIAamp DNA micro kit with a smaller elution volume for low input material (i.e., blood samples and samples with $<10^6$ cells). Prior to isolation of gDNA, thoroughly clean the bench and all required equipment with 70% ethanol followed by RNase away™ surface decontaminant. Set the heat block at 56 °C and perform all procedures according to the manufacturer's instructions. It is important to perform DNA isolations in a clean, low-copy working environment. To further prevent cross-contamination, it is recommended to perform a maximum of two rounds of DNA isolation (2×24 samples) each per day.

3.5.2 Barcode Amplification by Standard PCR

To have a sufficient amount of unbiased barcode sequences for next-generation sequencing, isolated gDNA is amplified by PCR in (at least) duplicate. Since the sequencing depth greatly exceeds the number of required reads for barcode analysis of a single sample, it is recommended to use uniquely indexed forward primers. This allows multiplexing multiple samples into one sequencing run, which is time saving and cost effective. An optimal design of uniquely indexed forward primers, which has been extensively described in a previous publication [35] will facilitate the de-multiplexing procedure.

Perform all steps in a clean, low-copy working environment, and preferably prepare the PCR master mix (**steps 2–5**) in a UV3 HEPA PCR workstation. Always include a positive and negative control to confirm accuracy of the PCR reaction.

1. Create a PCR preparation list, in which each PCR reaction tube has a corresponding sample of gDNA and an indexed forward primer (Fig. 2).
2. Dilute each indexed forward primer to a working concentration of 5 μM in nuclease-free water (*see* **Note 5**).
3. Prepare the PCR master mix for *n* samples in duplicate according to Table 2.
4. Transfer 16.5 μL of PCR master mix to a PCR reaction tube (or plate).

Column 1	Column 2	Column 3
1	9	17
eGFPfwd3_001tag	eGFPfwd3_015tag	eGFPfwd3_024tag
[Sample information]	**[Sample information]**	**[Sample information]**
[Sample information]	**[Sample information]**	**[Sample information]**
2	10	18
eGFPfwd3_002tag	eGFPfwd3_016tag	eGFPfwd3_026tag
[Sample information]	**[Sample information]**	**[Sample information]**
[Sample information]	**[Sample information]**	**[Sample information]**
3	11	19
eGFPfwd3_005tag	eGFPfwd3_018tag	eGFPfwd3_027tag
[Sample information]	**[Sample information]**	**[Sample information]**
[Sample information]	**[Sample information]**	**[Sample information]**
4	12	20
eGFPfwd3_006tag	eGFPfwd3_019tag	eGFPfwd3_028tag
[Sample information]	**[Sample information]**	**[Sample information]**
[Sample information]	**[Sample information]**	**[Sample information]**
5	13	21
eGFPfwd3_007tag	eGFPfwd3_020tag	eGFPfwd3_029tag
[Sample information]	**[Sample information]**	**[Sample information]**
[Sample information]	**[Sample information]**	**[Sample information]**
6	14	22
eGFPfwd3_008tag	eGFPfwd3_021tag	eGFPfwd3_030tag
[Sample information]	**[Sample information]**	**[Sample information]**
[Sample information]	**[Sample information]**	**[Sample information]**
7	15	Negative control
eGFPfwd3_009tag	eGFPfwd3_022tag	eGFPfwd3_030tag
[Sample information]	**[Sample information]**	
[Sample information]	**[Sample information]**	
8	16	Positive control
eGFPfwd3_011tag	eGFPfwd3_023tag	eGFPfwd3_030tag
[Sample information]	**[Sample information]**	
[Sample information]	**[Sample information]**	

Fig. 2 PCR preparation list. Outline of the first 22 samples, positive and negative control of a PCR preparation list. Depending on the number of samples, this list can be extended. Every four rows of a column refer to one sample, as indicated by the number in the first row. The second row refers to the unique indexed forward primer. The third and fourth rows can be used to provide sample information. It is recommended to aliquot the indexed forward primers in 8-tube strips or a 96-well format, according to the PCR preparation list. This will create a manageable experimental setup

Table 2
Preparation of a standard PCR master mix for barcode amplification

DreamTaq Green PCR Master Mix (2×)	10.0 μL
H_2O, nuclease free	6.45 μL
Universal reverse primer (100 μM)	0.05 μL

