Natural Barcodes for Longitudinal Single Cell Tracking of Leukemic and Immune Cell Dynamics

Livius Penter [1,2,3,4], Satyen H. Gohil [5,6] and Catherine J. Wu [1,2,3,7]*

[1] Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, United States, [2] Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, United States, [3] Harvard Medical School, Boston, MA, United States, [4] Department of Hematology, Oncology, and Tumorimmunology, Campus Virchow Klinikum, Berlin, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany, [5] Department of Academic Haematology, University College London Cancer Institute, London, United Kingdom, [6] Department of Haematology, University College London Hospitals NHS Foundation Trust, London, United Kingdom, [7] Department of Medicine, Brigham and Women's Hospital, Boston, MA, United States

Blood malignancies provide unique opportunities for longitudinal tracking of disease evolution following therapeutic bottlenecks and for the monitoring of changes in anti-tumor immunity. The expanding development of multi-modal single-cell sequencing technologies affords newer platforms to elucidate the mechanisms underlying these processes at unprecedented resolution. Furthermore, the identification of molecular events that can serve as in-vivo barcodes now facilitate the tracking of the trajectories of malignant and of immune cell populations over time within primary human samples, as these permit unambiguous identification of the clonal lineage of cell populations within heterogeneous phenotypes. Here, we provide an overview of the potential for chromosomal copy number changes, somatic nuclear and mitochondrial DNA mutations, single nucleotide polymorphisms, and T and B cell receptor sequences to serve as personal natural barcodes and review technical implementations in single-cell analysis workflows. Applications of these methodologies include the study of acquired therapeutic resistance and the dissection of donor- and host cellular interactions in the context of allogeneic hematopoietic stem cell transplantation.

Keywords: copy number variants (CNV), somatic nuclear mutation, mitochondrial DNA mutation, single nucleotide polymorphism, B cell receptor sequence, T cell receptor sequence, allogeneic hematopoietic stem cell transplantation (allo-HCT), single-cell sequencing

INTRODUCTION

After decades of research, cancer has remained a formidable enemy with relapse as an all too frequent outcome despite advances in treatment approaches. Deeper elucidation of the underlying malignant cell states and reprogrammed immune circuits will help to overcome current therapeutic limitations and improve long-term outcomes. Single-cell sequencing technologies are providing capabilities to understand cellular states at unprecedented depth, which has greatly accelerated our understanding of hematopoiesis (1–7) and haematological malignancies (8–14). Certainly, the dense cellular sampling of biospecimens, afforded by single-cell epigenomics and transcriptomics,

has allowed the inference of trajectories of cell differentiation from primary human samples and has deepened our understanding of the disruption between physiologic and malignant states (15).

A growing area of interest is the use of molecular barcodes to corroborate these insights experimentally. Technologies that introduce artificial barcodes such as fluorescence-based labeling, viral barcoding, Cre-Lox-based approaches and CRISPR-Cas9 genome editing enable prospective lineage tracing (16, 17) and permit the characterization of the phenotype of defined genotypes in great detail (18, 19). However, these studies are limited by the experimental models they employ. In addition, the introduction of barcodes may lead to unintended perturbation of cell states and off-target effects, thus skewing the studied phenotypes.

An alternative approach is to leverage naturally occurring molecular barcodes to perform retrospective *in-vivo* lineage tracing in unaltered primary human cells. This approach has already proven to be a powerful strategy for deconvoluting clonal cancer fractions from bulk sequencing data through the analysis of evolving variant allele frequencies of single nucleotide variants (20, 21). When coupled with single-cell analyses, the use of natural barcodes provides the potential for delineating the phylogeny of cell populations at much greater resolution, and depending on the technology may link the genotypes of individual cells to distinct phenotypic cell states (22).

Here we review the growing number of single-cell methodologies for exploiting such natural barcodes. As examples, we describe the utility of T or B cell receptor (TCR, BCR) sequences and copy number changes, somatic nuclear and mitochondrial DNA (mtDNA) mutations for lineage tracing. We further discuss specific opportunities in the setting of allogeneic hematopoietic stem cell transplantation (HSCT).

TOOL KITS FOR LONGITUDINAL TRACKING OF DISEASE EVOLUTION AT THE SINGLE-CELL LEVEL

Efforts to utilize natural barcodes to elucidate the developmental history of individual cells have preceded the availability of current single-cell sequencing modalities. Despite their low throughput, these early attempts laid the groundwork for our understanding how allelic variants can serve to track cell populations longitudinally over months or years at single-cell resolution. They have included techniques such as red blood cell phenotyping (23, 24), fluorescence *in-situ* hybridization (FISH) using probes specific for X and Y chromosomes and cell type-specific staining (25, 26), analysis of sex-linked electrophoretic variants of glucose-6-phosphate dehydrogenase (27) or clinical karyotyping (28). Although these techniques are amenable to single-cell sequencing approaches (29), they are mostly constrained to specific contexts that limits their wider applicability.

Physiologic Natural Barcodes
The TCR and BCR are examples of physiologic molecules with high diversity which lend themselves as natural barcodes that can be used for lineage tracing purposes (**Figure 1A**). Each unique TCR arises from a combination of α/β or γ/δ chains with highly variable sequences due to V(D)J rearrangement and junctional diversification (30). Single-cell sequencing of TCR now enables phenotyping of individual T cell clones (31, 32) and permits tracking of antigen-specific T cells across tissue compartments and following therapeutic interventions such as immune checkpoint blockade or vaccination (33–35). Further, TCR is used to phenotype malignant T cells, for example in angioimmunoblastic (36) or cutaneous T cell lymphoma (37, 38).

Compared to T cells, B cells additionally undergo affinity maturation through somatic hypermutation (SHM), which renders the BCR repertoire even more dynamic (**Figure 1B**) (39, 40). Besides applications to understanding immunoglobulin responses to viral or tumor-specific antigens (41–43), the ongoing changes in BCR enable lineage tracing of post-germinal center B cell malignancies that have undergone V(D)J rearrangement and SHM (44). Sequencing of follicular lymphoma (45), DLBCL (46) and multiple myeloma (47) has demonstrated intraclonal BCR heterogeneity that can shed light on clonal evolution following malignant transformation. Further, BCR sequences can be used to dissect disease pathogenesis (48, 49) or mechanisms underlying differential therapeutic sensitivity (50) of B cell malignancies.

Several RNA-based single-cell TCR/BCR sequencing platforms exist that differ in the length of covered sequence and error rate. 5' short read droplet-based sequencing yields CDR3 sequences by amplifying cDNA using primers specific for TCR and immunoglobulin constant regions, but is unable to cover most of the V region. This approach currently has the highest throughput and can resolve phenotypes of TCR/BCR clonotypes with detailed resolution. Smart-seq2 with computational reconstruction of TCR/BCR sequences leads to better coverage, but has a higher cost per sample and lower throughput (51–53). These limitations are being improved with the development of smart-seq3 (54, 55). Long-read sequencing can provide full-length TCR/BCR sequences in a large number of cells, although with higher error rates (56, 57). These technologies are therefore suited for studies that require information on regions outside the CDR3 sequence. Finally, cost-effective targeted approaches with multiplexed PCR can be an option for analysis of rare cells or little starting material (47, 58). Different analysis pipelines for these platforms have been developed that include cellranger (5' scRNA-seq with VD(J) enrichment) (59), MiXCR (60) or TRUST4 (61).

