

# Leveraging single-cell sequencing to unravel intratumour heterogeneity and tumour evolution in human cancers

Amy L Bowes<sup>1,2</sup> , Maxime Tarabichi<sup>1,3</sup> , Nischalan Pillay<sup>2,4\*</sup>  and Peter Van Loo<sup>1,5,6\*</sup> 

<sup>1</sup> Cancer Genomics Group, The Francis Crick Institute, London, UK

<sup>2</sup> Sarcoma Biology and Genomics Group, UCL Cancer Institute, London, UK

<sup>3</sup> Institute for Interdisciplinary Research, Université Libre de Bruxelles, Brussels, Belgium

<sup>4</sup> Department of Histopathology, The Royal National Orthopaedic Hospital NHS Trust, London, UK

<sup>5</sup> Department of Genetics, The University of Texas MD Anderson Cancer Centre, Houston, TX, USA

<sup>6</sup> Department of Genomic Medicine, The University of Texas MD Anderson Cancer Centre, Houston, TX, USA

\*Correspondence to: N Pillay, UCL Cancer Institute, Paul O’Gorman Building, 72 Huntley Street, London WC1E 6BT, UK. E-mail: [n.pillay@ucl.ac.uk](mailto:n.pillay@ucl.ac.uk) or P Van Loo, Department of Genetics, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Unit 1010, Houston, TX 77030, USA. E-mail: [pvanloo@mdanderson.org](mailto:pvanloo@mdanderson.org)

## Abstract

Intratumour heterogeneity (ITH) and tumour evolution are well-documented phenomena in human cancers. While the advent of next-generation sequencing technologies has facilitated the large-scale capture of genomic data, the field of single-cell genomics is nascent but rapidly advancing and generating many new insights into the complex molecular mechanisms of tumour biology. In this review, we provide an overview of current single-cell DNA sequencing technologies, exploring how recent methodological advancements have enumerated new insights into ITH and tumour evolution. Areas highlighted include the potential power of single-cell genome sequencing studies to explore evolutionary dynamics contributing to tumourigenesis through to progression, metastasis, and therapy resistance. We also explore the use of *in situ* sequencing technologies to study ITH in a spatial context, as well as examining the use of single-cell genomics to perform lineage tracing in both normal and malignant tissues. Finally, we consider the use of multimodal single-cell sequencing technologies. Taken together, it is hoped that these many facets of single-cell genome sequencing will improve our understanding of tumourigenesis, progression, and lethality in cancer, leading to the development of novel therapies.

© 2022 The Authors. *The Journal of Pathology* published by John Wiley & Sons Ltd on behalf of The Pathological Society of Great Britain and Ireland.

**Keywords:** clone; subclone; intratumour heterogeneity (ITH); tumour evolution; copy number aberrations (CNAs); tumour phylogeny; single-cell DNA sequencing

Received 28 February 2022; Revised 12 April 2022; Accepted 13 April 2022

*Conflict of interest statement:* NP is an associate editor of *The Journal of Pathology*. The other authors declare no conflicts of interest.

## Introduction

In 1890, the German pathologist David Paul Hanseemann was one of the first to observe asymmetrical nuclear divisions and failed chromosomal segregation in epithelial cancers [1,2]. Subsequently, Hanseemann coined the term ‘anaplasia’ to describe the reversion of differentiation and ensuing morphological nuclear changes associated with atypical mitoses in cancer cells [1,2]. Contemporaneously, further debate on the origin of tumours was fuelled by the husband-and-wife team of Theodor and Marcella Boveri, who elegantly demonstrated the relevance of chromosome mis-segregation in heredity and cancer development [2,3]. The Boveri’s proposed that human cancers originate from a single cell with an abnormality of its chromosomal constitution, which is passed onto all descendant cells due to rapid cell

proliferation [2,3]. Thus, the chromosome theory of cancer was born, which has been corroborated and expanded by a wealth of succeeding molecular studies investigating the genetic basis of tumour development and progression [4–6].

Cancer is a disease of the genome and with the unprecedented resolution afforded by high-depth whole-genome sequencing (WGS) we have discovered a repertoire of cancer-causing mutations, including single nucleotide variants (SNVs), copy number aberrations (CNAs), small insertions and deletions (termed indels) and structural variants (SVs). While most of these mutations can be classified as ‘innocent’, or unaf-fecting, passenger events, the acquisition of somatic driver mutations confers a selective advantage in some tumour cells enabling them to clonally expand and out-compete their neighbours [7]. Cancers subsequently develop through Darwinian evolutionary processes in

which complete selective sweeps result in an entire population of clonally related cells [8]. The cell of origin that prompted the last complete clonal sweep is called the most recent common ancestor (MRCA) from which all other cancer cells existing in a tumour are descendants. Later in tumour evolution, additional driver mutations may result in incomplete or locally complete clonal expansions [9]. As a result, tumours often contain several subclones that harbour unique subsets of mutations frequently conferring distinctive phenotypic features, such as treatment resistance and the ability to metastasise [10,11] (Figure 1). Table 1 lists the key concepts in tumour evolution.

With the advent of massively parallel sequencing, it has become possible to comprehensively catalogue virtually every clonal somatic mutation in a given cancer sample [6]. Most genomic studies rely on DNA sequencing data extracted from millions of cells, with the resulting signals containing intermixed mutations from different tumour subclones, as well as normal cells. Although these methods can provide important information regarding distinct tumour subclones [10,12], inferring rare populations that could lead to disease relapse or therapeutic resistance remains challenging. This intratumour heterogeneity (ITH) can have adverse consequences for the patient. For example, tumours with complex subclonal structures are often more aggressive, and more likely to metastasise due to an increased ability to adapt to local or systemic factors [13–15].

Single-cell sequencing provides a unique opportunity to study ITH and tumour evolution in greater detail when compared to bulk sequencing, often facilitating earlier MRCA inferences. In addition, single-cell DNA sequencing complements a key role carried out by the diagnostic pathologist for almost 100 years, i.e. the nuclear grading of tumours. On an almost daily basis pathologists use several nuclear grading systems for tumours, such as the Nottingham criteria for breast cancer [16] and the WHO/ISUP grading criteria for renal cell carcinoma [17,18]. For decades, these nuclear grading systems have been used for prognostication and potentially act as surrogate markers for aberrant genomic processes in tumour cells, most likely including chromosomal aneuploidies, whole-genome doubling events, or additional large-scale, complex SVs [19,20]. As these nuclear phenotypic changes are strongly correlated with tumour aggression and poor patient survival [21,22], single-cell sequencing represents a unique opportunity to probe the mutational processes underlying nuclear pleomorphism in cancer in greater detail [21,22].

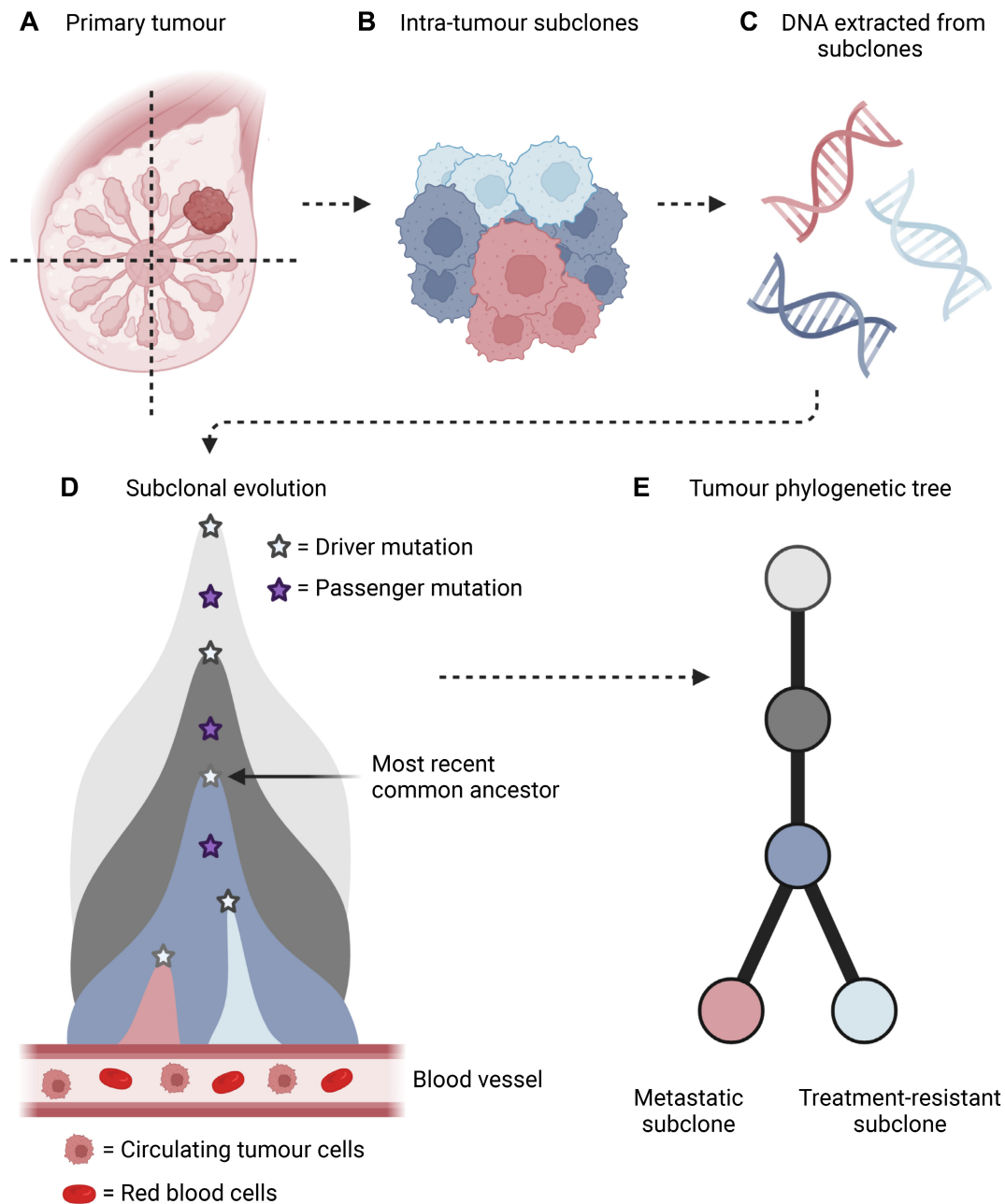
There are many excellent reviews of single-cell transcriptomics primarily aimed at deciphering the tumour microenvironment and ITH [23–25]. Single-cell genome profiling is less well explored. In this review, we describe current single-cell genome sequencing technologies used to investigate clonal dynamics and infer the evolutionary histories of cancer, providing insights into their application and how they can be leveraged to

