



# Single-Cell mRNA Sequencing in Cancer Research: Integrating the Genomic Fingerprint

Sören Müller<sup>1,2</sup> and Aaron Diaz<sup>1,2\*</sup>

<sup>1</sup> Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA, United States, <sup>2</sup> Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco, San Francisco, CA, United States

Critical cancer mutations are often regional and mosaic, confounding the efficacy of targeted therapeutics. Single cell mRNA sequencing (scRNA-seq) has enabled unprecedented studies of intra-tumor heterogeneity and its role in cancer progression, metastasis, and treatment resistance. When coupled with DNA sequencing, scRNA-seq allows one to infer the *in vivo* impact of genomic alterations on gene expression. This combination can be used to reliably distinguish neoplastic from non-neoplastic cells, to correlate paracrine-signaling pathways between neoplastic cells and stroma, and to map expression signatures to inferred clones and phylogenies. Here we review recent advances in scRNA-seq, with a special focus on cancer. We discuss the challenges and prospects of combining scRNA-seq with DNA sequencing to assess intra-tumor heterogeneity.

**Keywords:** cancer genomics, single-cell sequencing, tumor microenvironment, cancer phylogenetics, cancer stem cells

## OPEN ACCESS

### Edited by:

Ashani Weeraratna,  
Wistar Institute, United States

### Reviewed by:

Philipp Kaldis,  
Agency for Science, Technology  
and Research (A\*STAR), Singapore  
Jorge Melendez-Zajgla,  
National Institute of Genomic  
Medicine, Mexico

### \*Correspondence:

Aaron Diaz  
aaron.diaz@ucsf.edu;  
aad1974@gmail.com

### Specialty section:

This article was submitted to  
Cancer Genetics,  
a section of the journal  
Frontiers in Genetics

**Received:** 20 March 2017

**Accepted:** 18 May 2017

**Published:** 31 May 2017

### Citation:

Müller S and Diaz A (2017)  
Single-Cell mRNA Sequencing  
in Cancer Research: Integrating  
the Genomic Fingerprint.  
Front. Genet. 8:73.  
doi: 10.3389/fgene.2017.00073

## BACKGROUND

Next-generation sequencing (NGS) based studies have identified critical genetic alterations in a variety of malignancies (Brennan et al., 2013; Hoadley et al., 2014; Bai et al., 2015; Furnari et al., 2015; Ceccarelli et al., 2016; Cancer Genome Atlas Research Network, et al., 2016; Wang J. et al., 2016). However, relatively few targeted therapeutics are curative. Intra-tumor heterogeneity has emerged as an essential parameter confounding the delivery of a complete treatment (Saunders et al., 2012). Assessing tumor heterogeneity from bulk RNA or DNA extractions is limited to either inter-tumor comparisons (Cancer Genome Atlas Research Network, 2008; Müller et al., 2015), or comparisons across a small number of stereotactic biopsies (Gerlinger et al., 2012). Obtaining multi-region biopsies during complex cancer surgeries represents a major challenge. Moreover, inferring the contribution of specific tumor sub-clones and/or stromal cell-types from these data is a computationally difficult task, with a degree of uncertainty (El-Kebir et al., 2015; Turajlic et al., 2015).

The past decade has seen rapid advances in protocols for the faithful reverse-transcription (RT) and amplification of RNA from individual cells (Tang et al., 2009; Zong et al., 2012; Chapman et al., 2015). Microfluidic and other methods for single-cell isolation and library preparation have brought high-throughput single-cell RNA-sequencing (scRNA-seq) into the mainstream (Patel et al., 2014; Ting et al., 2014; Kim et al., 2015; Min et al., 2015; Hou et al., 2016; Li et al., 2016; Tirosh et al., 2016b; Gerber et al., 2017; Winterhoff et al., 2017). The limitations and uses of these novel

data are still being defined. However, most state-of-the-art algorithms for analyzing sequencing data were not designed with single-cell studies in mind (Stegle et al., 2015; Bacher and Kendziorski, 2016).

Computational biologists are racing to keep up (Tanay and Regev, 2017). Novel analysis tools for single-cell cancer studies are being rapidly developed (Garmire et al., 2016). Already, scRNA-seq has led to groundbreaking insights into clonal tumor evolution (Navin et al., 2011), metastatic dissemination (Lawson et al., 2015), the development of chemo-resistance (Kim et al., 2016), and interactions between tumor and stromal cells (Choi et al., 2015). In this review, we summarize current advances in the acquisition and analysis of scRNA-seq data from samples of tumor tissue. We focus on the integration of orthogonal assays and future directions for scRNA-seq in cancer research.

## SINGLE-CELL RNA SEQUENCING IN COMPLEX TUMOR-TISSUE

With only approximately 1pg of RNA in a single cell, and a median transcript abundance of fewer than 100 copies per gene, unbiased library-generation from such a small amount of starting material is challenging (Macaulay and Voet, 2014). The technical limitations of library generation for single-cell RNA-seq include non-uniform transcript coverage (3' bias), and non-linear library amplification. Strategies are being actively developed to minimize these effects (Kolodziejczyk et al., 2015).

Two recent publications compared protocols for single-cell library-generation (Svensson et al., 2016; Ziegenhain et al., 2017). Both papers compared sensitivity and accuracy across protocols, using synthetic-RNA spike-in controls as a gold-standard. The advantage of using a gold-standard is the ability to assess not just sensitivity (which could be inferred without spike-in controls), but also accuracy. If cDNA copy-number accurately reflects mRNA abundance in single cells, then these data can be used quantitatively to compare expression within and between individual cells. The caveat of these studies is that synthetic-RNA spike-in controls are subject to library-preparation effects, treatment effects and other technical biases not observed when cloning cDNA from tissue-derived RNA (Risso et al., 2014).

Nonetheless, these two studies found that Smart-Seq2 was the most sensitive method. For example, almost twice as many genes per cell were detected via Smart-Seq2 when compared to Drop-seq, given similar sequencing depths. These two approaches represent two ends of the spectrum, in terms of trade-offs between transcriptome coverage and number of cells profiled. Methods that use standard oligo-dT primers (e.g., Smart-seq2), as compared to mRNA-capture beads (e.g., Drop-seq), have greater transcriptome coverage in terms of the number of distinct genes-sequenced, and greater coverage of the 5' end of individual transcripts. The latter is especially useful when studying expressed mutations in cancer samples. Protocols like Smart-seq2 are typically applied in multi-well plate or

microfluidic-chip based platforms that have a throughput of hundreds of cells. Droplet-based methods capture thousands of cells at a time.

Not surprisingly, batch effects have been observed when comparing batches of cells captured in separate assays (Tung et al., 2017). Specifically, the proportion of measured genes typically accounts for the major proportion of observed variability between batches (Hicks et al., under review). Best practices of experimental design, such as randomized blocking, are advised whenever possible. Statistical methods can also be used to adjust for batch effects *a posteriori* (Finak et al., 2015).