Copy Number Changes
Chromosomal copy number variations (CNV) can provide robust signals that are detectable with RNA- and DNA-based sequencing platforms (**Figure 1C**). The high prevalence of CNV changes in blood malignancies and the genomic instability associated with relapse following therapy (62–65) render them as useful barcodes for lineage tracing and phylogenetic dissection of subclones within malignant cell populations. Examples of recurrent CNV events include *del*(5q) and *del*(7q) in myeloid disease or *tri*(12) and *del*(17p) in CLL. A strength of lineage tracing based on CNV changes is the shallow sequencing depth required for detection (66),
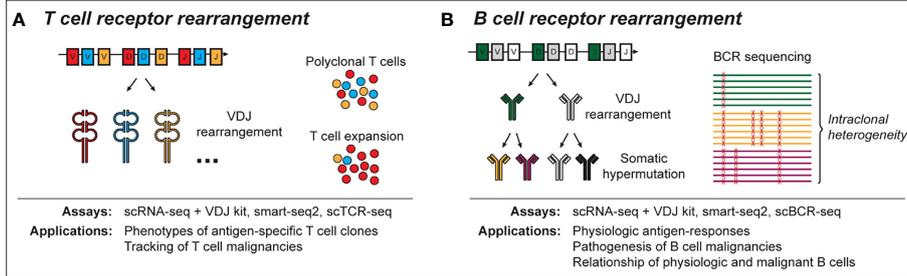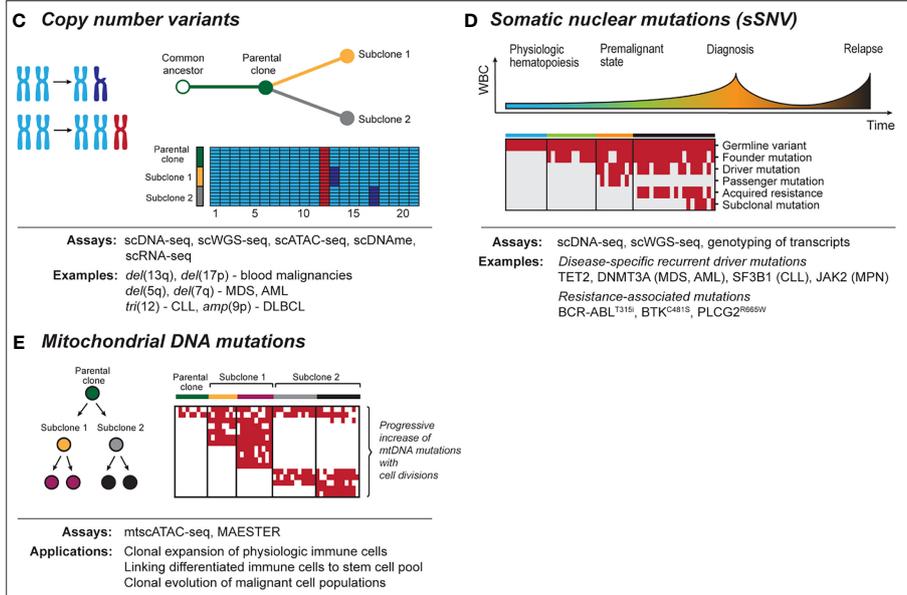
FIGURE 1 | Tool kits for lineage tracing with single-cell sequencing. Physiologic barcodes **(A)** T cell receptor (TCR) sequencing detects clonal expansion of T cells based on VDJ rearrangement and junctional diversification. TCR can help to identify and phenotype antigen-specific T cells or track malignant T cells. **(B)** B cell receptor (BCR) sequences arise due to VDJ rearrangement and ongoing somatic hypermutation. BCR can provide insight into physiological antigen responses and intraclonal heterogeneity in post-germinal B cell malignancies. Both BCR and TCR are best read out with RNA-based single-cell platforms due to the high number of BCR/TCR mRNA templates per cell. Acquired barcodes **(C)** Chromosomal copy number variants (CNV) are common aberrations in blood malignancies. CNV can provide robust signals with DNA- and RNA-based sequencing platforms that allow to dissect subclonal structure of cancer and can be detected using most single-cell sequencing platforms. **(D)** Somatic nuclear mutations (single somatic-nucleotide variant, sSNV) can track clonal evolution in cancer longitudinally. Different classes of somatic mutations are distinguished such as germline variants, cancer initiating mutations, mutations associated with therapeutic resistance and sporadic mutations unrelated to the disease pathogenesis. **(E)** Mitochondrial DNA mutations are progressively acquired as cells divide. This allows to link physiologic or malignant cells to a common ancestor and to resolve phylogeny. As coverage of mitochondrial transcripts tends to be incomplete, they can be best read out from DNA-based sequencing platforms. scDNA-seq, single-cell DNA sequencing; scWGS-seq, single-cell whole genome sequencing; scATAC-seq, single-cell Assay for Transposase-Accessible Chromatin with high-throughput sequencing; scDNAme, single-cell sequencing of DNA methylation; scRNA-seq, single-cell RNA sequencing; mtscATAC-seq, mitochondrial scATAC-seq; scBCR-seq, single-cell B cell receptor sequencing; scTCR-seq, single-cell T cell receptor sequencing.

which allows greater numbers of cells to be analyzed. Limitations for the utility of CNV changes for lineage tracing are either the common absence of chromosomal aberrations as in diseases such as AML with normal karyotype (67) or their omnipresence within monoclonal populations, if they are early founding events.

DNA-based single-cell assays can detect CNV in targeted regions of interest (scDNA-seq) or globally across the entire nuclear genome [e.g. single-cell whole genome (scWGS-seq) (68–70), assay for transposase-accessible chromatin using sequencing (scATAC-seq) (71, 72) or single-cell sequencing of DNA methylation (scDNAme) (73)]. Targeted approaches enable

detection of CNV changes at higher coverage and better cost efficiency, but require design of primers for these target regions, which can be accomplished either through large panels for recurrent genetic events or personalized solutions based on previous analyses (9, 74). While scDNA-seq and scWGS-seq profiles are unable to natively identify cell types and thus require either purification of cell populations prior to sequencing or additional detection of cell surface marker expression, scATAC-seq and scDNAme have the advantage of providing combined information on CNV and cell state, for example through capture of accessible chromatin (75). A major technical hurdle is the fact that detection of subchromosomal

CNV is prone to false positive calls due to effects such as unequal coverage of genomic regions or PCR amplification bias. To address this, several computational methods have been developed to overcome these challenges (72, 76, 77).

For transcriptomic data, the tool inferCNV, which is based on comparison of read coverage between target and reference cell types, has gained wide usage (78, 79). The main limitations of inference of CNV from transcriptomic data are cell type-specific expression profiles and the unequal gene coverage biased towards the 3' or 5' end, which can lead to false-positive results, especially for smaller CNV regions (80). It is therefore recommended to establish ground truth knowledge of CNV changes using orthogonal technologies such as whole exome sequencing (WES) or clinical karyotyping. Other tools for inferring CNV from single-cell transcriptomes include HoneyBADGER (81), scCNAutils (82), CaSpER (83), DENDRO (84) or CopyKAT (85).

Somatic Nuclear DNA Mutations

Together with CNV, single somatic-nucleotide variants (sSNVs) of nuclear DNA have been at the center of numerous efforts to understand cancer evolution using bulk sequencing approaches (86). Although somatic mutations are less frequent in blood malignancies compared to solid tumors (87), they often can be linked to altered gene function implicated in tumorigenesis such as mutations in *TET2* (88), *DNMT3A* (89) (MDS/AML), *SF3B1* (90, 91) (CLL, MDS) or *JAK2* (92) (MPN) and acquired therapeutic resistance (BCR-ABL^{T315i}, BTK^{C481S}, PLCG2^{R665W}) (93, 94) (**Figure 1D**). While bulk sequencing infers clonal structures bioinformatically, single cell sequencing can directly measure coexistence of mutations within individual cells. Currently, two major strategies for identification of somatic mutations exist. These include either targeted scDNA-seq with a primary focus on variant calling or genotyping of single cell transcriptomics. The advantage of scDNA-seq is the ability to genotype dozens of loci in thousands of single cells, which provides very high resolution for tracking of mutation dynamics across cell compartments and time. RNA-based approaches are less efficient but can establish links between altered cell states and somatic mutations, for example through deconvolution of differential gene expression between mutated and non-mutated cells within the same cluster (8, 95). Importantly, both approaches are prone to allelic dropout, an inherent limitation of PCR amplification from limited starting material (96).

For targeted scDNA-seq platforms, primer panels for recurrent somatic mutations facilitate identification of these disease-specific variants (9, 22, 97). However, this approach is unable to target personal mutations and less frequent variants (98, 99) that would otherwise substantially increase the resolution of leukemic evolution (100, 101). This shortcoming can be addressed either through screening of personal mutations with WES and subsequent targeted single-cell sequencing (102, 103), or with unbiased scWGS-seq. Similar to the analysis of CNV, a critical step in the identification of somatic nuclear mutations from scWGS-seq involves stringent filtering of false positive results (104). Mutation calling can be performed with the Tapestri pipeline, which is based on the variant caller GATK4 (105).

The calling of somatic nuclear mutations from short-read transcriptomic sequencing libraries is conceptually feasible but in actuality faces several challenges. Due to the 3' or 5' bias and shallow coverage of individual transcripts, somatic mutations can only be reliably called in a minority of cells (<5%) from unmodified single-cell gene expression profiles. Targeted amplification of regions of interest can increase the coverage of mutated loci and enables mutation calling in a higher percentage of cells, ranging from <10 to >50% depending on the underlying sequencing platform, such as 3' or 5' bias for short-read sequencing, the location of the mutation within the transcript and its expression level (8, 75, 106). For processing of amplicons, custom solutions have been developed (8, 95). At the present time, this approach works best for loci close to the end of the cDNA template, while for mutations that are more distant, long-read sequencing can increase their coverage (56, 95).

Mitochondrial DNA Mutations

While CNV and somatic nuclear mutations are well-established approaches to lineage tracing, recently it has been recognized that mitochondrial DNA (mtDNA) mutations have the potential to serve as phylogenetic barcodes (**Figure 1E**) (107, 108). This provides opportunities for lineage-tracing at increased resolution within cell populations defined by the same set of somatic mutations or even in absence of such genetic events. Compared to somatic nuclear mutations, mtDNA mutations have various advantages. Mitochondria can replicate independently of the cell cycle, and thus mtDNA may be present in hundreds of copies per cell. In combination with the small size (~16.6 kB), mtDNA can therefore be sequenced with high coverage at single-cell resolution. Finally, mtDNA has a considerably higher mutation rate compared to genomic DNA (109, 110).

In principle, mtDNA mutation can be read out with the same approaches that apply to somatic nuclear mutations, however due to the unique biology of mtDNA, specific tools are being developed: mtscATAC-seq is a multi-omics platform that provides the feasibility of integrating information on cell state, CNV and mtDNA mutations of the entire mtDNA genome at single-cell resolution. For these studies, mgatk is an optimized variant caller which utilizes strand concordance of forward and reverse reads as well as mean variance ratio to enrich for true-positive mtDNA mutations (77).