disentangle ITH. We also discuss opportunities to explore ITH in a spatial context using *in situ* sequencing technologies.

## Single cells, ITH, and tumour evolution

Pathologists had long observed that invasive cancers within the colon or breast were likely to develop from premalignant lesions, such as adenomatous polyps with dysplasia or ductal carcinoma *in situ* (DCIS), respectively. First, a step-wise accumulation of genetic alterations was established in colorectal cancer (CRC) by analysing individual lesions at different histological stages of progression [4,5], as well as applying micro-alleleotyping techniques in both adenomatous polyps and invasive CRCs [26]. Not only did these studies confirm the order of acquired mutations necessary for CRC development, but they also highlighted substantial allelic heterogeneity between different tissue regions of the same tumour. Similarly, analysis of mutations in known CRC-causing genes revealed that several distinct subpopulations coexist within both benign adenomatous polyps and invasive CRCs [27,28]. Heterogeneous patterns of allelic losses were also observed in synchronous DCIS and invasive breast carcinomas, confirming the occurrence of genetic divergence during the clonal evolution of breast cancer [29]. Later, studies combining more advanced molecular genetic analyses in breast cancer, such as HER2 fluorescence *in situ* hybridisation (FISH) and comparative genomic hybridisation (CGH) strategies to explore subclonal chromosomal gains and losses also reported substantial ITH [30–33]. The biological and clinical context for studying genetic tumour heterogeneity and evolution from bulk sequencing data have been reviewed elsewhere [34–36].

Many recent advances in single-cell sequencing technologies build on this earlier work, providing the ability to examine the subclonal architecture and evolutionary dynamics of human cancers at a significantly enhanced resolution (Figure 2). For example, WGS of single tumour cells in two triple-negative (ER-, PR-, HER2-) breast cancers revealed the existence of distinct, highly aneuploid, subclonal tumour cell populations that shared partially related but divergent copy number profiles [37]. In one case, a single clonal expansion from the primary breast cancer had seeded the patient's liver metastasis with surprisingly little further copy number evolution. Paradoxically, in both primary breast cancer patients a subpopulation of genetically diverse pseudodiploid tumour cells was also identified that did not appear to travel to metastatic sites. Taken as a whole, these results suggest that tumours grow by iterative clonal expansions, which is entirely consistent with Peter Nowell's theory of cancer evolution published almost four decades ago [9]. Additional single-cell DNA sequencing studies also demonstrated significant heterogeneity between large-scale CNAs in primary breast cancers



**Figure 1.** The subclonal architecture of human cancers. (A,B) A primary tumour may contain multiple subclones. (C) Subsequently, the DNA extracted from the tumour will contain intermixed signals from all subclones. (D) Following genetic sequencing, somatic mutations are called, including driver and passenger mutations. Somatic mutations belonging to the MRCA can be identified in all tumour subclones reflecting the last complete clonal sweep. The emergence of new mutations may give rise to additional tumour subclones with further phenotypic features. (E) Somatic mutations can also be leveraged to infer tumour phylogenies. Figure prepared with BioRender, agreement number CM23SAMGRR.

with some subclonal CNAs occurring in genes known to be associated with metastasis, as well as altered therapeutic responses [38,39].

Building on this work, topographic single-cell sequencing enables the preservation of important spatial information *via* a combination of laser capture microdissection (LCM) and WGS of isolated single tumour cells [40]. When applied to synchronous DCIS and invasive breast carcinomas, it became apparent that the necessary genomic changes needed for invasion had taken place in ductal units before single aberrant cells

had even escaped the confinement of their basement membrane. Intriguingly, single-cell DNA sequencing revealed that one or more subclonal populations had comigrated through the basement membrane to establish the primary tumour, which also mirrored the results of two previous studies investigating multiclonal invasion in primary breast cancers [41,42]. Thus, the somatic evolution theory of cancer has been upheld and expanded by multiple single-cell DNA sequencing studies that have greatly improved our understanding of tumour biology.

Table 1. Key concepts in tumour evolution.

Concept	Definition
Branching tumour evolution	Tumour clones diverge from the MRCA and evolve in parallel resulting in multiple clonal lineages.
Clone	A lineage of cells descended from the MRCA that inherited the genotype of the MRCA.
Linear tumour evolution	A linear, stepwise accumulation of driver mutations instigating selective sweeps.
Most recent common ancestor (MRCA)	The MRCA is the most recent cell that spawned a set of cells and is a term often used to refer to the genotype of that ancestor cell.
Punctuated tumour evolution	Many genomic aberrations are acquired in a short time burst, often at the earliest stages of tumour evolution.
Subclone	A clone that is a descendant of the MRCA, but has developed additional genomic alterations. These mutations are only present in a subset of tumour cells.

### Overview of single-cell isolation and sequencing strategies

It should be stressed that obtaining high-quality, single-cell sequencing data is dependent upon overcoming several technical challenges, including proficient single-cell isolation, successful DNA extraction (plus or minus DNA amplification), efficacious DNA sequencing, and careful interpretation of the sequencing results, considering any technical biases that may have arisen during the first three steps. In the next section, we will briefly discuss different experimental strategies that can be employed to isolate and sequence single tumour cells (Figure 3).

Isolating single cells from tissue samples with good quality nucleic acids is dependent on appropriate tissue sampling and storage methods. It is critical that tissue dissociation protocols are optimised for tissues of

interest to generate high-quality single-cell suspensions [43–45]. The more commonly used methods to isolate single cells include fluorescence-activated cell sorting (FACS), or micromanipulation either *via* manual (i.e. mouth-pipetting and the CellTram apparatus) or semiautomated techniques (i.e. di-electrophoretic DEParray system or CellSearch for circulating tumour cell [CTC] isolation) [45]. Dependent on the research question, single-cell isolation and FACS using specific phenotypic markers for cells of interest should ideally be performed on fresh tissues [46]. However, in translational research settings this is often not possible due to logistical challenges in obtaining fresh tumour samples. A reasonable compromise is the isolation of single nuclei suspensions from frozen tissues [37,47]. Once the cell population of interest has been successfully identified, single cells are then captured into individual reaction chambers, such as oil droplets, nano-wells, or plates [44–46]. It is important to emphasise that the method for isolating single cells determines the throughput of the experiment (i.e. how many cells), as well as determining the phenotypic information captured. In addition, digestion of solid tissues for single cells comes at a cost of losing spatial information, which could obscure the relationships between cell transcription outputs, as well as interactions between neighbouring tumour cells and the tumour microenvironment. The advantages and disadvantages of different single-cell isolation strategies are summarised in Table 2.

