A Track of the Clones: New developments in cellular barcoding

*Anne-Marie Lyne[1], *David G Kent[2,3], *Elisa Laurenti[2,3], Kerstin Cornils[4], *Ingmar Glauche[5], and *Leila Perié[1]

[1]Institute Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France
[2]Wellcome MRC Stem Cell Institute, University of Cambridge, Hills Road, Cambridge, CB2 0XY, United Kingdom
[3]Department of Haematology, University of Cambridge, CB2 0XY, United Kingdom
[4]UK Hamburg-Eppendorf, Hamburg, Germany
[5]TU Dresden, Dresden, Germany
*These authors contributed equally

Address correspondence:

Leila Perié Institute Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France
leila.perie@curie.fr
Tel: +33 1 56 24 63 60
or
David G. Kent, Wellcome MRC Stem Cell Institute, University of Cambridge, Cambridge, CB2 0AH, United Kingdom
dgk23@cam.ac.uk
Tel: +44 1223 762130

Summary

International experts from multiple disciplines gathered at Homerton College in Cambridge, UK from September 12-14, 2018 to consider recent advances and emerging opportunities in the clonal tracking of hematopoiesis in one of a series of StemCellMathLab workshops. The group included thirty-five participants with experience in the fields of theoretical and experimental aspects of clonal tracking, and ranged from PhD students to senior professors. Data from a variety of model systems as well as from clinical gene therapy trials were discussed alongside strategies for data analysis and sharing, as well as challenges arising due to underlying assumptions in data interpretation and communication. Recognizing the power of this technology underpinned a group consensus of a need for improved mechanisms for sharing data and analytical protocols to maintain reproducibility and rigor in its application to complex tissues.

Introduction

Methods to unambiguously mark individual cells and follow their clonal progeny over extended periods of time can provide unprecedented insights into the organizational principles operative in tissue homeostasis and regeneration. The potential of clonal tracking was initially demonstrated by studies of glucose-6-phosphate dehydrogenase (G-6-PD) cellular mosaicism (Fialkow 1980, Abkowitz, Ott et al. 1985). More recently, viral integration site analysis (IS) (Jordan and Lemischka 1990),(Schmidt, Glimm et al. 2001),(Biasco, Pellin et al. 2016) and integration of unique genetic "barcode" sequences (Gerrits, Dykstra et al. 2010),(Naik, Perie et al. 2013),(Lu, Neff et al. 2011),(Cornils, Thielecke et al. 2014) have been complemented by inducible transposon systems (Sun, Ramos et al. 2014),(Rodriguez-Fraticelli, Wolock et al. 2018) and the recombination of multiple segments separated by loxP sites to generate uniquely identifiable barcodes *in vivo* (Pei, Feyerabend et al. 2017). Much of this work has focused on elucidating the process of hematopoiesis and in this context has recently played a major role in revealing significant heterogeneity in populations previously considered as homogeneous in their potential. Clonal tracking has also helped to elucidate the dynamics of naïve versus post-transplantation hematopoiesis, and is now stimulating new ideas about hematopoietic differentiation processes. Rapid technological developments and reduced costs of genomic sequencing are also now promoting the introduction of new approaches in cellular barcoding with anticipated increased impact on understanding mechanisms controlling hematopoiesis *in vivo*.

As with any new technology, different experimental applications have led to a spectrum of solutions. These often involve different assumptions that impact the interpretation of derived datasets. As an example, variations in experimental protocols for DNA barcoding and related analytical tools have limited the comparison of results from related studies. The wealth of data present in clonal tracking studies also generates new challenges on quality control, data management and bioinformatic analysis. In addition, addressing system-wide questions regarding cell intermediates and population dynamics often requires the use of advanced mathematical and computational methods that can pose challenges for biologists who may be unfamiliar with the underlying assumptions and applicability of specific approaches. In particular, reconstruction of

differentiation pathways can impose several analytical and numerical challenges, especially when sampling depth is limited.

