Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

Computational approaches for inferring tumor evolution from single-cell genomic data

Hamim Zafar[1,2], Nicholas Navin[2,3], Luay Nakhleh[1] and Ken Chen[2]

Abstract

Genomic heterogeneity in tumors results from mutations and selection of high-fitness single cells, the operational components of evolution. Precise knowledge about mutational heterogeneity and evolutionary trajectory of a tumor can provide useful insights into predicting cancer progression and designing personalized treatment. The rapidly advancing field of single-cell genomics provides an opportunity to study tumor heterogeneity and evolution at the ultimate level of resolution. In this review, we present an overview of the state-of-the-art single-cell DNA sequencing methods, technical errors that are inherent in the resulting large-scale datasets, and computational methods to overcome these errors. Finally, we discuss the computational and mathematical approaches for understanding intratumor heterogeneity and cancer evolution at the resolution of a single cell.

Addresses

[1] Department of Computer Science, Rice University, Houston, TX, USA
[2] Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
[3] Department of Genetics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Corresponding author: Chen, Ken (kchen3@mdanderson.org)

Current Opinion in Systems Biology 2018, **7**:16–25

This review comes from a themed issue on **Genomics and epigenomics (2018)**

Edited by **Raul Rabadan**

For a complete overview see the Issue and the Editorial

Available online 6 December 2017

https://doi.org/10.1016/j.coisb.2017.11.008

2452-3100/© 2017 Elsevier Ltd. All rights reserved.

Keywords

Single-cell, Genomics, DNA sequencing, Variant detection, Phylogenetics, Intra-tumor heterogeneity, Tumor evolution.

Introduction

Cancer is a disease emerging from a single cell in the somatic tissue and is driven by a complex interplay of somatic mutations, copy number alterations (CNAs) and chromosomal rearrangements [1,2]. As a tumor progresses, diverse genomic aberrations give rise to genetically heterogeneous subpopulations (clones) of cells interacting with each other in a Darwinian framework of mutations, fitness and selection [3–5]. Intratumor heterogeneity (ITH) complicates the diagnosis and treatment of cancer patients and causes relapse and drug resistance [6–8]. The emergence of next-generation sequencing (NGS) technologies enabled a thorough analysis of tumor heterogeneity through the generation of large-scale quantitative genomic datasets [9–11]. However, despite these advances, a comprehensive understanding of ITH has proved elusive thus far [12,13].

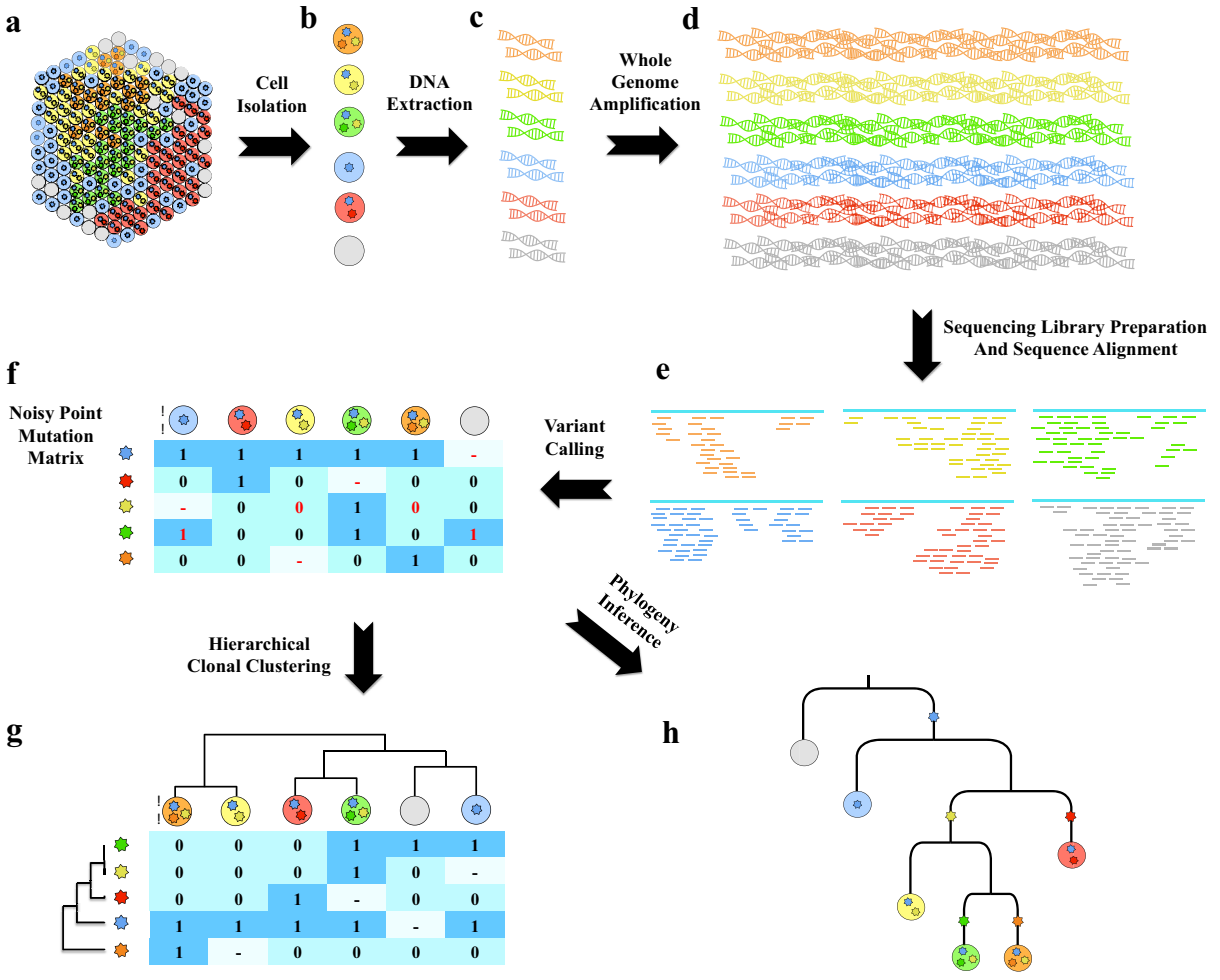
Bulk high-throughput sequencing has been the technology of choice for studying heterogeneity and tumor evolution [14,15]. Subpopulations are computationally inferred [16–22] from variant allele frequencies (VAFs) of mutations detected in bulk DNA that consists of an admixture of DNA from millions of cells in a cancer tissue. VAFs, however, provide a noisy signal for deconvoluting heterogeneity [23,24] and cannot reliably reconstruct rare subclones, or subclones having similar frequencies in the tumor mass. The single-sample approach of bulk sequencing is augmented in multi-region sequencing through which multiple samples obtained from different geographical regions of a tumor are analyzed [25–28]. Although multi-region sequencing can reveal geographically segregated subpopulations, resolving spatially intermixed subclones remains difficult and this approach still relies on deconvolution of subclones for phylogeny inference [29].

The emergence of single-cell DNA sequencing (SCS) technologies has enabled sequencing of individual cancer cells, providing the highest-resolution of the mutational histories of cancer [23,30]. SCS aims to further our knowledge of different aspects of cancer biology including resolving clonal substructure, tracing tumor evolution, identifying rare subclones and understanding the role of cancer microenvironment in tumor progression [23,24,31]. In this review, we discuss the state of the art of SCS technologies, technical challenges and computational approaches to overcome those, and finally, approaches for understanding ITH and tumor evolution from SCS data.