A significant challenge when working with single-cell DNA is the limited yield of nucleic acid that is obtained. Each human diploid cell only contains approximately 7 pg of genomic DNA, which until recently has been below the limits of effective library preparation [63]. This can be overcome with whole-genome amplification (WGA), but this procedure results in several biases and technical artefacts, including a nonuniform genome coverage, as well as allelic imbalances or allelic dropouts,

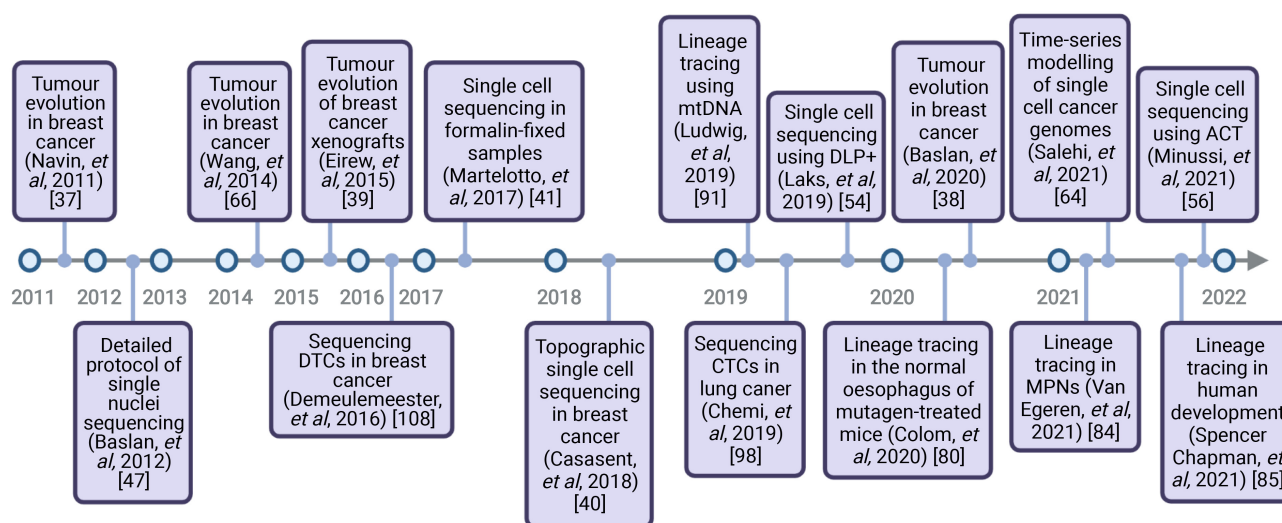
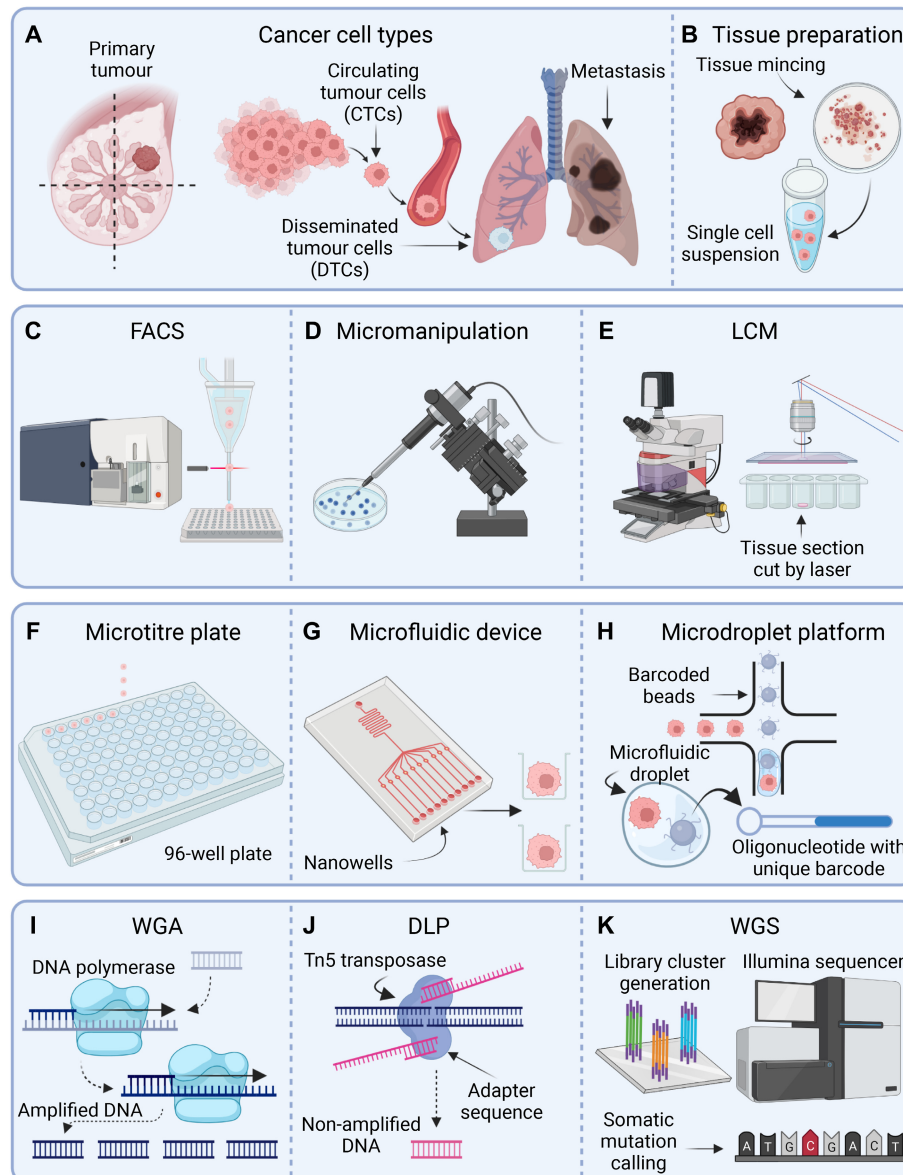


Figure 2. A timeline of single-cell DNA sequencing studies profiling ITH and tumour evolution. Acoustic cell tagmentation (ACT), colorectal cancer (CRC), circulating tumour cells (CTCs), disseminated tumour cells (DTCs), direct library preparation (DLP), mitochondrial DNA (mtDNA), and myeloproliferative neoplasms (MPNs). Figure prepared with BioRender, agreement number LS23SAQ0V3.



**Figure 3.** A summary of single-cell isolation and sequencing strategies. (A) Single-cell DNA sequencing can be performed on multiple cell types, including single cells isolated from the primary tumour or metastatic lesions, as well as circulating tumour cells (CTCs) and disseminated tumour cells (DTCs). (B) For some single-cell isolation methods, the tumour will first need to be dissociated to produce a single-cell suspension. (C–E) Three common methods used for isolating single cells include fluorescent activated cell sorting (FACS), micromanipulation, and laser capture microdissection (LCM). Only LCM techniques can preserve spatial information. (F–H) When working with single-cell solutions, cells are often isolated into individual reaction chambers prior to single-cell sequencing, including microtiter plates, nano-wells in microfluidic devices, or microdroplets. (I) In many workflows, DNA amplification using DNA polymerases is required prior to single-cell DNA sequencing. (J) Direct library preparation (DLP) techniques use a Tn5 transposase enzyme that simultaneously fragments and ligates synthetic oligonucleotides to single-cell DNA without the need for preamplification. These oligonucleotides contain sequencing adapters and indexing barcodes. (K) Indexed single-cell genomic libraries are pooled and sent for WGS, often using Illumina-based sequencing platforms that use library cluster generation techniques. Following sequencing, genetic mutations are identified. Figure prepared with BioRender, agreement number JK23SAS0N2.

whereby a particular allele is preferentially amplified or not amplified at all, respectively [64–68]. More recently, a direct library preparation (DLP) method has been introduced whereby indexed genomic libraries are constructed directly from single-cell DNA without the need for preamplification [53,55]. In short, DLP techniques use a hyperactive variant of the Tn5 transposase enzyme to simultaneously fragment double-stranded DNA and ligate synthetic oligonucleotides, including

sequencing adapters and single-cell barcodes. Barcoded genomic libraries can then be pooled for multiplex sequencing. DLP techniques offer substantial gains over traditional WGA methods, including greater coverage uniformity, improved mappability rates, reduced duplicate reads, and a more reliable detection of single-cell copy number profiles. The accurate measurement of single-cell genotypes using DLP combined with high-resolution microscopy images of individual cells, a

Table 2. The advantages and disadvantages of single-cell isolation strategies.