The *StemCellMathLab* workshop, which took place at Homerton College, Cambridge, UK on September 12th to 14th, focused on the central issue of how best to produce and make use of quantitative clonal data. There was clear consensus on the need for the development and sharing of field-standard tools, as well as for the compulsory deposition of cellular barcoding datasets in an online repository.

What's in a name? That which we call a cell state by any other name...

Description and reporting of experimental findings are usually language-based to convey the experimental approaches and results, as well as to define an underlying framework. Such descriptions commonly mean different things to different people. For example, the term "hematopoietic stem cells" has been variably used to refer to *populations* of cells that display certain functional activities (regeneration after irradiation), or a common set of cell surface markers, or more generally to the concept of a self-maintaining multipotent cell population. Similarly, a "clone" can relate to an entire organism (e.g., "Dolly" the sheep), the entire progeny arising from a uniquely marked founder cell or simply to a population of cells that share certain genetic aberrations (e.g., a mutant clone in cancer). However, the term "clone" becomes unambiguous if the unique mark identifying the founder cell is specified, be it a DNA barcode, a particular mutation, or a single cell from which a clone *in vitro* has been derived. Going forward, precise definitions will be critical to avoid confusion and facilitate replication of findings.

This issue is particularly important to analyses of the stages of hematopoietic cell differentiation, and hence to the majority of clonal tracking studies. Historically, the structure of this process has been greatly informed by flow cytometric isolation of cell populations with decreasing lineage and proliferative activity (Bryder, Rossi et al. 2006),(Eaves 2015),(Wahlestedt and Bryder 2017). The prospective isolation of these cell populations has also enhanced reproducibility in various clinical applications. However, a number of exceptions have emerged that argue against the concept of describing this process in terms of discrete and intrinsically homogeneous cell populations. Heterogeneity in cell function (Sieburg, Cho et al. 2006),(Dykstra, Kent et al. 2007) and molecular profiling (Paul, Arkin et al. 2015),(Wilson, Kent et al. 2015), as well as evidence of many intermediate molecular states (Nestorowa, Hamey et al. 2016),(Cabezas-Wallscheid, Buettner et al. 2017), now suggest that much of hematopoietic cell differentiation has features of a continuous process in which the molecular state (and phenotypic profile) of a cell changes more dynamically than originally envisaged. This makes snapshot measurements of enriched but non-homogeneous populations difficult to interpret, and the field clearly lacks mathematical tools to deal with a continuum or variable paradigms. This is a key area in which modelling and inference should advance future understanding of hematopoiesis.

Discussion of these issues led to a convergence on two areas. The first focused on strategies that would improve clarity about cell population descriptions; for example, by provision of flow cytometric gating details, and results of measurements of the proportion of cells in a given

population that displayed assigned functional properties. The second focused on the tools needed to collect and model dynamically changing properties of cells.

I'll believe it when I see it (and can do it)...

A recurring theme was the need to improve reproducibility, robustness and utility of datasets in the clonal tracking field. Difficulties in accessing and re-analyzing published datasets were noted and a frequent lack of published information about viral barcode libraries and the composition of barcoded samples was also felt to be a significant deficiency, recognizing a potential exception for barcoding data obtained on clinical subjects. There was thus much support for the concept of a central online repository being established that could accept standard format barcoding data as this would enable such information for most experimental studies to become a standard publication requirement, as is already the case for other types of large datasets.

Specific recommendations proposed to address these issues were as follows:

1. Provide all data necessary for the re-analysis of the major claims in the publication, in both raw and processed format.
2. Provide an informative description of the data and how they were produced.
3. Provide a clear description of the bioinformatics pipeline used, as well as of any modelling or statistical inference procedures.
4. Ensure the accessibility of all necessary software, including the specific code to reproduce tables, figures and other analyses.

The range of experimental and analytical methods used in clonal tracking makes it impossible to produce comprehensive standards for reporting barcode data. Nevertheless, minimum requirements for the implementation and reporting of barcoding experiments could be introduced; e.g., barcode detection levels, data filters used, numbers of barcoded and transplanted cells, and details of whether absolute or fractional numbers were used for reporting. Furthermore, for IS and *in vivo* genomic recombination barcoding, where the overlap between observed barcodes in different samples is often very low, attempts to circumvent this problem could be clearly described.

