Published in final edited form as:

Biochim Biophys Acta. 2017 April ; 1867(2): 109–126. doi:10.1016/j.bbcan.2017.03.001.

A population genetics perspective on the determinants of intra-tumor heterogeneity

Zheng Hu^{1,2}, Ruping Sun^{1,2}, and Christina Curtis^{1,2,*}

¹Departments of Medicine and Genetics, Stanford University School of Medicine, Stanford, CA 94305 USA

²Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA 94305 USA

Abstract

Cancer results from the acquisition of somatic alterations in a microevolutionary process that typically occurs over many years, much of which is occult. Understanding the evolutionary dynamics that are operative at different stages of progression in individual tumors might inform the earlier detection, diagnosis, and treatment of cancer. Although these processes cannot be directly observed, the resultant spatiotemporal patterns of genetic variation amongst tumor cells encode their evolutionary histories. Such intra-tumor heterogeneity is pervasive not only at the genomic level, but also at the transcriptomic, phenotypic, and cellular levels. Given the implications for precision medicine, the accurate quantification of heterogeneity within and between tumors has become a major focus of current research. In this review, we provide a population genetics perspective on the determinants of intra-tumor heterogeneity and approaches to quantify genetic diversity. We summarize evidence for different modes of evolution based on recent cancer genome sequencing studies and discuss emerging evolutionary strategies to therapeutically exploit tumor heterogeneity.

1. Introduction

Cancer results from the acquisition of alterations during somatic cell division in a microevolutionary process that typically occurs over decades, much of which is occult and with extended clinically latent intervals. For example, more than half of all detectable somatic mutations in colorectal cancers occur prior to transformation [1, 2]. A set of initiating (epi)genetic events (so called drivers) provide a selective fitness advantage (resulting in higher proliferation or lower apoptosis for example) in the target cell relative to other pre-malignant cells, resulting in clonal expansion. These initiating events in the founding tumor cell are often specific to certain tissue types or cells of origin as is the case for *APC* in colon and other gastrointestinal tumors [3–5] or *DNMT3* in leukemias [6]. Hence, the fitness benefit that the mutation confers may depend on the microenvironment or

*Corresponding author: Christina Curtis, Ph.D. (cncurtis@stanford.edu).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

cellular context in which it occurs. However, a single mutation is seldom sufficient to evoke a fully malignant phenotype and the pre-malignant clones likely maintain functionality until additional “hits” accrue, eventually leading to transformation and uncontrolled cell growth (Figure 1).

Given that tumor cells continuously accrue mutations and are subject to ever changing microenvironments, genetic and phenotypic heterogeneity is expected. Indeed, according to the clonal evolution theory, heterogeneity is attributed to continued genetic and heritable epigenetic alterations and selection in diverse microenvironments within tumors [7–9]. Hence mutation rates, genetic drift, population structure, microenvironment and selection impact the extent of tumor heterogeneity. As tumors progress, expanding into an environment that is resource limited, cells that are better able to replicate, utilize energy, and migrate will have a competitive advantage and these ‘hallmarks’ are noted in most advanced cancers [10]. At this stage, the initiating genetic lesions may no longer ensure cell survival. Hence, it is essential to develop an understanding of the phenotypic consequences of putative oncogenes and tumor suppressors in a context-dependent fashion.

Although intra-tumor heterogeneity (ITH) has been appreciated for decades [11], the advent of high-throughput technologies has enabled the characterization of (epi)genomic, transcriptomic, phenotypic, and cellular heterogeneity at enhanced resolution [12–25]. The presence of pervasive ITH poses significant challenges for precision medicine. For example, sampling bias due to solid tumor spatial structure within lesions can obscure the interpretation of genomic profiles for patient stratification and therapeutic decision-making. Elevated ITH may also be associated with disease progression [26] and poor prognosis [27, 28]. Given the numerous clinical implications, the accurate quantification of heterogeneity within and between tumors from the same patient is a major focus of current research. However, to date, these analyses have largely been descriptive in nature.