An overview of single-cell DNA sequencing methods

Figure 1 illustrates the steps of a single-cell DNA sequencing study. The first step in producing high-quality SCS data is the isolation of individual cells. Early experiments used techniques such as serial [32] or microwell dilution [33], micropipetting [34], laser-

Figure 1



Overview of single-cell DNA sequencing analysis: from isolation of single cells from a tissue to inference of subclones and phylogeny. (a) Illustration of a heterogeneous tumor tissue, different colors of the cells signify the membership of the cells to different subclones. The mutations present in each cell are represented by the small stars in it. (b–h) steps performed to conduct a heterogeneity or phylogeny analysis of single-cell DNA sequencing data. (b) First, single cells are isolated from tissue and (c) DNA is extracted from each cell. (d) Whole genome amplification (WGA) is performed on DNA extracted from each cell to produce the amount of DNA required for constructing a sequencing library for each individual cell. Various WGA methods are summarized in Refs. [23,75]. (e) Whole-exome or targeted sequencing libraries are constructed depending on the need of the study and sequenced reads are aligned to a reference genome. (f) Variant-calling [82,84,88] is performed on the sequencing library of single cells. Only single-nucleotide variant (SNV) profiles are illustrated here. Copy-number profiles for each cell can also be obtained and utilized in the subsequent steps. (g) Subclones are inferred by clustering [40,93] the cells into different populations. (h) Tumor phylogeny is inferred computationally [108,109,115] from the SNV profiles of single cells. The phylogeny also shows the order of mutations during the evolutionary history of the tumor.

capture microdissection (LCM) [35] to isolate cells from a solid tissue. Several methods [36,37] opted for isolation of single nuclei that remain intact in frozen samples. Later, flow-assisted cell sorting (FACS) [38,39] and microfluidics-based approaches [40] resulted in higher throughput. Scalability to thousands of cells came from barcoding methods [41,42] and single-nucleus DNA repair enabled sequencing of formalin-fixed paraffin-embedded (FFPE) tumor samples [43]. Commercial systems such as CellSearch [44], Magsweeper [45], DEP-Array system [46], CellCelector [47] have been used for the more challenging task of isolating circulating tumor cells (CTCs) and disseminated tumor cells (DTCs).

SCS was made possible by the development whole-genome amplification (WGA) methods that can amplify the 6 pg of DNA in a single-cell genome with a factor of 10^3 to 10^9 [48] to meet the amount of DNA (nanograms-micrograms) required for constructing a sequencing library. Three broad categories (PCR-based, isothermal and hybrid) of WGA methods exist with different advantages and limitations [48–50] (Table 1).

The technical artifacts associated with WGA methods limit the application of SCS. Use of G2/M cells [63] or performing cell lysis and DNA denaturation on ice [64] has improved some of these technical problems. Subsequently, multiplexing approaches coupled with

Table 1

Overview of whole genome amplification (WGA) methods.

Amplification method	DOP-PCR [38,54–56]	MDA [57–63]	MALBAC [34]
Type of amplification	PCR-based	isothermal	Hybrid
Advantages[a]	Uniform coverage	High physical coverage	Uniform coverage
Associated technical artifacts	High FP error, ADO	FP error, ADO, chimeric molecules	High FP error, ADO
Disadvantages[a]	Low physical coverage Over-amplification of focal genomic regions	Non-uniform coverage	Low physical coverage Higher false-positive rate
Application	CNA analysis	SNV analysis	CNA analysis

ADO: allelic dropout; CNA: copy number alterations; DOP-PCR: degenerate oligonucleotide primed PCR; FP: False positive; MALBAC: multiple annealing and looping-based amplification cycles; MDA: multiple displacement amplification; PCR: polymerase chain reaction; SNV: single nucleotide variant.
[a] Advantages and disadvantages are based on [31,48,51–53].

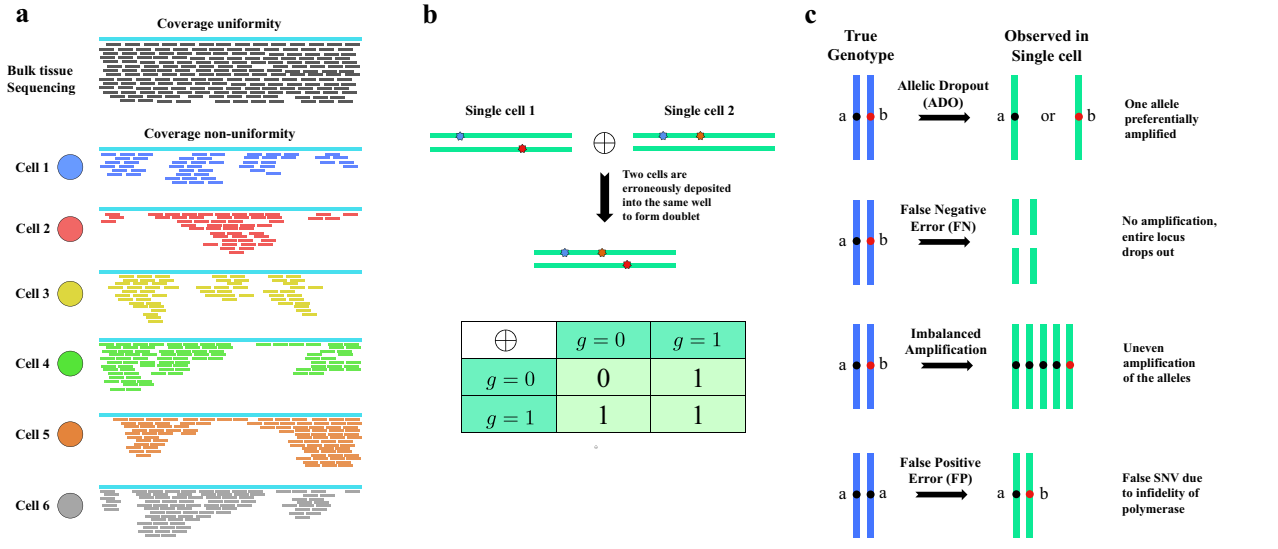
nanolitre volume reactions with microfluidics has led to higher scalability and lower bias [65–68]. WGA based on transposon insertion and *in vitro* transcription [69] has resulted in linear amplification, higher genome coverage suitable for SNV analysis. A direct library preparation (DLP) protocol [70] combines microfluidics with transposition reactions to increase throughput and decrease WGA artifacts such as pile-up regions for copy number analysis. Transposase-based combinatorial indexing [71] is an innovative approach that utilizes two rounds of tagmentation-based barcoding and pooling to vastly scale SCS to thousands of cells using flow-sorting. These methods have begun to address technical errors and throughput issues associated with SCS, however much progress is still needed to further improve these experimental approaches.

Single-cell sequencing errors

Different technical artifacts introduced during the single-cell DNA sequencing workflow may introduce noise into the datasets, confounding bioinformatics analysis (Figure 2). Inadvertent isolation of DNA from multiple cells violates the basic assumption of the methods designed for analyzing single-cell data resulting in spurious biological conclusions [72]. Specifically, presence of 'cell doublets' is a persisting error (ranging from 1% [38,42,56] to 10% [60,61,73]), in which more than one cell is deposited into the same well for analysis.

