



# The Impact of Heterogeneity on Single-Cell Sequencing

Samantha L. Goldman<sup>1,2</sup>, Matthew MacKay<sup>1,2</sup>, Ebrahim Afshinnekoo<sup>1,2,3</sup>, Ari M. Melnick<sup>4</sup>, Shuxiu Wu<sup>5,6</sup> and Christopher E. Mason<sup>1,2,3,7\*</sup>

<sup>1</sup> Department of Physiology and Biophysics, Weill Cornell Medical College, New York, NY, United States, <sup>2</sup> The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY, United States, <sup>3</sup> WorldQuant Initiative for Quantitative Prediction, Weill Cornell Medicine, New York, NY, United States, <sup>4</sup> Department of Medicine, Weill Cornell Medicine, New York, NY, United States, <sup>5</sup> Hangzhou Cancer Institute, Hangzhou Cancer Hospital, Hangzhou, China, <sup>6</sup> Department of Radiation Oncology, Hangzhou Cancer Hospital, Hangzhou, China, <sup>7</sup> The Feil Family Brain and Mind Research Institute, New York, NY, United States

## OPEN ACCESS

### Edited by:

Xinghua Victor Pan,  
Southern Medical University, China

### Reviewed by:

Saheli Sarkar,  
Northeastern University, United States  
Guangshuai Jia,  
Max-Planck-Institut für Herz- und  
Lungenforschung, Germany

### \*Correspondence:

Christopher E. Mason  
chm2042@med.cornell.edu

### Specialty section:

This article was submitted to  
Genomic Assay Technology,  
a section of the journal  
Frontiers in Genetics

Received: 20 July 2018

Accepted: 09 January 2019

Published: 01 March 2019

### Citation:

Goldman SL, MacKay M,  
Afshinnekoo E, Melnick AM, Wu S  
and Mason CE (2019) The Impact  
of Heterogeneity on Single-Cell  
Sequencing. *Front. Genet.* 10:8.  
doi: 10.3389/fgene.2019.00008

The importance of diversity and cellular specialization is clear for many reasons, from population-level diversification, to improved resiliency to unforeseen stresses, to unique functions within metazoan organisms during development and differentiation. However, the level of cellular heterogeneity is just now becoming clear through the integration of genome-wide analyses and more cost effective Next Generation Sequencing (NGS). With easy access to single-cell NGS (scNGS), new opportunities exist to examine different levels of gene expression and somatic mutational heterogeneity, but these assays can generate yottabyte scale data. Here, we model the importance of heterogeneity for large-scale analysis of scNGS data, with a focus on the utilization in oncology and other diseases, providing a guide to aid in sample size and experimental design.

**Keywords:** single-cell sequencing, heterogeneity, scRNA-seq, NGS, RNA, single cells

## INTRODUCTION

It has been well-documented, both theoretically (Elsasser, 1984) and experimentally, that nearly all cellular systems are heterogeneous (Altschuler and Wu, 2010). Heterogeneity may arise for a number of different reasons, and at many different levels, in order to improve survival and functionality. Both single-celled and multicellular organisms employ population-level survival strategies such as bet-hedging in order to achieve a better chance of survival when faced with new stresses though having a diverse population (Grimbergen et al., 2015). At a single-organism level, diversity further enables the existence of specialization and, within metazoan organisms, differentiation (Hadjantonakis and Arias, 2016).

Cellular heterogeneity can be measured in several different ways, most commonly via genomic, epigenomic, transcriptomic, and proteomic studies. However, the level of heterogeneity at one level of expression or regulation may not be the same at another level. Cells within a given person have nearly identical genomes, yet through specific modifications throughout development and disease, may generate many distinct cell types with unique expression profiles. Even the genome itself may be specifically rewired to generate increased genetic diversity within specific cell types, most notably B- and T-cells through V(D)J recombination. Uncovering the true diversity of cells is crucial to better understand cellular communication and responsibility within both healthy and disease states. It is now well understood that differentiation throughout development allows for

the necessary cellular specialization required for complex multicellular system function. Further, specific epigenomic modifications allow for this precise differentiation which inevitably results in the cascade of cellular diversity present in humans, and also is important in cancer (Li et al., 2014, 2016).

Next generation sequencing (NGS) is continuously being used more and more due to its rapidly decreased costs and ability to generate a large amount of data (Mason et al., 2014), with new data sets even being generated in zero gravity (McIntyre et al., 2016; Castro-Wallace et al., 2017). Within bulk-NGS analyses, many, typically hundreds of thousands to millions, of cells are analyzed at once. This generates an averaged picture of a given population of cells, and thus majority of our understanding of different cell and tissue types comes from the analysis of bulk experimentation which may underestimate the true heterogeneity of cells. Bulk-NGS is simply ill-equipped to address some important questions revolving around cellular heterogeneity. Single-cell NGS (scNGS) attempts to resolve issues facing bulk-NGS through the ability to relate sequences to a given cell, across the genetic, transcriptomic, epigenomic, and proteomic levels. This approach reduces the issue of data generalization which is prevalent in some bulk-NGS studies. However, scNGS is not without its faults. One of the main issues with scNGS is its cost and, though it has considerably decreased in recent years, it is still a large factor when designing experimentations, as well as technical issues and challenges in sensitivity. Here, we will outline the importance of cellular heterogeneity, assess factors of scNGS heterogeneity, and provide a practical sample size guide to aid in experimental design.

## THE IMPORTANCE OF CELLULAR HETEROGENEITY

Having a heterogeneous (i.e., diverse) population is beneficial for cellular systems for the same reason why it is beneficial for there to be variation among many organisms in a single species – bet-hedging (Beaumont et al., 2009). Bet-hedging is a population-level survival strategy in which less-fit individuals are maintained in a population as a precaution; if the environment were to drastically change, the originally less-fit organisms may be adapted to the new environment, thereby assuring the survival of the population (Grimbergen et al., 2015). In an ever-changing environment, a population has a greater overall fitness if there is greater diversity. In this way, the evolution adaptation of all cellular systems can be modeled in terms of Darwinian evolution.