While human tumor evolution cannot be directly observed, spatiotemporal patterns of genetic variation amongst tumor cells are stably inherited during cell division, and hence surreptitiously encode their ancestries. As such, the resultant patterns of ITH can be ‘read’ via multi-region sequencing (MRS) of spatially and/or temporally separated tumor regions or single-cell sequencing (SCS) and used to reconstruct the evolutionary relationships amongst the distinct cell populations (clones) that comprise a tumor, as has now been performed in a variety of tumor types [13, 16–19, 21, 22, 25, 29–34]. Intriguingly, these and other cancer sequencing studies suggest that distinct modes of evolution are operative during tumor progression and in different tumor types. However, this has yet to be systematically evaluated. More generally, while the elements of somatic evolution have been defined [7, 8, 35] and include mutation, genetic drift, migration, population structure, and selection, the evolutionary dynamics that are operative at different stages of progression in individual tumors are poorly characterized. A quantitative understanding of tumor dynamics has the potential to define cancers’ evolutionary playbook and might hold clues to as to how to better prevent, detect, and treat cancers. For example, the earlier the initial clonal expansion is detected, the less diverse and likely less fit the cell population, and the better the prognosis. Hence, knowledge of the initiating events and their relative temporal ordering may inform strategies for earlier intervention. Likewise, computational and mathematical

modeling can provide insights into mechanisms of progression and enable the inference of patient-specific tumor dynamics. Such information may ultimately inform the design of rational treatment strategies.

In this review, we provide a population genetic perspective on the origins of tumor heterogeneity and summarize approaches to quantify genetic diversity, drawing from established theory for studying genetic variation within and among natural populations in light of the forces of mutation, genetic drift, selection, and demography. We summarize evidence for different modes of clonal evolution based on recent cancer genome sequencing studies and the potential clinical relevance of the resultant heterogeneity. Finally, we discuss emerging evolutionary strategies to therapeutically exploit tumor heterogeneity and steer clonal dynamics.

2. Genetic and non-genetic tumor heterogeneity

2.1 Genomic heterogeneity within and between tumors

Conventionally, the distinct stages of tumor initiation and subsequent growth have been assumed to proceed in a ‘linear’ sequential fashion in which successively more fit ‘driver’ events arise within a clone, resulting in the replacement or clonal succession of less fit clones through selective sweeps [8,36]. While the sequential model may accurately describe tumor initiation, it is not known whether this applies to most solid tumors. Sequencing of established primary tumors has revealed extensive ITH in diverse tumors as demonstrated via multi-region profiling of clear-cell renal cell carcinoma (ccRCC) [13,18], colorectal cancer (CRC) [21,37], glioblastoma (GBM) [16,22], non-small cell lung cancer (NSCLC) [30], lung adenocarcinoma [29], cervical [38] and ovarian cancer [17], where somatic alterations are often restricted to distinct tumor regions. Several of these studies have reported evolutionary trajectories in which distinct driver alterations are ubiquitous (public/clonal) and therefore located on the trunk of the tree, whereas other drivers are heterogeneous (private/subclonal) and restricted to certain branches of a phylogenetic tree.

As a result of these and other studies, ITH is now accepted as an inherent feature of malignancy. The observation that many tumors exhibit genetic variegation has therapeutic implications since actionable “driver” mutations may be suboptimal targets if they are restricted to subclonal branches of the tumor’s ancestral tree. Moreover, ITH itself may represent a potential prognostic biomarker [26, 39]. There are numerous excellent reviews on ITH [15, 40–42], hence we do not attempt to comprehensively cover this vast literature here, but rather to introduce relevant concepts for quantifying, interpreting and exploiting ITH within an evolutionary framework. A crucial concept for understanding cancer genomic data is that clonal (public) mutations correspond to events that accumulated in the founding tumor cell lineage during initiation and prior to transformation such that they are present in all cells of the tumor (Figure 1). These include both pathogenic alterations that contribute to disease progression and benign passengers. Subclonal alterations occurring after transformation are by definition present in a subset of the tumor cell population and hence considered *private*. They may be restricted to certain tumor regions or more pervasive, but as these events initially occur in a single cell, they only become detectable after subsequent rounds of cell division or as a result of a clonal expansion.