The detection of mtDNA mutations from single-cell transcriptomic profiles is challenging as the coverage of mitochondrial transcripts is typically heterogeneous and insufficient to call mutations for large parts of the mtDNA genome, with a tendency for better coverage using full-length RNA-sequencing (108, 111). The MAESTER protocol overcomes these hurdles through a 65-primer multiplexed PCR amplification of mitochondrial transcripts to achieve sufficient coverage for mutation calling (112). By excluding UMIs with fewer than 3 reads and through consensus calling of reads from the same UMI, the variant caller optimized for MAESTER (maegtk) reduces false-positive mtDNA mutations deriving from PCR errors, and paves the way for more consistent calling of mtDNA mutations.

While studies on mtDNA mutations in blood malignancies are still in their infancy, it is already clear that they (i) are able to define subclonal structure within monoclonal cell populations,

(ii) remain stable over time in absence of strong selective pressure and (iii) change in frequency following therapeutic bottlenecks (75, 77, 106). Further, there is evidence that mtDNA and somatic nuclear mutations provide complementary information on clonal evolution in leukemia (106). However, many open questions remain such as whether mtDNA mutations provide selective advantage for example through alteration of oxidative phosphorylation (113, 114) or whether they are sufficiently sensitive and specific for the tracking of malignant clones in the context of low disease burden such as minimal residual disease.

In the non-malignant context, mtDNA mutations offer a possibility to track lineage within cell populations that lack natural barcodes such as monocytes or natural killer cells and enable to link differentiated cells to their progenitor and stem cell populations (115). There are also examples of how mtDNA mutations can track clonal expansion of T cells and further subdivide T cells with the same T cell receptor sequence, potentially providing an avenue to dissect dynamics within a T cell clone (75, 108).
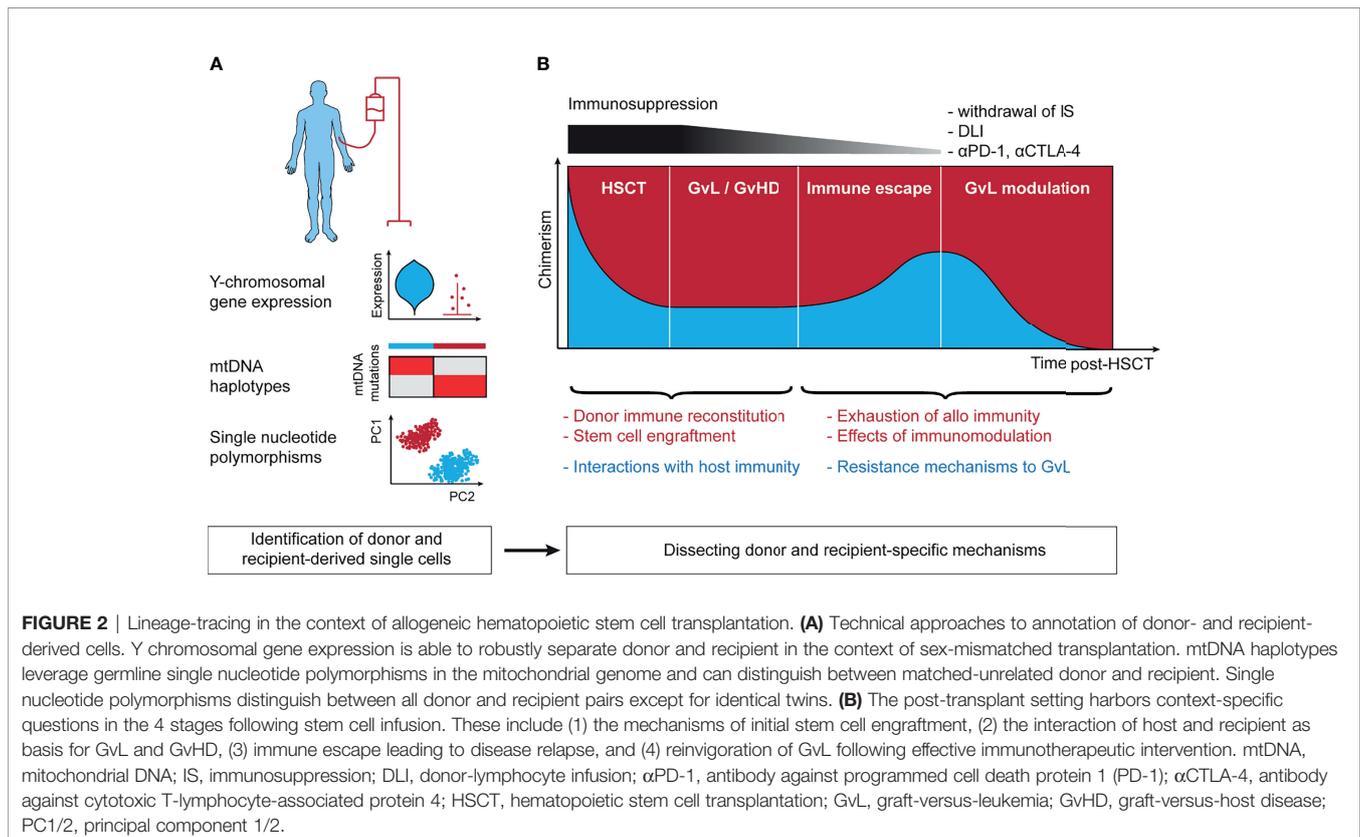
LINEAGE-TRACING IN THE CONTEXT OF ALLOGENEIC STEM CELL TRANSPLANTATION

HSCT is arguably the longest standing and the most successful immunotherapy with the graft-versus-leukemia/lymphoma (GvL) effect at its core (116). Research into target antigens of GvL and processes such as donor reconstitution or immunomodulation post-HSCT have advanced our understanding of fundamental immune processes (117–119). Despite these insights, many questions remain unanswered such as the exact mechanisms of response to immune modulation post-HSCT like donor-lymphocyte infusion (DLI) or interactions between donor and host under conditions of mixed chimerism. Lineage-tracing approaches are an opportunity to address these questions, leveraging the coexistence of donor- and recipient-derived cells but also require robust annotation of these two populations (**Figure 2A**).

Sex Mismatch

In the case of sex-mismatched HSCT (120), annotation of cells to donor and recipient can be based on expression of Y chromosomal genes including *RPS4Y1* and *DDX3Y* or genes implicated in X-inactivation such as *XIST* (121). As scRNA-seq data are typically sparse, with dropout of gene expression, higher accuracy can be expected from annotation using multiple genes. The strength of this approach is that no additional sequencing is needed and that it allows for annotation of even rare cells through a simple but robust analysis. Examples include tracking of circulating host tissue-resident T cells following myeloablative conditioning (122) or donor-derived immune cells following solid organ transplant (123). A limitation is that sex mismatch only occurs in a subset of transplants.

FIGURE 2 | Lineage-tracing in the context of allogeneic hematopoietic stem cell transplantation. **(A)** Technical approaches to annotation of donor- and recipient-derived cells. Y chromosomal gene expression is able to robustly separate donor and recipient in the context of sex-mismatched transplantation. mtDNA haplotypes leverage germline single nucleotide polymorphisms in the mitochondrial genome and can distinguish between matched-unrelated donor and recipient. Single nucleotide polymorphisms distinguish between all donor and recipient pairs except for identical twins. **(B)** The post-transplant setting harbors context-specific questions in the 4 stages following stem cell infusion. These include (1) the mechanisms of initial stem cell engraftment, (2) the interaction of host and recipient as basis for GvL and GvHD, (3) immune escape leading to disease relapse, and (4) reinvigoration of GvL following effective immunotherapeutic intervention. mtDNA, mitochondrial DNA; IS, immunosuppression; DLI, donor-lymphocyte infusion; αPD-1, antibody against programmed cell death protein 1 (PD-1); αCTLA-4, antibody against cytotoxic T-lymphocyte-associated protein 4; HSCT, hematopoietic stem cell transplantation; GvL, graft-versus-leukemia; GvHD, graft-versus-host disease; PC1/2, principal component 1/2.

Single Nucleotide Polymorphisms

Germline single nucleotide polymorphisms (SNP) can be used to distinguish donor and recipient unambiguously. Due to the large number of more than 1 million SNPs in the human genome (124), even siblings, except for identical twins, differ in their set of SNPs, thus allowing to identify the origin of single-cells from sparse scRNA-seq data with shallow coverage of individual loci. This can be achieved either through annotation with a genotype reference obtained using WES of purified donor or recipient-derived cells (125, 126) or with reference-free approaches that are based on statistical modeling. Tools that implement the latter strategy for scRNA-seq were originally developed for deconvolution of samples from different donors but can be used for assignment of donor and host at single-cell resolution (127–130). For scWGS-seq and scATAC-seq data, a similar approach is possible (131, 132), but has not been implemented yet.

mtDNA Haplotypes

Reference-free SNP-based annotation of donor- and recipient-derived cells from sparse single-cell data is based on clustering of similar cells and therefore accurate annotation of rare cells can be challenging. A possible alternative is to utilize mtDNA haplotypes for this purpose that arise from SNPs in the mitochondrial genome (133). From a technical standpoint this approach has the advantage that owing to the small size of the mtDNA genome and its high density of SNPs, the mtDNA haplotype is very informative and allows to unambiguously annotate the cell origin even from sparse single-cell data. This can be achieved either using the mtscATAC-seq protocol or with targeted amplification from scRNA-seq libraries (see Section "*Mitochondrial DNA Mutations*"). The only limitation is that due to the matrilineal inheritance of mitochondria, this approach is mainly useable for transplants from matched-unrelated or certain haploidentical donors.