There are many causes of cellular heterogeneity. Firstly, populations of cells will naturally contain individuals that develop random mutations. These unique subclones can become significant portions of the population if that mutation confers a selective advantage and proliferates. However, not all cellular heterogeneity is genetic. Rather, much heterogeneity is phenotypic, and is frequently expressed in transcriptomes that vary from cell to cell. This heterogeneity can arise via external or internal factors. Extrinsic heterogeneity can lead to phenotypic plasticity in response to an environmental change, and only affects the part of the population that is exposed to

the causative environment (Huang, 2009). It can also include variables such as cell-cycle stage and cell size (Singh and Soltani, 2013). Intrinsic heterogeneity is a more nuanced phenomenon, and is a result of stochastic events, such as gene expression noise (Huang, 2009), rather than a changing intracellular environment (Elowitz et al., 2002).

Because of stochastic gene fluctuation, there are varying levels of protein abundance in different cells in a population at any given time. This is most easily visualized via flow cytometry, which yields a bell-shaped curve (Brock et al., 2009). Stochastic gene expression may have its evolutionary advantages, as well. In the same way that populations of cells maintain random mutations in bet-hedging, populations of clonal, unicellular organisms may maintain variation via stochastic gene expression to ensure overall survival (Raj and van Oudenaarden, 2008). Although stochastic gene expression is a significant contributor to heterogeneity, it is not the only cause. The sub-state of any given genome/cell depends on a number of factors, including epigenetics, alternative splicing sites, post-translational modifications, and sometimes even microbial interactions (Shabaan et al., 2018). These processes are not always stochastic, and can therefore lead to “directed” heterogeneity, instead of the more random “non-directed” heterogeneity of stochastic gene expression (Chang and Marshall, 2017).

Interestingly, non-genetic, cellular heterogeneity also plays an important role in development. Early in the developmental process, before the small population of cells is beginning to differentiate, these cells are theoretically identical. However, as the cells begin to differentiate, they display non-genetic heterogeneity. The body of research on the role of heterogeneity in development is largely focused on transcriptional heterogeneity (Griffiths et al., 2018), which is a driver of differentiation of pluripotent stem cells. More recent work has also shown that RNA modifications, called the epitranscriptome (Saletore et al., 2012), can also lead to differential response of human cells to both disease and infection (Gokhale et al., 2016; Vu et al., 2017). Also, some transcriptional sub-states are heritable through several generations of cell divisions. Signaling factors, developmental regulators, and chromatin regulators contribute to transcriptional heterogeneity in stem cells (Kumar et al., 2014). “Directed” heterogeneity has been shown to lead the process behind the development of a body plan in *Drosophila melanogaster* (Chang and Marshall, 2017).

Even after development, all human tissue systems experience some level of differentiation. This allows cells to specialize, leading to a more flexible biological system. This principle has been most notably studied in the nervous and immune systems. In the central nervous system, for instance, there are dozens of different types of neurons. Subsets of these neurons form the myriad different regions within the brain (Emery and Barres, 2008). One phenotypic hallmark of heterogeneity in the nervous system, for example, is the distribution of mitochondria within the neuron. This heterogeneity is exhibited both regionally within the brain (e.g., brain regions that require more energy are composed of neurons with more mitochondria) (Dubinsky, 2009) and within individual neurons. This distribution differs greatly depending on the immediate

and current needs of the neuron, and is regulated by a complex system of proteins (Course and Wang, 2016). In the immune system, monocytes, macrophages (Gordon and Taylor, 2005), B-cells, and T-cells show heterogeneity. As an example, T-cell heterogeneity is essential for an effective immune response, since subtle differences in T-cell receptors (TCRs) enable the identification and elimination of foreign invaders (Durlanik and Thiel, 2015). However, in autoimmune disease, faulty TCR diversification can result in the improper identification of “self” as an invader resulting in normal tissue destruction.

Different diseases leverage heterogeneity to their advantage. A “survival of the fittest” model for cellular heterogeneity can be applied not only to populations of single-celled organisms, but also to tumors. Cancer cells continuously acquire and pass down genetic and epigenetic modifications to subsequent generations of cancer cells resulting in heterogeneity. These genetic mutations and epigenetic shifts may further lead to changes in fitness (Li et al., 2016). Cancer cells are often exposed to hostile environments, such as chemotherapy and radiation, during treatment (Afshinnkoo and Mason, 2016). Through bet-hedging, and therefore maintenance of a heterogeneous population, the chance of resistance or relapse from treatment is dramatically increased. As these cancer cells are all in the same small environment and are all competing for the same limited resources, there are complex interactions between different subclones that further reinforce these Darwinian relationships (Tabassum and Polyak, 2015). Cancer cells can be further driven into a “survival of the fittest” scenario via treatment with a chemotherapeutic drug, as this may lead to the selection for cancer-variants that are resistant to the drug. Over time, this could lead to chemotherapeutic resistance within the whole tumor (Dagogo-Jack and Shaw, 2017), as well as tumor subtypes (Shih et al., 2017). Indeed, it has been shown that a single tumor biopsy dramatically underrepresents the genetic diversity present within an entire tumor (Gerlinger et al., 2012). However, heterogeneity is not only clinically relevant in regards to chemotherapy. Immunotherapies can also be profoundly impacted by heterogeneity. Liver cancer-targeted immunotherapy is designed around tumor-infiltrating T-cells. Through the use of single-cell RNA sequencing, 11 tumor-infiltrating T-cell sub-states have been identified. Each of these sub-states has a unique profile of up- and downregulated genes, which may impact the efficacy of any immunotherapies (Zheng et al., 2017).

Intratumoral heterogeneity has been extensively studied through single-cell sequencing methods. For example, single-cell RNA sequencing has revealed significant heterogeneity in primary glioblastomas (Patel et al., 2014). Additionally, increased levels of heterogeneity in these tumors was inversely correlated with survival, indicating that intratumor heterogeneity should be an essential clinical factor, including events from DNA transposition (Henssen et al., 2017). Metastatic melanoma is also highly transcriptionally heterogeneous, and this heterogeneity is multifaceted; it is associated with a number of factors, including cell cycle stage, location, and chemotherapeutic resistance (Tirosch et al., 2016). The use of RNA sequencing here is key, as transcriptomics captures fine details of non-genetic heterogeneity

that other sequencing methods may have missed. Shifting of cellular heterogeneity is not just a hallmark of cancer, but of many other diseases, but here we will focus on the relevance for cancer.

