TRACKING HEMATOPOIETIC CLONES WITH EXPRESSED LENTIVIRAL BARCODES

Gabriele Casirati | Boston Children's Hospital and Harvard Medical School
Cell Discovery Network

A fluorescent lentiviral barcode library can enable precise dissection of the origin of transplanted blood cells and connection to their clonal output. Here we explain how to use lentiviral labeling strategy for in vivo clonal tracking across whole-graft measurements, sorted blood lineages, and single-cell transcriptomics.

Why follow clones *in vivo* during immune selection

We are developing epitope editing as an approach to enrich gene-engineered hematopoietic stem and progenitor cells (HSPCs), modified either *ex vivo* or directly *in vivo*, without chemo- or radiotherapy through selective immunotherapy [1]. By altering an antibody-binding site on proteins such as the stem cell antigen KIT, we can protect engineered cells while depleting their unedited counterparts. However, increasing the proportion of engineered cells alone does not necessarily expand the number of repopulating long-term hematopoietic stem cells (LT-HSCs) that sustain blood production. Selection could create a clonal bottleneck, with hematopoiesis ultimately maintained by a relatively small number of the starting HSPC population, which constitutes a potential risk factor for hematopoietic stress and the development of secondary malignancies. We therefore need to determine whether immune selection preserves broad clonal diversity or favors the disproportionate contribution of few LT-HSCs. This question is central to evaluating durable reconstitution and detecting unintended clonal dominance. To address this need, we developed lentiviral libraries combining inherited genomic barcodes with distinct fluorescent reporters. These libraries allow the distinction of multiple starting populations and quantify barcode-defined clonal diversity within each population and can be multiplexed with CRISPR gene editing (e.g. primer or base editing). Barcode recovery from bulk samples or sorted hematopoietic lineages reveals how individual tagged clones contribute to reconstitution during anti-KIT antibody treatment or other immune perturbations. The expressed barcode design also provides a route to connect clonal identity with single-cell transcriptional states. Together, these readouts support the evaluation of clonal preservation, repopulation dynamics and the translational safety of emerging genome-engineering strategies.

Step 1: Select a fluorescent barcode library

We generated fluorescent barcoded lentiviral libraries that express either mNeonGreen (mNG), mTagBFP (mtBFP) or mCherry, allowing researchers to distinguish up to three starting populations by flow cytometry while tracking clonal diversity independently within each population (**Fig.1a**). The reporter combination may be chosen to suit individual experimental design (for instance, multiparameter flow cytometry panels) and use established reagents and PCR protocols for cell labeling and subsequent barcode recovery. Each construct contains an hPGK-driven, constitutive fluorescent reporter with a random barcode in its 3′ untranslated region. The barcode comprises seven blocks of four random bases separated by fixed anchors [1,2]. The theoretical barcode diversity (~10e16) was designed to be vastly larger than actual library complexity (~10e5), to allow for unambiguous barcode discrimination. Unique flanking regions allow for selective amplification of each library (e.g. mNG vs mtBFP). Following lentiviral integration, the barcode is inherited by daughter cells, enabling clonal tracking through genomic DNA sequencing. Its inclusion in the expressed reporter transcript also provides a route to recover clonal identity by RNA sequencing alongside single-cell or bulk gene-expression profiles. To support this dual readout, we cloned the reporter cassette antisense to the lentiviral backbone and incorporated a full SV40 polyadenylation signal. This architecture places the barcode close to the reporter transcript's polyA, enabling 3′ end capture by oligo(dT) beads.

Step 2: Label starting HSPC populations

Assign a separate fluorescent library to each experimental HSPC population **(Fig.1b)**, transduce cells with appropriate vector amounts (e.g. MOI=30 with cyclosporin H as transduction enhancer), then proceed to humanization of immunodeficient mouse models (e.g. NSG or NBSGW). Preserve a fraction of transduced cells as liquid culture to record fluorescent marking and editing efficiency. Collecting an input gDNA aliquot allows the assessment of barcode representation before transplant (this should be normalized on the analyzed cell numbers and/or input gDNA used for PCR). High transduction efficiencies (>50-70%) can improve coverage and still provide clonal diversity correlation in bulk readouts, but they may inflate the number of retrieved barcodes per input cells or confound single cell readouts (due to potential barcode dropout), if lineage tracing analyses are performed. After in vivo hematopoietic reconstitution, constitutively expressed fluorescent proteins report the population origin and – if linked to a specific CRISPR edit – the editing condition. Mice can then receive treatment according to the corresponding condition, e.g. monoclonal antibodies aimed at HSPCs antigens, to deplete host progenitors and allow non-genotoxic hematopoietic replacement or select epitope edited

populations in competitive transplant designs (**Fig.1c**). Cell-of-origin and selection outcomes may be monitored in vivo by serial peripheral blood collection and flow analyses.