Opportunities for Understanding HSCT With Single-Cell Sequencing

As a complex immunotherapy, there are many open questions regarding HSCT than can be divided by the 4 stages of the post-transplant setting: (1) engraftment (134) and immune reconstitution (135) following HSCT, (2) GvL and GvHD when stable engraftment has been achieved (136), (3) immune escape mechanisms leading to relapse (137), and (4) immune modulation approaches to reinstate effective GvL (117) (**Figure 2B**). Single-cell sequencing offers windows into gaining broad understanding of these processes in detail and can potentially leverage all lineage-tracing approaches discussed in this text. Although progress in these directions has been made (75, 100, 138), the envisioned efforts are still in their infancy.

Single-cell transcriptomics are currently underway to describe immune cell populations implicated in GvHD such as B cells (139), regulatory T cells (140), or cytotoxic CD4[+] T cells (141). In the latter example, the authors tracked expansion of a mutation in the *mTOR* gene within donor-derived T cells which they associate with persistent immune activation and GvHD. Lineage-tracing approaches have the potential to deepen the understanding of GvHD by dissecting the contribution of donor and recipient-derived T cells to GvHD target

organs. An example of this approach demonstrated that tissue-resident memory T cells can retain a large fraction of host cells despite full systemic chimerism (122).

More generally, conditions of mixed chimerism are an area where high-throughput distinguishing of donor and recipient will be able to address questions that so far have been difficult to answer, for example whether recipient-derived cells persist in specific T cell subsets such as CD8[+] T cells of patients with aplastic anemia (142). Similarly, lineage-tracing will allow the understanding of the engraftment of hematopoietic stem cells in more detail, for example in the context of donor clonal hematopoiesis of indeterminate potential (CHIP) where somatic mutations can serve as barcodes (143).

Finally, mechanisms that underly disease relapse following HSCT (75, 100, 144) and response to therapies that reinstate GvL through DLI (145) or checkpoint immune blockade (119, 146) are increasingly characterized using single-cell sequencing. Lineage-tracing has many potential applications in these studies, for example finer dissection of intraclonal evolution (147) or tracking of exhausted T cells before and after immunomodulation, as has been demonstrated in a large-scale characterization of bone marrow-derived T cell states following effective DLI (145, 148).

DISCUSSION

In-vivo lineage tracing approaches using natural barcodes are seeing breathtaking progress and technological advances of single-cell sequencing are enabling studies that seemed inconceivable only a few years ago. While technical hurdles are being removed, new questions are becoming relevant. With multi-omics sequencing platforms pushing the boundaries of what is possible, these technologies are also associated with unprecedented costs per analyzed sample, thereby leading to focus on select samples. Undoubtedly, high-resolution and multi-modal analyses of individual samples are exciting, yet true advances will depend on applying these novel technologies on well-designed clinical cohorts with longitudinal sampling, for example in the setting of clinical trials that aim to answer specific biological questions. With the prospect of single-cell sequencing achieving hundreds of thousands of cells per sample throughput, clinical applications that leverage longitudinal lineage-tracing approaches such as the measuring of minimal residual disease become a possibility.

AUTHOR CONTRIBUTIONS

LP designed the figures and wrote the manuscript. SG provided critical input. CW supervised and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

1. Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, et al. Human Haematopoietic Stem Cell Lineage Commitment Is a Continuous Process. *Nat Cell Biol* (2017) 19(4):271–81. doi: 10.1038/ncb3493

2. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. The Human Cell Atlas. *Elife* (2017) 6:e27041. doi: 10.7554/eLife.27041

3. Karamitros D, Stoilova B, Aboukhalil Z, Hamey F, Reinisch A, Samitsch M, et al. Single-Cell Analysis Reveals the Continuum of Human Lympho-Myeloid Progenitor Cells. *Nat Immunol* (2018) 19(1):85–97. doi: 10.1038/s41590-017-0001-2

4. Pellin D, Loperfido M, Baricordi C, Wolock SL, Montepeloso A, Weinberg OK, et al. A Comprehensive Single Cell Transcriptional Landscape of Human Hematopoietic Progenitors. *Nat Commun* (2019) 10(1):2395. doi: 10.1038/s41467-019-10291-0

5. Park J-E, Botting RA, Domínguez Conde C, Popescu D-M, Lavaert M, Kunz DJ, et al. A Cell Atlas of Human Thymic Development Defines T Cell Repertoire Formation. *Science* (2020) 367(6480):eaay3224. doi: 10.1126/science.aay3224

6. Le J, Park JE, Ha VL, Luong A, Branciamore S, Rodin AS, et al. Single-Cell RNA-Seq Mapping of Human Thymopoiesis Reveals Lineage Specification Trajectories and a Commitment Spectrum in T Cell Development. *Immunity* (2020) 52(6):1105–18.e9. doi: 10.1016/j.immuni.2020.05.010

7. King HW, Orban N, Riches JC, Clear AJ, Warnes G, Teichmann SA, et al. Single-Cell Analysis of Human B Cell Maturation Predicts How Antibody Class Switching Shapes Selection Dynamics. *Sci Immunol* (2021) 6(56):eabe6291. doi: 10.1126/sciimmunol.abe6291

8. van Galen P, Hovestadt V, Wadsworth Ii MH, Hughes TK, Griffin GK, Battaglia S, et al. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell* (2019) 176(6):1265–81.e24. doi: 10.1016/j.cell.2019.01.031

9. Miles LA, Bowman RL, Merlinsky TR, Csete IS, Ooi AT, Durruthy-Durruthy R, et al. Single-Cell Mutation Analysis of Clonal Evolution in Myeloid Malignancies. *Nature* (2020) 587(7834):477–82. doi: 10.1038/s41586-020-2864-x

10. Giustacchini A, Thongjuea S, Barkas N, Woll PS, Povinelli BJ, Booth CAG, et al. Single-Cell Transcriptomics Uncovers Distinct Molecular Signatures of Stem Cells in Chronic Myeloid Leukemia. *Nat Med* (2017) 23(6):692–702. doi: 10.1038/nm.4336

11. Gawad C, Koh W, Quake SR. Dissecting the Clonal Origins of Childhood Acute Lymphoblastic Leukemia by Single-Cell Genomics. *Proc Natl Acad Sci USA* (2014) 111(50):17947–52. doi: 10.1073/pnas.1420822111

12. Wiseman DH, Baker SM, Dongre AV, Gurashi K, Storer JA, Somervaille TC, et al. Chronic Myelomonocytic Leukemia Stem Cell Transcriptomes Anticipate Disease Morphology and Outcome. *EBioMedicine* (2020) 58:102904. doi: 10.1016/j.ebiom.2020.102904

13. Gaiti F, Chaligne R, Gu H, Brand RM, Kothen-Hill S, Schulman RC, et al. Epigenetic Evolution and Lineage Histories of Chronic Lymphocytic Leukaemia. *Nature* (2019) 569(7757):576–80. doi: 10.1038/s41586-019-1198-z

14. Zavidij O, Haradhvala NJ, Mouhieddine TH, Sklavenitis-Pistofidis R, Cai S, Reidy M, et al. Single-Cell RNA Sequencing Reveals Compromised Immune Microenvironment in Precursor Stages of Multiple Myeloma. *Nat Cancer* (2020) 1(5):493–506. doi: 10.1038/s43018-020-0053-3

15. Kester L, van Oudenaarden A. Single-Cell Transcriptomics Meets Lineage Tracing. *Cell Stem Cell* (2018) 23(2):166–79. doi: 10.1016/j.stem.2018.04.014

16. Wagner DE, Klein AM. Lineage Tracing Meets Single-Cell Omics: Opportunities and Challenges. *Nat Rev Genet* (2020) 21(7):410–27. doi: 10.1038/s41576-020-0223-2

17. Kebschull JM, Zador AM. Cellular Barcoding: Lineage Tracing, Screening and Beyond. *Nat Methods* (2018) 15(11):871–9. doi: 10.1038/s41592-018-0185-x

18. McKenna A, Gagnon JA. Recording Development With Single Cell Dynamic Lineage Tracing. *Development* (2019) 146(12):dev169730. doi: 10.1242/dev.169730

19. Woodworth MB, Girskis KM, Walsh CA. Building a Lineage From Single Cells: Genetic Techniques for Cell Lineage Tracking. *Nat Rev Genet* (2017) 18(4):230–44. doi: 10.1038/nrg.2016.159

20. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The Life History of 21 Breast Cancers. *Cell* (2012) 149(5):994–1007. doi: 10.1016/j.cell.2012.04.023

21. Landau DA, Sun C, Rosebrock D, Herman SEM, Fein J, Sivina M, et al. The Evolutionary Landscape of Chronic Lymphocytic Leukemia Treated With Ibrutinib Targeted Therapy. *Nat Commun* (2017) 8(1):2185. doi: 10.1038/s41467-017-02329-y

22. Morita K, Wang F, Jahn K, Hu T, Tanaka T, Sasaki Y, et al. Clonal Evolution of Acute Myeloid Leukemia Revealed by High-Throughput Single-Cell Genomics. *Nat Commun* (2020) 11(1):5327. doi: 10.1038/s41467-020-19119-8

23. Schattenberg A, Witte TD, Salden M, Vet J, Dijk BV, Smeets D, et al. Mixed Hematopoietic Chimerism After Allogeneic Transplantation With Lymphocyte-Depleted Bone Marrow Is Not Associated With a Higher Incidence of Relapse. *Blood* (1989) 73(5):1367–72. doi: 10.1182/blood.V73.5.1367.bloodjournal7351367