Required amounts for one PCR reaction

Table 3
Thermal cycling conditions for a standard PCR reaction

Step	Temperature (°C)	Time	Number of cycles
Initial denaturation	95	2 min	1
Denaturation	95	30 s	
Annealing	58	30 s	35
Extension	72	1 min	
Final extension	72	10 min	1

5. Add 1 μL of the indexed forward primer (5 μM) to the corresponding PCR reaction tube (or plate).
6. Add 2.5 μL of gDNA to the corresponding PCR reaction tube or plate (total volume 20 μL/tube or plate).
7. Perform a 35-cycle PCR reaction with heated lid according to the thermal cycling conditions in Table 3 (*see* **Note** 7).
8. Confirm amplification of the barcode sequence by running the PCR product on a 1.5% agarose gel. The percentage of the gel is dependent on the expected product size.
9. Transfer the gel tray into the electrophoresis tank, fill with TAE buffer (1×), and remove the well combs from the gel.
10. Load 5 μL of ladder into the gel. The choice of ladder depends on the expected product size.
11. Combine duplicate PCR reactions and load 10 μL into the gel.
12. Run the gel at 90 V for ~30 min. Extend running time if required.
13. Visualize the gel using a UV transilluminator. Samples that show a clear PCR product can be stored at −20 °C. The samples that do not show a clear PCR product can be repeated or subjected to nested PCR.

3.5.3 Barcode Amplification by Nested PCR

For samples with low copy numbers of the barcode sequence, the standard PCR amplification as described in the previous paragraph might not be potent enough to produce sufficient amounts of product for sequencing. To circumvent this problem, cells can be monoclonally expanded [14]. Although this is an option for HSCs, it is more challenging for patient-derived B-ALL cells, as these cells are difficult to maintain in vitro. While it is tempting to increase the number of cycles of the barcode PCR reaction, it is not recommended to perform more than 40 cycles. An increase in the number of cycles will decrease the fidelity of the DNA polymerase, resulting in mispriming of the primers, with subsequent increase in the number of mutations and nonspecific PCR products. Nested PCR is an alternative approach to the standard PCR procedure, that will

increase the yield without jeopardizing the product specificity [36]. In this method, two rounds of thermal cycling are performed with two sets of PCR primers. The first set of primers flank the barcode sequence and the priming region of the second set of primers. The second set of primers are the indexed forward primer and the universal reverse primer, identical to the primers used in the standard PCR procedure.

To perform a successful nested PCR, it is important to first determine the optimal balance between the concentrations of the first and second set of primers. In general, the product from the first PCR reaction is used as input material for the second (nested) PCR reaction, without cleanup step. When the concentrations of the first set of primers are too high, they will interfere with the second PCR reaction, creating three additional products besides the target product. Therefore, the concentration of the first set of primers should be titrated and be lower compared to the second set of primers. In addition, variable amounts of input DNA for the second PCR reaction can be tested (Fig. 3).

Perform all steps for the first PCR reaction and the preparation of the PCR master mix for the second PCR reaction in a clean, low-copy working environment. Preferably prepare the PCR master mix (**steps 1–3**) in a UV3 HEPA PCR workstation. Make sure to prepare the second PCR master mix before handling high-copy material. Always include a positive and negative control to confirm the accuracy of the PCR reaction.

1. Prepare working concentrations (w.c.) for the first primer set in nuclease-free H_2O:
   (a) 5 μM w.c.: 1 μL forward primer (100 μM) + 1 μL reverse primer (100 μM) + 18 μL H_2O
   (b) 2.5 μM w.c.: 10 μL of 5 μM w.c. + 10 μL H_2O
   (c) 1.25 μM w.c.: 10 μL of 2.5 μM w.c. + 10 μL H_2O
   (d) 0.63 μM w.c.: 10 μL of 1.25 μM w.c. + 10 μL H_2O
   (e) 0.31 μM w.c.: 10 μL of 0.63 μM w.c. + 10 μL H_2O
2. Prepare the first PCR reaction master mix for *n* samples in duplicate for each primer set concentration according to Table 4.
3. Transfer 17.5 μL of PCR master mix to a PCR reaction tube.
4. Titrate a concentration of gDNA, derived from barcoded cells, that is comparable to the expected experimental sample (e.g., 25 ng/μL) in nuclease-free H_2O.
5. To this aim, first transfer 1×10^6 barcoded cells to an Eppendorf tube (*see* **Note 6**).
6. Centrifuge the cells at $900 \times g$ for 5 min at 5 °C.
7. Remove supernatant.