Step 3: Analyze whole graft or lineage resolved barcode composition

For a whole-graft view, recover gDNA from bulk bone marrow, spleen or blood. To ask whether individual tagged lineages contribute to distinct compartments, Lin-CD34+ progenitors, myeloid cells and B cells maybe be FACS-sorted before DNA extraction (**Fig.2a-b**). Record recovered cell numbers and gDNA input to normalize the number of retrieved barcodes on input material, as absolute counts can affect the results (e.g. 100 ng input correspond to approx. 15,000 human cells). Amplify the barcode region using the reporter-specific PCR1 primers, which also introduce partial Illumina adapters, followed by common PCR2 primers [2]. The seven random bases in the forward PCR1 primer form a unique molecular identifier (UMI), used to identify PCR duplicates and help control amplification bias. Keep this technical tag separate from the inherited 28-base lineage barcode throughout analysis. In our published workflow, reads were quality filtered using Flexbar (v3.5.0; Q30), annotated for the PCR UMI, and assigned to the mNG or mtBFP library using a six-base library identifier. Reads were then screened for consistent reporter assignment and the expected barcode structure, including its fixed anchors. To consolidate variants arising from sequencing or PCR errors, lineage barcodes were clustered using DBSCAN with a Levenshtein-distance threshold of eight. UMIs supported by only one read were excluded. Barcode abundance was estimated using both total read counts and the number of distinct associated UMIs]. Use the resulting barcode-abundance table to assess clonal composition. Our analysis calculated barcode richness, Shannon and Simpson diversity, and Pielou's evenness. Shannon diversity captures both the number of detected barcodes and their relative contributions; evenness describes how equally those contributions are distributed [1]. Compare matched populations and lineages at comparable sampling depth, inspect dominant barcodes individually, and assess whether conclusions remain consistent across filtering thresholds. Barcode counts are not a direct census of stem cells: multiple integrations, reuse of the same library barcode in independent cells, and PCR sampling can affect the relationship between recovered tags and contributing cells. The full analysis code and detailed processing workflow are available upon request by contacting the authors.