There was also agreement that analysis pipelines would benefit from clearer descriptions and improved accessibility of codes used. A major barrier to the reproducibility of barcoding data analysis is that, unlike other high-dimensional data types such as single-cell RNA-seq, the clonal tracking field does not have many applicable open source R packages. Some workshop participants expressed interest in creating such R packages, and/or making existing ones available. Such a step forward would not constitute a definitive analysis pipeline, but would make many of the unavoidable decisions involved in the preliminary processing of barcoding data more transparent, e.g., choices on read thresholds and barcode sequence error correction. This approach could also provide methods for data visualization and enable new methods to be more rapidly disseminated. Box 1 summarizes key recommendations for data and code provision.

An example of this kind of toolbox, comprising several user-friendly and customizable R packages, has been recently developed (Thielecke et al, under revision) and can be downloaded at https://cran.r-project.org/web/packages/genBaRcode/index.html. The workflow starts with barcode extraction from the raw sequencing files, independently of the known barcode library. After matching to the library allowing insertions and deletions in the reads, an error-correction procedure merges highly similar barcodes, assuming that small sequence differences are most likely the result of PCR errors (Thielecke, Aranyossy et al. 2017). The resulting barcode counts can be visualized in multiple ways, including histograms, graphical networks and tree-like structures. The complete R code is available allowing customization of the package.

As barcoding data starts to be combined with other functional measurements; e.g., scRNA-seq (Spanjaard, Hu et al. 2018),(Wagner, Weinreb et al. 2018),(Alemany, Florescu et al. 2018),(Raj, Wagner et al. 2018), or cell cycle measurements (Griessinger, Vargaftig et al. 2018), visualization of data becomes increasingly challenging. The chosen method will depend on data type and question(s) being posed, but figures that are informative and easily interpretable will reduce current challenges. A common toolbox (e.g., a set of dedicated R packages) for people to work from would thus be a major boon for the field (see attached example).

Box 1: Practical guidelines for making data and code available

Code availability

A. Open access code repository e.g. GitHub
B. Provision of code to generate each figure appearing in a paper (together with the source data where feasible)
C. Pre-processed data files deposited together with the code

Data availability

A. Deposit data in an online repository
B. Raw data should be made available as FASTQ file or counts matrix
C. Reference file of the sequencing of the plasmid library should be made available (where possible, construction details of the library provided in a publication)
D. Metadata should be unambiguous and allow for identification of each sample (“column” in the counts matrix) without additional information, as well as, sequence file id, animal, tissue, time, sort details and other relevant information

Human blood cell production barcoding experiments: making the most of gene therapy trials

The reliance of human hematopoietic cell biology on retrospective functional transplantation assays has traditionally posed major limitations on understanding the workings of this system. Animal models, where experimental bone marrow transplantation is now well established, have therefore become the mainstay of the field. For assessing the functional activity of human cells, researchers have generally relied on *in vitro* assays and xenotransplantation into immunodeficient

animals, or have drawn inferences from primate transplant models, each of which has drawbacks. Fortuitously, gene therapy trials are increasing in number and are now providing unique access to the clonal dynamics of human hematopoiesis in various clinical contexts. Although the main concerns in gene therapy trials are patient safety and treatment efficacy, associated analyses provide an unprecedented source of data about short and long-term clonal contributions and their stable and changing diversity of progeny outputs in humans (Aiuti, Biasco et al. 2013),(Biasco, Pellin et al. 2016).

Historically, these data relied primarily on the identification of viral integration sites (VIS). Recently developed methods (Zhou, Bonner et al. 2014) have shown all VIS in a sample can now be recovered, although previous percentages recovered were much lower (60-80%) (Schmidt, Schwarzwaelder et al. 2007),(Kustikova, Baum et al. 2008),(Gabriel, Eckenberg et al. 2009). Nevertheless, quantification of clone numbers based on sequencing reads is inherently imprecise (Cornils, Bartholomae et al. 2013),(Brugman, Suerth et al. 2013). Methods used to quantify clonal data in humans likewise need to be described clearly to allow comparison across trials.