Cell isolation method	Description	Advantages	Disadvantages	Examples
Microtiter-plate based methods	Individual cells are deposited into wells within a plate.	Information regarding a cell's phenotype can be collected simultaneously <i>via</i> FACS. Individual microscopic images can be recorded enabling the detection of doublets and damaged cells, as well as those with specific morphological features.	Spatial information is lost when tissues are dissociated into a single-cell suspension. Low throughput method.	Suarez-Quian <i>et al</i> [48] Vermeulen <i>et al</i> [49] Leelatian <i>et al</i> [43] Casasent <i>et al</i> [40]
Microfluidic array methods	Microfluidic chips facilitate the isolation of single cells into flow channels, as well as performing library preparation reactions simultaneously.	Single cells can be isolated from small volumes. Microfluidic platforms offer a high single-cell throughput.	Often requires a uniform cell size. Not appropriate for the analysis of rare cell populations as 90% of cells captured are lost during the experimental process.	Hsiao <i>et al</i> [50] Altomare <i>et al</i> [51] Yu <i>et al</i> [52] Zahn <i>et al</i> [53] Tian <i>et al</i> [54] Laks <i>et al</i> [55]
Microfluidic droplet-based methods	Encapsulates single cells inside a nano-sized partition that encompasses an oil droplet together with an enzyme coated bead. The bead is loaded with enzymes, adapters and barcodes required to amplify DNA and construct single-cell sequencing libraries.	Offer the highest throughput for single-cell experiments. DNA from all single cells can be pooled and sequenced together on the same sequencing run due to unique barcodes. FACS can be used upstream to enrich for a cell population of interest.	Thousands of single cells are sequenced in a single run sometimes leading to insufficient genome coverage. Cell morphology, including information regarding nuclear size is lost.	Nishikawa <i>et al</i> [56] Szulwach <i>et al</i> [57] Cottinet <i>et al</i> [58] Keller <i>et al</i> [59]
Laser capture micro-dissection (LCM)	Applies a laser to tissue sections to isolate single cells of interest under direct microscopic visualisation.	LCM methods capture single cells whilst recording their spatial topography. LCM can enrich experiments for single-cells displaying a specific morphology.	Low throughput method.	Casasent <i>et al</i> [40]
Micro-manipulation	The isolation single cells in a solution under direct microscopic visualisation using glass mouth pipettes or robotic micromanipulators (i.e. CellTram). Automated methods such as CellSearch are also available.	Micromanipulation methods can enrich experiments for single cells displaying a specific morphology or phenotype using fluorescent markers.	Low throughput method. Technically challenging. Spatial information is lost when tissues are dissociated into a single-cell suspension.	Cristofanilli <i>et al</i> [60] de Bono <i>et al</i> [61] Navin <i>et al</i> [37] Demeulemeester <i>et al</i> [62]

method referred to as DLP+, has also enabled scientists to decipher the mutational landscape of single tumour cells, whilst simultaneously recording their morphological features, including nuclear characteristics [55]. For DLP and DLP+ techniques, single cells are isolated *via* custom-made microfluidics devices prior to performing direct tagmentation [55]. However, it is also possible to index individual nuclei that have been FACS-sorted into plates using well-specific primers. Although this approach does not require any specialised equipment and uses off-the-shelf tagmentation reagents, its throughput is limited due to each cell requiring its own polymerase chain reaction (PCR) reaction [53]. A technique called acoustic cell tagmentation (ACT) also uses Tn5 tagmentation of FACS-sorted nuclei, which is combined with acoustic liquid transfer technology to perform single-cell DNA sequencing at a single-molecule resolution [69]. Moreover, Tn5 tagmentation methodologies have also been applied on intact tissue sections to perform spatially resolved single-cell DNA sequencing

[70]. These novel, nonamplification, Tn5 tagmentation single-cell DNA sequencing technologies are summarised in Table 3.

### Reconstructing tumour phylogenies at the single-cell level to investigate clonal dynamics and ITH

Unlike single-cell transcriptomics, the analytical tools available for single-cell genomes are less mature. SNV calling is challenging and hugely dependent on the upstream amplification method [64–67,71]. More amenable is single-cell copy number profiling using WGS, which can range from  $0.1 \times$  to  $20 \times$  or more in sequencing coverage. Where the identification of large-scale CNAs (>10 kb) is possible, the phylogenetic history of a tumour can be inferred. Briefly, for single-cell copy number analysis, mapped reads are assigned to a chromosome position and normalised for GC biases as well

Table 3. Summary of nonamplification, Tn5 tagmentation single-cell DNA sequencing technologies.

Nonamplification DNA sequencing technologies	Description	References
Acoustic cell tagmentation (ACT)	Combines FACS of single nuclei, Tn5 tagmentation and acoustic liquid transfer technology to perform high throughput single-cell DNA sequencing.	Minussi et al [69]
DLP	Uses Tn5 transposase enzymes to simultaneously fragment DNA and ligate synthetic oligonucleotides (adapters and index barcodes) to produce genomic libraries without the need for preamplification.	Zahn et al [53]
DLP+	As above (DLP) with the addition of high-resolution microscopy images of individual cells.	Laks et al [55]
Slide-DNA-Seq	A method that uses barcoded bead arrays combined with Tn5 tagmentation to perform spatially resolved DNA sequencing on intact tissue sections.	Zhao et al [70]

as additional amplification artefacts [37,72,73]. The relative read depth or number of sequencing reads aligning in each genomic portion are then segmented into regions of consistent copy number states. Integer copy numbers are then assigned to each genomic segment, after which single cells with similar copy number profiles can be grouped into clusters or placed onto a hypothetical evolutionary tree to infer phylogenetic relationships. While hierarchical clustering approaches based on either chromosomal breakpoints [55] or copy numbers can be used to build a tree [69], better phylogenetic models based on parsimony [74] or fully Bayesian interpretation of the data are being actively developed.

Technological advances in single-cell sequencing and computational workflows now permit the detection of additional mutation types outside of CNAs at a much higher resolution, including SNVs and SVs, facilitating new insights into tumour evolution [69,75–77]. However, for phylogeny reconstruction, the combination of matching bulk and single-cell DNA sequencing data is still recommended, as this enables the interpretation of genomic aberrations at a single-cell resolution, whilst still maintaining high-quality mutation calls from the bulk [78].

In theory, library preparations avoiding the biases of amplification allow for a pseudo-bulk to be obtained from pooled single-cell reads, reducing sequencing costs. In a recent study, single-cell DNA sequencing using DLP+ was used to measure clonal genotypes in ~50,000 cells derived from cell lines, xenografts, and diagnostic samples [55]. By aggregating single cells that share similar copy number profiles, clonal genotypes were calculated at a single nucleotide resolution, facilitating key phylogenetic inferences. In addition, analysis of matched genomic and single-cell imaging measurements revealed correlations between cellular morphology and genome ploidy states, as well as clone-specific chromosomal aneuploidies. In a separate study also using DLP+ methodologies, CNA-defined clonal fitness dynamics were determined over time in single tumour cells [77]. Using an hTERT immortalised breast epithelial cell line, the authors demonstrated that *TP53* mutations distributed tumour cell fitness across a larger number of clones with distinct CNAs when compared to wildtype cells. Cisplatin treatment also triggered CNA diversification in primary triple-negative breast cancer patient-derived xenografts, leading to the emergence of drug-resistant clones in previously defined

low-fitness tumour cells. Together, these results suggest that cancer clonal fitness is strongly correlated with CNA accumulation, the dynamics of which may influence the emergence of treatment-resistant subpopulations. It is also thought that the gradual accumulation of SNVs in tumour cells facilitate extensive clonal diversity and the divergence of specific tumour cell phenotypes, enabling single cells to survive selective pressures such as those induced by the immune system, hypoxia, or chemotherapy [79]. For example, via the use of a high-coverage, single-cell WGS technology called nuc-seq, aneuploid rearrangements were found to occur relatively early in tumour evolution, remaining highly stable throughout the clonal expansion of two primary breast carcinomas (one ER+ and the other triple-negative) [79]. In contrast, point mutations accumulated gradually in breast cancer cells, but led to widespread subclonal diversity. Thus, single-cell DNA sequencing is an excellent tool for studying somatic diversification in late tumour evolution.

While the importance of studying cancer in an evolutionary framework is now well recognised, relatively little is known about the spatial organisation of cancer subclones and how they might influence tumour growth and metastasis. Recently, the development of slide-DNA-seq, a method that captures spatially resolved DNA sequences from intact tissue sections at an almost single-cell resolution, has enabled scientists to better decipher the role of spatial genomics in cancer progression [70]. The authors applied slide-DNA-seq to a genetically engineered mouse model of lung adenocarcinoma and a primary human CRC, demonstrating that clonal populations are indeed confined to specific spatial regions with clone-specific transcriptomic changes reflecting divergent evolutionary paths. In addition, a relatively new single-cell RNA-based methodology called Base Specific *In Situ* Sequencing (BaSISS) was used to generate quantitative, histology-based maps of cancer subclones in two multifocal breast cancers [80]. Here, the authors found that coexistent genetic clones had distinct transcriptional, histological and immunological characteristics and that the patterns of spatial genetic heterogeneity were strongly influenced by resident tissue structures. Additional *in situ* mutation technologies such as Basescope are also in existence, which uses RNA *in situ* hybridisation to detect point mutations in tissue sections, providing valuable insight into the emergence of subclones in human cancers

[81]. Thus, as spatially resolved DNA and RNA sequencing is pushed to the single-cell level, these technologies will provide new opportunities to determine how cell-intrinsic and cell-extrinsic factors contribute to ITH and tumour evolution.