## ASSESSING HETEROGENEITY

Heterogeneity itself is a gradient which may be based on variable changes in the transcriptome or more permanent changes within the genome. Differences seen between cells may be temporal due to cell-cycle states, or spatial due to external stimuli (Dagogo-Jack and Shaw, 2017). Also, differences between cells may exist at any processing level of the cell, from the genome to transcriptome to proteome, or due to any additional modifications which may exist. With this in mind, it could be possible to define all cells as heterogeneous. However, two disparate cells might not behave functionally different, and their heterogeneity would therefore not be considered impactful (Altschuler and Wu, 2010). The overall assessment of cellular heterogeneity is therefore context-specific and the technologies used to assess cellular differences need to be considered carefully.

Proteomic and cell-marker classification has been historically used to discern cell types. Immunohistochemistry (IHC) can be used to distinguish immune cell types within healthy systems (Reuben et al., 2017b) or even the cancer subtyping such as HER2 expression within breast cancer (Potts et al., 2012). Surface markers help to distinguish cell types into broad classification, but this type of analysis required prior gene expression knowledge and specific antibody usage. Other approaches, such as whole genome sequencing (WGS), bisulfite sequencing, and RNA sequencing, allow for genome-wide analysis (Mason et al., 2017). Historically these techniques are done on heterogeneous tissue samples, generating an averaged picture of the tissue of interest (bulk-NGS). Although bulk-NGS has a tendency to generalize heterogeneity, certain biological understanding and computational modeling can mitigate this effect within genomic and epigenomic analyses.

Bulk-WGS can be directly used to assess the existence of subclonal mutations through the use of variant allele frequencies (VAFs). Through the modeling of VAFs and copy number changes, an understanding of the clonal architecture may be inferred from such bulk-NGS data. One such method, *Canopy*, uses a Bayesian analysis to identify subpopulations and build a phylogenetic tree detailing their likely evolutionary history (Jiang et al., 2016). Long read bulk TCR sequencing can also be used directly to assess clonal structures under the assumption that there is a unique V(D)J recombination per subclone. As such, the quantity of a given TCR gene can be directly related to the abundance of that subclone and the number of different TCR genes relates to the overall heterogeneity and diversity of the T-cell population. TCR sequencing has also been used, and has shown intratumoral heterogeneity in localized lung carcinomas, which may confer post-surgical recurrence (Reuben et al., 2017a). As epigenetics also plays a significant role in heterogeneity, bisulfite sequencing can be used to study patterns of DNA methylation and estimate clonality, such as with the algorithm methclone (Li et al., 2014). Bisulfite sequencing has also been

used to reveal heterogeneity in DNA methylation of the *MLH1* (a mismatch repair gene) promoter across several endometrial tumors (Varley et al., 2009).

While many bulk-NGS methods rely on mixture models of the VAFs to analyze small indels and point mutations, these methods often rely on the copy number of the gene in question, which can be altered in cancers, and are unable to relate multiple mutations which exist at low frequencies (Jiang et al., 2016). Additionally, bulk sequencing has a tendency to report what an “average” cell in a population would look like and for that reason would not be usable in the analysis of an all-or-nothing response (Altschuler and Wu, 2010). For example, *Xenopus* oocytes, have a binary response when signaled by progesterone to begin a process of maturation; they either mature or they do not (Ferrell and Machleder, 1998). In this case, looking at an average of two distinct oocyte subpopulations – one that has been signaled to mature and one that has not – would artificially yield a biologically impossible “mean oocyte” that has committed to maturation half-way (Altschuler and Wu, 2010).

There has been a significant effort within the field to quantitatively measure heterogeneity and relate it to a functional change. One approach to this is to quantify stochastic gene expression. This has been done through dividing stochastic gene expression into its intrinsic and extrinsic components via a two-color reporter experiment and deriving analytical formulas to measure each component of noise (Singh and Soltani, 2013). Systems have also been developed to quantify the individual contribution of unique processes to stochastic gene expression, and therefore to heterogeneity. For example, experimentally generated models have been used to quantify the individual contribution to chromatin dynamics in isogenic chicken-cell populations (Viñuelas et al., 2013). Also, shifted gene expression dynamics have been shown to drive cell fate choice for hematopoietic progenitors (Kleppe et al., 2017), induced pluripotent stem cells (iPSCs), and the mouse inner-cell mass during embryogenesis (Mojtahedi et al., 2016; Bargaje et al., 2017; Mohammed et al., 2017).

## UTILIZATION OF scNGS

To best understand cellular heterogeneity, single cells must be studied individually through the use of scNGS. Since assessing cellular co-occurrence is the main drawback of bulk-NGS, many studies have also been conducted to further elucidate clonal structures using single-cell DNaseq [including whole exome sequencing (WES) or WGS], bisulfite sequencing, and ATACseq (assay for transposable accessible chromatin, ATAC). Given the variability and importance of gene expression, sc-RNAseq is one of the most used single-cell sequencing techniques (Supplementary Table S1). Single-cell multi-omic analyses are also possible to uncover the true level of heterogeneity across expression levels within cells (Macaulay et al., 2017), which enable examination of the genome, transcriptome, and epigenome at once. scNGS has the ability to resolve noise in bulk-NGS through the additional ability to trace generated reads back to their cell of origin. Though, this added benefit comes at a steep monetary

cost, as single-cell sequencing is still much more expensive than more traditional bulk NGS given the need to sequence more (Supplementary Table S2). Also, subpopulations of cancer cells can be found by scATAC-seq, which has the power to identify specific chromatin motifs. Indeed, when combined with RNA-seq, it has been used to identify epigenetic plasticity between two cell subpopulations (Litzenburger et al., 2017).