An enormous amount of ambiguity surrounds many gene therapy datasets. This is due both to differences in the underlying diseases and technical differences in the methods used to quantify clonal abundance. Each trial is necessarily governed by different authorities with center- and disease-specific regulations that include limitations on the number of samples available for analysis. Provision of primary sequencing data might thus be limited by the use of proprietary vectors in addition to a patient's right to protect their genomic information. However, independent explicit descriptions of sample preparation, transduction and transplantation parameters, and details of data processing and quantification, sequencing platform used, IS calling, thresholds and noise filtering details are essential to enable independent interpretation of reported data. Creation and adoption of common standards to assess data and conclusions, developed by a group of specialists in this area, would thus be very helpful to the field.

Seeing the forest using the trees

With the advent of new experimental methods to track clones *in vivo* (e.g. (Sun, Ramos et al. 2014), (McKenna, Findlay et al. 2016), (Alemany, Florescu et al. 2018),(Raj, Wagner et al. 2018), (Spanjaard, Hu et al. 2018), (Wagner, Weinreb et al. 2018),(Maetzig, Morgan et al. 2018) and the rapid expansion in the size of datasets, many statistical and mathematical modelling approaches have been developed to conduct inferences from the data. These approaches are becoming extremely important to deriving information about the dynamics of hematopoiesis *in vivo*.

Improved reporting of model assumptions could permit more rigorous evaluation of current data and the validity of model-based conclusions from both a biological and mathematical perspective. As in the case of deciding which experimental assay to implement, the choice of a mathematical model is generally determined by a trade-off between its applicability to the biological situation and practical considerations, from the ease of equation manipulation (i.e., the availability of closed-form solutions or approximations) to the feasibility of parameter inference methods. From a theoretical point of view, these "practical considerations" usually result in assumptions that do

not fully match the experimental setting. It is thus crucial that these assumptions are made as explicit as they would be in a mathematical journal, so that the validity of the trade-off can be assessed. For example, most formalisms used to model barcoding data (e.g., branching processes, Markov chains, etc) assume the independence of division and cell fate decisions because, in practice, this is necessary in order to analyze the models. This is akin to the use of antibodies to measure cell identity, another practical assumption that has fallacies.

On the inference side, custom-made and complicated pipelines for parameter inference or model selection can also be a source of problems as they are not typically subject to the same level of scrutiny as other aspects of data analysis. Clearer descriptions of inference methodologies used would help to address this challenge.

In terms of specific enhancements to clonal tracking experiments, there was consensus that the addition of functional information to barcode counts would help to specify essential parameters of the model. For example, when inferring self-renewal, differentiation and death rates, measurements of cell cycle, apoptosis, and cellular state transition time can be hugely informative. This additional data is also important to embed fate decisions in the context of regulatory pathways. It was felt that this type of data will rapidly emerge, particularly as barcoding is combined with single-cell RNA-sequencing, and the time is ripe for developing the modelling/analysis strategies to make use of such data.

Looking toward the future of building mechanistic models describing hematopoiesis, approaches for reconstructing trees have been borrowed or adapted from the field of phylogenetics, particularly to analyze data from *in vivo* genomic recombination tracking methods and retrospective methods based on division-linked mutations (Shlush, Chapal-Ilani et al. 2012),(McKenna, Findlay et al. 2016),(Frieda, Linton et al. 2017),(Lee-Six, Obro et al. 2018),(Alemany, Florescu et al. 2018),(Raj, Wagner et al. 2018),(Spanjaard, Hu et al. 2018),(Griessinger, Vargaftig et al. 2018). While these approaches have not yet been robustly validated by, for example, combining them with virally-introduced barcoding approaches, experiments of this type would allow benchmarking of the various tree reconstruction methods.