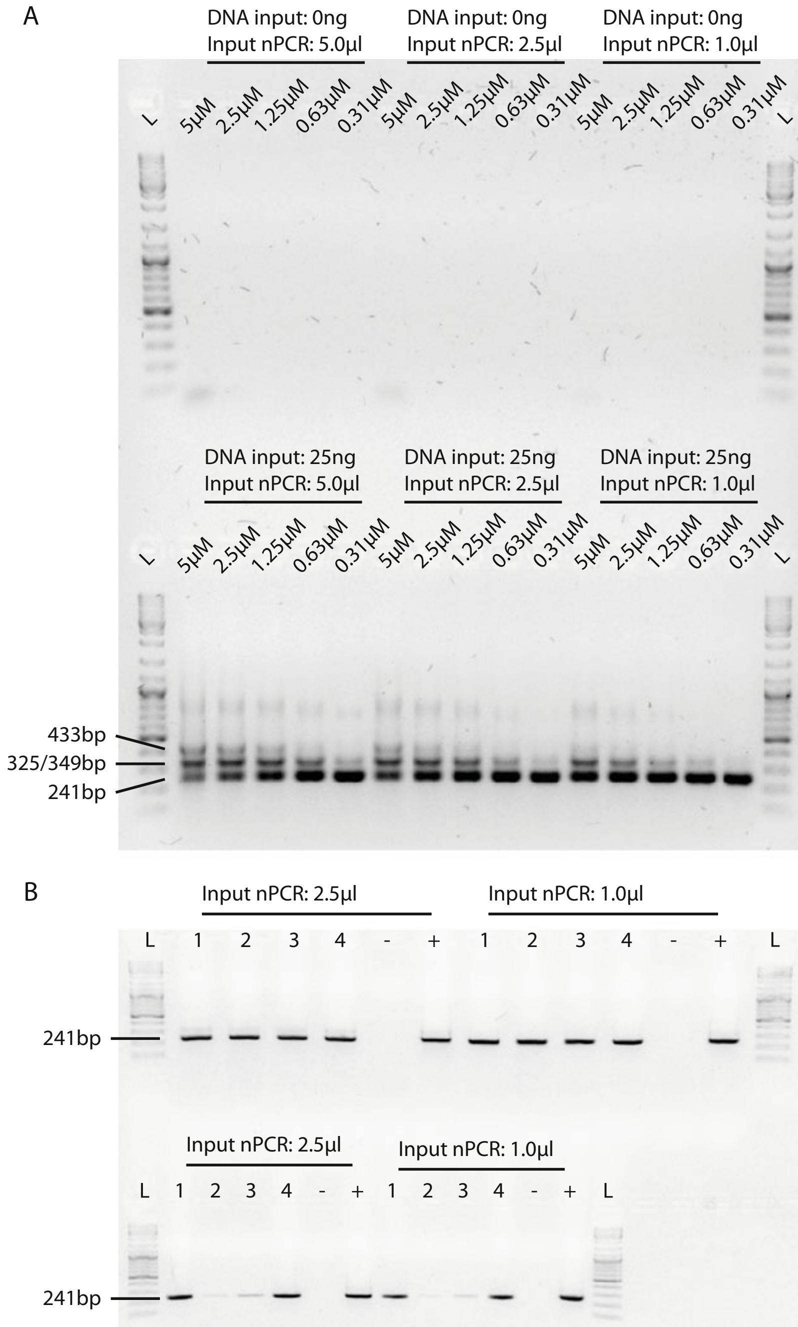

Fig. 3 Nested PCRs. (**a**) Primer titration for nested PCR. Isolated gDNA of barcoded cells is diluted to a concentration of 25 ng/µL. In the first thermal cycling round of nested PCR, 2.5 µL of 25 ng/µL DNA was used as input material and different concentrations of primers were used (5 µM, 2.5 µM, 1.25 µM, 0.63 µM, and 0.31 µM). As a negative control, nuclease-free H_2O (i.e., 0 ng/µL) was used as input material. After the first thermal cycling round of nested PCR, either 5 µL, 2.5 µL, or 1.0 µL was used as input material for the second thermal cycling round of nested PCR. Barcode sequence amplification was confirmed on an agarose gel. As expected, when the concentration of the first set of primers was too high, additional products were formed