WGA can add significant technical errors in SCS data, including coverage bias, allelic dropout (ADO), false-positive (FP) and false-negative (FN) errors [23]. ADOs are observed when one of the alleles in a

Figure 2



Technical errors in single-cell DNA sequencing data. (a) Coverage in single cells can vary from one cell to another compared to the coverage uniformity observed in bulk tissue sequencing data. (b) Doublets occur when two cells are erroneously deposited into the same well. Doublets result in merged genotype as illustrated here. Single cell 1 (blue and red mutations) and Single cell 2 (blue and orange mutations) created a doublet with three mutations (blue, red and orange). The matrix shows the binary operator to compute the expected genotype after combining two genotypes. (c) Errors introduced during WGA include: allelic dropout events, false-negative (FN) errors, uneven amplification of the alleles and false-positive (FP) errors. The technical errors have been reviewed in Refs. [23,75].

heterozygous mutation (ab) is unevenly amplified, resulting in a homozygous genotype (aa or bb). Infidelity of polymerase enzymes and deamination of cytosine bases introduce FP errors [23,69] that occur randomly across the genome [74]. FP errors can hinder the subclonal reconstruction, particularly when the number of FP errors exceeds the number of true somatic variants [75]. FN errors occur due to insufficient coverage when mutated loci are unamplified.

When analyzed, technical errors in SCS data must be systematically removed by filtering or carefully accounted for to distinguish signal from noise so that true biological variations can be identified and utilized in downstream analyses.

Variant calling from single cells

Detection of copy number variants from SCS data commonly involves a variable binning method where the genome is divided into bins and the read count in each bin represents whether the region is over- or under-represented compared to a diploid genome [38,56]. Loess normalization is applied for correcting bias due to GC content and circular binary segmentation (CBS) [76] is used to segment the copy number profiles. Specific algorithms account for technical artifacts introduced by WGA [77,78]. A recent study [70] extends a hidden Markov model-based approach, HMMcopy [79], for inferring single-cell copy number profiles. The study in Ref. [80] compares performance of HMM-based and CBS-based approaches for detecting megabase-scale somatic CNAs from single cells and also introduces an approach for robust, specific detection of CNAs exceeding 5 Mb. Another HMM-based approach [34] has been used by Ref. [81] to infer CNAs from primary tumor cells and CTCs. Most of the CNA detection methods rely on coverage uniformity across the genome, rendering DOP-PCR and MALBAC more suitable than MDA.

SCS has major applications for uncovering SNVs in individual cells to understand tumor heterogeneity at base-pair resolution. Earlier SCS studies [60–62] relied on traditional NGS variant callers for SNV calling. However, standard variant callers [82–84] have difficulty in addressing the unique error profiles of SCS data and therefore report a large number of FP and FN calls. Custom filters [63], consensus-based approaches (variants observed in more than one cell) [85] and reference bulk samples [40] have been used to remove FPs. The Bayes calibration for SNV calling in Ref. [86] employs allelic dropout (ADO) in a naïve statistical model. The BayesHammer approach [87] proposes clustering of reads to account for non-uniform coverage. A SCS specific SNV caller, Monovar [88], accounts for FP and ADO errors while quantifying the likelihood of the underlying genotype states in a single cell. Multi-sample probabilistic approach to overcome coverage non-uniformity and statistical modeling of the WGA errors allows Monovar to outperform bulk SNV callers. Recently introduced SCcaller [64] extends the model of MuTect [84] and reduces the number of FPs in somatic mutation calling from SCS. However, being a single-sample approach, SCcaller is likely to have issues from coverage non-uniformity. An important problem in SCS data is to impute mutations that do not have sufficient coverage in a single cell. Information from neighboring sites as well as from the same site of other cells can be incorporated to impute these missing data. A unified framework that uses both bulk sample and single cells can potentially improve single-cell SNV calling. Phylogenetic relationship between the cells can also be utilized in SNV calling [72].

Subclonal reconstruction from single cells

Variants detected from single cells are used to infer clonal subpopulations. Dimensionality reduction techniques such as PCA [89] and multidimensional scaling [90] have been used to infer monoclonality [60] or polyclonality [63] of a tumor. Hierarchical clustering has been applied on CNV profiles [70,91] as well as SNV profiles [62,63,86,92] to uncover the clonal composition in a tumor. Failure to account for errors in variant calling can result in spurious clustering. To overcome the limitations of naïve approaches such as hierarchical clustering that weigh error rates equally [40], employs a binomial mixture model for clustering cells and reports varying numbers of subclones for 6 AML patients. A variational Bayes inference method, SCG [93], extends the clustering in Ref. [40] to incorporate errors due to doublets and ADO events. SCG was used to infer the clonal genotypes of 1680 single cells for three high-grade serous ovarian cancer patients [94]. A non-parametric Bayesian clustering method, ddClone [95], unifies SCS and bulk data in a probabilistic framework for improved clonal cluster inference. However, the clustering approaches in Refs. [40,93,95] are phylogeny-naïve and the genealogical relationship between the clusters are not utilized.

Reconstruction of phylogeny from single cells

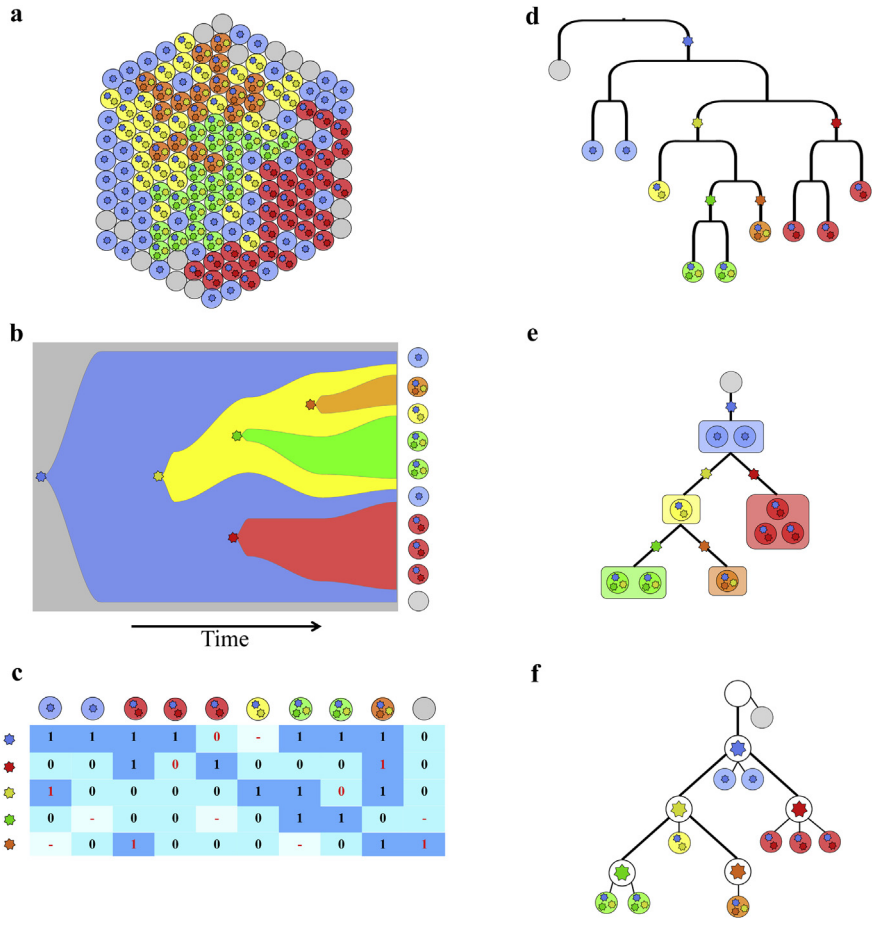
One of the major applications of SCS is to study tumor evolution via the inference of phylogeny, a binary genealogical tree along which the tumor cells evolve. Even though concepts borrowed from population genetics such as selection and fitness are useful in the context of tumor evolution [96], many concepts (e.g., meiotic recombination, sexual selection) do not apply to tumors [4]. The presence of technical artifacts further inhibits a straightforward use of classical phylogeny inference methods on SCS datasets. As a result, development of phylogeny inference algorithms has become an independent field within single-cell bioinformatics.