The effect of tissue dissociation on the efficiency of single-cell cDNA-library generation remains poorly understood. Some cell-isolation protocols for scRNA-seq may be biased toward certain cell types. For example, microfluidic platforms for automated library construction use chips which are graded to isolate cells of a given size (Müller et al., 2016). Biases present in droplet-based scRNA-seq platforms, for or against certain cell types, have not yet been fully investigated. Tumor disassociation protocols often involve cell selection by straining and/or density gradients (Venteicher et al., 2017). Fluorescence-activated cell sorting approaches to cell isolation followed by library preparation via Smart-seq2 provide perhaps the most flexible approach to apply scRNA-seq a specific, tumor-infiltrating cell-type of interest.

With the advent of droplet-based methods, there has been a trend to sequence more cells at lower coverage. This leads to a lower library-complexity per cell, and gives rise to the question: how many cells are required to obtain representative results from scRNA-seq data? As little as 50 cells have been shown to be sufficient to achieve a per-gene coefficient-of-variation that is comparable to a standard bulk RNA-seq experiment when sequencing a cell line (Shapiro et al., 2013). In another recent scRNA-seq study, only five cells from a patient-derived xenograft were required to represent 70% of the genes found in a bulk extraction (Kim et al., 2015), and robust transcriptome-wide correlations between single-cell and bulk experiments were observed when the sample sizes were increased to 35–50 cells. However, in both examples, cells were derived from relatively homogeneous populations.

Sample-size estimation in complex tissue, such as biopsies of patient tumors with a high degree of stromal infiltrate, remains an open problem. Given the wide range in cellular heterogeneity across cancer types, a one-size-fits-all recommendation as to sample size is likely impossible. However, techniques from capture statistics can be used to estimate sample sizes *ad hoc*, from pilot studies (Daley and Smith, 2013). Standards for sequencing depth per cell and methods to assess single-cell library complexity are beginning to emerge (Daley and Smith, 2014; Wu et al., 2014; Grün and Van Oudenaarden, 2015; Bacher and Kendziorski, 2016; Diaz et al., 2016). The majority of genes expressed in a cell are detected at a read-depth of 250,000–500,000 reads (Wu et al., 2014; Bacher and Kendziorski, 2016). If the goal is to survey cell diversity in an unbiased fashion, classify cell types by expression profile, and infer the proportions of each cell type, then even 50,000 reads per cell have been shown to be sufficient (Pollen et al., 2014). On the other hand, greater depth of

coverage per cell is required to rigorously distinguish neoplastic from stromal cells, or to triage cells by the presence or absence of expressed mutations. We now discuss how low sequencing depth, low cDNA library complexity, and other technical factors impact the ability to fully integrate DNA sequencing with scRNA-seq.

## QUANTIFYING EXPRESSED MUTATIONS IN scRNA-seq

In principle, single-nucleotide variants (SNVs) and small insertions/deletions (INDELs) in expressed regions can be detected in scRNA-seq. In contrast to the detection of SNVs from exome sequencing (exome-seq), there are additional challenges inherent to quantifying SNVs in scRNA-seq (Piskol et al., 2013). Calling SNVs *de novo* from RNA sequencing (RNA-seq) is challenging, even from deeply sequenced bulk-RNA extractions. Variability in gene expression and allele-specific expression contribute significantly to the error rate (Castel et al., 2015). For scRNA-seq, these challenges are magnified by low coverage. Some scRNA-seq library prep protocols also impart additional coverage bias toward the 3' end of the gene (Chapman et al., 2015), contributing to the dropout rate in SNV quantification in SNVs near the 5' end. The most robust approaches to quantifying SNVs in single cells have integrated orthogonal data, to classify cells based on expressed mutations that were called first from DNA sequencing. For example, two recent studies combine scRNA-seq with exome-seq to map transcriptional signatures to inferred clones.

Kim et al. (2015) studied the effect of intra-tumor heterogeneity on anti-cancer drug-response using scRNA-seq and bulk exome-seq of patient-derived xenograft (PDX) tumor cells from a lung-adenocarcinoma patient. In a novel demonstration of the possibilities of single-cell data-integration, they correlated the presence of a KRAS mutation in individual cells to an expression signature characteristic of RAS/MAPK pathway activation. The study also revealed the technical limitations of quantifying SNVs in scRNA-seq. From more than 1,000 somatic SNVs identified via exome-seq, only 50 were expressed in more than three cells. Nonetheless, they did quantify a set of highly prevalent mutations affecting known oncogenes.

In another study, here of oligodendroglioma (Tirosh et al., 2016b), Tirosh and colleagues identified stem-like cells as the main source of tumor proliferation and the apex of a developmental hierarchy. To distinguish malignant from non-malignant cells, they developed a strategy to quantify the sensitivity of scRNA-seq in detecting somatic SNVs. The authors compare the variant-allele frequencies (VAFs) observed in exome-seq to the cellular frequencies of expressed mutations found in scRNA-seq. On average, somatic SNVs called from exome-seq could be validated in only 1.3% of the expected fraction of cells. Not surprisingly, the sensitivity of detection in scRNA-seq was positively correlated with gene expression levels. Ultimately, the authors found that they had much greater sensitivity in quantifying large-scale copy-number variants (CNVs), than they had with SNVs.

Large-scale CNVs are proving to be a genomic alteration that can be robustly quantified both in exome-seq (Alerting et al., 2012; Zack et al., 2013; Wang et al., 2015; Witkiewicz et al., 2015) and scRNA-seq (Patel et al., 2014; Müller et al., 2016; Tirosh et al., 2016a,b). While the expression level of an individual gene may be stochastically up- or down-regulated independent from its DNA copy-number, tumor/normal exome-seq read-count fold-changes correlate with single-cell expression-trendlines over megabase-scale regions (Peña-Llopis and Brugarolas, 2013; Hou et al., 2016; Müller et al., 2016). Moreover, by using a scRNA-seq data set from a relevant non-malignant tissue as a normal control, the error rate in quantifying the presence/absence of large-scale CNVs (called from exome-seq) in individual cells (assessed by scRNA-seq) can be rigorously controlled (Müller et al., 2016). It's worth noting that large-scale CNVs are in principle detectable based on estimates of gene abundance alone, sequencing the entirety of each mRNA transcript is therefore not required. When large numbers of cells are sequenced simultaneously, cost-reduction strategies such as sequencing only the 3' end of each mRNA are often employed. While most expressed SNVs and INDELs would be lost with 3' sequencing, it is entirely compatible with large-scale CNV detection. All in all, for researchers who want to use scRNA-seq with heterogenous tumor samples, where neoplastic cells must be reliably separated from stromal and immune cells, augmenting scRNA-seq with exome-seq is a cost-effective strategy for achieving specificity while producing versatile data.

## FILTERING AND CLASSIFYING STROMAL AND IMMUNE CELLS FROM WHOLE-TUMOR scRNA-seq

While bulk RNA-seq experiments can only estimate the fraction of stromal and immune cells (Yoshihara et al., 2013; Becht et al., 2016), scRNA-seq gives information about the identity of every cell sequenced (Wagner et al., 2016). Neoplastic cells can often be distinguished from stromal/immune cells via a clustering of gene expression profiles (Satija et al., 2015). However, some degree of stochastic mixing inevitably occurs when clustering cells by gene expression. Neoplastic cells can also express genes typically associated with immune cells, further adding to the ambiguity of classification via clustering alone (Patel et al., 2014).