24. Schaap N, Schattenberg A, Bär B, Mensink E, Man AD, van Kessel AG, et al. Red Blood Cell Phenotyping Is a Sensitive Technique for Monitoring Chronic Myeloid Leukemia Patients After T-Cell-Depleted Bone Marrow Transplantation and After Donor Leucocyte Infusion. *Br J Haematology* (2000) 108(1):116–25. doi: 10.1046/j.1365-2141.2000.01803.x

25. Thiele J, Wickenhauser C, Kvasnicka HM, Varus E, Schneider C, Müller H, et al. Mixed Chimerism of Erythro- and Megakaryopoiesis Following Allogeneic Bone Marrow Transplantation. *Acta Haematol* (2003) 109(4):176–83. doi: 10.1159/000070966

26. Bielorai B, Trakhtenbrot L, Amariglio N, Rothman R, Tabori U, Dallal I, et al. Multilineage Hematopoietic Engraftment After Allogeneic Peripheral Blood Stem Cell Transplantation Without Conditioning in SCID Patients. *Bone Marrow Transplant* (2004) 34(4):317–20. doi: 10.1038/sj.bmt.1704565

27. Linder D, Gartler SM. Glucose-6-Phosphate Dehydrogenase Mosaicism: Utilization as a Cell Marker in the Study of Leiomyomas. *Science* (1965) 150(3692):67–9. doi: 10.1126/science.150.3692.67

28. Martin CL, Warburton D. Detection of Chromosomal Aberrations in Clinical Practice: From Karyotype to Genome Sequence. *Annu Rev Genomics Hum Genet* (2015) 16(1):309–26. doi: 10.1146/annurev-genom-090413-025346

29. Xia C, Fan J, Emanuel G, Hao J, Zhuang X. Spatial Transcriptome Profiling by MERFISH Reveals Subcellular RNA Compartmentalization and Cell Cycle-Dependent Gene Expression. *PNAS* (2019) 116(39):19490–9. doi: 10.1073/pnas.1912459116

30. Cabaniols J-P, Fazilleau N, Casrouge A, Kourilsky P, Kanellopoulos JM. Most α/β T Cell Receptor Diversity Is Due to Terminal Deoxynucleotidyl Transferase. *J Exp Med* (2001) 194(9):1385–90. doi: 10.1084/jem.194.9.1385

31. Penter L, Dietze K, Ritter J, Cobo MFL, Garmshausen J, Aigner F, et al. Localization-Associated Immune Phenotypes of Clonally Expanded Tumor-Infiltrating T Cells and Distribution of Their Target Antigens in Rectal Cancer. *OncoImmunology* (2019) 8(6):e1586409. doi: 10.1080/2162402X.2019.1586409

32. Penter L, Dietze K, Bullinger L, Westermann J, Rahn H-P, Hansmann L. FACS Single Cell Index Sorting Is Highly Reliable and Determines Immune Phenotypes of Clonally Expanded T Cells. *Eur J Immunol* (2018) 48(7):1248–50. doi: 10.1002/eji.201847507

33. Oliveira G, Stromhaug K, Klaeger S, Kula T, Frederick DT, Le PM, et al. Phenotype, Specificity and Avidity of Antitumour CD8+ T Cells in Melanoma. *Nature* (2021) 596(7870):119–25. doi: 10.1038/s41586-021-03704-y

34. Hu Z, Leet DE, Allesøe RL, Oliveira G, Li S, Luoma AM, et al. Personal Neoantigen Vaccines Induce Persistent Memory T Cell Responses and Epitope Spreading in Patients With Melanoma. *Nat Med* (2021) 27(3):515–25. doi: 10.1038/s41591-020-01206-4

35. Zhang S-Q, Ma K-Y, Schonnesen AA, Zhang M, He C, Sun E, et al. High-Throughput Determination of the Antigen Specificities of T Cell Receptors in Single Cells. *Nat Biotechnol* (2018) 36(12):1156–9. doi: 10.1038/nbt.4282

36. Yao W, Wu F, Zhang W, Chuang S, Thompson JS, Chen Z, et al. Angioimmunoblastic T-Cell Lymphoma Contains Multiple Clonal T-Cell Populations Derived From a Common TET2 Mutant Progenitor Cell. *J Pathol* (2020) 250(3):346–57. doi: 10.1002/path.5376

37. Herrera A, Cheng A, Mimitou EP, Seffens A, George DD, Bar-Natan M, et al. Multimodal Single-Cell Analysis of Cutaneous T Cell Lymphoma Reveals Distinct Sub-Clonal Tissue-Dependent Signatures. *Blood* (2021) 138 (16):1456–64. doi: 10.1182/blood.2020009346

38. Borcherding N, Voigt AP, Liu V, Link BK, Zhang W, Jabbari A. Single-Cell Profiling of Cutaneous T-Cell Lymphoma Reveals Underlying Heterogeneity Associated With Disease Progression. *Clin Cancer Res* (2019) 25(10):2996–3005. doi: 10.1158/1078-0432.CCR-18-3309

39. Papavasiliou FN, Schatz DG. Somatic Hypermutation of Immunoglobulin Genes: Merging Mechanisms for Genetic Diversity. *Cell* (2002) 109(2):S35–44. doi: 10.1016/S0092-8674(02)00706-7

40. Mishra AK, Mariuzza RA. Insights Into the Structural Basis of Antibody Affinity Maturation From Next-Generation Sequencing. *Front Immunol* (2018) 9:117. doi: 10.3389/fimmu.2018.00117

41. Mathew NR, Jayanthan JK, Smirnov IV, Robinson JL, Axelsson H, Nakka SS, et al. Single-Cell BCR and Transcriptome Analysis After Influenza Infection Reveals Spatiotemporal Dynamics of Antigen-Specific B Cells. *Cell Rep* (2021) 35(12):109286. doi: 10.1016/j.celrep.2021.109286

42. Hu Q, Hong Y, Qi P, Lu G, Mai X, Xu S, et al. Atlas of Breast Cancer Infiltrated B-Lymphocytes Revealed by Paired Single-Cell RNA-Sequencing and Antigen Receptor Profiling. *Nat Commun* (2021) 12(1):2186. doi: 10.1038/s41467-021-22300-2

43. Chen J, Tan Y, Sun F, Hou L, Zhang C, Ge T, et al. Single-Cell Transcriptome and Antigen-Immunoglobin Analysis Reveals the Diversity of B Cells in Non-Small Cell Lung Cancer. *Genome Biol* (2020) 21(1):152. doi: 10.1186/s13059-020-02064-6

44. Mlynarczyk C, Fontán L, Melnick A. Germinal Center-Derived Lymphomas: The Darkest Side of Humoral Immunity. *Immunol Rev* (2019) 288(1):214–39. doi: 10.1111/imr.12755

45. Ottensmeier CH, Thompsett AR, Zhu D, Wilkins BS, Sweetenham JW, Stevenson FK. Analysis of VH Genes in Follicular and Diffuse Lymphoma Shows Ongoing Somatic Mutation and Multiple Isotype Transcripts in Early Disease With Changes During Disease Progression. *Blood* (1998) 91 (11):4292–9. doi: 10.1182/blood.V91.11.4292

46. Jiang Y, Redmond D, Nie K, Eng KW, Clozel T, Martin P, et al. Deep Sequencing Reveals Clonal Evolution Patterns and Mutation Events Associated With Relapse in B-Cell Lymphomas. *Genome Biol* (2014) 15 (8):432. doi: 10.1186/PREACCEPT-6612547881370092

47. Hansmann L, Han A, Penter L, Liedtke M, Davis MM. Clonal Expansion and Interrelatedness of Distinct B-Lineage Compartments in Multiple Myeloma Bone Marrow. *Cancer Immunol Res* (2017) 5(9):744–54. doi: 10.1158/2326-6066.CIR-17-0012

48. Zelenetz AD, Chen TT, Levy R. Histologic Transformation of Follicular Lymphoma to Diffuse Lymphoma Represents Tumor Progression by a Single Malignant B Cell. *J Exp Med* (1991) 173(1):197–207. doi: 10.1084/jem.173.1.197

49. Gawad C, Pepin F, Carlton VEH, Klinger M, Logan AC, Miklos DB, et al. Massive Evolution of the Immunoglobulin Heavy Chain Locus in Children With B Precursor Acute Lymphoblastic Leukemia. *Blood* (2012) 120 (22):4407–17. doi: 10.1182/blood-2012-05-429811

50. Choi S, Henderson MJ, Kwan E, Beesley AH, Sutton R, Bahar AY, et al. Relapse in Children With Acute Lymphoblastic Leukemia Involving Selection of a Preexisting Drug-Resistant Subclone. *Blood* (2007) 110 (2):632–9. doi: 10.1182/blood-2007-01-067785

51. Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-Seq2 for Sensitive Full-Length Transcriptome Profiling in Single Cells. *Nat Methods* (2013) 10(11):1096–8. doi: 10.1038/nmeth.2639

52. Rizzetto S, Koppstein DNP, Samir J, Singh M, Reed JH, Cai CH, et al. B-Cell Receptor Reconstruction From Single-Cell RNA-Seq With VDJPuzzle. *Bioinformatics* (2018) 34(16):2846–7. doi: 10.1093/bioinformatics/bty203