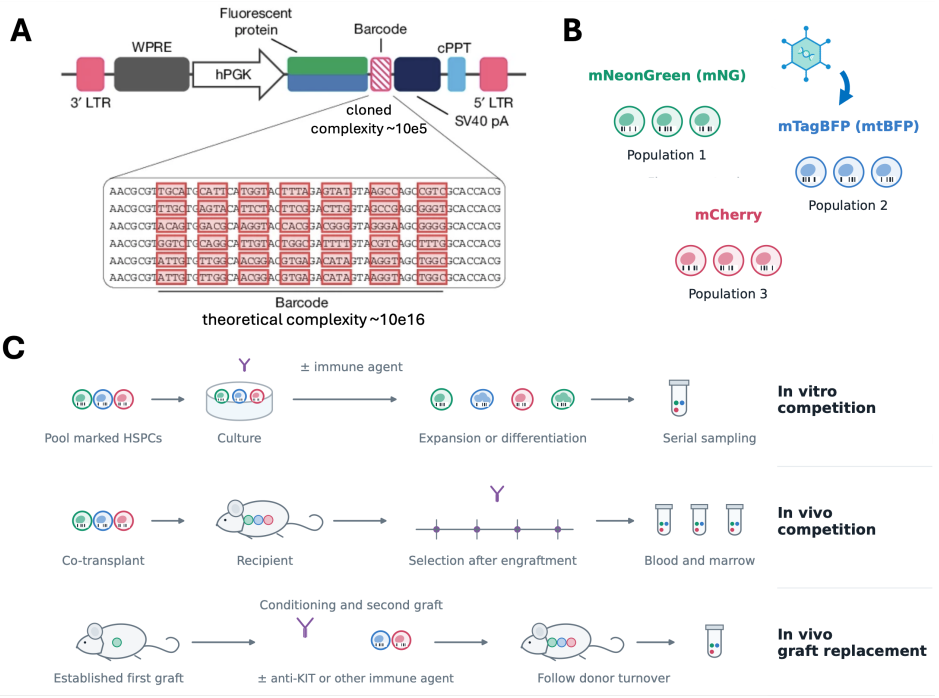


Figure 1 | Fluorescent barcoding enables population- and clone-resolved tracking. A Lentiviral vector design encoding for constitutive fluorescent protein(s) and expressed barcode in the 3'UTR. **B** Independently barcoded hematopoietic stem and progenitor cell (HSPC) populations expressing mNeonGreen, mTagBFP or mCherry can be distinguished by multiparameter flow cytometry while inherited lentiviral barcodes identify their progeny. **C** After in vitro culture or transplantation and immune selection in vivo, barcodes can be recovered from bulk genomic DNA or sorted hematopoietic subsets to assess clonal contributions. The expressed reporter design also provides a route to barcode recovery from bulk RNA or 3′ single-cell RNA sequencing. RNA-based readouts are

Step 4: Extend analyses to single-cell RNA sequencing

The expressed barcode provides a route to connect clonal identity with each cell's transcriptional state. Its location near the fluorescent reporter transcript' s poly(A) tail was chosen to support recovery in oligo(dT)-based 3′ single-cell assays. A closely related precedent is LARRY/pLARRY, developed by Weinreb, Rodriguez-Fraticelli, Camargo and Klein, which combines inherited, expressed barcodes with single-cell transcriptomics to link cell state with clonal fate [6,7]. Viable cells from culture, blood or bone marrow should be collected and, where appropriate, enriched by FACS-sorting. These cells would enter a 3'scRNA-seq workflow, such as Chromium 10X, capturing reporter transcripts alongside endogenous mRNAs. A complementary barcode readout from the same amplified cDNA would recover the inherited tag while preserving its association with the platform's cell identifier and transcript UMI. After quality filtering and barcode error correction, these identifiers would connect clonal assignments to the gene-

expression matrix [7]. Overlaying clone labels on transcriptionally defined populations allows to examine lineage contributions, progenitor states and responses to immune selection within individual clones.

What the published experiments show

We applied fluorescent marking and genomic barcoding in two complementary transplantation models to distinguish enrichment of engineered cells from changes in their clonal diversity. In a competitive-selection experiment, mtBFP-marked HSPCs edited at KIT H378R and the BCL11A +58/+55 enhancers were transplanted alongside mNG-marked, AAVS1-edited controls. Starting six weeks after transplantation, mice received six 0.5 mg anti-KIT Fab-79D doses at intervals of two, five or ten days. Bone marrow was analyzed at week 18. Barcode sequencing of sorted CD33+ myeloid cells, CD19+ B cells and Lin−CD34+ progenitors resolved the effects of treatment within each starting population and lineage [1,4]. Fab-79D reduced control-pool barcode recovery in myeloid cells and progenitors. For example, mean barcode counts in the five-day dosing group fell from 1,301 to 632 in CD33+ cells and from 505 to 252 in Lin−CD34+ progenitors – approximately 50% reductions relative to mock-treated mice. Control progenitors also showed reduced Shannon diversity, whereas B-cell barcode representation was mostly unchanged, due to low KIT expression on B phenotype progenitors. The KIT epitope edited, mtBFP-marked population retained barcode richness, Shannon diversity and Pielou's evenness across the three compartments. Thus, enrichment of the engineered population was accompanied by preservation of its measured clonal diversity [1,4].

In hematopoietic replacement experiments (**Fig.2c**), we tested a more stringent setting: replacing an established human graft while conditioning antibody remained active. Mice bearing an mNG-marked first graft received two SR-1 doses, followed only 12 hours later by an mtBFP-marked replacement graft, without allowing time for mAb washout. SR-1 strongly depleted the first graft and robustly reduced its barcode richness and Shannon diversity (mNeonGreen library). Replacement cells (mTagBFP-marked) lacking KIT protection also showed markedly reduced barcode recovery. By comparison, epitope edited KIT D121L/BCL11A-edited replacement grafts retained mean barcode counts of 397 and 451 after the lower and higher conditioning doses, respectively, compared with 455 without conditioning, demonstrating no significant reduction, which extended to their Shannon diversity (**Fig.2d**) [1,5]. Secondary transplantation provided complementary functional evidence that protected grafts retained long-term, multilineage repopulating activity (**Fig.2e**). Together, these experiments support immune selection and graft replacement while preserving a diverse engineered population. This conclusion applies to the tested models and follow-up periods; the analyses do not exclude rare clonal effects or establish long-term clinical safety [1].