There are currently dozens of variations of techniques to study the genome, epigenome, transcriptome, and epitranscriptome of cells, and here, we focus on those most commonly in use (Supplementary Table S1). Each of these technologies has had a significant impact on numerous fields, including immunology, oncology, and microbiology. Because the scope of the benefits of single-cell analysis is so wide, there is tremendous pressure to advance the technologies in the field. This is evident in the dramatic increase in recent years in publications referencing single-cell technologies (Wang and Navin, 2015). These techniques are highly varied, from manual manipulation (Pan et al., 2013) to droplet microfluidics used for sc-WGS (Hosokawa et al., 2017) to the creation of an RNA-library (Hedlund and Deng, 2018), such as bisulfite sequencing, can also be used on the single-cell level (Clark et al., 2017). A novel approach that combines Raman spectroscopy with an algorithmic biomolecular component analysis (microRaman-BCA) allows for the profiling of single organelles from a cell. Because this technique does not destroy the cell during analysis, the study can be performed multiple times on the same cell, providing a better picture of heterogeneity over time (Kuzmin et al., 2017).

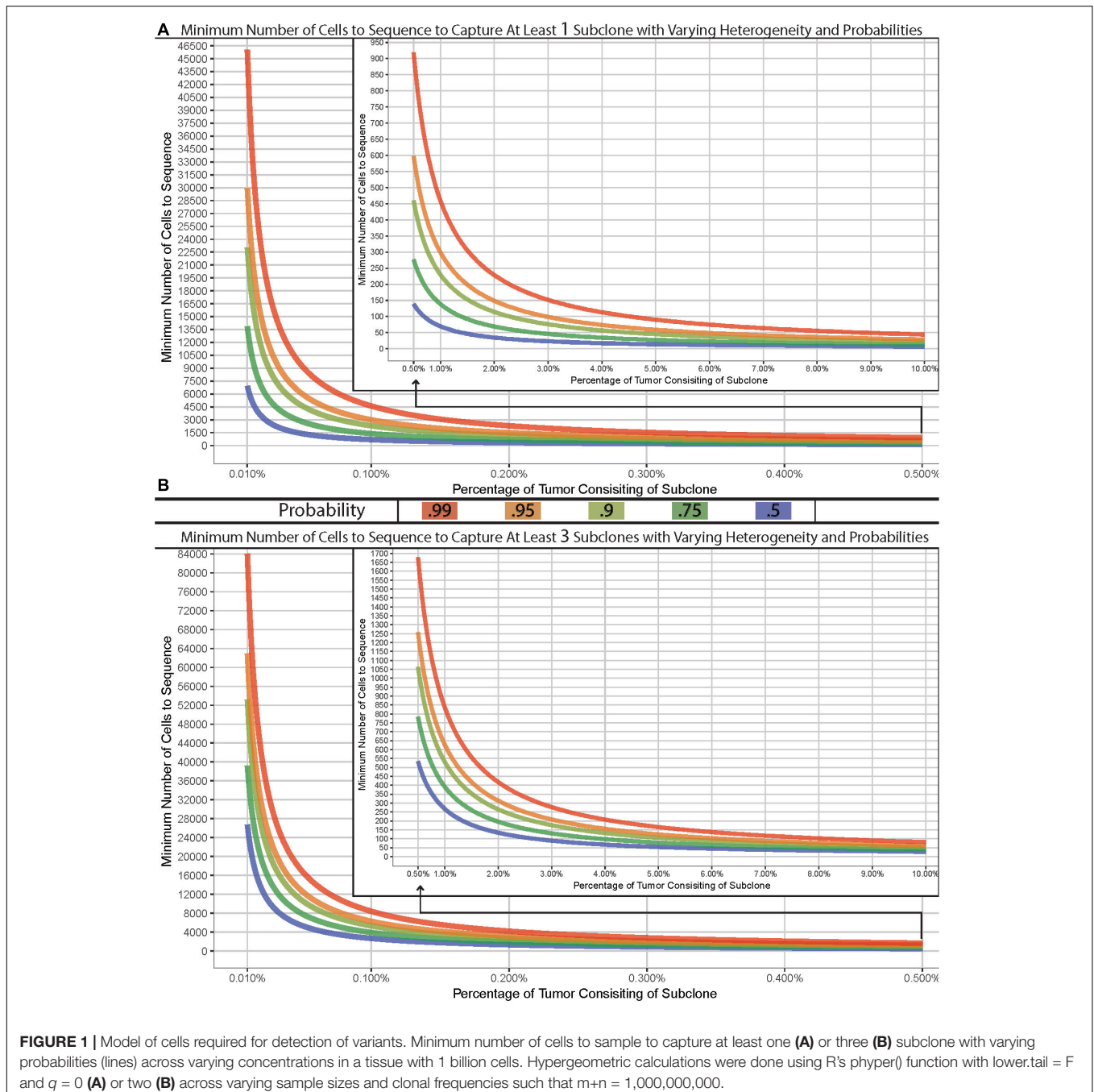
While much of the current knowledge of cellular heterogeneity is transcriptional, newer techniques such as single-cell epigenomics have tremendous potential to study heterogeneity (Hassan et al., 2017) and may be able to provide further insights into the characterization and mechanisms of heterogeneity (Clark et al., 2016). Several topics in epigenomics are best suited to study with single-cell methods, including the relationship between transcriptional heterogeneity and epigenetic heterogeneity, which may vary greatly from cell to cell. Another application of single-cell sequencing is to study tumor resistance and therapeutic response to decrease the chance of resistance or relapse. scNGS can be used to not only detect heterogeneous subclones within a tumor, but also to characterize these cells. Additionally, it can be used to characterize metastases and to create an effective treatment plan that minimizes the chance of chemotherapeutic resistance of specific subclones (Liang and Fu, 2017). In one study, analysis via deep whole-exome sequencing revealed that 75% of relapsed tumors in pediatric B-acute lymphoblastic leukemia were descendants of originally rare subclones (Ma et al., 2015). Given technical and sampling limitations, it is possible that resistant subclones existed within more patients. Although scNGS is currently expensive, treatment for cancer is often much more expensive. For this reason, any possible technique that could lead to a more effective therapy (even an expensive one like scRNA-seq) has clinical potential (Shalek and Benson, 2017).