As discussed in detail in subsequent sections, multiple factors influence the ability to reliably detect subclonal variants, including the sensitivity and specificity of an assay, tumor purity (or the proportion of tumor cells), sequencing coverage, as well as the nature of the sample and the mode of tumor evolution. To date, most large-scale cancer genome sequencing efforts have profiled a single bulk sample. This requires deconvolution to delineate the fraction of tumor cells harboring a given alteration, and poses limitations due to tissue architecture in solid tumors, which can result in the spatial segregation of subclones and therefore sampling bias. Sampling of multiple spatially separated bulk tumor regions or microdissected subregions such as individual glands can help to mitigate sampling bias and provide varying degrees of spatial resolution. However, these approaches still require deconvolution. In contrast, single cell sequencing enables direct measurement of clonal genotypes, but is not yet scalable to large numbers of cells and suffers from measurement errors. Each of these methods requires tissue dissociation so spatial information is lost if not recorded during sampling.

In situ analyses that retain tissue architecture have typically been limited to relatively low resolution techniques such as fluorescent in situ hybridization (FISH). However, these can be complemented by high-resolution spatial tumor sampling studies, where both approaches have revealed the intermingling of heterogeneous subclones. For example, analysis of EGFR copy number states in GBM via FISH exposed striking patterns of mosaicism [43]. Copy number heterogeneity was similarly reported for both EGFR and PDGFRA and found to be associated with differential growth factor responsiveness in GBM [44]. These findings were recapitulated by genome-wide copy number profiling of multiple tumor sectors from a cohort of 11 GBM patients [16]. Moreover, parallel transcriptional profiling revealed the presence of multiple molecular subgroups within a given patient, highlighting the challenges that ITH poses for precision medicine. In CRC, single gland neutral methylation tag profiling was consistent with the presence of numerous intermixed subclones [37]. More recently, multi-region, single gland and single cell analyses revealed extensive genetic variegation and subclone mixing at the copy number and mutational levels in both adenomas and carcinomas [21], as discussed in greater detail in *Section 4*. Computational modeling of spatial tumor growth and statistical inference on the genomic data revealed that most detectable ITH occurs early during tumor growth, long before the lesion is clinically evident. These findings were subsequently corroborated by reports that subclonal alterations originate early in colorectal adenomas/carcinomas [45–48]. Hence, ITH is already present in early lesions and continues to accrue during progression. Subclone mixing has been reported in other solid tumors, when detailed multi-region sampling was performed [49], whereas such patterns would be obscured by bulk profiling. Single cell sequencing has similarly revealed extreme levels of genetic heterogeneity [32, 50–53]. Intriguingly, early studies in GBM reported that genetically distinct cell subpopulations can cooperate to promote tumor growth, progression, and maintenance [54]. Several recent studies have similarly reported clonal cooperation in a breast cancer mouse model [55] and xenografts [56]. Collectively, these findings highlight the potential for a minor tumor cell population to promote ITH. They also suggest that interdependencies amongst subclones may represent potential therapeutic vulnerabilities.

Despite the extensive diversity, convergent evolution is apparent in some tumors, suggestive of constrained evolutionary trajectories. This is most notable in ccRCC, where on a background of clonal (truncal) inactivating mutations in the von Hippel Lindau (VHL) gene and loss of heterozygosity (LOH) of chromosome 3p, harboring the second copy of the VHL gene, spatially separated subclones were found to harbor distinct mutations in multiple members of the SWItch/Sucrose non Fermentable (SWI/SNF) chromatin remodeling complex, including *PBRM1*, *SETD2*, *ARID1A* and *SMARCA4* or even distinct mutations in the same gene [13, 18]. Convergent evolution of *EGFR* mutations/amplifications was also observed in GBM via single nucleus sequencing [57], and in chronic lymphocytic leukemia (CLL) [58]. These findings imply that alterations in these pathways are selectively advantageous and suggest genetic canalization towards a particular phenotype. Such patterns have only been noted in a minority of tumors examined to date, although the integration of multiple ‘omic’ features may unmask this. In the context of targeted therapies, which impose a highly specific selective pressure, convergent evolution of on-target point mutations in kinase domains and the activation of compensatory pathways have been observed [59, 60].