53. Zhang L, Yu X, Zheng L, Zhang Y, Li Y, Fang Q, et al. Lineage Tracking Reveals Dynamic Relationships of T Cells in Colorectal Cancer. *Nature* (2018) 564(7735):268–72. doi: 10.1038/s41586-018-0694-x

54. Hagemann-Jensen M, Ziegenhain C, Chen P, Ramsköld D, Hendriks G-J, Larsson AJM, et al. Single-Cell RNA Counting at Allele and Isoform Resolution Using Smart-Seq3. *Nat Biotechnol* (2020) 38(6):708–14. doi: 10.1038/s41587-020-0497-0

55. Hagemann-Jensen M, Ziegenhain C, Sandberg R. Scalable Full-Transcript Coverage Single Cell RNA Sequencing With Smart-Seq3xpress. *biorxiv* (2021). doi: 10.1101/2021.07.10.451889

56. Singh M, Al-Eryani G, Carswell S, Ferguson JM, Blackburn J, Barton K, et al. High-Throughput Targeted Long-Read Single Cell Sequencing Reveals the Clonal and Transcriptional Landscape of Lymphocytes. *Nat Commun* (2019) 10(1):3120. doi: 10.1038/s41467-019-11049-4

57. Byrne A, Beaudin AE, Olsen HE, Jain M, Cole C, Palmer T, et al. Nanopore Long-Read RNAseq Reveals Widespread Transcriptional Variation Among the Surface Receptors of Individual B Cells. *Nat Commun* (2017) 8:16027. doi: 10.1038/ncomms16027

58. Han A, Glanville J, Hansmann L, Davis MM. Linking T-Cell Receptor Sequence to Functional Phenotype at the Single-Cell Level. *Nat Biotechnol* (2014) 32(7):684–92. doi: 10.1038/nbt.2938

59. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively Parallel Digital Transcriptional Profiling of Single Cells. *Nat Commun* (2017) 8(1):14049. doi: 10.1038/ncomms14049

60. Bolotin DA, Poslavsky S, Davydov AN, Frenkel FE, Fanchi L, Zolotareva OI, et al. Antigen Receptor Repertoire Profiling From RNA-Seq Data. *Nat Biotechnol* (2017) 35(10):908–11. doi: 10.1038/nbt.3979

61. Song L, Cohen D, Ouyang Z, Cao Y, Hu X, Liu XS. TRUST4: Immune Repertoire Reconstruction From Bulk and Single-Cell RNA-Seq Data. *Nat Methods* (2021) 18(6):627–30. doi: 10.1038/s41592-021-01142-2

62. Ribera J, Zamora L, Morgades M, Mallo M, Solanes N, Batlle M, et al. Copy Number Profiling of Adult Relapsed B-Cell Precursor Acute Lymphoblastic Leukemia Reveals Potential Leukemia Progression Mechanisms. *Genes Chromosomes Cancer* (2017) 56(11):810–20. doi: 10.1002/gcc.22486

63. Bokemeyer A, Eckert C, Meyr F, Koerner G, von Stackelberg A, Ullmann R, et al. Copy Number Genome Alterations Are Associated With Treatment Response and Outcome in Relapsed Childhood ETV6/RUNX1-Positive Acute Lymphoblastic Leukemia. *Haematologica* (2014) 99(4):706–14. doi: 10.3324/haematol.2012.072470

64. Pasqualucci L, Khiabanian H, Fangazio M, Vasishtha M, Messina M, Holmes AB, et al. Genetics of Follicular Lymphoma Transformation. *Cell Rep* (2014) 6(1):130–40. doi: 10.1016/j.celrep.2013.12.027

65. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal Evolution in Relapsed Acute Myeloid Leukaemia Revealed by Whole-Genome Sequencing. *Nature* (2012) 481(7382):506–10. doi: 10.1038/nature10738

66. Fu Y, Zhang F, Zhang X, Yin J, Du M, Jiang M, et al. High-Throughput Single-Cell Whole-Genome Amplification Through Centrifugal Emulsification and eMDA. *Commun Biol* (2019) 2(1):1–10. doi: 10.1038/s42003-019-0401-y

67. Kohlmann A, Bullinger L, Thiede C, Schaich M, Schnittger S, Döhner K, et al. Gene Expression Profiling in AML With Normal Karyotype Can Predict Mutations for Molecular Markers and Allows Novel Insights Into Perturbed Biological Pathways. *Leukemia* (2010) 24(6):1216–20. doi: 10.1038/leu.2010.73

68. Laks E, McPherson A, Zahn H, Lai D, Steif A, Brimhall J, et al. Clonal Decomposition and DNA Replication States Defined by Scaled Single-Cell Genome Sequencing. *Cell* (2019) 179(5):1207–21.e22. doi: 10.1016/j.cell.2019.10.026

69. Zhang L, Dong X, Lee M, Maslov AY, Wang T, Vijg J. Single-Cell Whole-Genome Sequencing Reveals the Functional Landscape of Somatic Mutations in B Lymphocytes Across the Human Lifespan. *PNAS* (2019) 116(18):9014–9. doi: 10.1073/pnas.1902510116

70. Velazquez-Villarreal EI, Maheshwari S, Sorenson J, Fiddes IT, Kumar V, Yin Y, et al. Single-Cell Sequencing of Genomic DNA Resolves Sub-Clonal

Heterogeneity in a Melanoma Cell Line. *Commun Biol* (2020) 3(1):1–8. doi: 10.1038/s42003-020-1044-8

71. Chen X, Miragaia RJ, Natarajan KN, Teichmann SA. A Rapid and Robust Method for Single Cell Chromatin Accessibility Profiling. *Nat Commun* (2018) 9(1):5345. doi: 10.1038/s41467-018-07771-0

72. Satpathy AT, Granja JM, Yost KE, Qi Y, Meschi F, McDermott GP, et al. Massively Parallel Single-Cell Chromatin Landscapes of Human Immune Cell Development and Intratumoral T Cell Exhaustion. *Nat Biotechnol* (2019) 37(8):925–36. doi: 10.1038/s41587-019-0206-z

73. Chaligne R, Gaiti F, Silverbush D, Schiffman JS, Weisman HR, Kluegel L, et al. Epigenetic Encoding, Heritability and Plasticity of Glioma Transcriptional Cell States. *Nat Genet* (2021) 53(10):1469–79. doi: 10.1038/s41588-021-00927-7

74. Demaree B, Delley CL, Vasudevan HN, Peretz CAC, Ruff D, Smith CC, et al. Joint Profiling of DNA and Proteins in Single Cells to Dissect Genotype-Phenotype Associations in Leukemia. *Nat Commun* (2021) 12(1):1583. doi: 10.1038/s41467-021-21810-3

75. Penter L, Gohil SH, Lareau C, Ludwig LS, Parry EM, Huang T, et al. Longitudinal Single-Cell Dynamics of Chromatin Accessibility and Mitochondrial Mutations in Chronic Lymphocytic Leukemia Mirror Disease History. *Cancer Discov* (2021) 11:3048–63. doi: 10.1158/2159-8290.CD-21-0276

76. Mallory XF, Edrisi M, Navin N, Nakhleh L. Methods for Copy Number Aberration Detection From Single-Cell DNA-Sequencing Data. *Genome Biol* (2020) 21:208. doi: 10.1186/s13059-020-02119-8

77. Lareau CA, Ludwig LS, Muus C, Gohil SH, Zhao T, Chiang Z, et al. Massively Parallel Single-Cell Mitochondrial DNA Genotyping and Chromatin Profiling. *Nat Biotechnol* (2021) 39(4):451–61. doi: 10.1038/s41587-020-0645-6

78. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, et al. Single-Cell RNA-Seq Highlights Intratumoral Heterogeneity in Primary Glioblastoma. *Science* (2014) 344(6190):1396–401. doi: 10.1126/science.1254257

79. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the Multicellular Ecosystem of Metastatic Melanoma by Single-Cell RNA-Seq. *Science* (2016) 352(6282):189–96. doi: 10.1126/science.aad0501

80. Durante MA, Rodriguez DA, Kurtenbach S, Kuznetsov JN, Sanchez MI, Decatur CL, et al. Single-Cell Analysis Reveals New Evolutionary Complexity in Uveal Melanoma. *Nat Commun* (2020) 11:496. doi: 10.1038/s41467-019-14256-1

81. Fan J, Lee H-O, Lee S, Ryu D-E, Lee S, Xue C, et al. Linking Transcriptional and Genetic Tumor Heterogeneity Through Allele Analysis of Single-Cell RNA-Seq Data. *Genome Res* (2018) 28(8):1217–27. doi: 10.1101/gr.228080.117

82. Couturier CP, Ayyadhury S, Le PU, Nadaf J, Monlong J, Riva G, et al. Single-Cell RNA-Seq Reveals That Glioblastoma Recapitulates a Normal Neurodevelopmental Hierarchy. *Nat Commun* (2020) 11(1):3406. doi: 10.1038/s41467-020-17979-8

83. Serin Harmanci A, Harmanci AO, Zhou X. CaSpER Identifies and Visualizes CNV Events by Integrative Analysis of Single-Cell or Bulk RNA-Sequencing Data. *Nat Commun* (2020) 11(1):89. doi: 10.1038/s41467-019-13779-x

84. Zhou Z, Xu B, Minn A, Zhang NR. DENDRO: Genetic Heterogeneity Profiling and Subclone Detection by Single-Cell RNA Sequencing. *Genome Biol* (2020) 21(1):10. doi: 10.1186/s13059-019-1922-x