### Leveraging single-cell sequencing to perform lineage tracing in normal and malignant tissues

Recent studies of somatic mutations in normal human tissues have illuminated how preneoplastic and early cancer genomes are potentially shaped by a life-long process of mutational accumulation that occurs with normal ageing [82–87]. Simultaneously, innovations in single-cell genomics have permitted detailed investigations of the lineage commitment and differentiation of preneoplastic cells, improving our understanding of how cellular dynamics influence ageing and cancer development. In model organisms, most cell lineage approaches rely on engineered genetic labels tagging single cells with heritable markers, such as fluorescent reporter genes, Cre-mediated recombination, or CRISPR-based genetic scars [88–92]. Combining these tracing methods with single-cell sequencing approaches has enabled scientists to interrogate both lineage fates and phenotypes of single cells simultaneously. By performing transgenic lineage tracing in mutagen-treated mice, deep sequencing of the normal oesophageal surface epithelial revealed strong clonal competition over time, with multiple genes under positive selection, including *Trp53*, *Notch1*, and *Notch2* [93]. Intriguingly, when a mutant clone collides with another of a similar competitive fitness, mutant clones tended to revert to a homeostatic fate, suggesting a constraining neighbour-regulated fitness programme for cancer prevention in the normal, ageing oesophagus.

As the genetic manipulations required for lineage tracing can only be applied to animal or *in vitro* cancer models, lineage tracing in humans is restricted to retrospective approaches, including the detection of somatic mutations [94]. For example, the pathognomonic JAK2-V617F mutation underlies the manifestation of myeloproliferative neoplasms [95,96]. Therefore, lineage tracing of somatic mutation patterns in the *JAK2* gene of haematopoietic stem cells (HSC) has enabled scientists to identify when this mutation first occurred and how it influences cell behaviour [97]. Astonishingly, JAK2-V617F mutations were found to occur in a single HSC several decades prior to a malignant diagnosis in treatment-naïve patients, inducing selective advantages in HSCs. In line with these findings, lineage tracing of single-cell-derived haematopoietic colonies in healthy foetuses has shown that haematopoietic progenitors already acquire tens of somatic mutations by 18 weeks of gestation [98]. Similarly, WGS of ~1,000 clonal haematopoietic colonies from patients with myeloproliferative neoplasms demonstrated that the acquisition of

JAK2-V617F driver mutations often occurred early in life, including during the *in utero* period [99].

Taken further, lineage tracing using WGS and somatic mutation calling in normal tissues has the potential to transform our understanding of somatic mutagenesis, ageing, and cancer development. In one study using NanoSeq or nanorate sequencing, a duplex approach that sequences both strands of a DNA molecule to reduce sequencing errors, differentiated cells in the blood and colon displayed remarkably similar mutational loads when compared to their corresponding stem cells, despite blood cells having undergone considerably more cell divisions [100]. Similarly, postmitotic neurons were also found to have accumulated somatic mutations at a similar rate to those of mitotically active tissues despite, the absence of cell division. Lineage tracing has also been applied to investigate nongenetic drivers of cancer therapy failure and the acquisition of a drug-tolerant ‘persister’ state in some tumour cells [101]. For example, although most persister cells remain arrested in the presence of a drug, some may reenter the cell cycle despite continued treatment [102]. The application of Watermelon, a high-complexity expressed barcode lentiviral library for simultaneous tracing of clonal lineages, as well as proliferation and transcriptional states in single tumour cells, showed that persister cell proliferative capacity is associated with antioxidant gene upregulation and a shift to fatty acid oxidative metabolism, exposing new vulnerabilities in tumour cells that may be amenable to therapeutic intervention [101].

As technologies aiming to detect somatic mutations by WGS in single cells are still in their infancy and had previously been difficult to apply at scale, lineage-tracing studies using mitochondrial DNA (mtDNA) sequence variants have also been successfully applied to infer clonal relationships in normal tissues. In one study, lineage tracing of mtDNA mutations was used to observe the clonal evolutionary dynamics of stem cell populations within the human colon [103]. Recently, single-cell RNA and ATAC sequencing has also been used to track somatic mutations in mtDNA, providing the ability to relate clonal cell dynamics to gene expression and chromatin accessibility [104]. In addition, a new single-cell transposon-based WGA method, termed multiplexed end-tagging amplification of complementary strands (META-CS), has significantly improved single-cell SNV detection rates [105]. This technology takes advantage of the complementary strands of double-stranded DNA to filter out false-positives and achieves a high accuracy for detecting SNVs in single cells.

### Single-cell sequencing, CTCs, and disseminated tumour cells

The isolation and analysis of CTCs from peripheral blood represents an alternative strategy to investigate tumour progression and evolution [106]. In the last



decade, a wealth of studies has successfully enumerated the clinical relevance of studying CTC single-cell genomes, which have been comprehensively reviewed [106–108]. CTC status is known to be a prognostic marker in multiple human cancers, including nonsmall-cell lung cancer (NSCLC) and CRC [109,110]. For example, pulmonary venous CTCs were successfully detected in 48% of 100 NSCLC patients in the TRACERX study ([ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT01888601) NCT01888601), which was significantly correlated with lung-cancer-specific relapse [111]. In a single genomic case study of one NSCLC patient, genomic profiling of pulmonary venous CTCs revealed a 91% SNV mutational overlap with a metastatic lesion detected at 10 months when compared with a 79% SNV mutational overlap in the primary tumour. Although most CTCs are single cells, a small proportion circulate as a group of clustered cells comprising two or more nuclei [112]. CTC clustering is thought to be part of an adaptive mechanism that improves CTC survival and enhances their metastatic potential, as CTC clusters have been shown to metastasise up to 100 times more effectively when compared to their single-cell counterparts [113–115]. As CTCs and CTC clusters are thought to drive metastasis formation in many human cancers, the demand for therapies that specifically target these single or clustered tumour cells is extremely high. Thus, the analysis of CTCs and CTC clusters at a single-cell resolution could offer a unique, minimally invasive approach to characterise and monitor dynamic changes in ITH in cancer patients, as well as providing future avenues for therapeutic exploration.

CTCs can extravasate and inhabit distant organs, where they are referred to as disseminated tumour cells (DTCs), which may remain dormant for many years or provide a reservoir of malignant progenitor cells for distant metastases [116,117]. There is increasing evidence that primary tumours have systemic effects beyond seeding of DTCs; for example, *via* immunosuppression or extracellular matrix remodelling [118]. Therefore, DTCs could benefit in their niches from primary tumour-induced effects, which may be essential for early DTC colonisation [119]. Intriguingly, DTCs isolated prior to metastasis often lack the characteristic genomic alterations typical of the metastatic tumour, suggesting that DTCs need to acquire additional genomic alterations prior to metastasis formation or that they could help to prepare the metastatic niche for late-arriving DTCs [120]. In contrast, late DTCs may remain dormant until unknown signals stimulate reentry into the cell cycle and take over the formation of metastases. Importantly, the role of early, genomically immature DTCs in metastasis and tumour progression is still under intense debate. For example, tumour-specific truncal mutations were completely absent in the subset of DTCs isolated from the bone marrow of nonmetastatic breast cancer patients, suggesting that they did not arise from the MRCA [62]. Thus, the detailed genomic profiling of single DTCs is still required to decipher the exact roles of early and late disseminating cancer cells in metastasis formation.

## Correlating genotype with phenotype in single- cells

Single-cell RNA sequencing is a well-established, fundamental tool for characterising cell phenotypes, reviewed elsewhere [121–123]. Like single-cell DNA sequencing technologies, low-throughput, plate-based methods (i.e. Smart-seq2 [124] and CEL-Seq2 [125]), as well as high-throughput droplet [126,127] or nanowell-based techniques [128,129] are in existence, whereby specific barcodes in individual reaction chambers detect all the complementary DNA sequences generated from a single cell. The rapid growth in the scale and robustness of single-cell RNA sequencing protocols has paved the way to substantial scientific discoveries, including the commencement of an extensive international initiative called The Human Cell Atlas, aiming to build comprehensive RNA reference maps for all human cell types [130]. To understand the molecular processes leading to cell phenotypic differentiation more comprehensively, new technologies are required that simultaneously assay different types of molecules in the same cell (i.e. DNA and RNA). Such multi-omics approaches would offer a direct link between a cell's modified genome and its transcriptomic landscape. In addition, cell lineage trees generated from single-cell WGS data can be phenotypically annotated reflecting the types and states of cells that are present.

In an approach termed G&T-seq (genome and transcriptome sequencing), a single cell's polyadenylated (polyA) RNA can be separated from its genomic DNA using a biotinylated oligo-dT primer [131]. Thereafter, both a single-cell's genome and transcriptome are amplified in parallel followed by sequencing. This integrated method can be used to expose the diverse effects of genetic variation on transcript levels [131]. Additional single-cell multi-omic protocols aiming to sequence both the cell's DNA and RNA simultaneously have also been devised, including DR-Seq and direct nuclear tagmentation and RNA sequencing (DNTR-seq), both of which were able to link single-cell states to corresponding genetic alterations, enabling routine analysis of heterogeneous tumours and additional complex tissues [132,133].