Future Perspective

This workshop highlighted the potential that clonal tracking has to gain from close interactions between experimentalists and theoreticians. As datasets grow in size and complexity, it will be critical to develop increasingly robust assessment tools to establish ways that facilitate data interpretation and permit a broader understanding of hematopoiesis beyond single experiments. A more complete understanding of the journeys that primitive cells and their progeny may make to produce all types of mature cells throughout life will greatly inform efforts to generate these *ex vivo*. Being able to predict the clonal dynamics of a mutant clone relative to normal cells or being able to robustly assess the success, failure, or potential dangers of a graft in cellular and gene therapies may also be possible to realize in the future. Dynamic models of such processes will become more robust as more accurate and comprehensive information is contributed by

experimentalists, and the iterative process of building such models will also be facilitated by better (and more) sharing of data and associated analytical methods.

Author Contributions:
IG, AML, DGK, LP, and EL developed the workshop program and article structure, and co-wrote the first draft of the article. KC contributed substantially to the revised version.

Acknowledgements:
We thank the BBSRC (BB/R021465/1), the German Stem Cell Network, the Wellcome-MRC Cambridge Stem Cell Institute and the *Labex CelTisPhyBio* (No. ANR-10-LBX-0038) for funds to support this workshop and the conference staff at Homerton College. Dorit Ludwig (Dresden) and Martin Dawes (Cambridge) are particularly acknowledged for their help in organizing the workshop.

List of workshop participants and affiliation(s):

Adair, Jennifer E.	Fred Hutchinson Cancer Research Center, Seattle	USA
Baldow, Christoph	TU Dresden	Germany
Brugman, Martijn	GlaxoSmithKline	UK
Bystrykh, Leonid	ERIBA, Groningen	The Netherlands
Calabria, Andrea	SR-TIGET, Milan	Italy
Campbell, Peter	Wellcome Trust Sanger Institute, Hinxton	United Kingdom
Cesana, Daniela	San Raffaele Telethon Institute for Gene Therapy, Milan	Italy
Cornils, Kerstin	UK Hamburg-Eppendorf	Germany
Cosgrove, Jason	Institut Curie	France
Duffy, Ken	Maynooth University	Ireland
Dunbar, Cynthia	NIH, Bethesda	USA
Eaves, Connie	University of British Columbia, Vancouver	Canada
Glauche, Ingmar	TU Dresden	Germany
Gottgens, Bertie	University of Cambridge	United Kingdom
Greco, Alessandro	DKFZ Heidelberg	Germany
Guilloux, Agathe	Université d'Evry Val d'Essonne Paris-Saclay	France
Hammond, Colin	BC Cancer Research Centre, Vancouver	Canada
Höfer, Thomas	DKFZ Heidelberg	Germany
Kent, David	University of Cambridge	United Kingdom
Kucinski, Iwo	University of Cambridge, Stem Cell Institute	United Kingdom
Laurenti, Elisa	University of Cambridge	United Kingdom
Lin, Dawn	The Walter and Eliza Hall Institute of Medical Research	Australia
Lu, Rong	Keck School of Medicine, University of Southern California	USA
Lyne, Anne-Marie	Institut Curie, Paris	France
Minin, Vladimir	University of California, Irvine	USA
Pellin, Danilo	Harvard University	USA
Perié, Leïla	Institut Curie, Paris	France
Rodriguez-Fraticelli, Alejo	Harvard University	USA
Sawai, Catherine	University of Bordeaux	France
Sehl, Mary	David Geffen School of Medicine at UCLA	USA
Simons, Ben	University of Cambridge	United Kingdom
Six, Emmanuelle	Imagine Institute for Genetic Diseases, Paris	France
Spanjaard, Bastiaan	Max Delbrück Center for Molecular Medicine, Berlin	Germany

Wojtowicz, Edyta	Karolinska Institutet, Stockholm	Sweden
Xu, Jason	Duke University	USA

References

Abkowitz, J. L., R. L. Ott, J. M. Nakamura, L. Steinmann, P. J. Fialkow and J. W. Adamson (1985). "Feline glucose-6-phosphate dehydrogenase cellular mosaicism. Application to the study of retrovirus-induced pure red cell aplasia." J Clin Invest **75**(1): 133-140.