The inference of large-scale CNVs from scRNA-seq data has become one of the most reliable techniques to distinguish neoplastic from stromal/immune cells (Tirosh et al., 2016a). For example, Tirosh et al. (2016b) used the presence of the 1p/19q co-deletion in oligodendrogliomas [a hallmark of that disease (Yip et al., 2012)] to identify neoplastic cells. Of the approximately 7% of cells that lacked detectable CNVs, all expressed markers of microglia or oligodendrocytes, confirming their approach. A related computational technique that can be used to add support to inferred, large-scale CNVs uses the VAFs of heterozygous germline mutations. Changes in copy number will skew the observed VAFs of heterozygous germline SNVs. Analysis of germline SNV VAFs is integrated into state-of-the-art

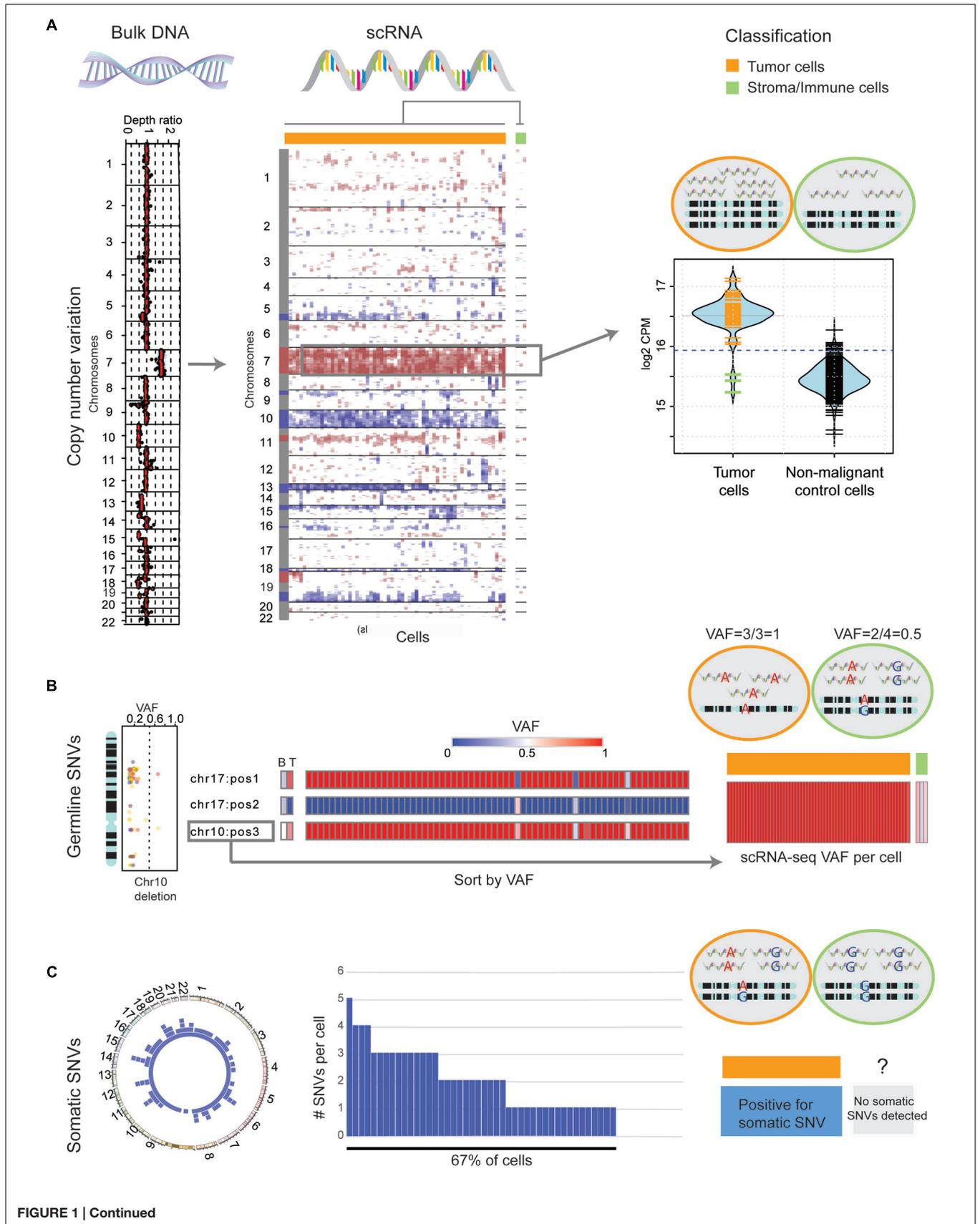


FIGURE 1 | Continued

**FIGURE 1** | Classification of genomic mutations in single cells distinguishes neoplastic cells from immune infiltrate. **(A)** Left: The depth ratio of exome-seq reads from bulk tumor and blood control (x-axis) along autosomes (y-axis) identifies large-scale CNVs in a primary GBM. Middle: The detected genomic CNVs are reflected in single cells (columns) from the same case after normalizing the mean expression, within windows of 100 adjacent genes, by the mean expression in a normal brain control (red: fold-change > 1, blue: fold-change < 1). Hierarchical clustering (complete linkage, Euclidean distance) reveals three cells lacking large-scale CNVs. Right: A comparison of total sequencing depth on chromosome 7, measured by the sum-total counts per million (CPM), in individual cells between the tumor biopsy and a normal brain control. The 5% significance level of the control distribution is indicated by dotted lines. **(B)** Left: The VAF of heterozygous germline mutations (x-axis) deviates from 0.5 in regions of copy number alterations, here chromosome 10 is given as an example. Middle: Three heterozygous germline SNVs change in VAF (0.5 in blood sample) between blood (B) and tumor (T) exome-seq. In RNA-seq of individual cells, only the reference (blue) or the variant allele (red) are observed. Three cells are outliers, expressing both alleles. Right: The presence of both alleles in these three cells verifies their previous classification as non-neoplastic based on CNVs. **(C)** Left: Circos-plot of somatic SNVs detected by Mutect from exome-seq, for all autosomes. Middle: Histogram of somatic SNVs (y-axis) detected in single cells (x-axis). Right: 67% of cells can be classified as tumor cells due to the presence of at least one somatic SNV that has been validated in exome-seq.

algorithms to detect large-scale CNVs from exome-seq data (Favero et al., 2015), but its utility has not yet been explored in scRNA-seq data. As opposed to somatic SNVs, germline SNVs have been shown to have less allelic bias (Li et al., 2015). This suggests that germline-SNV VAF analysis can provide additional evidence to confirm large-scale CNVs.