In an early approach [97], a reference-free k-mer-based method was proposed to infer tumor phylogeny that represented the evolution of copy number profiles. However, this approach was tested only on one of the earliest SCS datasets [38]. A recent study [91] inferred punctuated copy number evolution in breast cancer patients using parsimony-based classical phylogenetic method [98]. Distance-based phylogeny method, such as neighbor-joining (NJ) [99] has also been used to infer phylogeny in breast cancer [63] and convergent evolution from copy number profiles of primary and circulating tumor cells [81].

Early SCS studies [61,86] applied classical phylogenetic methods NJ [99] and UPGMA [100] for phylogeny inference from single-cell somatic SNV profiles. Classical maximum-likelihood-based approach and Bayesian phylogenetic approach [101] were used to infer phylogenies for AML patients [102] and in a study involving primary tumors and derived xenograft lines [103] respectively. However, none of these methods accounted for technical artifacts and might have resulted in spurious inferences.

Following the trend for bulk sequencing data, infinite-sites assumption, ISA (no site mutates more than once) has also been applied in the context of SCS to infer a perfect phylogeny [104]. As pointed in Ref. [105], three different representations of tumor phylogenies can be obtained from SCS data (Figure 3). Methods for single-cell phylogenetic inference [106–109] rely on probabilistic approaches to account for uncertainties due to

Figure 3



Tumor phylogeny inference from single-cell sequencing data. (a) Illustration of a heterogeneous tumor tissue that evolved according to the clonal expansion represented in (b). (b) Schematic representation of the evolution of clones with time. The clonal populations are visualized using fishplot [131]. Ten single cells (different colors represent membership to corresponding subpopulations) are sequenced from the tumor. The grey colored cell is an unmutated normal cell and the rests of the nine cells contain mutations denoted by stars. The cells evolved along the branches of a binary genealogical tree. (c) The noisy SNV profile of ten single cells. The red entries in the matrix denote FP, FN or missing entries (denoted by '-') that occur due to errors in SCS. (d–f) Types of trees inferred in single-cell tumor phylogenetics. (d) A phylogenetic tree with the single cells at the leaves. By focusing on the genealogical relationship between the cells it resembles the binary genealogical tree along which the cells evolve. SiFit [115] infers such a tree. (e) Clonal lineage tree that focuses on the genealogical relationship between the subclones. The nodes represent subclones (cluster of cells) and mutations are placed on the branches. OncoNEM [108] infers such a tree. (f) Mutation tree that shows chronological order of mutations. The internal nodes represent mutations and the cells are attached as leaves. SCITE [109] infers such a tree.

technical artifacts and infer one of the phylogeny representations with the underlying ISA. The first approach [106] constructs a directed weighted (posterior probability) graph based on the maximal posterior pairwise ordering for each pair of mutations and then infers the maximum spanning tree as the maximum-likelihood mutation tree. Empirical Bayes estimate of the prior probabilities of the mutation orders may result in a suboptimal tree [72]. BitPhylogeny [107] infers a clonal tree, where single cells are clustered into clones. BitPhylogeny uses a tree-structured stick-breaking process to define a prior distribution on the number of clones and employs the Markov chain Monte Carlo (MCMC) scheme of [110] to explore the search space of all trees with an arbitrary number of clones. However, the technical artifacts of SCS are insufficiently modeled in BitPhylogeny. More recent methods, OncoNEM [108] and SCITE [109] model both FPs and FNs in the likelihood calculation of a tree. However, OncoNEM infers a clonal tree and marginalizes over the placement of mutations along the edges, whereas SCITE infers a mutation tree and marginalizes over the attachment of cells to mutation nodes. OncoNEM relies on a greedy search and cell clustering to infer the maximum-likelihood tree. SCITE employs an MCMC algorithm to explore the tree space and can also infer the maximum-likelihood mutation tree. Both of these methods can also estimate error rates from the data assuming ISA. OncoNEM when applied on published datasets [60,62], corroborated the findings of the original studies. SCITE has been applied to reanalyze previous datasets [60,61,63] as well as to infer the phylogeny for new metastatic colorectal cancer datasets [111].

While ISA is a convenient assumption to restrict the search space, there is considerable ambiguity about how realistic it is in the context of tumor evolution. Copy number variants and aneuploidy are very common events in cancer and can violate ISA. Events such as chromosomal deletion and loss of heterozygosity (LOH) can result in a mutation loss [72,105], a possibility that is completely excluded by ISA. For bulk sequencing ovarian cancer data, a phylogenetic method based on stochastic Dollo process [112,113] (allows for mutation loss due to deletion and LOH) detected evidence for convergent evolution where different CNAs affected the same genomic locus [94]. Convergent evolution has been detected in other multi-region sequencing studies as well [25,28]. Considering these events during phylogeny reconstruction can uncover histories that are completely excluded by ISA. A statistical test on SCS SNV datasets shows a potential violation of ISA [114]. A new phylogeny reconstruction method, SiFit [115] introduces a finite-site evolution model to account for mutation recurrence and losses and reports higher-likelihood phylogenies for colon cancers [92,111] compared to ISA-based methods [108,109]. Another approach for accounting for convergent evolution is to use a directed acyclic graph that allows for the presence of confluent trajectories [116].

Conclusion & future directions

In conclusion, single-cell genomics is a promising new method that can improve many facets of cancer research, by illuminating tumor initiation, metastasis and therapy resistance. In the clinic, these tools are likely to have important applications in early detection, non-invasive monitoring and personalized therapy. However, significant challenges still remain and will need to be overcome before clinical applications can truly be realized. Even though the error rates of SCS datasets have continued to improve over the last five years, there is still much room for further improvement.

To date, subclonal or phylogeny reconstruction has been performed based on a single type of variant (CNA or SNV). Following the trend in bulk data, it is very important to consider them jointly. Novel modeling frameworks have to be developed given the evidence that CNAs and SNVs may follow different evolutionary model [96]. Such analysis will also require datasets where both CNAs and SNVs can be inferred from the same single cell. Even though initial attempts have already been made [69], more datasets are needed and will become more widely available as these methods become adopted by the research community.

There is considerable ambiguity regarding the number of single-cell samples that are required for delineating the clonal substructure of a tumor or its phylogenetic lineage. The obvious solution to sequence more cells comes at a higher cost with no guarantee to sample cells from rare subclones. Integrating bulk and single cells from the same tissue can provide complementary knowledge in such scenario. The daunting task of reconstructing phylogeny from a small number of mutations for targeted SCS datasets can be facilitated by an informed prior distribution (contributed by linear [117], branching [40,28,118], neutral [119,120] and punctuated [91,121] model of evolution) on the shape of the tumor tree in a Bayesian framework. To date, most computational approaches have treated variant calling, subclonal reconstruction and phylogeny inference from SCS separately, a unified approach to solve them as a single problem may reveal novel insights.