Additionally, subclones can communicate and interact with each other, leading to complex relationships that may only be fully elucidated via scNGS. Although some of these

interactions are neutral, they can also be positive (leading to a commensalistic/mutualistic relationships in which one or both of the subclones benefit), or negative (leading to competition between subclones, e.g.), and can contribute to the chemotherapeutic resistance of one or more subclones within a tumor (Tabassum and Polyak, 2015). For instance, one study demonstrated that various clonal lineages in a case of colorectal cancer responded differently to treatment with chemotherapy (Kreso et al., 2013). Additionally, there is evidence that parallel evolution of various subclones within a tumor can lead to

polyclonal resistance (Gerlinger et al., 2014). Additionally, intra-tumor heterogeneity makes it more difficult to precisely identify either histologically or genetically a tumor via a traditional biopsy (Tellez-Gabriel et al., 2016).

The implications of tumor heterogeneity in cancer evolution, clinical treatment, and tumoral spatial organization are not yet fully understood (Alizadeh et al., 2015), but scNGS provides a mechanism for beginning to unravel these relationships. Although heterogeneity makes the histological and genetic identity of a tumor more ambiguous, if the mechanisms



driving heterogeneity are further elucidated, they may lead to a better understanding of carcinogenesis (Gay et al., 2016). Moreover, data gathered from single-cell sequencing may help to clarify the methods of cancer progression and subclone resistance to chemotherapeutic treatment by sequencing both smaller transcripts and whole genomes in single cellular representatives of heterogeneous populations (Baslan and Hicks, 2017).

Interestingly, scNGS also has implications in lineage tracking in the development of differentiated tissues, as it may help to further clarify the developmental pathways involved in tissue differentiation (Kester and van Oudenaarden, 2018). As discussed, the nervous and immune systems are both well-studied examples of cellular systems that display cellular heterogeneity. For example, this technique can be used to study the central nervous system, and has the potential to not only molecularly classify various neurons or groups of neurons, but also to further study the molecular mechanisms behind, and possible therapies for, neurological diseases (Ofengeim et al., 2017). Indeed, this application can also be utilized to type sperm and oocytes, allowing for the confirmation and subsequent study of recombination events and polymorphisms in these haploids (Zhang et al., 1992).

## DESIGN OF scNGS EXPERIMENTS

One of the key questions in planning the methodology of a single-cell study is how many cells to sequence. Sequencing more cells enables a greater representation of the cells in a population, giving a more accurate model of the diversity of subclones. The number of single-cells sequenced in a study has scaled exponentially with the development of new technologies. In 2009, for example, only one cell could be sequenced at a time. By 2017, however, the technology has advanced enough to permit the analysis of hundreds of thousands of cells at once (Svensson et al., 2018) and the possibility to generate exobytes and even yottabytes of data in the future.

Many complexities exist with scNGS analyses and need to be carefully considered. Other work have covered the specific differences, benefits, and drawbacks between the various scNGS protocols (Kanter and Kalisky, 2015; Clark et al., 2016; Haque et al., 2017; Liang and Fu, 2017). Previous data have shown that the best scNGS technology should be used for a given hypothesis, in tandem with a proper experimental design for the number of cells. Due to this, the required number of cells necessary to address a given question or tissue model will largely vary depending on the overall hypothesis. However, the question of “how many cells should I sequence” can be simplified to how many cells do you need to sample in order to capture at least one subclonal cell. The chance of sampling a subclone from a tissue of interest depending on the subclonal prevalence, the size of the tissue, and the size of the sample. Therefore, this question can be modeled using the hypergeometric distribution with varying degrees of probability (Figure 1A). It is common within sc-NGS analysis to require multiple cells to contain a given phenotype, and therefore may be more appropriate to ask the

question of “how many cells should I sample to capture at least three subclonal cells” (Figure 1B).

We have built a model to demonstrate the number of cells required for a sampling design can widely vary. As an example, if the goal was to sample a tissue which has 1 billion cells for a previously undefined stem-cell which exist at a population of 0.01%, you would have a 99% chance of sampling at least one stem-cell if you analyzed approximately 46,000 cells. However, to truly characterize and identify this subclonal population or to detect a lower threshold, the number of cells required could easily reach, or even surpass, 100,000 depending on tissue size (Figure 1B). Given the recent advances in scNGS and decreases in costs, this is now possible to do. Such a design – while completely impossible 5 years ago – should be strongly considered when designing experimentations today.

## THE FUTURE OF SINGLE-CELL ANALYSES

While single-cell sequencing has many advantages, it certainly is not a perfect technique. There are many different techniques for obtaining single-cell sequencing data and single-cell whole genome sequencing (sc-WGS), and each of these methods presents its own unique strengths and weaknesses. Multiple displacement amplification (MDA) and other PCR-based sequencing techniques often experience significant amplification bias (de Bourcy et al., 2014; Ahsanuddin et al., 2017). This could lead to incorrect interpretation of the prevalence and diversity of certain genes. Nonetheless, thanks to the breakthroughs in scNGS, the long-sought goal of sequencing of single cells is possible. This has created significant opportunities for advancement in the study of heterogeneity, especially as it applies to cancer. While it may be necessary to sample thousands or even millions of cells to encounter a unique subclone at low prevalence within a large tissue, sequencing continues to get cheaper, and thus scNGS will continue to open up many new research directions into the mechanisms of heterogeneity study variation on cell-by-cell resolution.

## AUTHOR CONTRIBUTIONS

CM and SG conceived and designed the study. CM, SG, and MM analyzed the data. SG, MM, EA, AM, and CM wrote the paper. All authors, reviewed, edited, and approved the manuscript.

## FUNDING

This work was supported by funding from the Irma T. Hirschl and Monique Weill-Caulier Charitable Trusts, Bert L and N Kuggie Vallee Foundation, the WorldQuant Foundation, The Pershing Square Sohn Cancer Research Alliance, NASA (NNX14AH50G and NNX17AB26G), the National Institutes of Health (R25EB020393, R01NS076465, R01AI125416, R01ES021006, 1R21AI129851, and 1R01MH117406), the Bill and Melinda Gates Foundation (OPP1151054).

## ACKNOWLEDGMENTS

We would like to thank the Epigenomics Core Facility at Weill Cornell Medicine, as well as the Starr Cancer Consortium (I9-A9-071).