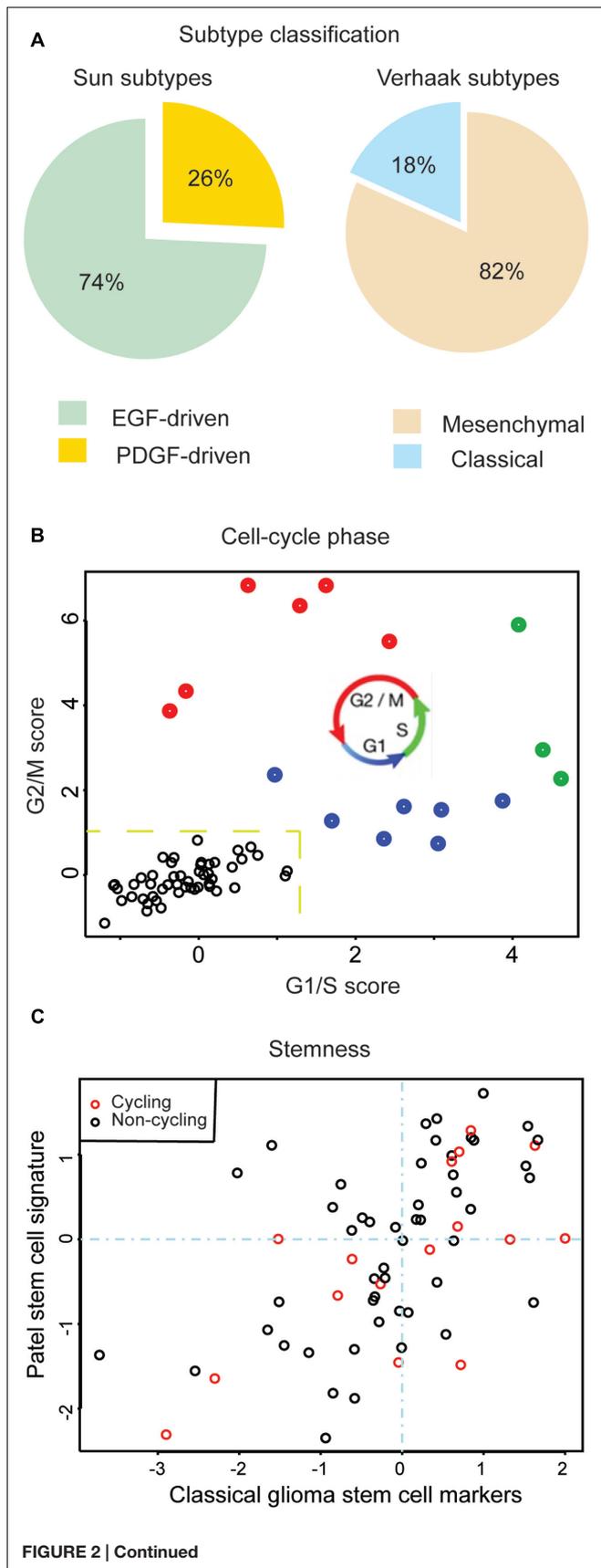
Integrating an auxiliary exome-seq experiment provides a cost-effective way to rigorously separate neoplastic from stromal and immune cells, in scRNA-seq data. In this context, we propose separating cells based on four sources of evidence: (1) large-scale CNVs that are observed in both platforms; (2) the VAFs of germline SNVs, compared between platforms; (3) somatic SNVs found in both platforms; and (4) a clustering of scRNA-seq transcriptional profiles. As an example, we apply the above criterion to previously published scRNA-seq and matched exome-seq from a primary human glioblastoma (GBM) biopsy, SF10360 (Müller et al., 2016). Exome-seq revealed large-scale CNVs common in GBM, including a gain of chromosome 7 and a loss of chromosome 10 (Li et al., 2012). Both occurred with high VAF. Plotting gene expression, in sliding windows of 100 adjacent genes and normalized by a non-malignant brain control (Darmanis et al., 2015), indicates the presence of these two mutations in all but three cells (Figure 1A, middle). We previously described an approach to rigorously classify the presence of large-scale somatic CNVs in single cells, by comparison to a set of non-malignant control cells (Müller et al., 2016). These three cells show no evidence of CNVs, based on that method (Figure 1A, right). Next, consider heterozygous germline SNVs with differences in VAF between blood and tumor exome-seq. Cells harboring heterozygous germline SNVs in regions of copy-number loss should only express either the reference or the variant allele, thus providing further support for single-cell CNV calls. Three germline SNVs, two on chromosome 10 and one on chromosome 17 fulfill these criteria (Figure 1B, left). While there is only one allele found in putative neoplastic cells, the three cells which lack clonal, large-scale CNVs express both the reference and germline variants (Figure 1B, middle). Furthermore, of the somatic SNVs identified in exome-seq (Figure 1C, left), 67% of cells express at least one (Figure 1C, middle). Cells not classified as neoplastic (based on large-scale CNV and germline-SNV analysis) are devoid of somatic SNVs (Figure 1C, right), further confirming their status as non-neoplastic cells. Finally, hierarchical clustering in the space of GBM marker-genes as well

as tumor-associated-macrophage markers reveals two clusters of cells (Supplementary Figure S1). The 3 putative non-malignant cells clustered separately and express high levels of macrophage/microglia markers. Taken together, we can classify these three cells as non-neoplastic, infiltrating immune cells based on our four criteria.

## ACCESSING INTRA-TUMOR HETEROGENEITY

Large-scale molecular profiling has identified prognostic cancer-subtypes based on transcriptional signatures (Brennan et al., 2013; Cancer Genome Atlas Research Network, 2013, 2014a,b; Bass et al., 2014; Cancer Genome Atlas Research Network, et al., 2016; Wang Q. et al., 2016). However, recent scRNA-seq studies have revealed that most tumors are a heterogeneous composition of cells conforming to multiple subtypes (Figure 2A) (Patel et al., 2014; Müller et al., 2016). Since a variety of genomic alterations are detectable in scRNA-seq data, scRNA-seq can be used to analyze intra-tumor heterogeneity at both the transcriptional and mutational levels simultaneously. This is useful for studying how intra-tumor heterogeneity arises in the first place. Several groups have begun to use scRNA-seq data to address the fundamental question of how tumors propagate through cellular hierarchies (Müller et al., 2016; Tirosh et al., 2016b; Woodworth et al., 2017).

In the cancer stem-cell model, a small population of stem-like cells gives rise to differentiated, phenotypically diverse progeny with limited proliferative potential (Ghaffari, 2011). Assuming that the majority of these cancer stem-cells persist in a slow-cycling or quiescent state, as observed in some cancers (Dembinski and Krauss, 2010; Chen et al., 2012), the genetic diversity of the tumor is largely explained by the genetic diversity within the stem-cell population. In the model of clonal evolution, those acquired mutations which provide a selective advantage will expand (Greaves and Maley, 2012). These two models are not strictly contradictory. The progeny of cancer stem-cells may retain proliferative potential and thereby contribute additional mutations. If cancers follow the stem-cell model, clonal evolution, or a mixture of both, or if this even depends on the cancer type currently remains an open question (Shackleton et al., 2009). ScRNA-seq is uniquely suited to address this challenge. Two recent studies have performed this type of integrated analysis, both in glioma.