Finally, integrating multiple data types [122] will likely be powerful in elucidating our understanding of cancer biology. Information about DNA and RNA from the same cell can greatly improve in the validation of findings as genotypic changes can be correlated with phenotypic changes [30,75]. Due to technological advances, such multiomic analyses have recently been demonstrated [123–125] and are providing useful insights into the heterogeneity and evolution of cancer.

Simultaneous analysis of DNA, RNA and proteome [126,127], transcriptome and epigenome [128], transcriptome and methylome [129], chromatin state and methylome [130] can reveal different dimensions of genomic landscape culminating a more comprehensive understanding of cancer progression.

Funding

The study was supported by the National Cancer Institute (grant R01 CA172652 to KC), the NCI-Designated cancer center support grant to MD Anderson cancer center (P30 CA016672), and the Andrew Sabin Family Foundation. This work was supported by grants to NN from NCI (1RO1CA169244-01) and the Chan-Zuckerberg Foundation (HCA-A-1704-01668).

References

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Stratton MR, Campbell PJ, Futreal PA: **The cancer genome**. *Nature* 2009, **458**:719–724.
2. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW: **Cancer genome landscapes**. *Science* 2013, **339**: 1546–1558.
3. Nowell PC: **The clonal evolution of tumor cell populations**. *Science* 1976, **194**:23–28.
4. Merlo LMF, Pepper JW, Reid BJ, Maley CC: **Cancer as an evolutionary and ecological process**. *Nat Rev Cancer* 2006, **6**: 924–935.
5. Yates LR, Campbell PJ: **Evolution of the cancer genome**. *Nat Rev Genet* 2012, **13**:795–806.
6. Greaves M, Maley CC: **Clonal evolution in cancer**. *Nature* 2012, **481**:306–313.
7. Gillies RJ, Verduzco D, Gatenby RA: **Evolutionary dynamics of carcinogenesis and why targeted therapy does not work**. *Nat Rev Cancer* 2012, **12**:487–493.
8. Burrell RA, McGranahan N, Bartek J, Swanton C: **The causes and consequences of genetic heterogeneity in cancer evolution**. *Nature* 2013, **501**:338–345.
9. McLendon R, Friedman A, Bigner D, Van Meir EG, Brat DJ, Mastrogianakis GM, Olson JJ, *et al.*: **Comprehensive genomic characterization defines human glioblastoma genes and core pathways**. *Nature* 2008, **455**:1061–1068.
10. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, Raine K, Jones D, Marshall J, Ramakrishna M, *et al.*: **The life history of 21 breast cancers**. *Cell* 2012, **149**:994–1007.
11. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, *et al.*: **Mutational landscape and significance across 12 major cancer types**. *Nature* 2013, **502**:333–339.
12. Glickman MS, Sawyers CL: **Converting cancer therapies into cures: lessons from infectious diseases**. *Cell* 2012, **148**: 1089–1098.
13. Korolev KS, Xavier JB, Gore J: **Turning ecology and evolution against cancer**. *Nat Rev Cancer* 2014, **14**:371–380.
14. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, Turashvili G, Ding J, Tse K, Haffari G, *et al.*: **The clonal and mutational evolution spectrum of primary triple-negative breast cancers**. *Nature* 2012, **486**:395–399.
15. Landau DA, Tausch E, Taylor-Weiner AN, Stewart C, Reiter JG, Bahlo J, Kluth S, Bozic I, Lawrence M, Böttcher S, *et al.*: **Mutations driving CLL and their evolution in progression and relapse**. *Nature* 2015, **526**:525–530.
16. Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, Ha G, Aparicio S, Bouchard-Côté A, Shah SP: **PyClone: statistical inference of clonal population structure in cancer**. *Nat Methods* 2014, **11**:396–398.
17. Miller CA, White BS, Dees ND, Griffith M, Welch JS, Griffith OL, Vij R, Tomasson MH, Graubert TA, Walter MJ, *et al.*: **SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution**. *PLoS Comput Biol* 2014, **10**.
18. El-Kebir M, Oesper L, Acheson-Field H, Raphael BJ: **Reconstruction of clonal trees and tumor composition from multi-sample sequencing data**. *Bioinformatics* 2015:i62–i70.
19. Deshwar AG, Vembu S, Yung CK, Jang G, Stein L, Morris Q: **PhyloWGS: reconstructing subclonal composition and evolution from whole-genome sequencing of tumors**. *Genome Biol* 2015, **16**:35.
20. El-Kebir M, Satas G, Oesper L, Raphael BJ: **Inferring the mutational history of a tumor using multi-state perfect phylogeny mixtures**. *Cell Syst* 2016, **3**:43–53.
21. Jiang Y, Qiu Y, Minn AJ, Zhang NR: **Assessing intratumor heterogeneity and tracking longitudinal and spatial clonal evolutionary history by next-generation sequencing**. *Proc Natl Acad Sci* 2016, **113**:E5528–E5537.
22. Reiter JG, Makohon-Moore AP, Gerold JM, Bozic I, Chatterjee K, Iacobuzio-Donahue CA, Vogelstein B, Nowak MA: **Reconstructing metastatic seeding patterns of human cancers**. *Nat Commun* 2017, **8**:14114.
23. Navin NE: **Cancer genomics: one cell at a time**. *Genome Biol* 2014, **15**:452.
24. Van Loo P, Voet T: **Single cell analysis of cancer genomes**. *Curr Opin Genet Dev* 2014, **24**:82–91.
25. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, *et al.*: **Intratumor heterogeneity and branched evolution revealed by multiregion sequencing**. *N Engl J Med* 2012, **366**: 883–892.
26. Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I, Fisher R, McGranahan N, Matthews N, Santos CR, *et al.*: **Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing**. *Nat Genet* 2014, **46**:225–233.
27. Zhang J, Fujimoto J, Zhang J, Wedge DC, Song X, Zhang J, Seth S, Chow C-W, Cao Y, Gumbs C, *et al.*: **Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing**. *Science* 2014, **346**:256–259.
28. Yates LR, Gerstung M, Knappskog S, Desmedt C, Gundem G, Van Loo P, Aas T, Alexandrov LB, Larsimont D, Davies H, *et al.*: **Subclonal diversification of primary breast cancer revealed by multiregion sequencing**. *Nat Med* 2015, **21**:751–759.
29. Alves JM, Prieto T, Posada D: **Multiregional tumor trees are not phylogenies**. *Trends Cancer* 2017, **3**:546–550.
30. Baslan T, Hicks J: **Unravelling biology and shifting paradigms in cancer with single-cell sequencing**. *Nat Rev Cancer* 2017, **17**:557–569.
31. Wang Y, Navin NE: **Advances and applications of single-cell sequencing technologies**. *Mol Cell* 2015, **58**:598–609.
32. Ham RG: **Clonal growth of mammalian cells in a chemically defined, synthetic medium**. *Proc Natl Acad Sci U S A* 1965, **53**: 288–293.
33. Gole J, Gore A, Richards A, Chiu YJ, Fung HL, Bushman D, Chiang HI, Chun J, Lo YH, Zhang K: **Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microwells**. *Nat Biotechnol* 2013, **31**: 1126–1132.

34. Zong C, Lu S, Chapman AR, Xie XS: **Genome-wide detection of single-nucleotide and copy-number variations of a single human cell**. *Science* 2012, **338**:1622–1626.

35. Nakamura N, Ruebel K, Jin L, Qian X, Zhang H, Lloyd RV: **Laser capture microdissection for analysis of single cells**. *Methods Mol Med* 2007, **132**:11–18.