Aiuti, A., L. Biasco, S. Scaramuzza, F. Ferrua, M. P. Cicalese, C. Baricordi, F. Dionisio, A. Calabria, S. Giannelli, M. C. Castiello, M. Bosticardo, C. Evangelio, A. Assanelli, M. Casiraghi, S. Di Nunzio, L. Callegaro, C. Benati, P. Rizzardi, D. Pellin, C. Di Serio, M. Schmidt, C. Von Kalle, J. Gardner, N. Mehta, V. Neduva, D. J. Dow, A. Galy, R. Miniero, A. Finocchi, A. Metin, P. P. Banerjee, J. S. Orange, S. Galimberti, M. G. Valsecchi, A. Biffi, E. Montini, A. Villa, F. Ciceri, M. G. Roncarolo and L. Naldini (2013). "Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome." Science **341**(6148): 1233151.

Alemany, A., M. Florescu, C. S. Baron, J. Peterson-Maduro and A. van Oudenaarden (2018). "Whole-organism clone tracing using single-cell sequencing." Nature **556**(7699): 108.

Biasco, L., D. Pellin, S. Scala, F. Dionisio, L. Basso-Ricci, L. Leonardelli, S. Scaramuzza, C. Baricordi, F. Ferrua, M. P. Cicalese, S. Giannelli, V. Neduva, D. J. Dow, M. Schmidt, C. Von Kalle, M. G. Roncarolo, F. Ciceri, P. Vicard, E. Wit, C. Di Serio, L. Naldini and A. Aiuti (2016). "In Vivo Tracking of Human Hematopoiesis Reveals Patterns of Clonal Dynamics during Early and Steady-State Reconstitution Phases." Cell Stem Cell **19**(1): 107-119.

Brugman, M. H., J. D. Suerth, M. Rothe, S. Suerbaum, A. Schambach, U. Modlich, O. Kustikova and C. Baum (2013). "Evaluating a ligation-mediated PCR and pyrosequencing method for the detection of clonal contribution in polyclonal retrovirally transduced samples." Human gene therapy methods **24**(2): 68-79.

Bryder, D., D. J. Rossi and I. L. Weissman (2006). "Hematopoietic stem cells: the paradigmatic tissue-specific stem cell." Am J Pathol **169**(2): 338-346.

Cabezas-Wallscheid, N., F. Buettner, P. Sommerkamp, D. Klimmeck, L. Ladel, F. B. Thalheimer, D. Pastor-Flores, L. P. Roma, S. Renders, P. Zeisberger, A. Przybylla, K. Schonberger, R. Scognamiglio, S. Altamura, C. M. Florian, M. Fawaz, D. Vonficht, M. Tesio, P. Collier, D. Pavlinic, H. Geiger, T. Schroeder, V. Benes, T. P. Dick, M. A. Rieger, O. Stegle and A. Trumpp (2017). "Vitamin A-Retinoic Acid Signaling Regulates Hematopoietic Stem Cell Dormancy." Cell **169**(5): 807-823 e819.

Cornils, K., C. C. Bartholomae, L. Thielecke, C. Lange, A. Arens, I. Glauche, U. Mock, K. Riecken, S. Gerdes and C. von Kalle (2013). "Comparative clonal analysis of reconstitution kinetics after transplantation of hematopoietic stem cells gene marked with a lentiviral SIN or a γ-retroviral LTR vector." Experimental hematology **41**(1): 28-38. e23.

Cornils, K., L. Thielecke, S. Huser, M. Forgber, M. Thomaschewski, N. Kleist, K. Hussein, K. Riecken, T. Volz, S. Gerdes, I. Glauche, A. Dahl, M. Dandri, I. Roeder and B. Fehse (2014). "Multiplexing clonality: combining RGB marking and genetic barcoding." Nucleic Acids Res **42**(7): e56.

Dykstra, B., D. Kent, M. Bowie, L. McCaffrey, M. Hamilton, K. Lyons, S. J. Lee, R. Brinkman and C. Eaves (2007). "Long-term propagation of distinct hematopoietic differentiation programs in vivo." Cell Stem Cell **1**(2): 218-229.