Table 4
Preparation of master mix for the first thermal cycling round of nested PCR. Required amounts are for one PCR reaction

DreamTaq Green PCR Master Mix (2×)	10.0 μL
H_2O, nuclease free	5.5 μL
Forward and reverse primers (5 μM, 2.5 μM, 1.25 μM, 0.63 μM, 0.31 μM)	2.0 μL

Table 5
Thermal cycling conditions for nested PCR reaction

Step	Temperature (°C)	Time	Number of cycles
Initial denaturation	95	2 min	1
Denaturation	95	30 s	
Annealing	58	30 s	25
Extension	72	1 min	
Final extension	72	10 min	1

8. Isolate gDNA according to Qiagen DNeasy Blood and Tissue kit.
9. Determine gDNA concentration by Nanodrop 2000 spectrophotometer.
10. Dilute gDNA concentration to 25 ng/μL in elution buffer.
11. Add 2.5 μL of isolated gDNA of known concentration to the corresponding PCR reaction tube.
12. Perform a 25-cycle PCR reaction according to the thermal cycling conditions in Table 5 (*see* **Note** 7).
13. Before handling high-copy material, prepare the second PCR reaction master mix for *n* samples in duplicate for each primer

Fig. 3 (continued) (products: 433 bp, 325 bp, and 349 bp). When using either 0.63 μM or 0.31 μM of primer in the first thermal cycling round and 2.5 μL or 1.0 μL as input for the second thermal cycling round, the generation of additional products was reduced. (**b**) Isolated gDNA of experimental samples to confirm the nested PCR protocol. The best barcode amplifications were obtained when 0.63 μM of primer was used as input in the first thermal cycling round (upper row). Both 2.5 μL and 1.0 μL as inputs for the second thermal cycling round showed clear bands for all four samples and the positive control. As expected, the negative control did not show a band. However, when the primer concentration was reduced to 0.31 μM (lower row), barcode amplification was suboptimal for samples 2 and 3. To prevent PCR bias, we decided to continue with 2.5 μL. Abbreviations: negative control (–), positive control (+), and DNA Ladder Mix (L)

Table 6
Preparations of master mix for the second thermal cycling round of nested PCR. Required amounts for one PCR reaction

DreamTaq Green PCR Master Mix (2×)	10.0 μL
H_2O, nuclease free	8.0 μL – *x* μL
Universal reverse primer (5 μM)	1.0 μL

set concentration according to Table 6. Variable amounts of input *x* (e.g., 1.0 μL, 2.5 μL, and 5.0 μL) for the second PCR reaction can be tested.

14. Transfer 19.0 μL – *x* of the second PCR reaction master mix to a PCR reaction tube.
15. Add 1.0 μL of indexed forward primer (5 μM) and store at 4 °C.
16. Continue with the second PCR reaction, which should be performed in high-copy environment. Add variable amount *x* (i.e., 1.0 μL, 2.5 μL, or 5.0 μL) of the first PCR reaction to the corresponding PCR reaction tubes of the second PCR reaction.
17. Perform another 25-cycle PCR reaction according to the thermal cycling conditions in Table 5 (*see* **Note** 7).
18. Confirm amplification of the barcode sequence by running the PCR product on a 1.5% agarose gel. Determine the optimal concentration for the first set of primers, which are the lower concentration(s) that primarily produce one product of expected size. An example of an output is given in Fig. 3a.
19. Repeat this procedure with a selection of experimental samples to confirm its functionality. To this aim, use the lower concentration(s) for the first set of primers. An example of an output is given in Fig. 3b.
20. Once the optimal primer concentrations and input DNA are confirmed, all samples that failed standard PCR can be subjected to nested PCR.