36. Tomasson MH: **Cancer stem cells: a guide for skeptics**. *J Cell Biochem* 2009, **106**:745–749.

37. Cristofanilli M, Budd GT, Ellis MJ, Stopeck A, Matera J, Miller MC, Reuben JM, Doyle GV, Allard WJ, Terstappen LWMM, *et al.*: **Circulating tumor cells, disease progression, and survival in metastatic breast cancer**. *N Engl J Med* 2004, **351**:781–791.

38. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D, *et al.*: **Tumour evolution inferred by single-cell sequencing**. *Nature* 2011, **472**:90–94.

39. Potter NE, Ermini L, Papaemmanuil E, Cazzaniga G, Vijayaraghavan G, Titley I, Ford A, Campbell P, Kearney L, Greaves M: **Single-Cell mutational profiling and clonal phylogeny in cancer**. *Genome Res* 2013, **23**:2115–2125.

40. Gawad C, Koh W, Quake SR: **Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics**. *Proc Natl Acad Sci U S A* 2014, **111**:17947–17952.

41. Baslan T, Kendall J, Ward B, Cox H, Leotta A, Rodgers L, Riggs M, D'Italia S, Sun G, Yong M, *et al.*: **Optimizing sparse sequencing of single cells for highly multiplex copy number profiling**. *Genome Res* 2015, **125**:714–724.

42. Leung ML, Wang Y, Kim C, Gao R, Jiang J, Sei E, Navin NE: **Highly multiplexed targeted DNA sequencing from single nuclei**. *Nat Protoc* 2016, **11**:214–235.

43. •• Martelotto LG, Baslan T, Kendall J, Geyer FC, Burke KA, Spraggon L, Piscuoglio S, Chadalavada K, Nanjangud G, Ng CKY, *et al.*: **Whole-genome single-cell copy number profiling from formalin-fixed paraffin-embedded samples**. *Nat Med* 2017, **23**:376–385.

Developed and validated a first approach for whole-genome single-cell copy number profiling from formalin-fixed paraffin-embedded samples and extended the application of single-cell genomics beyond 'fresh/frozen' samples.

44. Yu M, Stott S, Toner M, Maheswaran S, Haber DA: **Circulating tumor cells: approaches to isolation and characterization**. *J Cell Biol* 2011, **192**:373–382.

45. Powell AA, Talasaz AAH, Zhang H, Coram MA, Reddy A, Deng G, Telli ML, Advani RH, Carlson RW, Mollick JA, *et al.*: **Single cell profiling of Circulating tumor cells: transcriptional heterogeneity and diversity from breast cancer cell lines**. *PLoS One* 2012, **7**.

46. Altomare L, Borgatti M, Medoro G, Manaresi N, Tartagni M, Guerrieri R, Gambari R: **Levitation and movement of human tumor cells using a printed circuit board device based on software-controlled dielectrophoresis**. *Biotechnol Bioeng* 2003, **82**:474–479.

47. Choi JH, Ogunniyi AO, Du M, Du M, Kretschmann M, Eberhardt J, Love JC: **Development and optimization of a process for automated recovery of single cells identified by microengraving**. *Biotechnol Prog* 2010, **26**:888–895.

48. De Bourcy CFA, De Vlaminck I, Kanbar JN, Wang J, Gawad C, Quake SR: **A quantitative comparison of single-cell whole genome amplification methods**. *PLoS One* 2014, **9**.

49. Hou Y, Wu K, Shi X, Li F, Song L, Wu H, Dean M, Li G, Tsang S, Jiang R, *et al.*: **Comparison of variations detection between whole-genome amplification methods used in single-cell resequencing**. *Gigascience* 2015, **4**:37.

50. Huang L, Ma F, Chapman A, Lu S, Xie XS: **Single-cell whole-genome amplification and sequencing: methodology and applications**. *Annu Rev Genomics Hum Genet* 2015, **16**:79–102.

51. Ning L, Li Z, Wang G, Hu W, Hou Q, Tong Y, Zhang M, Chen Y, Qin L, Chen X, *et al.*: **Quantitative assessment of single-cell whole genome amplification methods for detecting copy number variation using hippocampal neurons**. *Sci Rep* 2015, **5**:11415.

52. Garvin T, Aboukhalil R, Kendall J, Baslan T, Atwal GS, Hicks J, Wigler M, Schatz MC: **Interactive analysis and assessment of single-cell copy-number variations**. *Nat Methods* 2015, **12**:1058–1060.

53. Börgstrom E, Paterlini M, Mold JE, Frisen J, Lundeberg J: **Comparison of whole genome amplification techniques for human single cell exome sequencing**. *PLoS One* 2017:12.

54. Telenius H, Carter NP, Bebb CE, Nordenskjöld M, Ponder BAJ, Tunnacliffe A: **Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer**. *Genomics* 1992, **13**:718–725.

55. Arneson N, Hughes S, Houlston R, Done S: **Whole-genome amplification by degenerate oligonucleotide primed PCR (DOP-PCR)**. *Cold Spring Harb Protoc* 2008, **3**.

56. Baslan T, Kendall J, Rodgers L, Cox H, Riggs M, Stepansky A, Troge J, Ravi K, Esposito D, Lakshmi B, *et al.*: **Genome-wide copy number analysis of single cells**. *Nat Protoc* 2012, **7**:1024–1041.

57. Dean FB, Nelson JR, Giesler TL, Lasken RS: **Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification**. *Genome Res* 2001, **11**:1095–1099.

58. Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, *et al.*: **Comprehensive human genome amplification using multiple displacement amplification**. *Proc Natl Acad Sci* 2002, **99**:5261–5266.

59. Zhang D, Brandwein M, Hsuih T, Li HB: **Ramification amplification: a novel isothermal DNA amplification method**. *Mol Diagn* 2001, **6**:141–150.

60. Hou Y, Song L, Zhu P, Zhang B, Tao Y, Xu X, Li F, Wu K, Liang J, Shao D, *et al.*: **Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm**. *Cell* 2012, **148**:873–885.

61. Xu X, Hou Y, Yin X, Bao L, Tang A, Song L, Li F, Tsang S, Wu K, Wu H, *et al.*: **Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor**. *Cell* 2012, **148**:886–895.

62. Li Y, Xu X, Song L, Hou Y, Li Z, Tsang S, Li F, Im KM, Wu K, Wu H, *et al.*: **Single-cell sequencing analysis characterizes common and cell-lineage-specific mutations in a muscle-invasive bladder cancer**. *Gigascience* 2012, **1**:12.

63. Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, Chen K, Scheet P, Vattathil S, Liang H, *et al.*: **Clonal evolution in breast cancer revealed by single nucleus genome sequencing**. *Nature* 2014, **512**:155–160.

64. • Dong X, Zhang L, Milholland B, Lee M, Maslov AY, Wang T, Vijg J: **Accurate identification of single-nucleotide variants in whole-genome-amplified single cells**. *Nat Methods* 2017, **14**:491–493.

Combination of custom WGA and improved variant caller resulted in lowering FP errors in somatic SNV detection

65. • Leung K, Klaus A, Lin BK, Laks E, Biele J, Lai D, Bashashati A, Huang Y-F, Aniba R, Moksa M, *et al.*: **Robust high-performance nanoliter-volume single-cell multiple displacement amplification on planar substrates**. *Proc Natl Acad Sci U S A* 2016, **113**:8484–8489.

Performed high-throughput single-cell multiple displacement amplification in nanoliter volume.