Most genomic profiling studies to date have focused on primary tumors, whereas relatively fewer have examined genetic divergence between paired primaries and metastases or recurrences arising at local or distant sites after initial therapy. This is due, in part, to the challenge in obtaining biopsies from distant metastatic sites, which are sometimes inaccessible and the time lag between initial diagnosis and disease recurrence, wherein procedures may be performed at different centers. Given that therapy can alter a tumor’s evolutionary trajectory, it is important that the genomic data are interpreted in this context. Several studies have comprehensively profiled paired primary and recurrent disease in glioma [19], GBM [22, 61], ovarian cancer [62], AML [63], ALL [64], neuroblastoma [65].

Genomic concordance between paired primaries and metastases has been investigated in diverse solid tumors at the level of somatic single nucleotide variants (SSNVs) [13, 66–73], structural variants (SVs) [74], copy number aberrations (CNAs) [66, 67, 75, 76] [73] and loss of heterozygosity (LOH) [77, 78] with varying degrees of genomic divergence reported. The inferences that can be drawn from these efforts depend on the study design, platform(s) used, depth of sequencing (when relevant), and extent of clinical and treatment information. For example, studies that are targeted in nature preclude the unbiased discovery of global genomic similarities and differences. Sampling bias and tumor spatial structure, as well as intervening therapy can confound the interpretation of patterns of heterogeneity between lesions. Hence, all of these factors should be reported, even if they cannot practically be controlled for.

The genomic characterization of multiple metastases from the same patient has been limited to a handful of studies [18, 34, 79, 80], which tend to suggest greater similarities amongst metastases than between the primary and metastasis. For example, in a large study of paired primaries and metastases (n=86), Brastianos et al. [34] reported that spatially and temporally separated brain metastases were often genetically homogenous, whereas 53% of cases harbored potentially clinically informative alterations in brain metastases that were not detected in the matched primary. The authors also reported that extra-cranial metastases were often distinct from the brain metastases.

Although multi-region sampling is essential for the accurate reconstruction of tumor evolutionary histories [21, 81], few studies have performed MRS of both the primary and metastasis and this has generally been restricted to one or a few cases [47, 73, 82]. Hence, there remains a need to characterize larger patient cohorts at a genome-wide level in order to delineate concordance between primaries and metastases, as well as the timing of metastasis and patterns of seeding.

2.2 Sources of tumor heterogeneity

Cancer is a disease of the genome, resulting from both endogenous and environmental damage to DNA, which can be propagated during normal cell division. In this review, we focus on *genetic heterogeneity* in cancer and the evolutionary forces that shape these patterns. However, tumor heterogeneity manifests not only at the genomic level, but at the cellular, phenotypic and microenvironmental levels and they jointly influence one another.

2.2.1 Mutational heterogeneity—As much of tumor progression is occult, by the time the lesion is clinically evident (and composed of tens of millions to billions of cells), tens to hundreds of thousands of mutations can have accrued. Comprehensive cancer genome sequencing studies have revealed that the aberration landscape is vaster than previously thought with different genes mutated in individuals with the same histological tumor type [36, 83, 84]. As such while a handful of genes are recurrently mutated at high frequency in the population, the vast majority of alterations occur infrequently, resulting in the so-called “long-tail” distribution of putative driver genes. Considerable (> 1000-fold) variation in the frequency of non-synonymous mutations was observed across tumor types with pediatric cancers exhibiting the lowest mutational burden, and melanoma and lung cancer the highest. In some cases, the elevated mutation frequency could be explained by exposures to known carcinogens, namely UV radiation in the case of melanoma and tobacco smoke in lung cancer. However, even amongst tumors of the same type, mutation frequencies varied multiple orders of magnitude (0.01 – 100/Mb), highlighting diversity in somatic mutation rates.

2.2.2 Genomic instability—It has been more than 40 years since Loeb et al. [85] proposed the mutator phenotype hypothesis, which posits that errors during DNA replication promote malignant progression [86]. Two years later, Nowell’s perspective piece on the clonal evolution of cancer proposed that heterogeneity in cancer results from increased genetic instability during disease progression [7]. Cancer genome sequencing efforts have revealed genomic instability resulting from point mutations (SSNVs), CNAs, and SVs, each of which can contribute to elevated genetic diversity. For example, chromosomal instability (CIN) leading to CNAs and SVs [87, 88] results in cells in which the mechanisms that ensure the fidelity of chromosome segregation during cell division are compromised. As such, chromosomal aberrations can dramatically impact cellular fitness [89, 90], which further propagates aneuploidy and facilitates rapid evolution [91].