85. Gao R, Bai S, Henderson YC, Lin Y, Schalck A, Yan Y, et al. Delineating Copy Number and Clonal Substructure in Human Tumors From Single-Cell Transcriptomes. *Nat Biotechnol* (2021) 39(5):599–608. doi: 10.1038/s41587-020-00795-2

86. Black JRM, McGranahan N. Genetic and Non-Genetic Clonal Diversity in Cancer Evolution. *Nat Rev Cancer* (2021) 21(6):379–92. doi: 10.1038/s41568-021-00336-2

87. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The Repertoire of Mutational Signatures in Human Cancer. *Nature* (2020) 578(7793):94–101. doi: 10.1038/s41586-020-1943-3

88. Delhommeau F, Dupont S, Della Valle V, James C, Trannoy S, Massé A, et al. Mutation in TET2 in Myeloid Cancers. *N Engl J Med* (2009) 360(22):2289–301. doi: 10.1056/NEJMoa0810069

89. Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, et al. DNMT3A Mutations in Acute Myeloid Leukemia. *N Engl J Med* (2010) 363 (25):2424–33. doi: 10.1056/NEJMoa1005143

90. Wan Y, Wu CJ. SF3B1 Mutations in Chronic Lymphocytic Leukemia. *Blood* (2013) 121(23):4627–34. doi: 10.1182/blood-2013-02-427641

91. Malcovati L, Stevenson K, Papaemmanuil E, Neuberg D, Bejar R, Boultwood J, et al. SF3B1-Mutant MDS as a Distinct Disease Subtype: A Proposal From the International Working Group for the Prognosis of MDS. *Blood* (2020) 136(2):157–70. doi: 10.1182/blood.2020004850

92. Levine RL, Wadleigh M, Cools J, Ebert BL, Wernig G, Huntly BJP, et al. Activating Mutation in the Tyrosine Kinase JAK2 in Polycythemia Vera, Essential Thrombocythemia, and Myeloid Metaplasia With Myelofibrosis. *Cancer Cell* (2005) 7(4):387–97. doi: 10.1016/j.ccr.2005.03.023

93. O'Hare T, Eide CA, Deininger MWN. Bcr-Abl Kinase Domain Mutations, Drug Resistance, and the Road to a Cure for Chronic Myeloid Leukemia. *Blood* (2007) 110(7):2242–9. doi: 10.1182/blood-2007-03-066936

94. Puła B, Gołos A, Górniak P, Jamroziak K. Overcoming Ibrutinib Resistance in Chronic Lymphocytic Leukemia. *Cancers (Basel)* (2019) 11(12):1834. doi: 10.3390/cancers11121834

95. Nam AS, Kim K-T, Chaligne R, Izzo F, Ang C, Taylor J, et al. Somatic Mutations and Cell Identity Linked by Genotyping of Transcriptomes. *Nature* (2019) 571(7765):355–60. doi: 10.1038/s41586-019-1367-0

96. Wang C, Schroeder KB, Rosenberg NA. A Maximum-Likelihood Method to Correct for Allelic Dropout in Microsatellite Data With No Replicate Genotypes. *Genetics* (2012) 192(2):651–69. doi: 10.1534/genetics.112.139519

97. Albertí-Servera L, Demeyer S, Govaerts I, Swings T, De Bie J, Gielen O, et al. Single-Cell DNA Amplicon Sequencing Reveals Clonal Heterogeneity and Evolution in T-Cell Acute Lymphoblastic Leukemia. *Blood* (2021) 137 (6):801–11. doi: 10.1182/blood.2020006996

98. Landau DA, Tausch E, Taylor-Weiner AN, Stewart C, Reiter JG, Bahlo J, et al. Mutations Driving CLL and Their Evolution in Progression and Relapse. *Nature* (2015) 526(7574):525–30. doi: 10.1038/nature15395

99. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* (2016) 374(23):2209–21. doi: 10.1056/NEJMoa1516192

100. Bachireddy P, Ennis C, Nguyen VN, Gohil SH, Clement K, Shukla SA, et al. Distinct Evolutionary Paths in Chronic Lymphocytic Leukemia During Resistance to the Graft-Versus-Leukemia Effect. *Sci Trans Med* (2020) 12 (561):eabb7661. doi: 10.1126/scitranslmed.abb7661

101. Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, et al. Evolution and Impact of Subclonal Mutations in Chronic Lymphocytic Leukemia. *Cell* (2013) 152(4):714–26. doi: 10.1016/j.cell.2013.01.019

102. Dillon LW, Ghannam J, Nosiri C, Gui G, Goswami M, Calvo KR, et al. Personalized Single-Cell Proteogenomics to Distinguish Acute Myeloid Leukemia From Nonmalignant Clonal Hematopoiesis. *Blood Cancer Discov* (2021) 2(4):319–25. doi: 10.1158/2643-3230.BCD-21-0046

103. Wang L, Fan J, Francis JM, Georghiou G, Hergert S, Li S, et al. Integrated Single-Cell Genetic and Transcriptional Analysis Suggests Novel Drivers of Chronic Lymphocytic Leukemia. *Genome Res* (2017) 27(8):1300–11. doi: 10.1101/gr.217331.116

104. Luquette LJ, Bohrson CL, Sherman MA, Park PJ. Identification of Somatic Mutations in Single Cell DNA-Seq Using a Spatial Model of Allelic Imbalance. *Nat Commun* (2019) 10(1):3908. doi: 10.1038/s41467-019-11857-8

105. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr Protoc Bioinf* (2013) 43(1):11.10.1–11.10.33. doi: 10.1002/0471250953.bi1110s43

106. Velten L, Story BA, Hernández-Malmierca P, Raffel S, Leonce DR, Milbank J, et al. Identification of Leukemic and Pre-Leukemic Stem Cells by Clonal Tracking From Single-Cell Transcriptomics. *Nat Commun* (2021) 12 (1):1366. doi: 10.1038/s41467-021-21650-1

107. Xu J, Nuno K, Litzenburger UM, Qi Y, Corces MR, Majeti R, et al. Single-Cell Lineage Tracing by Endogenous Mutations Enriched in Transposase Accessible Mitochondrial DNA. *Elife* (2019) 8:e45105. doi: 10.7554/eLife.45105

108. Ludwig LS, Lareau CA, Ulirsch JC, Christian E, Muus C, Li LH, et al. Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics. *Cell* (2019) 176(6):1325–1339.e22. doi: 10.1016/j.cell.2019.01.022

109. Lawless C, Greaves L, Reeve AK, Turnbull DM, Vincent AE. The Rise and Rise of Mitochondrial DNA Mutations. *Open Biol* (2020) 10(5):200061. doi: 10.1098/rsob.200061

110. Stewart JB, Chinnery PF. Extreme Heterogeneity of Human Mitochondrial DNA From Organelles to Populations. *Nat Rev Genet* (2021) 22(2):106–18. doi: 10.1038/s41576-020-00284-x

111. Marshall AS, Jones NS. Discovering Cellular Mitochondrial Heteroplasmy Heterogeneity With Single Cell RNA and ATAC Sequencing. *Biol (Basel)* (2021) 10(6):503. doi: 10.3390/biology10060503

112. Miller TE, Lareau CA, Verga JA, Ssozi D, Ludwig LS, Farran CE, et al. Mitochondrial Variant Enrichment From High-Throughput Single-Cell RNA-Seq Resolves Clonal Populations. *biorxiv* (2021). doi: 10.1101/2021.03.08.434450

113. Smith ALM, Whitehall JC, Bradshaw C, Gay D, Robertson F, Blain AP, et al. Age-Associated Mitochondrial DNA Mutations Cause Metabolic Remodeling That Contributes to Accelerated Intestinal Tumorigenesis. *Nat Cancer* (2020) 1(10):976–89. doi: 10.1038/s43018-020-00112-5

114. Schöpf B, Weissensteiner H, Schäfer G, Fazzini F, Charoentong P, Naschberger A, et al. OXPHOS Remodeling in High-Grade Prostate Cancer Involves mtDNA Mutations and Increased Succinate Oxidation. *Nat Commun* (2020) 11(1):1487. doi: 10.1038/s41467-020-15237-5

115. Lareau CA, Ludwig LS, Sankaran VG. Longitudinal Assessment of Clonal Mosaicism in Human Hematopoiesis *via* Mitochondrial Mutation Tracking. *Blood Adv* (2019) 3(24):4161–5. doi: 10.1182/bloodadvances.2019001196

116. Cieri N, Maurer K, Wu CJ. 60 Years Young: The Evolving Role of Allogeneic Hematopoietic Stem Cell Transplantation in Cancer Immunotherapy. *Cancer Res* (2021) 81(17):4373–84. doi: 10.1158/0008-5472.CAN-21-0301

117. Penter L, Wu CJ. Personal Tumor Antigens in Blood Malignancies: Genomics-Directed Identification and Targeting. *J Clin Invest* (2020) 130(4):1595–607. doi: 10.1172/JCI129209

118. Penter L, Zhang Y, Savell A, Huang T, Cieri N, Thrash EM, et al. Molecular and Cellular Features of CTLA-4 Blockade for Relapsed Myeloid Malignancies After Transplantation. *Blood* (2021) 137(23):3212–7. doi: 10.1182/blood.2021010867