66. Fu Y, Li C, Lu S, Zhou W, Tang F, Xie XS, Huang Y: **Uniform and accurate single-cell sequencing based on emulsion whole-genome amplification**. *Proc Natl Acad Sci U S A* 2015, **112**:11923–11928.

67. Sidore AM, Lan F, Lim SW, Abate AR: **Enhanced sequencing coverage with digital droplet multiple displacement amplification**. *Nucleic Acids Res* 2015, **44**.

68. Hosokawa M, Nishikawa Y, Kogawa M, Takeyama H: **Massively parallel whole genome amplification for single-cell sequencing using droplet microfluidics**. *Sci Rep* 2017, **7**:5199.

69. •• Chen C, Xing D, Tan L, Li H, Zhou G, Huang L, Xie XS: **Single-cell whole-genome analyses by linear amplification via transposon insertion (LIANTI)**. *Science* 2017, **356**:189–194.
Combination of transpostition and *in vitro* transcription resulted in linear amplification and coverage uniformity suitable for variant detection.

70. •• Zahn H, Steif A, Laks E, Eirew P, VanInsberghe M, Shah SP, Aparicio S, Hansen CL: **Scalable whole-genome single-cell library preparation without preamplification**. *Nat Methods* 2017, **14**:167–173.
Introduced a scalable direct library preparation (DLP) protocol based on nanoliter-volume transposition reactions for single-cell whole genome library preparation.

71. •• Vitak SA, Torkenczy KA, Rosenkrantz JL, Fields AJ, Christiansen L, Wong MH, Carbone L, Steemers FJ, Adey A: **Sequencing thousands of single-cell genomes with combinatorial indexing**. *Nat Methods* 2017, **14**:302–308.
Utilizes two rounds of tagmentation-based barcoding and pooling along with transposition reactions to scale single-cell libraries to thousands of cells.

72. Kuipers J, Jahn K, Beerenwinkel N: **Advances in understanding tumour evolution through single-cell sequencing**. *Biochim Biophys Acta Rev Cancer* 2017, **1867**:127–138.

73. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, *et al.*: **Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets**. *Cell* 2015, **161**: 1202–1214.

74. Leung ML, Wang Y, Waters J, Navin NE: **SNES: single nucleus exome sequencing**. *Genome Biol* 2015, **16**:55.

75. Omics S, Gawad C, Koh W, Quake SR: **Single-cell genome sequencing: current state of the science**. *Nat Rev Genet* 2016, **17**:175–188.

76. Olshen AB, Venkatraman ES, Lucito R, Wigler M: **Circular binary segmentation for the analysis of array-based DNA copy number data**. *Biostatistics* 2004, **5**:557–572.

77. Cheng J, Vanneste E, Konings P, Voet T, Vermeesch JR, Moreau Y: **Single-cell copy number variation detection**. *Genome Biol* 2011, **12**:R80.

78. Zhang C, Zhang C, Chen S, Yin X, Pan X, Lin G, Tan Y, Tan K, Xu Z, Hu P, *et al.*: **A single cell level based method for copy number variation analysis by low coverage massively parallel sequencing**. *PLoS One* 2013, **8**.

79. Ha G, Roth A, Lai D, Bashashati A, Ding J, Goya R, Giuliany R, Rosner J, Oloumi A, Shumansky K, *et al.*: **Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer**. *Genome Res* 2012, **22**: 1995–2007.

80. Knouse KA, Wu J, Amon A: **Assessment of megabase-scale somatic copy number variation using single cell sequencing**. *Genome Res* 2016, https://doi.org/10.1101/gr.198937.115.

81. • Gao Y, Ni X, Guo H, Su Z, Ba Y, Tong Z, Guo Z, Yao X, Chen X, Yin J, *et al.*: **Single-cell sequencing deciphers a convergent evolution of copy number alterations from primary to circulating tumor cells**. *Genome Res* 2017, **27**:1312–1322.
Reports a convergent evolution of copy number variants in circulating tumor cells.

82. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, *et al.*: **A framework for variation discovery and genotyping using next-generation DNA sequencing data**. *Nat Genet* 2011, **43**: 491–498.

83. Li H: **A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data**. *Bioinformatics* 2011, **27**:2987–2993.

84. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G: **Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples**. *Nat Biotechnol* 2013, **31**: 213–219.

85. Zhang C-Z, Adalsteinsson VA, Francis J, Cornils H, Jung J, Maire C, Ligon KL, Meyerson M, Love JC: **Calibrating genomic and allelic coverage bias in single-cell sequencing**. *Nat Commun* 2015, **6**:6822.

86. Yu C, Yu J, Yao X, Wu WK, Lu Y, Tang S, Li X, Bao L, Li X, Hou Y, *et al.*: **Discovery of biclonal origin and a novel oncogene SLC12A5 in colon cancer by single-cell sequencing**. *Cell Res* 2014, **24**:701–712.

87. Nikolenko SI, Korobeynikov AI, Alekseyev MA: **BayesHammer: bayesian clustering for error correction in single-cell sequencing**. *BMC Genom* 2013, **14**:1–11.

88. • Zafar H, Wang Y, Nakhleh L, Navin N, Chen K: **Monovar: single-nucleotide variant detection in single cells**. *Nat Methods* 2016, **13**:505–507.
Statistical variant calling method that specifically accounts for WGA errors and employs multisample approach to deal with coverage bias.

89. Jolliffe IT: **Principal component analysis**. *J Am Stat Assoc* 2002, **98**:487.

90. Hout MC, Papesh MH, Goldinger SD: **Multidimensional scaling**. *Wiley Interdiscip Rev Cogn Sci* 2013, **4**:93–103.

91. •• Gao R, Davis A, McDonald TO, Sei E, Shi X, Wang Y, Tsai P-C, Casasent A, Waters J, Zhang H, *et al.*: **Punctuated copy number evolution and clonal stasis in triple-negative breast cancer**. *Nat Genet* 2016, **48**:1119–1130.
Highly-multiplexed single-cell sequencing of thousands of cells for copy number profiling revealed a punctuated evolution of copy number variants.

92. Wu H, Zhang X-Y, Hu Z, Hou Q, Zhang H, Li Y, Li S, Yue J, Jiang Z, Weissman SM, *et al.*: **Evolution and heterogeneity of non-hereditary colorectal cancer revealed by single-cell exome sequencing**. *Oncogene* 2017, **36**:2857–2867.

93. • Roth A, McPherson A, Laks E, Biele J, Yap D, Wan A, Smith MA, Nielsen CB, McAlpine JN, Aparicio S, *et al.*: **Clonal genotype and population structure inference from single-cell tumor sequencing**. *Nat Methods* 2016, **13**:573–576.
A variational Bayes method for clustering single cells into clones. Also accounts for doublets and ADO during clustering.

94. McPherson A, Roth A, Laks E, Masud T, Bashashati A, Zhang AW, Ha G, Biele J, Yap D, Wan A, *et al.*: **Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer**. *Nat Genet* 2016, **48**(7):758–767.

95. • Salehi S, Steif A, Roth A, Aparicio S, Bouchard-Côté A, Shah SP: **ddClone: joint statistical inference of clonal populations from single cell and bulk tumour sequencing data**. *Genome Biol* 2017, **18**:44.
Combined bulk and SCS data in a unified framework for statistical inference of clonal composition of a tumor.

96. Davis A, Gao R, Navin N: **Tumor evolution: linear, branching, neutral or punctuated?** *Biochim Biophys Acta Rev Cancer* 2017, **1867**:151–161.