Another form of genome instability, known as microsatellite instability, results from DNA slippage events at simple tandem repeat elements present throughout the genome, and is common in tumors with defective DNA mismatch repair (MMR). Although microsatellite

instability (MSI) has been appreciated for over two decades beginning with its discovery in familial CRC patients [92], the advent of next-generation sequencing (NGS) has revealed fresh insights into its functional implications [93]. As the mechanisms and consequences of genome instability in relation to tumor heterogeneity have recently been reviewed [40, 94], here we focus on core concepts and points relevant to an understanding of the population dynamics amongst tumor cells.

A particularly intriguing aspect of genomic instability is that it appears to be associated with cellular fitness tradeoffs [89, 95–97]. For example, aneuploidy has been shown to both initiate and inhibit tumor progression depending on the extent and context [98]. On the one hand, aneuploidy appears to be associated with tumor cell fitness advantages. For example, in CRCs whole genome doubling events are commonly thought to precede the acquisition of a CIN phenotype and to endow the tumor with *tolerance* to additional genome instability, thereby driving further evolution [97]. Indeed, genome doubling is associated with worse relapse free survival in colorectal [97] and ovarian [99] cancers. CIN has also been associated with poor prognosis [100]. On the other hand overexpression of *MAD2*, a gene contributing to chromosome instability (CIN), usually leads to cell growth arrest or death, suggesting that it is disadvantageous to cells [90]. These apparent contradictions may be explained in part by the findings of Laughney et al. [89] who demonstrated that cancer cells have evolved to exist within a narrow range of chromosome missegregation rates that optimize phenotypic heterogeneity and clonal survival, suggesting that CIN is under selective constraints.

Mutational burden has also been hypothesized to achieve saturation, where additional mutations are lethal, resulting in “error catastrophe”, that render the tumor more sensitive to insult and therefore confer favorable outcome [95]. Indeed, higher total mutation burden does not universally associate with worse prognosis. This is perhaps best exemplified in CRC, where mismatch repair (MMR) deficiency results in a hypermutator phenotype and excess (100 fold) SSNVs, and yet MMR-deficient microsatellite unstable (MSI-high) tumors tend to have favorable prognosis compared to microsatellite stable (MSS) tumors [101].

Given that genome instability has been established as an important determinant of genetic diversity [40, 102], it is critical to understand how tumors respond to these insults, and how they propagate diversity as this may inform strategies to limit ITH. While a common feature of many cancers, variable levels of mutational burden and genomic instability are observed within and between tumor types [83, 103]. In adult tumors, relatively genomically (copy number) stable subgroups of breast [104] and gastric cancers [105] have been reported. Moreover, some pediatric cancers have remarkably “quiet” genomes, as is the case for synovial sarcoma, which is commonly driven by the SS18-SSX oncogenic fusion, resulting in disruption of the SWI/SNF (BAF) chromatin regulatory complex [106]. Thus, while genome instability is a critical source of tumor heterogeneity in many cancers, it does not appear to be universally essential for disease progression.

2.2.3 Cellular, phenotypic and microenvironmental heterogeneity—Tumor cell heterogeneity can manifest at the level of cellular morphology, proliferation kinetics, cell motility, and cell metabolism amongst others [107, 108]. These differences can be attributed

to both genetic and non-genetic variation, where the latter includes epigenetic and transcriptional changes, as well as microenvironmental factors. For example, epigenetic changes can shape cellular phenotypes and modulate cellular plasticity [109, 110]. Phenotypic heterogeneity provides cancer cells with immense plasticity to cope with environmental stress during tumor progression in the face of limited resources and hence may also be important for allowing tumor cells to escape therapeutic insult. Indeed, phenotypic heterogeneity is now recognized as a major source of drug resistance [111, 112]. Early work demonstrated the presence of reversible drug tolerance associated with specific chromatin states and the re-establishment of heterogeneity of the putative cancer stem cell (CSC) surface marker CD133 [109]. Subsequently, stem-like characteristics were reported to contribute to the stabilization of drug-tolerant populations for extended time periods [111]. These and other studies point to the fact that high-dose cytotoxic therapies can induce pathogenic cell state transitions in the residual tumor cell population, wherein they revert to a primordial (stem-like) developmental state [111, 112]. Intriguingly, other studies have demonstrated phenotypic divergence between distant metastases relative to primary breast tumors and lymph node metastases [113], consistent with the notion that elevated diversity in advanced stage disease contributes to treatment resistance.