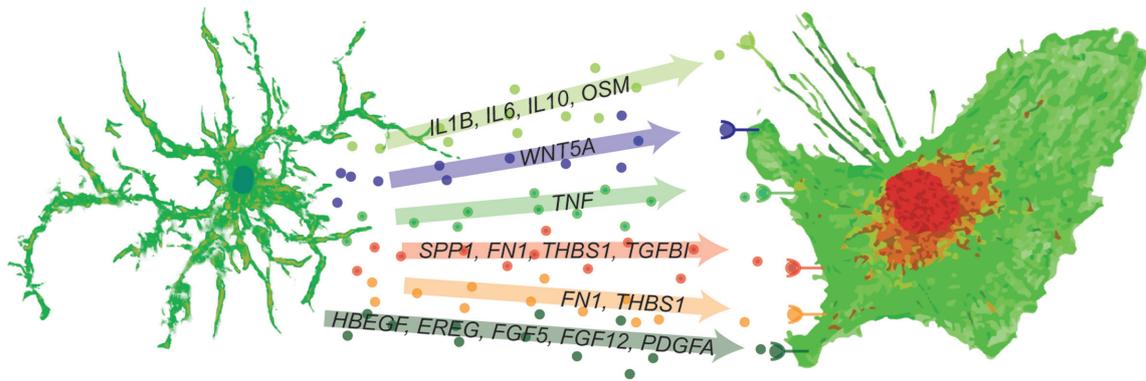


**FIGURE 2 |** Assessments of intra-tumor heterogeneity made possible by scRNA-seq. **(A)** Percentage of single cells associated to a given GBM subtype. Adapted from Müller et al. (2016). **(B)** Estimation of cycling cells. An average of G1/S and G2/M scores  $> 1.2$  classifies cells as cycling (labeled in red, green, or blue) or non-cycling (labeled in black). **(C)** Comparison of stem-like expression signatures for individual cells, based on marker genes canonical to GBM stem-cells: *CD44*, *CD133* (*PROM1*), *NES*, *KLF4*, *MYC*, *NANOG*, *STAT3*, *SOX2*, *MET* (x-axis) and marker genes published by Patel et al. (2014) (y-axis).

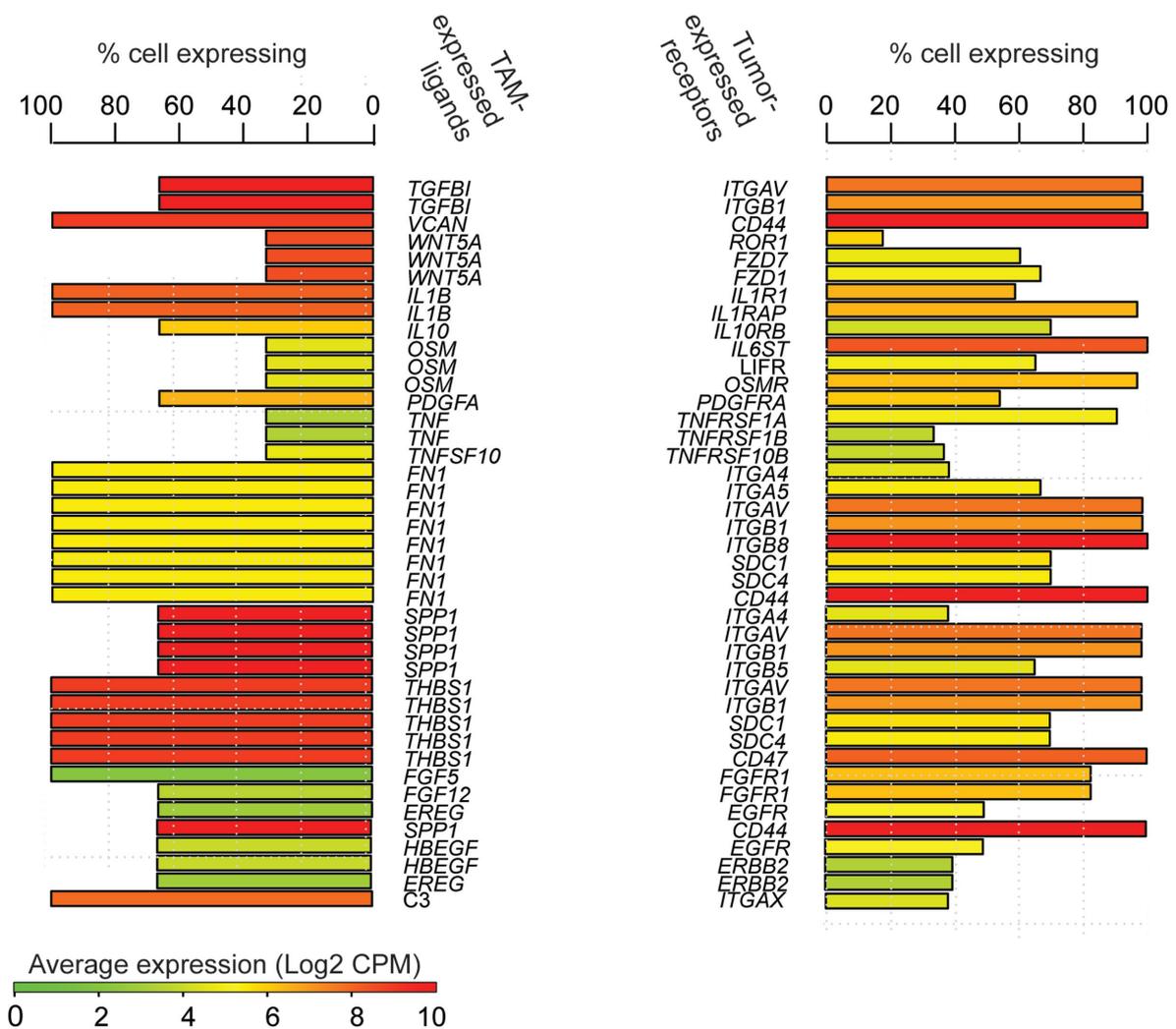
Working with high-grade glioblastomas, Müller et al. (2016) first identified large-scale CNVs from exome-seq data and then classified individual cells according to the presence or absence of these alterations via scRNA-seq. Using standard phylogenetic approaches, they then organized cells into mutational hierarchies. They found that these hierarchies correlated with transcriptional hierarchies of cell-types found in the developing brain. Tirosh et al. (2016b) took a complementary perspective and first organized their low-grade glioma scRNA-seq data based on hierarchies of transcriptional phenotypes, corresponding to stem cells and their differentiated progeny. They then cross-referenced validated, expressed mutations. In contrast to Müller et al. (2016) they found that their transcriptional and mutational hierarchies were largely uncorrelated. While in Müller et al. (2016) found that differentiated cell types more frequently harbored sub-clonal mutations than stem-like cells, Tirosh et al. (2016b) found that sub-clonal mutations occurred with equal frequency in both stem-like and differentiated populations. The interpretation of Tirosh et al. (2016b) was that in their low-grade gliomas proliferation was restricted to stem-like cells. By contrast, the data of Müller et al. (2016) support an expansion in high-grade glioblastoma of proliferative cell-types that do not have a stem-like transcriptional signature, but rather the mRNA profile of an oligodendrocyte progenitor or migrating neuroblast. An expansion of transit-amplifying, proliferative cell-types in high-grade glioblastoma, relative to low-grade glioma, is also supported by a cell-cycle analysis of the scRNA-seq expression signatures. For example, in the glioblastoma case SF10360 described in Müller et al. (2016) cycling cells can be immediately identified and classified by cell-cycle stage (Figure 2B). Cycling glioblastoma cells are frequently depleted of both the glioma-stemness genes identified by Patel et al. (2014), as well as classical glioma stem cell markers (Figure 2C) (Bradshaw et al., 2016). This type of analysis, where cells are separated based on genomic alterations, transcriptional phenotypes (e.g., stem-like expression pattern), or cell state (e.g., cycling cells), demonstrates the versatility of scRNA-seq data.