## REFERENCES

- Afshinnekoo, A., and Mason, C. E. (2016). Epigenetic therapy in a new era of medicine: creating and integrating molecular profiles of patients. *Ann. Transl. Med.* 4:436. doi: 10.21037/atm.2016.11.19
- Ahsanuddin, S., Afshinnekoo, E., Gandara, J., Hakyemezoglu, M., Bezdian, D., Minot, S., et al. (2017). Assessment of Repli-G Multiple Displacement Whole Genome Amplification (WGA) for Metagenomic Sequencing. *J. Biomol. Tech.* 28, 46–55. doi: 10.7171/jbt.17-2801-008
- Alizadeh, A. A., Aranda, V., Bardelli, A., Blanpain, C., Bock, C., Borowski, C., et al. (2015). Toward understanding and exploiting tumor heterogeneity. *Nat. Med.* 21, 846–853. doi: 10.1038/nm.3915
- Altschuler, S. J., and Wu, L. F. (2010). Cellular heterogeneity: do differences make a difference? *Cell* 141, 559–563. doi: 10.1016/j.cell.2010.04.033
- Bargaje, R., Trachana, K., Shelton, M. N., McGinnis, C. S., Zhou, J. X., Chadick, C., et al. (2017). Cell population structure prior to bifurcation predicts efficiency of directed differentiation in human induced pluripotent cells. *Proc. Natl. Acad. Sci. U.S.A.* 114, 2271–2276. doi: 10.1073/pnas.1621412114
- Baslan, T., and Hicks, J. (2017). Unravelling biology and shifting paradigms in cancer with single-cell sequencing. *Nat. Rev. Clin. Oncol.* 13, 557–569. doi: 10.1038/nrco.2017.58
- Beaumont, H. J. E., Gallie, J., Kost, C., Ferguson, G. C., and Rainey, P. B. (2009). Experimental evolution of bet hedging. *Nature* 462, 90–93. doi: 10.1038/nature08504
- Brock, A., Chang, H., and Huang, S. (2009). Non-genetic heterogeneity — a mutation-independent driving force for the somatic evolution of tumours. *Nat. Rev. Genet.* 10, 336–342. doi: 10.1038/nrg2556
- Castro-Wallace, S. L., Chiu, C. Y., John, K. K., Stahl, S. E., Rubins, K. H., McIntyre, A. B. R. (2017). Nanopore DNA sequencing and genome assembly on the international space station. *Nat. Sci. Data* 7:18022. doi: 10.1038/s41598-017-18364-0
- Chang, A. Y., and Marshall, W. F. (2017). Organelles – understanding noise and heterogeneity in cell biology at an intermediate scale. *J. Cell Sci.* 130, 819–826. doi: 10.1242/jcs.181024
- Clark, S. J., Lee, H. J., Smallwood, S. A., Kelsey, G., and Reik, W. (2016). Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. *Genome Biol.* 17:72. doi: 10.1186/s13059-016-0944-x
- Clark, S. J., Smallwood, S. A., Lee, H. J., Krueger, F., Reik, W., and Kelsey, G. (2017). Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nat. Protoc.* 12, 534–547. doi: 10.1038/nprot.2016.187
- Course, M. M., and Wang, X. (2016). Transporting mitochondria in neurons. *F1000Res* 5:F1000 Faculty Rev-1735. doi: 10.12688/f1000research.7864.1
- Dagogo-Jack, I., and Shaw, A. T. (2017). Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* 15, 81–94. doi: 10.1038/nrclinonc.2017.166
- de Bourcy, C. F. A., De Vlaminck, I., Kanbar, J. N., Wang, J., Gawad, C., and Quake, S. R. (2014). A quantitative comparison of single-cell whole genome amplification methods. *PLoS One* 9:e105585. doi: 10.1371/journal.pone.0105585
- Dubinsky, J. M. (2009). Heterogeneity of nervous system mitochondria: location, location, location! *Exp. Neurol.* 218, 293–307. doi: 10.1016/j.expneurol.2009.05.020
- Durlanik, S., and Thiel, A. (2015). Requirement of immune system heterogeneity for protective immunity. *Vaccine* 33, 5308–5312. doi: 10.1016/j.vaccine.2015.05.096
- Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science* 297, 1183–1186. doi: 10.1126/science.1070919
- Elsasser, W. M. (1984). Outline of a theory of cellular heterogeneity. *Proc. Natl. Acad. Sci. U.S.A.* 81, 5126–5129. doi: 10.1073/pnas.81.16.5126
- Emery, B., and Barres, B. A. (2008). Unlocking CNS cell type heterogeneity. *Cell* 135, 596–598. doi: 10.1016/j.cell.2008.10.031
- Ferrell, J. E. Jr., and Machleder, E. M. (1998). The biochemical basis of an all-or-none cell fate switch in *Xenopus oocytes*. 280, 895–898. doi: 10.1126/science.280.5365.895
- Gay, L., Baker, A.-M., and Graham, T. A. (2016). Tumour Cell Heterogeneity. *F1000Res* 5:F1000 Faculty Rev-238. doi: 10.12688/f1000research.7210.1
- Gerlinger, M., McGranahan, N., Dewhurst, S. M., Burrell, R. A., Tomlinson, I., and Swanton, C. (2014). Cancer: evolution within a lifetime. *Annu. Rev. Genet.* 48, 215–236. doi: 10.1146/annurev-genet-120213-092314
- Gerlinger, M., Rowan, A. J., Horswell, S., Math, M., Larkin, J., Endesfelder, D., et al. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* 366, 883–892. doi: 10.1056/NEJMoa1113205
- Gokhale, N. S., McIntyre, A. B. R., McFadden, M. J., Roder, A. E., Kennedy, E. M., Gandara, J. A., et al. (2016). N6-methyladenosine in Flaviviridae viral RNA genomes regulates infection. *Cell Host Microbe* 20, 654–665. doi: 10.1016/j.chom.2016.09.015
- Gordon, S., and Taylor, P. R. (2005). Monocyte and macrophage heterogeneity. *Nat. Rev. Immunol.* 5:953. doi: 10.1038/nri1733
- Griffiths, J. A., Scialdone, A., and Marioni, J. C. (2018). Using single-cell genomics to understand developmental processes and cell fate decisions. *Mol. Syst. Biol.* 14:e8046. doi: 10.15252/msb.20178046
- Grimbergen, A. J., Siebring, J., Solopova, A., and Kuipers, O. P. (2015). Microbial bet-hedging: the power of being different. *Curr. Opin. Microbiol.* 25, 67–72. doi: 10.1016/j.mib.2015.04.008
- Hadjantonakis, A., and Arias, A. (2016). Single-cell approaches: pandora's box of developmental mechanisms. *Dev. Cell* 38, 574–578. doi: 10.1016/j.devcel.2016.09.012
- Haque, A., Engel, J., Teichmann, S. A., and Lönnberg, T. (2017). A practical guide to single-cell RNA-seq for biomedical research and clinical applications. *Genome Med.* 9:75.
- Hassan, C., Afshinnekoo, E., Wu, S., and Mason, C. E. (2017). Genetic and epigenetic heterogeneity and the impact on cancer relapse. *Exp. Hematol.* 54, 26–30. doi: 10.1016/j.exphem.2017.07.002
- Hedlund, E., and Deng, Q. (2018). Single-cell RNA sequencing: technical advancements and biological applications. *Mol. Aspects Med.* 59, 36–46. doi: 10.1016/j.mam.2017.07.003
- Hensen, A. G., Koche, R., Zhuang, J., Jiang, E., Reed, C., Eisenberg, A., et al. (2017). Human PGBD5 DNA transposase promotes site-specific oncogenic mutations in rhabdoid tumors. *Nat. Genet.* 49, 1005–1014. doi: 10.1038/ng.3866
- Hosokawa, M., Nishikawa, Y., Kogawa, M., and Takeyama, H. (2017). Massively parallel whole genome amplification for single-cell sequencing using droplet microfluidics. *Sci. Rep.* 7:5199. doi: 10.1038/s41598-017-05436-4
- Huang, S. (2009). Non-genetic heterogeneity of cells in development: more than just noise. *Development* 136, 3853–3862. doi: 10.1242/dev.035139
- Jiang, Y., Qiu, Y., Minn, A. J., and Zhang, N. R. (2016). Assessing intratumor heterogeneity and tracking longitudinal and spatial clonal evolutionary history by next-generation sequencing. *Proc. Natl. Acad. Sci. U.S.A.* 113, E5528–E5537. doi: 10.1073/pnas.1522203113
- Kanter, I., and Kalisky, T. (2015). Single cell transcriptomics: methods and applications. *Front. Oncol.* 5:53. doi: 10.3389/fonc.2015.00053
- Kester, L., and van Oudenaarden, A. (2018). Single-cell transcriptomics meets lineage tracing. *Cell Stem Cell.* 23, 166–179. doi: 10.1016/j.stem.2018.04.014
- Kleppe, M., Spitzer, M. H., Li, S., Hill, C. E., Dong, L., Papalex, E., et al. (2017). Jak1 Integrates Cytokine Sensing to Regulate Hematopoietic Stem Cell Function and