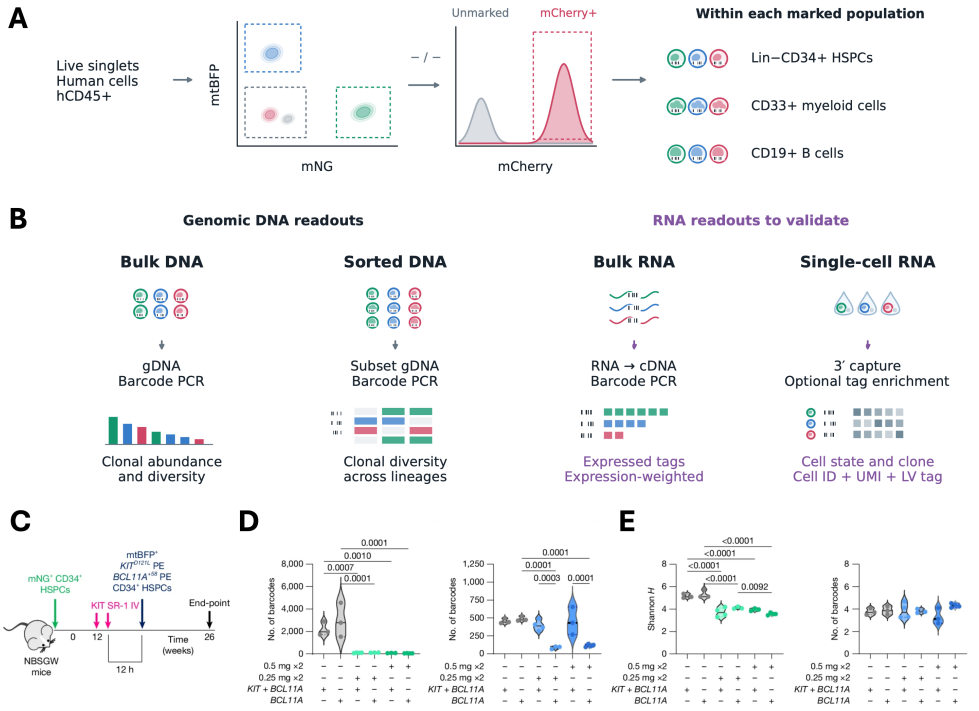


Figure 2 | Genomic barcodes reveal preservation of graft diversity during anti-KIT selection. A Lentiviral library marked cells can be separated by flow, isolating either specific lineages and/or fluorescent library, to assess population-specific barcode composition. **B** The multi-color genomic barcode library allows for multiple experimental designs, including bulk DNA and RNA, sorted subpopulation analyses or single-cell RNA seq with direct polyA capture through 3'end oligo-d(T) beads. **C** Experimental design for immune-based hematopoietic replacement with graft-specific tracking of clonal diversity. **D** Number of retrieved barcodes from bulk BM cells, either from the first graft (mNeonGreen marked) or the transplant-derived hematopoiesis (mTagBFP marked) according to editing and treatment condition. **E** Shannon H index from the same samples and conditions as (D).

Public reagents and resources

The peer-reviewed paper, primer table, quality-control figures and underlying figure source data are publicly available online [1–5]. This vignette reports the analysis rationale and experimental processes

involved in studying clonal diversity in humanized hemato-chimeric models in vivo, with multiple competing stem cell populations, with or without immunotherapeutic pressure. All code, materials, including plasmid libraries, lentiviral vectors and detailed experimental protocols to generate them are available upon request to the authors (general inquiries, reagents, protocols: gabriele.casirati@childrens.harvard.edu, code and analysis: danilo.pellin@childrens.harvard.edu).

Linked publications data and reference code

[1] Casirati G, Cosentino A, Freschi M, et al. Non-genotoxic transplantation and in vivo selection through epitope editing. Nature 657, 262–274 (2026).

[2] Supplementary Tables 1–11. Table 2 contains reporter-specific barcode PCR1 primers and common PCR2 primers; Table 3 provides antibodies and gating information.

[3] Supplementary Figure 1. Read retention and distance-based barcode quality control for sorted fractions and bulk bone marrow.

[4] Source Data Fig. 2. Published sorted-lineage barcode richness and diversity measurements, including the values replotted in Figure 2A–C.

[5] Source Data Fig. 5. Published bulk bone marrow barcode richness and diversity measurements, including the values replotted in Figure 2D,E.

[6] Weinreb C, Rodriguez-Fraticelli A, Camargo FD, Klein AM. Lineage tracing on transcriptional landscapes links state to fate during differentiation. Science 367, eaaw3381 (2020).

[7] LARRY public reference code. An established example of linking expressed barcodes to single-cell profiles. Its parsing and capture design require adaptation and validation for the vectors described here.