## PREDICTING OF INTERACTIONS BETWEEN THE TUMOR AND THE MICROENVIRONMENT

Tumor-infiltrating stromal and immune cells contribute significantly to tumor heterogeneity (Augsten, 2014). While computational models for predicting tumor-stroma crosstalk from bulk-extraction sequencing experiments are under



Y Interleukin receptors Y Frizzled receptors Y TNF receptor Y Integrins Y Syndecans Y RTKs



**FIGURE 3 |** Inference of TAM-tumor crosstalk from scRNA-seq. Genes encoding ligands robustly expressed by at least 20% of TAMs with an average expression >2 CPM are paired with genes encoding their cognate receptors that are expressed in tumor cells. Each row represents a potential tumor-TAM interaction, bars represent the percentage of cells expressing each mRNA, colors indicate mean expression across cells.

development (Hackl et al., 2016), scRNA-seq also provides a powerful tool to infer paracrine-signaling networks. For example, in glioma, tumor associated macrophages/microglia (TAMs) are the most abundant immune infiltrate and can reach up to 30% of the total tumor mass (Cretu et al., 2005). By simply cross-referencing gene expression levels in single TAMs and neoplastic cells sequenced from SF10360 (Müller et al., 2016), with the receptor-ligand pairs from CCExplorer (Choi et al., 2015), one can infer a myriad of potential crosstalk (Figure 3). Here we see that TAMs express a variety of growth factors and growth-promoting cytokines, while neoplastic cells from the same sample express their cognate receptors. ScRNA-seq thus provides a powerful hypothesis-generating mechanism for paracrine-signaling studies.

## CONCLUSION

Recent advances in scRNA-seq have led to novel insights in cancer development, progression, metastasis, and drug-resistance, that were previously “veiled” by the mixing of cells intrinsic to standard bulk-sequencing experiments. Still, a variety of challenges go hand in hand with this rapid progress. For example, reliably distinguishing between neoplastic and infiltrating stromal/immune cells requires more than an analysis of transcriptional profiles alone. Analysis of expressed SNVs, CNVs, and other mutations from scRNA-seq can be used to filter stromal from neoplastic cells, and to map gene-expression signatures to putative tumor sub-clones. While most cancer scRNA-seq studies to date have focused on tumor cells, applications of scRNA-seq to paracrine-signaling studies of the tumor microenvironment are an exciting frontier. Therefore, scRNA-seq is a powerful tool for understanding the molecular

processes that govern one of the most difficult diseases of our time: Cancer.

## AUTHOR CONTRIBUTIONS

SM and AD wrote the manuscript. SM collected literature and generated figures with input from AD.

## FUNDING

This work has been supported by a Shurl and Kay Curci Foundation Research Grant, a UCSF Brain Tumor SPORE Career Development Award (P50-CA097257-13:7017), and a gift from the Dabbieri Family to AD.

## ACKNOWLEDGMENT

We would like to thank Ara Cho from our lab, who contributed to the literature acquisition.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fgene.2017.00073/full#supplementary-material>

**FIGURE S1** | Expression profiles support mutation-based cell classification. Expression of immune and tumor marker-genes (rows) in individual cells (columns). Clusters obtained via unsupervised hierarchical clustering are indicated by black bars.

## REFERENCES

- Alerting, E., Krumm, N., Sudmant, P. H., Ko, A., O’Roak, B. J., Malig, M., et al. (2012). Copy number variation detection and genotyping from exome sequence data. *Genome Res.* 22, 1525–1532. doi: 10.1101/gr.138115.112
- Augsten, M. (2014). Cancer-associated fibroblasts as another polarized cell type of the tumor microenvironment. *Front. Oncol.* 4:62. doi: 10.3389/fonc.2014.00062
- Bacher, R., and Kendzierski, C. (2016). Design and computational analysis of single-cell RNA-sequencing experiments. *Genome Biol.* 17:63. doi: 10.1186/s13059-016-0927-y
- Bai, H., Harmanc, A. S., Erson-Omay, E. Z., Li, J., Cokun, S., Simon, M., et al. (2015). Integrated genomic characterization of IDH1-mutant glioma malignant progression. *Nat. Genet.* 48, 59–66. doi: 10.1038/ng.3457
- Bass, A. J., Thorsson, V., Shmulevich, I., Reynolds, S. M., Miller, M., Bernard, B., et al. (2014). Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 513, 202–209. doi: 10.1038/nature13480
- Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 17:218.
- Bradshaw, A., Wickremsekera, A., Tan, S. T., Peng, L., Davis, P. F., and Itinteang, T. (2016). Cancer stem cell hierarchy in glioblastoma multiforme. *Front. Surg.* 3:21. doi: 10.3389/fsurg.2016.00021
- Brennan, C. W., Verhaak, R. G. W., McKenna, A., Campos, B., Noushmehr, H., Salama, S. R., et al. (2013). The somatic genomic landscape of glioblastoma. *Cell* 155, 462–477. doi: 10.1016/j.cell.2013.09.034
- Cancer Genome Atlas Research Network (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455, 1061–1068. doi: 10.1038/nature07385
- Cancer Genome Atlas Research Network (2013). Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 499, 43–49. doi: 10.1038/nature12222
- Cancer Genome Atlas Research Network (2014a). Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* 507, 315–322. doi: 10.1038/nature12965
- Cancer Genome Atlas Research Network (2014b). Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511, 543–550. doi: 10.1038/nature13385
- Cancer Genome Atlas Research Network, Linehan, W. M., Spellman, P. T., Ricketts, C. J., Creighton, C. J., Fei, S. S., et al. (2016). Comprehensive molecular characterization of papillary renal-cell carcinoma. *N. Engl. J. Med.* 374, 135–145. doi: 10.1056/NEJMoa1505917
- Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E., and Lappalainen, T. (2015). Tools and best practices for data processing in allelic expression analysis. *Genome Biol.* 16:195. doi: 10.1186/s13059-015-0762-6
- Ceccarelli, M., Barthel, F. P., Malta, T. M., Sabedot, T. S., Salama, S. R., Murray, B. A., et al. (2016). Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* 164, 550–563. doi: 10.1016/j.cell.2015.12.028
- Chapman, A. R., He, Z., Lu, S., Yong, J., Tan, L., Tang, F., et al. (2015). Single cell transcriptome amplification with MALBAC. *PLoS ONE* 10:e0120889. doi: 10.1371/journal.pone.0120889