97. Subramanian A, Schwartz R: **Reference-free inference of tumor phylogenies from single-cell sequencing data**. *BMC Genomics* 2015, **16**:S7.

98. Schliep KP: **Phangorn: phylogenetic analysis in R**. *Bioinformatics* 2011, **27**:592–593.

99. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees**. *Mol Biol Evol* 1987, **4**: 406–425.

100. Sokal R, Michener C: **A statistical method for evaluating systematic relationships**. *Univ Kans Sci Bull* 1958, **2**:1409–1438.

101. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP: **Mrbayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space**. *Syst Biol* 2012, **61**: 539–542.

102. Hughes AEO, Magrini V, Demeter R, Miller CA, Fulton R, Fulton LL, Eades WC, Elliott K, Heath S, Westervelt P, *et al.*: **Clonal architecture of secondary acute myeloid leukemia defined by single-cell sequencing**. *PLoS Genet* 2014, **10**.

103. Eirew P, Steif A, Khattra J, Ha G, Yap D, Farahani H, Gelmon K, Chia S, Mar C, Wan A, *et al.*: **Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution**. *Nature* 2014, **518**:422–426.

104. Gusfield D: *Algorithms on strings, trees and sequences: computer science and computational biology*. Cambridge University Press; 1997.

105. Davis A, Navin NE: **Computing tumor trees from single cells**. *Genome Biol* 2016, **17**:113.

106. Kim KI, Simon R: **Using single cell sequencing data to model the evolutionary history of a tumor**. *BMC Bioinforma* 2014, **15**:27.

107. Yuan K, Sakoparnig T, Markowetz F, Beerenwinkel N: **BitPhylogeny: a probabilistic framework for reconstructing intra-tumor phylogenies**. *Genome Biol* 2015, **16**:36.

108. • Ross EM, Markowetz F: **OncoNEM: inferring tumor evolution from single-cell sequencing data**. *Genome Biol* 2016, **17**:69.
Infers a clonal tree from single-cell SNV profiles under infinite-sites assumption. Also infers error rates in datasets.

109. • Jahn K, Kuipers J, Beerenwinkel N: **Tree inference for single-cell data**. *Genome Biol* 2016, **17**:86.
Infers a mutation tree from single-cell SNV profiles under infinite-sites assumption. Also infers error rates in datasets.

110. Adams RP, Ghahramani Z, Jordan MI: *Tree-structured stick breaking for hierarchical data*. Advances in Neural Information Processing Systems (NIPS). 2010:19–27.

111. Leung ML, Davis A, Gao R, Casasent A, Wang Y, Sei E, Sanchez E, Maru D, Kopetz S, Navin NE: **Single cell DNA sequencing reveals a late-dissemination model in metastatic colorectal cancer**. *Genome Res* 2017, https://doi.org/10.1101/gr.209973.116.

112. Alekseyenko AV, Lee CJ, Suchard MA: **Wagner and Dollo: a stochastic duet by composing two parsimonious solos**. *Syst Biol* 2008, **57**:772–784.

113. Ryder RJ, Nicholls GK: **Missing data in a stochastic Dollo model for binary trait data, and its application to the dating of Proto-Indo-European**. *J R Stat Soc Ser C Appl Stat* 2011, **60**:71–92.

114. Kuipers J, Jahn K, Raphael BJ, Beerenwinkel N: **A statistical test on single-cell data reveals widespread recurrent mutations in tumor evolution**. *bioRxiv* 2016, https://doi.org/10.1101/094722.

115. • Zafar H, Tzen A, Navin N, Chen K, Nakhleh L: **SiFit: inferring tumor trees from single-cell sequencing data under finite-sites models**. *Genome Biol* 2017, https://doi.org/10.1186/s13059-017-1311-2.
Employs a finite-sites model of evolution to infer tumor phylogeny and error rates from single-cell SNV profiles.

116. Ramazzotti D, Graudenzi A, De Sano L, Caravagna G: **Learning mutational graphs of individual tumor evolution from multi-sample sequencing data**. *bioRxiv* 2017, https://doi.org/10.1101/132183.

117. Fearon ER, Vogelstein B: **A genetic model for colorectal tumorigenesis**. *Cell* 1990, **61**:759–767.

118. Kim TM, Jung SH, An CH, Lee SH, Baek IP, Kim MS, Park SW, Rhee JK, Lee SH, Chung YJ: **Subclonal genomic architectures of primary and metastatic colorectal cancer based on intratumoral genetic heterogeneity**. *Clin Cancer Res* 2015, **21**:4461–4472.

119. Ling S, Hu Z, Yang Z, Yang F, Li Y, Lin P, Chen K, Dong L, Cao L, Tao Y, *et al.*: **Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution**. *Proc Natl Acad Sci* 2015, **112**:E6496–E6505.

120. Williams MJ, Werner B, Barnes CP, Graham TA, Sottoriva A: **Identification of neutral tumor evolution across cancer types**. *Nat Genet* 2016, **48**:238–244.

121. Sottoriva A, Kang H, Ma Z, Graham TA, Salomon MP, Zhao J, Marjoram P, Siegmund K, Press MF, Shibata D, *et al.*: **A Big Bang model of human colorectal tumor growth**. *Nat Genet* 2015, **47**:209–216.

122. Macaulay IC, Ponting CP, Voet T: **Single-cell multiomics: multiple measurements from single cells**. *Trends Genet* 2017, **33**:155–168.

123. Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, Teng MJ, Goolam M, Saurat N, Coupland P, Shirley LM, *et al.*: **G&T-seq: parallel sequencing of single-cell genomes and transcriptomes**. *Nat Methods* 2015, **12**:519–522.

124. Dey SS, Kester L, Spanjaard B, Bienko M, van Oudenaarden A: **Integrated genome and transcriptome sequencing of the same cell**. *Nat Biotechnol* 2015, **33**:285–289.

125. •• Wang L, Fan J, Zhang C, Francis JM, Georghiou G, Li S, Gambe R, Zhou CW, Yang C, Cin PD, *et al.*: **Integrated single cell genetic and transcriptional analysis suggests novel private drivers of chronic lymphocytic leukemia**. *Genome Res* 2017, https://doi.org/10.1101/gr.217331.116.27.
Studied genotype-phenotype relationship in chronic lymphocytic leukemia by integrating DNA and RNA information from thousands of cells.

126. Spitzer MH, Nolan GP: **Mass cytometry: single cells, many features**. *Cell* 2016, **165**:780–791.

127. Miller MA, Weissleder R: **Imaging of anticancer drug action in single cells**. *Nat Rev Cancer* 2017, **17**:399–414.

128. Angermueller C, Clark SJ, Lee HJ, Macaulay IC, Teng MJ, Hu TX, Krueger F, Smallwood SA, Ponting CP, Voet T, *et al.*: **Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity**. *Nat Methods* 2016, **13**:229–232.

129. Hu Y, Huang K, An Q, Du G, Hu G, Xue J, Zhu X, Wang C-Y, Xue Z, Fan G: **Simultaneous profiling of transcriptome and DNA methylome from a single cell**. *Genome Biol* 2016, **17**:88.

130. Guo F, Li L, Li J, Wu X, Hu B, Zhu P, Wen L, Tang F: **Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells**. *Cell Res* 2017, **27**:967–988.

131. Miller CA, McMichael J, Dang HX, Maher CA, Ding L, Ley TJ, Mardis ER, Wilson RK: **Visualizing tumor evolution with the fishplot package for R**. *BMC Genomics* 2016, **17**:880.