Eaves, C. J. (2015). "Hematopoietic stem cells: concepts, definitions, and the new reality." Blood **125**(17): 2605-2613.

Fialkow, P. J. (1980). Clonal and stem cell origin of blood cell neoplasms. . Contemporary Hematology/Oncology. J. Lobue, A. S. Gordon, R. Silber and F. M. Muggia, Plenum Publishing; New York, NY**:** 1-46.
Frieda, K. L., J. M. Linton, S. Hormoz, J. Choi, K. K. Chow, Z. S. Singer, M. W. Budde, M. B. Elowitz and L. Cai (2017). "Synthetic recording and in situ readout of lineage information in single cells." Nature **541**(7635): 107-111.
Gabriel, R., R. Eckenberg, A. Paruzynski, C. C. Bartholomae, A. Nowrouzi, A. Arens, S. J. Howe, A. Recchia, C. Cattoglio and W. Wang (2009). "Comprehensive genomic access to vector integration in clinical gene therapy." Nature medicine **15**(12): 1431.
Gerrits, A., B. Dykstra, O. J. Kalmykowa, K. Klauke, E. Verovskaya, M. J. Broekhuis, G. de Haan and L. V. Bystrykh (2010). "Cellular barcoding tool for clonal analysis in the hematopoietic system." Blood **115**(13): 2610-2618.
Griessinger, E., J. Vargaftig, S. Horswell, D. C. Taussig, J. Gribben and D. Bonnet (2018). "Acute myeloid leukemia xenograft success prediction: Saving time." Experimental hematology **59**: 66-71. e64.
Jordan, C. T. and I. R. Lemischka (1990). "Clonal and systemic analysis of long-term hematopoiesis in the mouse." Genes Dev **4**(2): 220-232.
Kustikova, O. S., C. Baum and B. Fehse (2008). Retroviral integration site analysis in hematopoietic stem cells. Hematopoietic Stem Cell Protocols, Springer**:** 255-267.
Lee-Six, H., N. F. Obro, M. S. Shepherd, S. Grossmann, K. Dawson, M. Belmonte, R. J. Osborne, B. J. P. Huntly, I. Martincorena, E. Anderson, L. O'Neill, M. R. Stratton, E. Laurenti, A. R. Green, D. G. Kent and P. J. Campbell (2018). "Population dynamics of normal human blood inferred from somatic mutations." Nature **561**(7724): 473-478.
Lu, R., N. F. Neff, S. R. Quake and I. L. Weissman (2011). "Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding." Nat Biotechnol **29**(10): 928-933.
Maetzig, T., M. Morgan and A. Schambach (2018). "Fluorescent genetic barcoding for cellular multiplex analyses." Experimental hematology.
McKenna, A., G. M. Findlay, J. A. Gagnon, M. S. Horwitz, A. F. Schier and J. Shendure (2016). "Whole-organism lineage tracing by combinatorial and cumulative genome editing." Science **353**(6298): aaf7907.
Naik, S. H., L. Perie, E. Swart, C. Gerlach, N. van Rooij, R. J. de Boer and T. N. Schumacher (2013). "Diverse and heritable lineage imprinting of early haematopoietic progenitors." Nature **496**(7444): 229-232.
Nestorowa, S., F. K. Hamey, B. Pijuan Sala, E. Diamanti, M. Shepherd, E. Laurenti, N. K. Wilson, D. G. Kent and B. Gottgens (2016). "A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation." Blood **128**(8): e20-31.
Paul, F., Y. Arkin, A. Giladi, D. A. Jaitin, E. Kenigsberg, H. Keren-Shaul, D. Winter, D. Lara-Astiaso, M. Gury, A. Weiner, E. David, N. Cohen, F. K. Lauridsen, S. Haas, A. Schlitzer, A. Mildner, F. Ginhoux, S. Jung, A. Trumpp, B. T. Porse, A. Tanay and I. Amit (2015). "Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors." Cell **163**(7): 1663-1677.
Pei, W., T. B. Feyerabend, J. Rossler, X. Wang, D. Postrach, K. Busch, I. Rode, K. Klapproth, N. Dietlein, C. Quedenau, W. Chen, S. Sauer, S. Wolf, T. Hofer and H. R. Rodewald (2017). "Polylox barcoding reveals haematopoietic stem cell fates realized in vivo." Nature **548**(7668): 456-460.
Raj, B., D. E. Wagner, A. McKenna, S. Pandey, A. M. Klein, J. Shendure, J. A. Gagnon and A. F. Schier (2018). "Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain." Nature biotechnology.
Rodriguez-Fraticelli, A. E., S. L. Wolock, C. S. Weinreb, R. Panero, S. H. Patel, M. Jankovic, J. Sun, R. A. Calogero, A. M. Klein and F. D. Camargo (2018). "Clonal analysis of lineage fate in native haematopoiesis." Nature **553**(7687): 212.