- Chen, J., Li, Y., Yu, T.-S., McKay, R. M., Burns, D. K., Kernie, S. G., et al. (2012). A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature* 488, 522–526. doi: 10.1038/nature11287
- Choi, H., Sheng, J., Gao, D., Li, F., Durran, A., Ryu, S., et al. (2015). Transcriptome analysis of individual stromal cell populations identifies stroma-tumor crosstalk in mouse lung cancer model. *Cell Rep.* 10, 1187–1201. doi: 10.1016/j.celrep.2015.01.040
- Cretu, A., Fotos, J. S., Little, B. W., and Galileo, D. S. (2005). Human and rat glioma growth, invasion, and vascularization in a novel chick embryo brain tumor model. *Clin. Exp. Metastasis* 22, 225–236. doi: 10.1007/s10585-005-7889-x
- Daley, T., and Smith, A. D. (2013). Predicting the molecular complexity of sequencing libraries. *Nat. Methods* 10, 325–327. doi: 10.1038/nmeth.2375
- Daley, T., and Smith, A. D. (2014). Modeling genome coverage in single cell sequencing. *Bioinformatics* 30, 3159–3165. doi: 10.1093/bioinformatics/btu540
- Darmanis, S., Sloan, S. A., Zhang, Y., Enge, M., Caneda, C., Shuer, L. M., et al. (2015). A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. U.S.A.* 112, 7285–7290. doi: 10.1073/pnas.1507125112
- Dembinski, J. L., and Krauss, S. (2010). A distinct slow-cycling cancer stem-like subpopulation of pancreatic adenocarcinoma cells is maintained in Vivo. *Cancers* 2, 2011–2025. doi: 10.3390/cancers2042011
- Diaz, A., Liu, S. J., Sandoval, C., Pollen, A., Nowakowski, T. J., Lim, D. A., et al. (2016). SCell: integrated analysis of single-cell RNA-seq data. *Bioinformatics* 32, 2219–2220. doi: 10.1093/bioinformatics/btw201
- El-Kebir, M., Oesper, L., Acheson-Field, H., and Raphael, B. J. (2015). Reconstruction of clonal trees and tumor composition from multi-sample sequencing data. *Bioinformatics* 31, i62–i70. doi: 10.1093/bioinformatics/btv261
- Favero, F., Joshi, T., Marquard, A. M., Birkbak, N. J., Krzystanek, M., Li, Q., et al. (2015). Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann. Oncol.* 26, 64–70. doi: 10.1093/annonc/mdl479
- Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A. K., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* 16:278. doi: 10.1186/s13059-015-0844-5
- Furnari, F. B., Cloughesy, T. F., Cavenee, W. K., and Mischel, P. S. (2015). Heterogeneity of epidermal growth factor receptor signalling networks in glioblastoma. *Nat. Rev. Cancer* 15, 302–310. doi: 10.1038/nrc3918
- Garmire, L., Poirion, O. B., Zhu, X., and Ching, T. (2016). Single-cell transcriptomics bioinformatics and computational challenges. *Front. Genet.* 7:163.
- Gerber, T., Willscher, E., Loeffler-Wirth, H., Hopp, L., Schadendorf, D., Schartl, M., et al. (2017). Mapping heterogeneity in patient-derived melanoma cultures by single-cell RNA-seq. *Oncotarget* 8, 846–862. doi: 10.18632/oncotarget.13666
- Gerlinger, M., Rowan, A. J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., et al. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* 366, 883–892. doi: 10.1056/NEJMoa1113205
- Ghaffari, S. (2011). Cancer, stem cells and cancer stem cells: old ideas, new developments. *F1000 Med. Rep.* 3:23. doi: 10.3410/M3-23
- Greaves, M., and Maley, C. C. (2012). Clonal evolution in cancer. *Nature* 481, 306–313. doi: 10.1038/nature10762
- Grün, D., and Van Oudenaarden, A. (2015). Design and analysis of single-cell sequencing experiments. *Cell* 163, 799–810. doi: 10.1016/j.cell.2015.10.039
- Hackl, H., Charoentong, P., Finotello, F., and Trajanoski, Z. (2016). Computational genomics tools for dissecting tumour-immune cell interactions. *Nat. Rev. Genet.* 17, 441–458. doi: 10.1038/nrg.2016.67
- Hoadley, K. A., Yau, C., Wolf, D. M., Cherniack, A. D., Tamborero, D., Ng, S., et al. (2014). Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* 158, 929–944. doi: 10.1016/j.cell.2014.06.049
- Hou, Y., Guo, H., Cao, C., Li, X., Hu, B., Zhu, P., et al. (2016). Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res.* 26, 304–319. doi: 10.1038/cr.2016.23
- Kim, K.-T., Lee, H. W., Lee, H.-O., Kim, S. C., Seo, Y. J., Chung, W., et al. (2015). Single-cell mRNA sequencing identifies subclonal heterogeneity in anti-cancer drug responses of lung adenocarcinoma cells. *Genome Biol.* 16:127. doi: 10.1186/s13059-015-0692-3
- Kim, K.-T., Lee, H. W., Lee, H.-O., Song, H. J., Jeong, D. E., Shin, S., et al. (2016). Application of single-cell RNA sequencing in optimizing a combinatorial therapeutic strategy in metastatic renal cell carcinoma. *Genome Biol.* 17:80. doi: 10.1186/s13059-016-0945-9
- Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C., and Teichmann, S. A. (2015). The technology and biology of single-cell RNA sequencing. *Mol. Cell* 58, 610–620. doi: 10.1016/j.molcel.2015.04.005
- Lawson, D. A., Bhakta, N. R., Kessenbrock, K., Prummel, K. D., Yu, Y., Takai, K., et al. (2015). Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature* 526, 131–135. doi: 10.1038/nature15260
- Li, B., Senbabaoglu, Y., Peng, W., Yang, M.-L., Xu, J., and Li, J. Z. (2012). Genomic estimates of aneuploid content in glioblastoma multiforme and improved classification. *Clin. Cancer Res.* 18, 5595–5605. doi: 10.1158/1078-0432.CCR-12-1427
- 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
- Li, W., Calder, R. B., Mar, J. C., and Vijg, J. (2015). Single-cell transcriptogenomics reveals transcriptional exclusion of ENU-mutated alleles. *Mutat. Res.* 772, 55–62. doi: 10.1016/j.mrfmm.2015.01.002
- Macaulay, I. C., and Voet, T. (2014). Single cell genomics: advances and future perspectives. *PLoS Genet.* 10:e1004126. doi: 10.1371/journal.pgen.1004126
- Min, J. W., Kim, W. J., Han, J. A., Jung, Y. J., Kim, K. T., Park, W. Y., et al. (2015). Identification of distinct tumor subpopulations in lung adenocarcinoma via single-cell RNA-seq. *PLoS ONE* 10:e0135817. doi: 10.1371/journal.pone.0135817
- Müller, S., Liu, S. J., Di Lullo, E., Malatesta, M., Pollen, A. A., Nowakowski, T. J., et al. (2016). Single-cell sequencing maps gene expression to mutational phylogenies in PDGF- and EGF-driven gliomas. *Mol. Syst. Biol.* 12, 889. doi: 10.15252/msb.20166969
- Müller, S., Raulefs, S., Bruns, P., Afonso-Grunz, F., Plötner, A., Thermann, R., et al. (2015). Next-generation sequencing reveals novel differentially regulated mRNAs, lncRNAs, miRNAs, sdrRNAs and a piRNA in pancreatic cancer. *Mol. Cancer* 14, 94. doi: 10.1186/s12943-015-0358-5
- Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., et al. (2011). Tumour evolution inferred by single-cell sequencing. *Nature* 472, 90–94. doi: 10.1038/nature09807
- 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
- Peña-Llopis, S., and Brugarolas, J. (2013). Simultaneous isolation of high-quality DNA, RNA, miRNA and proteins from tissues for genomic applications. *Nat. Protoc.* 8, 2240–2255. doi: 10.1038/nprot.2013.141
- Piskol, R., Ramaswami, G., and Li, J. B. (2013). Reliable identification of genomic variants from RNA-seq data. *Am. J. Hum. Genet.* 93, 641–651. doi: 10.1016/j.ajhg.2013.08.008
- Pollen, A. A., Nowakowski, T. J., Shuga, J., Wang, X., Leyrat, A. A., Lui, J. H., et al. (2014). Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.* 32, 1053–1058. doi: 10.1038/nbt.2967
- Risso, D., Ngai, J., Speed, T. P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* 32, 896–902. doi: 10.1038/nbt.2931
- Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33, 495–502. doi: 10.1038/nbt.3192
- Saunders, N. A., Simpson, F., Thompson, E. W., Hill, M. M., Endo-Munoz, L., Leggett, G., et al. (2012). Role of intratumoural heterogeneity in cancer drug resistance: molecular and clinical perspectives. *EMBO Mol. Med.* 4, 675–684. doi: 10.1002/emmm.201101131
- Shackleton, M., Quintana, E., Fearon, E. R., and Morrison, S. J. (2009). Heterogeneity in cancer: cancer stem cells versus clonal evolution. *Cell* 138, 822–829. doi: 10.1016/j.cell.2009.08.017
- Shapiro, E., Biezuner, T., and Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* 14, 618–630. doi: 10.1038/nrg3542