The tumour microenvironment and nongenetic adaptive mechanisms also impact cell phenotypes and tumour progression [134]. New technologies such as *Visium* [135], Nanostring DSP [136], and slide-RNA-seq [137] permit spatially resolved transcriptomics at a single-cell or near single-cell resolution, capturing *in situ* RNA sequences from intact tissue sections. These techniques enable the accurate spatial analysis of gene expression, the results of which can be easily integrated with large-scale single-cell RNA sequencing datasets to facilitate cell phenotype examination in both normal and malignant tissues. It is hoped that future advances in spatially resolved single-cell RNA sequencing will provide new opportunities to explore ITH and tumour evolution, whilst considering the spatial locations of both normal and tumour cells.

## Conclusions and perspectives

The rate of development of single-cell sequencing technologies is increasing rapidly, providing the ability to resolve tumour phylogenies in unprecedented detail with less uncertainty when compared to bulk sequencing data. Further progress is likely to be achieved by refining existing single-cell sequencing techniques to improve their sensitivity and accuracy. As discussed above, non-amplification, Tn5 tagmentation single-cell DNA sequencing methodologies, such as DLP, DLP+, and ACT, are likely to replace older methods relying on the amplification of genomic DNA by PCR, thereby removing unwanted technical artefacts that could lead to false mutation calls. In addition, technological advances in microfluidic droplet and microfluidic array-based platforms will facilitate an increasingly high single-cell throughput, processing thousands of single cells in a single experimental run.

Linking cell morphology with genomic changes will become more frequent *via* the use of live cell imaging prior to single-cell sequencing. In addition, new technologies, such as slide-DNA-seq and slide-RNA-seq, will continue to improve in efficacy and resolution, permitting spatially resolved analyses, whilst preserving tissue architecture. Likewise, alternative methods that facilitate multiple omic layers, such as joint genomics and transcriptomics, are likely to increase in use, permitting the simultaneous assessment of both cell genotype and phenotype.

For accurate tumour evolution inference, the combination of matching bulk and single-cell DNA sequencing data is still recommended, as this enables a greater ability to resolve phylogenetic relationships. However, as single-cell DNA sequencing technologies improve their resolution, this may no longer be necessary in the future. Lastly, novel computational methods that permit ever-increasing complex analyses of somatic mutation acquisition in single cells, as well as clustering multiple types of somatic mutations between thousands of single cells will be required. It is hoped that the advancement of single-cell sequencing technologies will greatly improve our understanding of ITH and tumour evolution, leading to novel biological insights and the development of innovative therapies for cancer patients.

## Acknowledgements

AB, MT, and PVL are supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001202), the UK Medical Research Council (FC001202), and the Wellcome Trust (FC001202). MT is a postdoctoral researcher of the F.R.S.-FNRS. PVL is a Winton Group Leader in recognition of the Winton Charitable Foundation's support towards the establishment of the Francis Crick Institute. PVL is a CPRIT Scholar in Cancer Research and acknowledges CPRIT grant support (RR210006).

NP holds a Cancer Research UK Clinician Scientist fellowship (Award no: 18387). We thank Jonas Demeulemeester for constructive criticism of the article. Figures were created using [BioRender.com](https://www.biorender.com).

## Author Contributions Statement

AB, PVL and NP conceived the concept and designed the article. AB performed the literature search and drafted the article and figures. AB, MT, PVL and NP were responsible for reviewing and editing the article.

## References

- Bignold LP, Coghlan B, Jersmann H. David Paul Hanseman: chromosomes and the origin of the cancerous features of tumor cells. *Cell Oncol* 2009; **31**: 61.
- Hardy PA, Zacharias H. Reappraisal of the Hanseman-Boveri hypothesis on the origin of tumors. *Cell Biol Int* 2005; **29**: 983–992.
- Satzinger H. Theodor and Marcella Boveri: chromosomes and cytoplasm in heredity and development. *Nat Rev Genet* 2008; **9**: 231–238.
- Vogelstein B, Fearon ER, Hamilton SR, *et al*. Genetic alterations during colorectal-tumor development. *N Engl J Med* 1988; **319**: 525–532.
- Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990; **61**: 759–767.
- ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* 2020; **578**: 82–93.
- Yates LR, Campbell PJ. Evolution of the cancer genome. *Nat Rev Genet* 2012; **13**: 795–806.
- Davis A, Gao R, Navin N. Tumor evolution: linear, branching, neutral or punctuated? *Biochim Biophys Acta Rev Cancer* 2017; **1867**: 151–161.
- Nowell PC. The clonal evolution of tumor cell populations. *Science* 1976; **194**: 23–28.
- Nik-Zainal S, Van Loo P, Wedge DC, *et al*. The life history of 21 breast cancers. *Cell* 2012; **149**: 994–1007.
- Dentro SC, Leshchiner I, Haase K, *et al*. Characterizing genetic intratumor heterogeneity across 2,658 human cancer genomes. *Cell* 2021; **184**: 2239–2254.e39.
- Carter SL, Cibulskis K, Helman E, *et al*. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* 2012; **30**: 413–421.
- Andor N, Graham TA, Jansen M, *et al*. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat Med* 2016; **22**: 105–113.
- Espiritu SMG, Liu LY, Rubanova Y, *et al*. The evolutionary landscape of localized prostate cancers drives clinical aggression. *Cell* 2018; **173**: 1003–1013.e15.
- Jamal-Hanjani M, Wilson GA, McGranahan N, *et al*. Tracking the evolution of non-small-cell lung cancer. *N Engl J Med* 2017; **376**: 2109–2121.
- Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 1991; **19**: 403–410.
- Delahunt B, McKenney JK, Lohse CM, *et al*. A novel grading system for clear cell renal cell carcinoma incorporating tumor necrosis. *Am J Surg Pathol* 2013; **37**: 311–322.