119. Penter L, Gohil SH, Huang T, Thrash EM, Schmidt D, Li S, et al. Coevolving JAK2V617F+ Relapsed AML and Donor T Cells With PD-1 Blockade After Stem Cell Transplantation: An Index Case. *Blood Adv* (2021) 5(22):4701–9. doi: 10.1182/bloodadvances.2021004335

120. Randolph SSB, Gooley TA, Warren EH, Appelbaum FR, Riddell SR. Female Donors Contribute to a Selective Graft-Versus-Leukemia Effect in Male Recipients of HLA-Matched, Related Hematopoietic Stem Cell Transplants. *Blood* (2004) 103(1):347–52. doi: 10.1182/blood-2003-07-2603

121. Staedtler F, Hartmann N, Letzkus M, Bongiovanni S, Scherer A, Marc P, et al. Robust and Tissue-Independent Gender-Specific Transcript Biomarkers. *Biomarkers* (2013) 18(5):436–45. doi: 10.3109/1354750X.2013.811538

122. de Almeida GP, Lichtner P, Mädler S, Chu C-F, Zielinski CE. Transcriptional Profiling and Single-Cell Chimerism Analysis Identifies Human Tissue Resident T Cells in the Human Skin After Allogeneic Stem Cell Transplantation. *biorxiv* (2020). doi: 10.1101/2020.04.11.037101

123. Byrne AJ, Powell JE, O'Sullivan BJ, Ogger PP, Hoffland A, Cook J, et al. Dynamics of Human Monocytes and Airway Macrophages During Healthy Aging and After Transplant. *J Exp Med* (2020) 217(3):e20191236. doi: 10.1084/jem.20191236

124. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, et al. A Map of Human Genome Sequence Variation Containing 1.42 Million Single Nucleotide Polymorphisms. *Nature* (2001) 409(6822):928–33. doi: 10.1038/35057149

125. Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, McCarthy E, et al. Multiplexed Droplet Single-Cell RNA-Sequencing Using Natural Genetic Variation. *Nat Biotechnol* (2018) 36(1):89–94. doi: 10.1038/nbt.4042

126. Malone AF, Wu H, Fronick C, Fulton R, Gaut JP, Humphreys BD. Harnessing Expressed Single Nucleotide Variation and Single Cell RNA Sequencing To Define Immune Cell Chimerism in the Rejecting Kidney Transplant. *JASN* (2020) 31(9):1977–86. doi: 10.1681/ASN.2020030326

127. Heaton H, Talman AM, Knights A, Imaz M, Gaffney DJ, Durbin R, et al. Souporcell: Robust Clustering of Single-Cell RNA-Seq Data by Genotype Without Reference Genotypes. *Nat Methods* (2020) 17(6):615–20. doi: 10.1038/s41592-020-0820-1

128. Xu J, Falconer C, Nguyen Q, Crawford J, McKinnon BD, Mortlock S, et al. Genotype-Free Demultiplexing of Pooled Single-Cell RNA-Seq. *Genome Biol* (2019) 20(1):290. doi: 10.1186/s13059-019-1852-7

129. Huang Y, McCarthy DJ, Stegle O. Vireo: Bayesian Demultiplexing of Pooled Single-Cell RNA-Seq Data Without Genotype Reference. *Genome Biol* (2019) 20(1):273. doi: 10.1186/s13059-019-1865-2

130. McGinnis CS, Siegel DA, Xie G, Hartoularos G, Stone M, Ye CJ, et al. No Detectable Alloreactive Transcriptional Responses Under Standard Sample Preparation Conditions During Donor-Multiplexed Single-Cell RNA Sequencing of Peripheral Blood Mononuclear Cells. *BMC Biol* (2021) 19(1):10. doi: 10.1186/s12915-020-00941-x

131. Chan Y, Chan YK, Goodman DB, Guo X, Chavez A, Lim ET, et al. Enabling Multiplexed Testing of Pooled Donor Cells Through Whole-Genome Sequencing. *Genome Med* (2018) 10(1):31. doi: 10.1186/s13073-018-0541-6

132. Massarat AR, Sen A, Jaureguy J, Tyndale ST, Fu Y, Erikson G, et al. Discovering Single Nucleotide Variants and Indels From Bulk and Single-Cell ATAC-Seq. *Nucleic Acids Res* (2021) 49(14):7986–94. doi: 10.1093/nar/gkab621

133. Mitchell SL, Goodloe R, Brown-Gentry K, Pendergrass SA, Murdock DG, Crawford DC. Characterization of Mitochondrial Haplogroups in a Large Population-Based Sample From the United States. *Hum Genet* (2014) 133(7):861–8. doi: 10.1007/s00439-014-1421-9

134. Hutt D. Engraftment, Graft Failure, and Rejection. In: Kenyon M, Babic A, editors. *The European Blood and Marrow Transplantation Textbook for Nurses: Under the Auspices of EBMT [Internet]*. Cham, Switzerland: Springer (2018). Available from: http://www.ncbi.nlm.nih.gov/books/NBK543659/.

135. Ogonek J, Kralj Juric M, Ghimire S, Varanasi PR, Holler E, Greinix H, et al. Immune Reconstitution After Allogeneic Hematopoietic Stem Cell Transplantation. *Front Immunol* (2016) 7:507. doi: 10.3389/fimmu.2016.00507

136. Negrin RS. Graft-Versus-Host Disease Versus Graft-Versus-Leukemia. *Hematol Am Soc Hematol Educ Program* (2015) 2015:225–30. doi: 10.1182/asheducation-2015.1.225

137. Zeiser R, Vago L. Mechanisms of Immune Escape After Allogeneic Hematopoietic Cell Transplantation. *Blood* (2019) 133(12):1290–7. doi: 10.1182/blood-2018-10-846824

138. Rodriguez-Fraticelli AE, Weinreb C, Wang S-W, Migueles RP, Jankovic M, Usart M, et al. Single-Cell Lineage Tracing Unveils a Role for TCF15 in Haematopoiesis. *Nature* (2020) 583(7817):585–9. doi: 10.1038/s41586-020-2503-6

139. Poe JC, Zhang D, Xie J, DiCioccio RA, Qin X, Fang J, et al. Single-Cell RNA-Seq Identifies Potentially Pathogenic B Cell Populations That Uniquely Circulate in Patients With Chronic Gvhd. *Blood* (2019) 134(Supplement_1):874–4. doi: 10.1182/blood-2019-130928

140. Luo Y, Xu C, Wang B, Niu Q, Su X, Bai Y, et al. Single-Cell Transcriptomic Analysis Reveals Disparate Effector Differentiation Pathways in Human Treg Compartment. *Nat Commun* (2021) 12(1):3913. doi: 10.1038/s41467-021-24213-6

141. Kim D, Park G, Huuhtanen J, Lundgren S, Khajuria RK, Hurtado AM, et al. Somatic mTOR Mutation in Clonally Expanded T Lymphocytes Associated With Chronic Graft Versus Host Disease. *Nat Commun* (2020) 11(1):2246. doi: 10.1038/s41467-020-16115-w

142. Grimaldi F, Potter V, Perez-Abellan P, Veluchamy JP, Atif M, Grain R, et al. Mixed T Cell Chimerism After Allogeneic Hematopoietic Stem Cell Transplantation for Severe Aplastic Anemia Using an Alemtuzumab-Containing Regimen Is Shaped by Persistence of Recipient CD8 T Cells. *Biol Blood Marrow Transplant* (2017) 23(2):293–9. doi: 10.1016/j.bbmt.2016.11.003

143. Frick M, Chan W, Arends CM, Hablesreiter R, Halik A, Heuser M, et al. Role of Donor Clonal Hematopoiesis in Allogeneic Hematopoietic Stem-Cell Transplantation. *J Clin Oncol* (2019) 37(5):375–85. doi: 10.1200/JCO.2018.79.2184

144. Ediriwickrema A, Aleshin A, Reiter JG, Corces MR, Köhnke T, Stafford M, et al. Single-Cell Mutational Profiling Enhances the Clinical Evaluation of AML MRD. *Blood Adv* (2020) 4(5):943–52. doi: 10.1182/bloodadvances.2019001181

145. Bachireddy P, Azizi E, Burdziak C, Nguyen VN, Ennis CS, Maurer K, et al. Mapping the Evolution of T Cell States During Response and Resistance to Adoptive Cellular Therapy. *Cell Rep* (2021) 37(6):109992. doi: 10.1016/j.celrep.2021.109992

146. Abbas HA, Hao D, Tomczak K, Barrodia P, Im JS, Reville PK, et al. Single-Cell Characterization of Acute Myeloid Leukemia and Its Microenvironment Following PD-1 Blockade Based Therapy. *Nat Commun* (2021) 12(1):6071. doi: 10.1101/2020.09.03.278309

147. Xu L, Durruthy-Durruthy R, Eastburn DJ, Pellegrino M, Shah O, Meyer E, et al. Clonal Evolution and Changes in Two AML Patients Detected With A Novel Single-Cell DNA Sequencing Platform. *Sci Rep* (2019) 9(1):11119. doi: 10.1038/s41598-019-47297-z

148. Noviello M, Manfredi F, Ruggiero E, Perini T, Oliveira G, Cortesi F, et al. Bone Marrow Central Memory and Memory Stem T-Cell Exhaustion in AML Patients Relapsing After HSCT. *Nat Commun* (2019) 10(1):1065. doi: 10.1038/s41467-019-08871-1

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.