3.5.4 PCR Product Cleanup of Pooled Samples

Prior to multiplexing and cleaning up samples, one can approximate the barcode complexity of each sample by Sanger sequencing. Depending on the outcome, one can then decide whether it is worth adding them to the high-throughput sequencing analysis [11].

The number of samples that can be combined into one sequencing run depends on the number of expected barcodes and the sequencing depth. When working with a library of 800 different

barcodes, we can multiplex 200–300 samples into a paired-end half-lane Illumina HiSeq 2500 sequencing run of approximately 75 million reads. This will result in a dataset of thousands of reads per sample, and a few hundred reads per barcode per sample, providing sufficient coverage. In an experimental design with a larger barcode library (or with reduced sequencing depth), multiplexing of fewer samples may be needed, or multiplexing might not be an option at all.

Equimolar amounts of 200–300 successfully amplified samples can be pooled together and purified using the QiaQuick PCR purification kit. Depending on the fragment size, a different purification kit might be required. When pooling samples, make sure not to exceed the maximum binding and column reservoir capacities. All procedures can be performed at room temperature and according to the manufacturer's instructions.

3.5.5 Quality Control of Purified Barcode Sequences

To determine the DNA concentration of the multiplexed sample (s), we use the QuDye dsDNA HS assay kit, which is suitable for the Qubit®. All steps are according to the manufacturer's instructions. Always calibrate the Qubit® when a newly prepared working solution is used. Determine the DNA concentration of each sample, dilute a fraction of each sample in elution buffer to ~1 ng/μL, and reconfirm the concentration. Next, determine the DNA concentration and quality of the diluted sample using the Agilent High Sensitivity DNA assay (Agilent 2100 Bioanalyzer). All steps are performed according to the manufacturer's instructions.

3.6 Data Processing

3.6.1 Pre-filtering

The first step is a quality control of the reads. Sequencing reads with a low quality score (Phred <30) are removed. To create a manageable dataset, all reads with the same sequence are collapsed into a single read, accompanied by its cumulative frequency. Single reads are removed, as it is assumed that they either are derived from a cell which lacks repopulating potential, or (perhaps more likely) result from a sequencing error.

The next step is to de-multiplex the data. To this aim, the 200–300 pooled samples are separated by looking for an exact match of 13 nucleotides, spanning both the index sequence and (part of) the adjacent primer. It is important to note that barcode sequences might be present in reads that lack an exact match with the index and adjacent primer. With the accurate development of indexed primers (i.e., indices of the same length and sufficient minimal distance) more elegant motif-search algorithms with error-correcting properties can be used to improve the recovery of sequencing reads [35].

Barcodes are detected by a matrix search algorithm. To this aim, the backbone of the barcode reference sequence is translated into a position-weight matrix (PWM) and/or position-specific scoring matrix (PSSM). Next, the script uses these matrices to

calculate a similarity score for each sequencing read. The similarity score is used to retrieve barcodes from sequencing reads by setting threshold that ensure that (1) in every read, only one barcode is returned, and (2) multiple mismatches in the barcode backbone are allowed (as specified by an empirically established level). Several packages in Python support matrix search, like Bio.motifs [37], MOODS [38], or Motility [39]. Please bear in mind that all the aforementioned motif finders cannot cope with indels.

3.6.2 Check for Multiple Integrations per Cell

Although the probability of having multiple barcode integrations in a single cell is limited with lower transduction efficiency, it is recommended to check for multiple integrations in the same cell, using linear regression analysis. To this aim, we compare the frequencies of one barcode to the frequencies of the other barcodes in all samples. When one cell contains multiple barcodes, we expect the slope and R^2 to be close to one ($p < 0.05$). If such cases are found, the number of counted barcodes will slightly overestimate the number of clones. In this case, it is better to remove such duplicates from the data to make barcode and clonal counts equal again.