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00008/full#supplementary-material>

- Stress Hematopoiesis. *Cell Stem Cell* 21, 48.e7–501.e7. doi: 10.1016/j.stem.2017.08.011
- Kreso, A., O'Brien, C. A., van Galen, P., Gan, O. I., Notta, F., Brown, A. M. K., et al. (2013). Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science* 339, 543–548. doi: 10.1126/science.1227670
- Kumar, R. M., Cahan, P., Shalek, A. K., Satija, R., Daleykeyser, A. J., Li, H., et al. (2014). Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature* 516, 56–61. doi: 10.1038/nature13920
- Kuzmin, A. N., Levchenko, S. M., Pliss, A., Qu, J., and Prasad, P. N. (2017). Molecular profiling of single organelles for quantitative analysis of cellular heterogeneity. *Sci. Rep.* 7:6512. doi: 10.1038/s41598-017-06936-z
- Li, S., Garrett-Bakelman, F., Perl, A. E., Luger, S. M., Zhang, C., To, B. L., et al. (2014). Dynamic evolution of clonal epialleles revealed by methclone. *Genome Biol.* 15:472. doi: 10.1186/s13059-014-0472-5
- Li, S., Garrett-Bakelman, F. E., Chung, S. S., Sanders, M. A., Hricik, T., Rapaport, F., et al. (2016). Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. *Nat. Med.* 22, 792–799. doi: 10.1038/nm.4125
- Liang, S.-B., and Fu, L.-W. (2017). Application of single-cell technology in cancer research. *Biotechnol. Adv.* 35, 443–449. doi: 10.1016/j.biotechadv.2017.04.001
- Litzenburger, U. M., Buenrostro, J. D., Wu, B., Shen, Y., Sheffield, N. C., Kalthira, A., et al. (2017). Single-cell epigenomic variability reveals functional cancer heterogeneity. *Genome Biol.* 18:15. doi: 10.1186/s13059-016-1133-7
- Ma, X., Edmonson, M., Yergeau, D., Muzny, D. M., Hampton, O. A., Rusch, M., et al. (2015). Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. *Nat. Commun.* 6:6604. doi: 10.1038/ncomms7604
- Macaulay, I. C., Ponting, C. P., and Voet, T. (2017). Single-Cell Multiomics: multiple measurements from single Cells. *Trends Genet.* 33, 155–168. doi: 10.1016/j.tig.2016.12.003
- Mason, C. E., Afshinnekoo, E., Tighe, S., Wu, S., and Levy, S. (2017). International standards for genomes, transcriptomes, and metagenomes. *J. Biomol. Tech.* 28, 8–18. doi: 10.7171/jbt.17-2801-006
- Mason, C. E., Porter, S., and Smith, T. (2014). Characterizing Multi-omic data in Systems Biology. *Adv. Exp. Med. Biol.* 799, 15–38. doi: 10.1007/978-1-4614-8778-4\_2
- McIntyre, A. B. R., Rizzardi, L., Yu, A. M., Alexander, N., Rosen, G. L., Botkin, D. J., et al. (2016). Nanopore sequencing in microgravity. *Nat. Partner. J. Microgravity* 2:16035. doi: 10.1038/npjimggrav.2016.35
- Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., et al. (2017). Single-cell landscape of transcriptional heterogeneity and cell fate decisions during mouse early gastrulation. *Cell Rep.* 20, 1215–1228. doi: 10.1016/j.celrep.2017.07.009
- Mojtahedi, M., Skupin, A., Zhou, J., Castaño, I. G., Leong-Quong, R. Y. Y., Chang, H., et al. (2016). Cell fate decision as high-dimensional critical state transition. *PLoS Biol.* 14:e2000640. doi: 10.1371/journal.pbio.2000640
- Ofengeim, D., Giagtzoglou, N., Huh, D., Zou, C., and Yuan, J. (2017). Single-cell RNA sequencing: unraveling the brain one cell at a time. *Trends Mol. Med.* 23, 563–576. doi: 10.1016/j.molmed.2017.04.006
- Pan, X., Durrett, R. E., Zhu, H., Tanaka, Y., Li, Y., Zi, X., et al. (2013). Two methods for full-length RNA-seq for low quantities of cells and single cells. *Proc. Natl. Acad. Sci. U.S.A.* 110, 594–599. doi: 10.1073/pnas.1217322109
- Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344, 1396–1401. doi: 10.1126/science.1254257
- Potts, S. J., Krueger, J. S., Landis, N. D., Eberhard, D. A., Young, G. D., Schmechel, S. C., et al. (2012). Evaluating tumor heterogeneity in immunohistochemistry-stained breast cancer tissue. *Lab. Invest. A J. Tech. Methods Pathol.* 92, 1342–1357. doi: 10.1038/labinvest.2012.91