18. Warren AY, Harrison D. WHO/ISUP classification, grading and pathological staging of renal cell carcinoma: standards and controversies. *World J Urol* 2018; **36**: 1913–1926.
19. Cancer Genome Atlas Research Network. Comprehensive and integrated genomic characterization of adult soft tissue sarcomas. *Cell* 2017; **171**: 950–965.e28.
20. Steele CD, Tarabichi M, Oukrif D, et al. Undifferentiated sarcomas develop through distinct evolutionary pathways. *Cancer Cell* 2019; **35**: 441–456.e8.
21. Sukov WR, Lohse CM, Leibovich BC, et al. Clinical and pathological features associated with prognosis in patients with papillary renal cell carcinoma. *J Urol* 2012; **187**: 54–59.
22. Delahunt B, Eble JN, Egevad L, et al. Grading of renal cell carcinoma. *Histopathology* 2019; **74**: 4–17.
23. Fan J, Slowikowski K, Zhang F. Single-cell transcriptomics in cancer: computational challenges and opportunities. *Exp Mol Med* 2020; **52**: 1452–1465.
24. Li L, Xiong F, Wang Y, et al. What are the applications of single-cell RNA sequencing in cancer research: a systematic review. *J Exp Clin Cancer Res* 2021; **40**: 163.
25. Sun G, Li Z, Rong D, et al. Single-cell RNA sequencing in cancer: applications, advances, and emerging challenges. *Mol Ther Oncolytics* 2021; **21**: 183–206.
26. Boland CR, Sato J, Appelman HD, et al. Microallelotyping defines the sequence and tempo of allelic losses at tumour suppressor gene loci during colorectal cancer progression. *Nat Med* 1995; **1**: 902–909.
27. Saraga E, Bautista D, Dorta G, et al. Genetic heterogeneity in sporadic colorectal adenomas. *J Pathol* 1997; **181**: 281–286.
28. Baisse B, Bouzourene H, Saraga EP, et al. Intratumor genetic heterogeneity in advanced human colorectal adenocarcinoma. *Int J Cancer* 2001; **93**: 346–352.
29. Fujii H, Marsh C, Cairns P, et al. Genetic divergence in the clonal evolution of breast cancer. *Cancer Res* 1996; **56**: 1493–1497.
30. Aubele M, Mattis A, Zitzelsberger H, et al. Intratumoral heterogeneity in breast carcinoma revealed by laser-microdissection and comparative genomic hybridization. *Cancer Genet Cytogenet* 1999; **110**: 94–102.
31. Lewis JT, Ketterling RP, Halling KC, et al. Analysis of intratumoral heterogeneity and amplification status in breast carcinomas with equivocal (2+) HER-2 immunostaining. *Am J Clin Pathol* 2005; **124**: 273–281.
32. Chang MC, Malowany JI, Mazurkiewicz J, et al. 'Genetic heterogeneity' in HER2/neu testing by fluorescence in situ hybridization: a study of 2,522 cases. *Mod Pathol* 2012; **25**: 683–688.
33. Navin N, Krasnitz A, Rodgers L, et al. Inferring tumor progression from genomic heterogeneity. *Genome Res* 2010; **20**: 68–80.
34. McGranahan N, Swanton C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell* 2015; **27**: 15–26.
35. Black JRM, McGranahan N. Genetic and non-genetic clonal diversity in cancer evolution. *Nat Rev Cancer* 2021; **21**: 379–392.
36. Fittall MW, Van Loo P. Translating insights into tumor evolution to clinical practice: promises and challenges. *Genome Med* 2019; **11**: 20.
37. Navin N, Kendall J, Troge J, et al. Tumour evolution inferred by single-cell sequencing. *Nature* 2011; **472**: 90–94.
38. Baslan T, Kendall J, Volyanskyy K, et al. Novel insights into breast cancer copy number genetic heterogeneity revealed by single-cell genome sequencing. *Elife* 2020; **9**: e51480.
39. Eirew P, Steif A, Khattra J, et al. Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* 2015; **518**: 422–426.
40. Casasent AK, Schalck A, Gao R, et al. Multiclonal invasion in breast tumors identified by topographic single-cell sequencing. *Cell* 2018; **172**: 205–217.e12.
41. Martelotto LG, Baslan T, Kendall J, et al. Whole-genome single-cell copy number profiling from formalin-fixed paraffin-embedded samples. *Nat Med* 2017; **23**: 376–385.
42. Yates LR, Gerstung M, Knappskog S, et al. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med* 2015; **21**: 751–759.
43. Leelatian N, Doxie DB, Greenplate AR, et al. Single cell analysis of human tissues and solid tumors with mass cytometry. *Cytometry B Clin Cytom* 2017; **92**: 68–78.
44. Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. *Nat Rev Genet* 2016; **17**: 175–188.
45. Chappell L, Russell AJC, Voet T. Single-cell (multi)omics technologies. *Annu Rev Genomics Hum Genet* 2018; **19**: 15–41.
46. Hu P, Zhang W, Xin H, et al. Single cell isolation and analysis. *Front Cell Dev Biol* 2016; **4**: 116.
47. Baslan T, Kendall J, Rodgers L, et al. Genome-wide copy number analysis of single-cells. *Nat Protoc* 2012; **7**: 1024–1041.
48. Suarez-Quian CA, Goldstein SR, Pohida T, et al. Laser capture microdissection of single-cells from complex tissues. *Biotechniques* 1999; **26**: 328–335.
49. Vermeulen L, Todaro M, de Sousa Mello F, et al. Single-cell cloning of colon cancer stem cells reveals a multi-lineage differentiation capacity. *Proc Natl Acad Sci U S A* 2008; **105**: 13427–13432.
50. Hsiao AP, Barbee KD, Huang X. Microfluidic device for capture and isolation of single-cells. *Proc SPIE Int Soc Opt Eng* 2010; **7759**: 77590W\_1.
51. Altomare L, Borgatti M, Medoro G, et al. Levitation and movement of human tumor cells using a printed circuit board device based on software-controlled dielectrophoresis. *Biotechnol Bioeng* 2003; **82**: 474–479.
52. Yu Z, Lu S, Huang Y. Microfluidic whole genome amplification device for single-cell sequencing. *Anal Chem* 2014; **86**: 9386–9390.
53. Zahn H, Steif A, Laks E, et al. Scalable whole-genome single-cell library preparation without preamplification. *Nat Methods* 2017; **14**: 167–173.
54. Tian HC, Benitez JJ, Craighead HG. Single cell on-chip whole genome amplification via micropillar arrays for reduced amplification bias. *PLoS One* 2018; **13**: e0191520.
55. Laks E, McPherson A, Zahn H, et al. Clonal decomposition and DNA replication states defined by scaled single-cell genome sequencing. *Cell* 2019; **179**: 1207–1221.e22.
56. Nishikawa Y, Hosokawa M, Maruyama T, et al. Monodisperse picoliter droplets for low-bias and contamination-free reactions in single-cell whole genome amplification. *PLoS One* 2015; **10**: e0138733.
57. Szulwach KE, Chen P, Wang X, et al. Single-cell genetic analysis using automated microfluidics to resolve somatic mosaicism. *PLoS One* 2015; **10**: e0135007.
58. Cottinet D, Condamine F, Bremond N, et al. Lineage tracking for probing heritable phenotypes at single-cell resolution. *PLoS One* 2016; **11**: e0152395.
59. Keller A, Tilleman L, Dziedzicka D, et al. Uncovering low-level mosaicism in human embryonic stem cells using high throughput single-cell shallow sequencing. *Sci Rep* 2019; **9**: 14844.
60. Cristofanilli M, Budd GT, Ellis MJ, et al. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N Engl J Med* 2004; **351**: 781–791.
61. de Bono JS, Scher HI, Montgomery RB, et al. Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. *Clin Cancer Res* 2008; **14**: 6302–6309.
62. Demeulemeester J, Kumar P, Møller EK, et al. Tracing the origin of disseminated tumor cells in breast cancer using single-cell sequencing. *Genome Biol* 2016; **17**: 250.
63. Van Loo P, Voet T. Single cell analysis of cancer genomes. *Curr Opin Genet Dev* 2014; **24**: 82–91.