3.6.3 Removing Sequencing Noise

Accurate assessment of clonal complexity in barcode deep sequencing data requires discrimination between "true" signals and "noise." Deep sequencing, but also PCR amplification, is known to introduce errors. The raw sequencing data therefore consist of a combination of real barcodes and some variations from those real barcodes, which differ by one or multiple nucleotides. One can visualize the degree of noise in a given dataset by simply plotting the histogram of unique read frequencies. Usually, this shows a biphasic distribution. One peak is centered at the very low read frequencies and likely represents sequencing noise. The other peak is centered far from the low frequencies, and represents mostly true barcodes. In practical terms, we perform two kinds of data filtering: the first filtering relies on uniqueness of the barcode sequence. With a proper library design, we expect that each barcode in the library differs from any other by more than 1 base. Therefore, sequence reads that only differ by one nucleotide (i.e., Hamming distance of one) are most likely a consequence of PCR or sequencing errors. To prevent an overestimation of the number of barcodes due to this type of error, the barcodes with a smaller frequency can either be eliminated or merged to the similar barcode with the higher frequency. Hamming distance can be calculated in Python using packages "python-Levenshtein" [40], "scipy.spatial.distances" [41], "skbio.sequence.distance" [42], or "DNABarcodes" [43] for R script. After this operation, we can check the histogram of the frequencies of the remaining barcodes. Ideally, the low-frequency peak should disappear. If it persists, however, we can either add an extra cutoff for absolute barcode read frequency

or increase the threshold for Hamming distance between barcodes (however, the second option should be experimentally justified). The second filtering strategy relies on biological relevance. In our experiments, we define barcodes as biologically relevant when (i) they are present in more than three individual samples and (ii) they have an overall sample frequency of ≥0.05% across all samples in the experiment. The logic is that very-low-frequency barcodes, which are observed incidentally, are likely not relevant contributors to hematopoiesis or leukemia. However, this decision must be made individually for each particular experiment in its biological context. Sometimes, even accidental barcodes could be biologically important and relevant.

3.6.4 Counting Clones

Counting all unique barcodes above technical thresholds (i.e., nominal counts) is only the first step in reporting the clonal complexity of a cell population. The more relevant clonal diversion of a population is the so-called Shannon count, which is described in detail elsewhere [11]. Briefly it works in two steps. First, we estimate the Shannon diversity index (also known as Shannon entropy or Shannon information index). Next we calculate the natural exponent of that index, which converts the dimensionless index into an actual measure of the number of barcodes. The Shannon count tells us what population of equally sized barcodes would give an equivalent diversity of the observed population of barcodes. Some authors irregularly use predictors (e.g., Chao predictor) for the estimation of the total number of barcodes in the library or cell population (seen and unseen in the real data). As recently reported, estimation of the nominal barcode population size from sequencing data is highly sensitive to arbitrary decisions in data filtering and therefore should be taken with great care [21].

4 Notes

1. Coating the T75 culture flasks with 0.1% gelatin improves HEK293FT cell adherence and viablity.
2. The use of siliconized Eppendorf tubes will improve sample recovery due to their low surface adhesion.
3. RetroNectin® can be reused up to 4–5 times. However, it is recommended to use fresh or one-time used RetroNectin® for patient-derived material.
4. Due to interpatient differences, it is difficult to determine up front the optimal volume of virus to transduce patient-derived B-ALL at a transduction efficiency of ~10%. Higher amounts of virus do not per definition result in higher transduction efficiencies, as these may be toxic to the cells.

5. It is recommended to aliquot the indexed forward primers in 8-tube strips or a 96-well plate format, according to the PCR preparation list, to create a manageable experimental setup (Fig. 2).
6. Low copy number of barcode sequences depends on the DNA concentration, but also on the chimerism levels. We optimized the primer concentrations of the set of primers for the first thermal cycling round based on low DNA concentration of 100% barcoded cells. It might be useful to optimize the primer concentrations for different fractions of barcoded cells, as this affects a successful barcode amplification as well.
7. The optimal annealing temperature is primer dependent and should be determined beforehand.

Acknowledgments

M. Belderbos was supported by research funding from the University Medical Center Groningen (Mandema Stipend), the European Society for Blood and Marrow Transplantation (research fellowship grant), and the Dutch Cancer Society (grants RUG 2014-6957 and RUG 2015-7964).