Schmidt, M., H. Glimm, N. Lemke, A. Muessig, C. Speckmann, S. Haas, P. Zickler, G. Hoffmann and C. Von Kalle (2001). "A model for the detection of clonality in marked hematopoietic stem cells." Ann N Y Acad Sci **938**: 146-155; discussion 155-146.
Schmidt, M., K. Schwarzwaelder, C. Bartholomae, K. Zaoui, C. Ball, I. Pilz, S. Braun, H. Glimm and C. Von Kalle (2007). "High-resolution insertion-site analysis by linear amplification–mediated PCR (LAM-PCR)." Nature Methods **4**(12): 1051.
Shlush, L. I., N. Chapal-Ilani, R. Adar, N. Pery, Y. Maruvka, A. Spiro, R. Shouval, J. M. Rowe, M. Tzukerman, D. Bercovich, S. Izraeli, G. Marcucci, C. D. Bloomfield, T. Zuckerman, K. Skorecki and E. Shapiro (2012). "Cell lineage analysis of acute leukemia relapse uncovers the role of replication-rate heterogeneity and microsatellite instability." Blood **120**(3): 603-612.
Sieburg, H. B., R. H. Cho, B. Dykstra, N. Uchida, C. J. Eaves and C. E. Muller-Sieburg (2006). "The hematopoietic stem compartment consists of a limited number of discrete stem cell subsets." Blood **107**(6): 2311-2316.
Spanjaard, B., B. Hu, N. Mitic, P. Olivares-Chauvet, S. Janjuha, N. Ninov and J. P. Junker (2018). "Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars." Nat Biotechnol **36**(5): 469-473.
Sun, J., A. Ramos, B. Chapman, J. B. Johnnidis, L. Le, Y. J. Ho, A. Klein, O. Hofmann and F. D. Camargo (2014). "Clonal dynamics of native haematopoiesis." Nature **514**(7522): 322-327.
Thielecke, L., T. Aranyossy, A. Dahl, R. Tiwari, I. Roeder, H. Geiger, B. Fehse, I. Glauche and K. Cornils (2017). "Limitations and challenges of genetic barcode quantification." Sci Rep **7**: 43249.
Wagner, D. E., C. Weinreb, Z. M. Collins, J. A. Briggs, S. G. Megason and A. M. Klein (2018). "Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo." Science **360**(6392): 981-987.
Wahlestedt, M. and D. Bryder (2017). "The slippery slope of hematopoietic stem cell aging." Experimental hematology **56**: 1-6.
Wilson, N. K., D. G. Kent, F. Buettner, M. Shehata, I. C. Macaulay, F. J. Calero-Nieto, M. Sanchez Castillo, C. A. Oedekoven, E. Diamanti, R. Schulte, C. P. Ponting, T. Voet, C. Caldas, J. Stingl, A. R. Green, F. J. Theis and B. Gottgens (2015). "Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations." Cell Stem Cell **16**(6): 712-724.
Zhou, S., M. A. Bonner, Y.-D. Wang, S. Rapp, S. S. De Ravin, H. L. Malech and B. P. Sorrentino (2014). "Quantitative shearing linear amplification polymerase chain reaction: an improved method for quantifying lentiviral vector insertion sites in transplanted hematopoietic cell systems." Human gene therapy methods **26**(1): 4-12.