- Stegle, O., Teichmann, S. A., and Marioni, J. C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* 16, 133–145. doi: 10.1038/nrg3833
- Svensson, V., Natarajan, K. N., Ly, L.-H., Miragaia, R. J., Labalette, C., Macaulay, I. C., et al. (2016). Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* 14, 381–387. doi: 10.1038/nmeth.4220
- Tanay, A., and Regev, A. (2017). Scaling single-cell genomics from phenomenology to mechanism. *Nature* 541, 331–338. doi: 10.1038/nature21350
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6, 377–382. doi: 10.1038/nmeth.1315
- Ting, D. T., Wittner, B. S., Ligorio, M., Vincent Jordan, N., Shah, A. M., Miyamoto, D. T., et al. (2014). Single-cell RNA sequencing identifies extracellular matrix gene expression by pancreatic circulating tumor cells. *Cell Rep.* 8, 1905–1918. doi: 10.1016/j.celrep.2014.08.029
- Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., et al. (2016a). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352, 189–196. doi: 10.1126/science.aad0501
- Tirosh, I., Venteicher, A. S., Hebert, C., Escalante, L. E., Patel, A. P., Yizhak, K., et al. (2016b). Single-cell RNA-seq supports a developmental hierarchy in human oligodendrogloma. *Nature* 539, 309–313. doi: 10.1038/nature20123
- Tung, P.-Y., Blischak, J. D., Hsiao, C. J., Knowles, D. A., Burnett, J. E., Pritchard, J. K., et al. (2017). Batch effects and the effective design of single-cell gene expression studies. *Sci. Rep.* 7:39921. doi: 10.1038/srep39921
- Turajlic, S., McGranahan, N., and Swanton, C. (2015). Inferring mutational timing and reconstructing tumour evolutionary histories. *Biochim. Biophys. Acta* 1855, 264–275. doi: 10.1016/j.bbcan.2015.03.005
- Venteicher, A. S., Tirosh, I., Hebert, C., Yizhak, K., Neftel, C., Filbin, M. G., et al. (2017). Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* 355:eaai8478. doi: 10.1126/science.aai8478
- Wagner, A., Regev, A., and Yosef, N. (2016). Revealing the vectors of cellular identity with single-cell genomics. *Nat. Biotechnol.* 34, 1145–1160. doi: 10.1038/nbt.3711
- Wang, J., Cazzato, E., Ladewig, E., Frattini, V., Rosenbloom, D. I. S., Zairis, S., et al. (2016). Clonal evolution of glioblastoma under therapy. *Nat. Genet.* 48, 768–776. doi: 10.1038/ng.3590
- Wang, L., Ni, X., Covington, K. R., Yang, B. Y., Shiu, J., Zhang, X., et al. (2015). Genomic profiling of Sézary syndrome identifies alterations of key T cell signaling and differentiation genes. *Nat. Genet.* 47, 1426–1434. doi: 10.1038/ng.3444
- Wang, Q., Hu, X., Hu, B., Muller, F., Kim, H., Squatrito, M., et al. (2016). Tumor evolution of glioma intrinsic gene expression subtype associates with immunological changes in the microenvironment. *Neuro Oncol* 18(Suppl. 6), vi202. doi: 10.1093/neuonc/now212.854
- Winterhoff, B. J., Maile, M., Mitra, A. K., Sebe, A., Bazzaro, M., Geller, M. A., et al. (2017). Single cell sequencing reveals heterogeneity within ovarian cancer epithelium and cancer associated stromal cells. *Gynecol. Oncol.* 144, 598–606. doi: 10.1016/j.ygyno.2017.01.015
- Witkiewicz, A. K., McMillan, E. A., Balaji, U., Baek, G., Lin, W.-C., Mansour, J., et al. (2015). Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nat. Commun.* 6:6744. doi: 10.1038/ncomms7744
- Woodworth, M. B., And, K. M. G., and Walsh, C. A. (2017). Building a lineage from single cells: genetic techniques for cell lineage tracking. *Nat. Rev. Genet.* 18, 230–244. doi: 10.1038/nrg.2016.159
- Wu, A. R., Neff, N. F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M. E., et al. (2014). Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* 11, 41–46. doi: 10.1038/nmeth.2694
- Yip, S., Butterfield, Y. S., Morozova, O., Chittaranjan, S., Blough, M. D., An, J., et al. (2012). Concurrent CIC mutations, IDH mutations, and 1p/19q loss distinguish oligodendrogliomas from other cancers. *J. Pathol.* 226, 7–16. doi: 10.1002/path.2995
- Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-García, W., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 4:2612. doi: 10.1038/ncomms3612
- Zack, T. I., Schumacher, S. E., Carter, S. L., Cherniack, A. D., Saksena, G., Tabak, B., et al. (2013). Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* 45, 1134–1140. doi: 10.1038/ng.2760
- Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., et al. (2017). Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* 65, 631–643. doi: 10.1016/j.molcel.2017.01.023
- Zong, C., Lu, S., Chapman, A., and Xie, X. S. (2012). Genome-wide detection of single nucleotide and copy number variations of a single human cell. *Science* 338, 1622–1626. doi: 10.1126/science.1229164

**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 © 2017 Müller and Diaz. 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) or licensor 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.