References

1. Greaves M, Maley CC (2012) Clonal evolution in cancer. Nature 481:306
2. Gawad C, Koh W, Quake SR (2014) Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. Proc Natl Acad Sci 111:17947–17952. https://doi.org/10.1073/pnas.1420822111
3. Loh ML, Mullighan CG (2012) Advances in the genetics of high-risk childhood B-progenitor acute lymphoblastic leukemia and juvenile myelomonocytic leukemia: implications for therapy. Clin Cancer Res 18:2754–2767. https://doi.org/10.1158/1078-0432.CCR-11-1936
4. Bardini M, Woll PS, Corral L et al (2014) Clonal variegation and dynamic competition of leukemia-initiating cells in infant acute lymphoblastic leukemia with MLL rearrangement. Leukemia 29:38
5. Mullighan CG, Phillips LA, Su X et al (2008) Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. Science 322:1377–1380. https://doi.org/10.1126/science.1164266
6. Ma X, Edmonson M, Yergeau D et al (2015) Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. Nat Commun 6:6604
7. Yang JJ, Bhojwani D, Yang W et al (2008) Genome-wide copy number profiling reveals molecular evolution from diagnosis to relapse in childhood acute lymphoblastic leukemia. Blood 112:4178 LP–4174183. https://doi.org/10.1182/blood-2008-06-165027
8. Roth A, Khattra J, Yap D et al (2014) PyClone: statistical inference of clonal population structure in cancer. Nat Methods 11:396
9. Opasic L, Zhou D, Werner B et al (2019) How many samples are needed to infer truly clonal mutations from heterogenous tumours? BMC Cancer 19:403. https://doi.org/10.1186/s12885-019-5597-1
10. Glauche I, Bystrykh L, Eaves C, Roeder I (2013) Stem cell clonality—theoretical concepts, experimental techniques, and clinical challenges. Blood Cells Mol Dis 50:232–240. https://doi.org/10.1016/j.bcmd.2013.01.007
11. Bystrykh LV, Belderbos ME (2016) Clonal analysis of cells with cellular barcoding: when numbers and sizes matter. In: Turksen K

(ed) Stem cell heterogeneity: methods and protocols. Springer, New York, NY, pp 57–89