64. Borgström E, Paterlini M, Mold JE, *et al.* Comparison of whole genome amplification techniques for human single-cell exome sequencing. *PLoS One* 2017; **12**: e0171566.
65. Chen M, Song P, Zou D, *et al.* Comparison of multiple displacement amplification (MDA) and multiple annealing and looping-based amplification cycles (MALBAC) in single-cell sequencing. *PLoS One* 2014; **9**: e114520.
66. de Bourcy CF, De Vlaminck I, Kanbar JN, *et al.* A quantitative comparison of single-cell whole genome amplification methods. *PLoS One* 2014; **9**: e105585.
67. Deleye L, Tilleman L, Vander Plaetsen AS, *et al.* Performance of four modern whole genome amplification methods for copy number variant detection in single-cells. *Sci Rep* 2017; **7**: 3422.
68. Ning L, Li Z, Wang G, *et al.* Quantitative assessment of single-cell whole genome amplification methods for detecting copy number variation using hippocampal neurons. *Sci Rep* 2015; **5**: 11415.
69. Minussi DC, Nicholson MD, Ye H, *et al.* Breast tumours maintain a reservoir of subclonal diversity during expansion. *Nature* 2021; **592**: 302–308.
70. Zhao T, Chiang ZD, Morriss JW, *et al.* Spatial genomics enables multi-modal study of clonal heterogeneity in tissues. *Nature* 2022; **601**: 85–91.
71. Sabina J, Leamon JH. Bias in whole genome amplification: causes and considerations. *Methods Mol Biol* 2015; **1347**: 15–41.
72. Garvin T, Aboukhalil R, Kendall J, *et al.* Interactive analysis and assessment of single-cell copy-number variations. *Nat Methods* 2015; **12**: 1058–1060.
73. Zaccaria S, Raphael BJ. Characterizing allele- and haplotype-specific copy numbers in single-cells with CHISEL. *Nat Biotechnol* 2021; **39**: 207–214.
74. Schwarz RF, Trinh A, Sipos B, *et al.* Phylogenetic quantification of intratumour heterogeneity. *PLoS Comput Biol* 2014; **10**: e1003535.
75. Voet T, Kumar P, Van Loo P, *et al.* Single-cell paired-end genome sequencing reveals structural variation per cell cycle. *Nucleic Acids Res* 2013; **41**: 6119–6138.
76. McPherson A, Roth A, Laks E, *et al.* Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nat Genet* 2016; **48**: 758–767.
77. Salehi S, Kabeer F, Ceglia N, *et al.* Clonal fitness inferred from time-series modelling of single-cell cancer genomes. *Nature* 2021; **595**: 585–590.
78. Tarabichi M, Salcedo A, Deshwar AG, *et al.* A practical guide to cancer subclonal reconstruction from DNA sequencing. *Nat Methods* 2021; **18**: 144–155.
79. Wang Y, Waters J, Leung ML, *et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* 2014; **512**: 155–160.
80. Lomakin A, Svedlund J, Strell C, *et al.* Spatial genomics maps the structure, character and evolution of cancer clones. *BioRxiv* 2021; doi.org/10.1101/2021.04.16.439912 [Not peer reviewed].
81. Baker AM, Huang W, Wang XM, *et al.* Robust RNA-based in situ mutation detection delineates colorectal cancer subclonal evolution. *Nat Commun* 2017; **8**: 1998.
82. Lawson ARJ, Abascal F, Coorens THH, *et al.* Extensive heterogeneity in somatic mutation and selection in the human bladder. *Science* 2020; **370**: 75–82.
83. Lee-Six H, Olafsson S, Ellis P, *et al.* The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* 2019; **574**: 532–537.
84. Moore L, Leongamornlert D, Coorens THH, *et al.* The mutational landscape of normal human endometrial epithelium. *Nature* 2020; **580**: 640–646.
85. Martincorena I, Roshan A, Gerstung M, *et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* 2015; **348**: 880–886.
86. Yokoyama A, Kakiuchi N, Yoshizato T, *et al.* Age-related remodeling of oesophageal epithelia by mutated cancer drivers. *Nature* 2019; **565**: 312–317.
87. Martincorena I, Fowler JC, Wabik A, *et al.* Somatic mutant clones colonize the human esophagus with age. *Science* 2018; **362**: 911–917.
88. Kester L, van Oudenaarden A. Single-cell transcriptomics meets lineage tracing. *Cell Stem Cell* 2018; **23**: 166–179.
89. Woodworth MB, Girsakis KM, Walsh CA. Building a lineage from single-cells: genetic techniques for cell lineage tracking. *Nat Rev Genet* 2017; **18**: 230–244.
90. McKenna A, Findlay GM, Gagnon JA, *et al.* Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* 2016; **353**: aaf7907.
91. Sun J, Ramos A, Chapman B, *et al.* Clonal dynamics of native haematopoiesis. *Nature* 2014; **514**: 322–327.
92. Yu VWC, Yusuf RZ, Oki T, *et al.* Epigenetic memory underlies cell-autonomous heterogeneous behavior of hematopoietic stem cells. *Cell* 2017; **168**: 944–945.
93. Colom B, Alcolea MP, Piedrafitra G, *et al.* Spatial competition shapes the dynamic mutational landscape of normal esophageal epithelium. *Nat Genet* 2020; **52**: 604–614.
94. Lodato MA, Woodworth MB, Lee S, *et al.* Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* 2015; **350**: 94–98.
95. Baxter EJ, Scott LM, Campbell PJ, *et al.* Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet* 2005; **365**: 1054–1061.
96. Levine RL, Wadleigh M, Cools J, *et al.* Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell* 2005; **7**: 387–397.
97. Van Egeren D, Escabi J, Nguyen M, *et al.* Reconstructing the lineage histories and differentiation trajectories of individual cancer cells in myeloproliferative neoplasms. *Cell Stem Cell* 2021; **28**: 514–523.e9.
98. Spencer Chapman M, Ranzoni AM, Myers B, *et al.* Lineage tracing of human development through somatic mutations. *Nature* 2021; **595**: 85–90.
99. Williams N, Lee J, Mitchell E, *et al.* Life histories of myeloproliferative neoplasms inferred from phylogenies. *Nature* 2022; **602**: 162–168.
100. Abascal F, Harvey LMR, Mitchell E, *et al.* Somatic mutation landscapes at single-molecule resolution. *Nature* 2021; **593**: 405–410.
101. Oren Y, Tsabar M, Cuoco MS, *et al.* Cycling cancer persister cells arise from lineages with distinct programs. *Nature* 2021; **596**: 576–582.
102. Vallette FM, Olivier C, Lézot F, *et al.* Dormant, quiescent, tolerant and persister cells: four synonyms for the same target in cancer. *Biochem Pharmacol* 2019; **162**: 169–176.
103. Baker AM, Cereser B, Melton S, *et al.* Quantification of crypt and stem cell evolution in the normal and neoplastic human colon. *Cell Rep* 2014; **8**: 940–947.
104. Ludwig LS, Lareau CA, Ulirsch JC, *et al.* Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* 2019; **176**: 1325–1339.e22.
105. Xing D, Tan L, Chang CH, *et al.* Accurate SNV detection in single-cells by transposon-based whole-genome amplification of complementary strands. *Proc Natl Acad Sci U S A* 2021; **118**: e2013106118.
106. Keller L, Pantel K. Unravelling tumour heterogeneity by single-cell profiling of circulating tumour cells. *Nat Rev Cancer* 2019; **19**: 553–567.
107. Lin D, Shen L, Luo M, *et al.* Circulating tumor cells: biology and clinical significance. *Signal Transduct Target Ther* 2021; **6**: 404.
108. Cabel L, Proudhon C, Gortais H, *et al.* Circulating tumor cells: clinical validity and utility. *Int J Clin Oncol* 2017; **22**: 421–430.

109. Crosbie PA, Shah R, Krysiak P, et al. Circulating tumor cells detected in the tumor-draining pulmonary vein are associated with disease recurrence after surgical resection of NSCLC. *J Thorac Oncol* 2016; **11**: 1793–1797.
110. Nanduri LK, Hissa B, Weitz J, et al. The prognostic role of circulating tumor cells in colorectal cancer. *Expert Rev Anticancer Ther* 2019; **19**: 1077–1088.
111. Chemi F, Rothwell DG, McGranahan N, et al. Pulmonary venous circulating tumor cell dissemination before tumor resection and disease relapse. *Nat Med* 2019; **25**: 1534–1539.
112. Schuster E, Taftaf R, Reduzzi C, et al. Better together: circulating tumor cell clustering in metastatic cancer. *Trends Cancer* 2021; **7**: 1020–1032.
113. Liu X, Taftaf R, Kawaguchi M, et al. Homophilic CD44 interactions mediate tumor cell aggregation and polyclonal metastasis in patient-derived breast cancer models. *Cancer Discov* 2019; **9**: 96–113.
114. Aceto N, Bardia A, Miyamoto DT, et al. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* 2014; **158**: 1110–1122.
115. Grillet F, Bayet E, Villeronce O, et al. Circulating tumour cells from patients with colorectal cancer have cancer stem cell hallmarks in ex vivo culture. *Gut* 2017; **66**: 1802–1810.
116. Aguirre-Ghiso JA. Models, mechanisms and clinical evidence for cancer dormancy. *Nat Rev Cancer* 2007; **7**: 834–846.
117. Banys M, Krawczyk N, Fehm T. The role and clinical relevance of disseminated tumor cells in breast cancer. *Cancers (Basel)* 2014; **6**: 143–152.
118. Alečković M, McAllister SS, Polyak K. Metastasis as a systemic disease: molecular insights and clinical implications. *Biochim Biophys Acta Rev Cancer* 2019; **1872**: 89–102.
119. Klein CA. Cancer progression and the invisible phase of metastatic colonization. *Nat Rev Cancer* 2020; **20**: 681–694.
120. Klein CA. Selection and adaptation during metastatic cancer progression. *Nature* 2013; **501**: 365–372.
121. Haque A, Engel J, Teichmann SA, et al. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med* 2017; **9**: 75.
122. Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* 2018; **50**: 1–14.
123. Nayak R, Hasija Y. A hitchhiker's guide to single-cell transcriptomics and data analysis pipelines. *Genomics* 2021; **113**: 606–619.
124. Picelli S, Björklund ÅK, Faridani OR, et al. Smart-seq2 for sensitive full-length transcriptome profiling in single-cells. *Nat Methods* 2013; **10**: 1096–1098.
125. Hashimshony T, Senderovich N, Avital G, et al. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol* 2016; **17**: 77.
126. Macosko EZ, Basu A, Satija R, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 2015; **161**: 1202–1214.
127. Klein AM, Mazutis L, Akartuna I, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 2015; **161**: 1187–1201.
128. Gierahn TM, Wadsworth MH 2nd, Hughes TK, et al. Seq-Well: portable, low-cost RNA sequencing of single-cells at high throughput. *Nat Methods* 2017; **14**: 395–398.
129. Han X, Wang R, Zhou Y, et al. Mapping the mouse cell atlas by microwell-seq. *Cell* 2018; **172**: 1091–1107.e17.
130. Regev A, Teichmann SA, Lander ES, et al. The human cell atlas. *Elife* 2017; **6**: e27041.
131. Macaulay IC, Haerty W, Kumar P, et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat Methods* 2015; **12**: 519–522.
132. Dey SS, Kester L, Spanjaard B, et al. Integrated genome and transcriptome sequencing of the same cell. *Nat Biotechnol* 2015; **33**: 285–289.
133. Zachariadis V, Cheng H, Andrews N, et al. A highly scalable method for joint whole-genome sequencing and gene-expression profiling of single-cells. *Mol Cell* 2020; **80**: 541–553.e5.
134. Duan Q, Zhang H, Zheng J, et al. Turning cold into hot: firing up the tumor microenvironment. *Trends Cancer* 2020; **6**: 605–618.
135. Fawcner-Corbett D, Antanaviciute A, Parikh K, et al. Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* 2021; **184**: 810–826.e23.
136. Decalf J, Albert ML, Ziai J. New tools for pathology: a user's review of a highly multiplexed method for in situ analysis of protein and RNA expression in tissue. *J Pathol* 2019; **247**: 650–661.
137. Rodriques SG, Stickels RR, Goeva A, et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 2019; **363**: 1463–1467.