- Raj, A., and van Oudenaarden, A. (2008). Stochastic gene expression and its consequences. *Cell* 135, 216–226. doi: 10.1016/j.cell.2008.09.050
- Reuben, A., Gittelman, R., Gao, J., Zhang, J., Yusko, E. C., Wu, C.-J., et al. (2017a). TCR repertoire intratumor heterogeneity in localized lung adenocarcinomas: an association with predicted neoantigen heterogeneity and postsurgical recurrence. *Cancer Discov.* 7, 1088–1097. doi: 10.1158/2159-8290.CD-17-0256
- Reuben, A., Spencer, C. N., Prieto, P. A., Gopalakrishnan, V., Reddy, S. M., Miller, J. P., et al. (2017b). Genomic and immune heterogeneity are associated with differential responses to therapy in melanoma. *NPJ Genom. Med.* 2:10. doi: 10.1038/s41525-017-0013-8
- Saletore, Y., Meyer, K., Korlach, J., Vilfan, I., Jaffrey, S., and Mason, C. E. (2012). The birth of the epitranscriptome: deciphering the function of RNA modifications. *Genome Biol.* 13:175. doi: 10.1186/gb-2012-13-10-175
- Shabaan, H., Westfall, D. A., Mohammad, R., Danko, D., Bezdan, D., Afshinnekoo, E., et al. (2018). The microbe directory: an annotated, searchable inventory of microbes' characteristics. *Gates Open Res.* 2:3. doi: 10.12688/gatesopenres.12772.1
- Shalek, A. K., and Benson, M. (2017). Single-cell analyses to tailor treatments. *Sci. Transl. Med.* 9:eaan4730. doi: 10.1126/scitranslmed.aan4730
- Shih, A., Meydan, C., Shank, K., Garrett-Bakelman, F., Ward, F., and Levine, R. (2017). Combination targeted therapy to disrupt aberrant oncogenic signaling and reverse epigenetic dysfunction in IDH2- and TET2-mutant acute myeloid leukemia. *Cancer Discov.* 7, 494–505. doi: 10.1158/2159-8290.CD-16-1049
- Singh, A., Soltani, M. (2013). Quantifying intrinsic and extrinsic variability in stochastic gene expression models. *PLoS One* 8:e84301. doi: 10.1371/journal.pone.0084301
- Svensson, V., Vento-Tormo, R., and Teichmann, S. A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* 13, 599–604. doi: 10.1038/nprot.2017.149
- Tabassum, D. P., and Polyak, K. (2015). Tumorigenesis: it takes a village. *Nat. Rev. Can.* 15, 473–483. doi: 10.1038/nrc3971
- Tellez-Gabriel, M., Ory, B., Lamoureux, F., Heymann, M.-F., and Heymann, D. (2016). Tumour heterogeneity: the key advantages of single-cell analysis. *Int. J. Mol. Sci.* 17 2142. doi: 10.3390/ijms17122142
- Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352, 189–196. doi: 10.1126/science.aad0501
- Varley, K. E., Mutch, D. G., Edmonston, T. B., Goodfellow, P. J., and Mitra, R. D. (2009). Intra-tumor heterogeneity of MLH1 promoter methylation revealed by deep single molecule bisulfite sequencing. *Nucleic Acids Res.* 37, 4603–4612. doi: 10.1093/nar/gkp457
- Viñuelas, J., Kaneko, G., Coulon, A., Vallin, E., Morin, V., Mejia-Pous, C., et al. (2013). Quantifying the contribution of chromatin dynamics to stochastic gene expression reveals long, locus-dependent periods between transcriptional bursts. *BMC Biol.* 11:15. doi: 10.1186/1741-7007-11-15
- Vu, L., Pickering, B. F., Cheng, Y., Zaccara, S., Nguyen, D., Minuesa, G., et al. (2017). The N6-methyladenosine (m6A)-forming enzyme METTL3 controls myeloid differentiation of normal and leukemia cells. *Nat. Med.* 23, 1369–1376. doi: 10.1038/nm.4416
- Wang, Y., and Navin, N. E. (2015). Advances and applications of single-cell sequencing technologies. *Mol. Cell* 58, 598–609. doi: 10.1016/j.molcel.2015.05.005
- Zhang, L., Cui, X., Schmitt, K., Hubert, R., Navidi, W., and Arnhem, N. (1992). Whole genome amplification from a single cell: implications for genetic analysis. *Proc. Natl. Acad. Sci. U.S.A.* 89, 5847–5851. doi: 10.1073/pnas.89.13.5847
- Zheng, C., Zheng, L., Yoo, J. K., Guo, H., Zhang, Y., Guo, X., et al. (2017). Landscape of Infiltrating T cells in liver cancer revealed by single-cell sequencing. *Cell* 169, 1342.e16–56.e16. doi: 10.1016/j.cell.2017.05.035

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Goldman, MacKay, Afshinnekoo, Melnick, Wu and Mason. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.