12. Bystrykh LV, Verovskaya E, Zwart E et al (2012) Counting stem cells: methodological constraints. Nat Methods 9:567
13. Golden JA, Fields-Berry SC, Cepko CL (1995) Construction and characterization of a highly complex retroviral library for lineage analysis. Proc Natl Acad Sci 92:5704 LP–5705708. https://doi.org/10.1073/pnas.92.12.5704
14. Gerrits A, Dykstra B, Kalmykowa OJ et al (2010) Cellular barcoding tool for clonal analysis in the hematopoietic system. Blood 115:2610 LP–2612618. https://doi.org/10.1182/blood-2009-06-229757
15. Schepers K, Swart E, van Heijst JWJ et al (2008) Dissecting T cell lineage relationships by cellular barcoding. J Exp Med 205:2309 LP–2302318. https://doi.org/10.1084/jem.20072462
16. Belderbos ME, Koster T, Ausema B et al (2017) Clonal selection and asymmetric distribution of human leukemia in murine xenografts revealed by cellular barcoding. Blood 129:3210 LP–3213220. https://doi.org/10.1182/blood-2016-12-758250
17. Wu C, Espinoza DA, Koelle SJ et al (2018) Geographic clonal tracking in macaques provides insights into HSPC migration and differentiation. J Exp Med 215:217–232. https://doi.org/10.1084/jem.20171341
18. Verovskaya E, Broekhuis MJC, Zwart E et al (2014) Asymmetry in skeletal distribution of mouse hematopoietic stem cell clones and their equilibration by mobilizing cytokines. J Exp Med 211:487 LP–487497. https://doi.org/10.1084/jem.20131804
19. Verovskaya E, Broekhuis MJC, Zwart E et al (2013) Heterogeneity of young and aged murine hematopoietic stem cells revealed by quantitative clonal analysis using cellular barcoding. Blood 122:523 LP–523532. https://doi.org/10.1182/blood-2013-01-481135
20. Elder A, Bomken S, Wilson I et al (2017) Abundant and equipotent founder cells establish and maintain acute lymphoblastic leukaemia. Leukemia 31:2577
21. Belderbos ME, Jacobs S, Koster T et al (2019) Donor-to-donor heterogeneity in the clonal dynamics of transplanted human cord blood stem cells in murine xenografts. Biol Blood Marrow Transplant 26:16. https://doi.org/10.1016/j.bbmt.2019.08.026
22. Thielecke L, Aranyossy T, Dahl A et al (2017) Limitations and challenges of genetic barcode quantification. Sci Rep 7:43249
23. Kester L, van Oudenaarden A (2018) Single-cell transcriptomics meets lineage tracing. Cell Stem Cell 23:166–179. https://doi.org/10.1016/j.stem.2018.04.014
24. Livet J, Weissman TA, Kang H et al (2007) Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. Nature 450:56–62. https://doi.org/10.1038/nature06293
25. Weissman TA, Pan YA (2015) Brainbow: new resources and emerging biological applications for multicolor genetic labeling and analysis. Genetics 199:293 LP–293306. https://doi.org/10.1534/genetics.114.172510
26. Weber TS, Dukes M, Miles DC et al (2016) Site-specific recombinatorics: in situ cellular barcoding with the Cre lox system. BMC Syst Biol 10:43. https://doi.org/10.1186/s12918-016-0290-3
27. Pei W, Feyerabend TB, Rössler J et al (2017) Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. Nature 548:456
28. Guernet A, Mungamuri SK, Cartier D et al (2016) CRISPR-barcoding for intratumor genetic heterogeneity modeling and functional analysis of oncogenic driver mutations. Mol Cell 63:526–538. https://doi.org/10.1016/j.molcel.2016.06.017
29. McKenna A, Findlay GM, Gagnon JA et al (2016) Whole-organism lineage tracing by combinatorial and cumulative genome editing. Science 353:aaf7907. https://doi.org/10.1126/science.aaf7907
30. Junker JP, Spanjaard B, Peterson-Maduro J et al (2016) Massively parallel whole-organism lineage tracing using CRISPR/Cas9 induced genetic scars. bioRxiv:56499. https://doi.org/10.1101/056499
31. Kalhor R, Kalhor K, Mejia L et al (2018) Developmental barcoding of whole mouse via homing CRISPR. Science 361:eaat9804. https://doi.org/10.1126/science.aat9804
32. Bystrykh LV, de Haan G, Verovskaya E (2014) In: Bunting KD, Qu C-K (eds) Barcoded vector libraries and retroviral or lentiviral barcoding of hematopoietic stem cells BT—hematopoietic stem cell protocols. Springer, New York, New York, NY, pp 345–360
33. Fehse B, Kustikova OS, Bubenheim M, Baum C (2004) Pois(s)on—it's a question of dose.... Gene Ther 11:879–881. https://doi.org/10.1038/sj.gt.3302270
34. Anon (1993) Removal of blood from laboratory mammals and birds. First report of the BVA/FRAME/RSPCA/UFAW joint working group on refinement. Lab Anim 27:1–22. https://doi.org/10.1258/002367793781082412

35. Bystrykh LV (2012) Generalized DNA barcode design based on hamming codes. PLoS One 7:e36852. https://doi.org/10.1371/journal.pone.0036852
36. Haff LA (1994) Improved quantitative PCR using nested primers. Genome Res 3:332–337
37. Biopython Biopython. https://biopython.org/. Accessed 30 Sep 2019
38. Korhonen J, Martinmäki P, Pizzi C et al (2009) MOODS: fast search for position weight matrix matches in DNA sequences. Bioinformatics 25:3181–3182. https://doi.org/10.1093/bioinformatics/btp554
39. GitHub—ctb/motility. https://github.com/ctb/motility. Accessed 2 Oct 2019
40. python-Levenshtein PyPI. https://pypi.org/project/python-Levenshtein/. Accessed 30 Sep 2019
41. SciPy.org — SciPy.org. https://scipy.org/. Accessed 2 Oct 2019
42. scikit-bio. http://scikit-bio.org/. Accessed 2 Oct 2019
43. Buschmann T, Bystrykh LV (2013) Levenshtein error-correcting barcodes for multiplexed DNA sequencing. BMC Bioinformatics 14:272