It is also evident that stochastic fluctuations can be functional, particularly in constrained conditions such as development [114–116], stress response [117], apoptosis [118] and cancer progression [119, 120]. Further, non-cell-autonomous factors can contribute to tumor growth and promote clonal interactions that may, in turn, lead to new phenotypes [56]. These non-genetic components occur on the backdrop of genetic variation and jointly shape a tumor's evolutionary trajectory. As tumors progress, irrespective of their tissue of origin, similar microenvironmental cues appear to arise as a result of spatial constraints, limited nutrients and vasculature, as well as activation of the immune system [121]. This is in accordance with the long-held view that advanced cancers exhibit similar defining hallmarks [10].

Microenvironment factors, including immune and stromal cells, acidification, oxygen, and geographical constraints also contribute to phenotypic heterogeneity within tumors, influencing signaling pathways and regulatory circuitry. For example, cellular adaptations to hypoxia and acidosis have been reported during breast tumor evolution [122] and invasiveness has been attributed to adaptive changes in the microenvironment [123, 124]. Additionally, cancer cells at the tumor periphery versus necrotic core can have different phenotypes resulting from regional variation in microenvironmental selective forces, where local cancer cell populations rapidly converge to the fittest phenotype given a stable environment [124].

In summary, the genetic, non-genetic, and environmental forces at play during a tumors long evolution shape the detectable patterns of ITH in a tumor sampled at the time of diagnosis. While studies to date have largely focused on genomic heterogeneity, in part due to greater ease in measuring it, new techniques to delineate the genotype to phenotype map in cancer and to probe tumor microenvironment *in situ* may offer opportunities to characterize their relationship.

3. Quantification of genetic diversity in cancer genomes

3.1 Practical challenges in quantifying ITH in tumor samples

The quantification of ITH from current NGS data is challenging for multiple reasons. From a practical perspective, the power to detect subclonal somatic single nucleotide variants (SSNVs) is limited in current bulk tumor WES studies, which often aim to achieve 80-100X coverage (with WGS often lower) [84]. This is particularly problematic when coupled with the low purity of many tumor samples, which contain a heterogeneous mixture of non-cancerous and cancerous cells. Ideally purity is estimated from adjacent hematoxylin and eosin (H&E) stained tissue sections prior to nucleic acid extraction to guide the target depth of coverage. When needed, flow sorting [125], single gland isolation [21, 37] or microdissection [126, 127] can be performed to enrich for tumor cells. As pathology based purity estimates may not agree with molecularly inferred estimates obtained using tools such as *PurBayes* [128], *ABSOLUTE* [99], or *TITAN* [129], it can also be beneficial to perform initial sequencing to determine how much additional coverage is needed to achieve a given power for detection. However, even then, genome sequence context, and high duplication rates can hinder uniform and optimal target coverage, the latter of which can be particularly problematic for formalin fixed paraffin embedded (FFPE) clinical specimens. Nonetheless, given the wealth of clinically annotated archival tissue samples, it is important to optimize both the study design and bioinformatics approaches for such specimens.

Perhaps even more problematic is the issue of sampling bias, which is inevitable in solid tumors defined by their unique tissue architecture and spatial structure. Multi-region sequencing (MRS) of spatially separated tumor regions can help mitigate the issue of sampling bias, and has become more common in recent years [13, 16–19, 21, 22, 25, 29, 30, 34]. Indeed, several recent studies have highlighted the need for MRS to accurately reconstruct tumor subclonal architecture [81, 130–133] and infer tumor evolutionary dynamics [21, 37]. From a practical perspective, it is not feasible to randomly sample the entire surgically resected lesion and in most cases, the location of the specimen within the tumor is not known. For biopsies, the specimen itself is limited and hence only provides a local sampling. Moreover, multiple factors can influence the extent of sampling bias, including spatial constraints, the mode of evolution, and the rates of migration, which are as of yet poorly characterized. Despite these complicating factors, it is evident that MRS aids the detection and quantification of ITH [21, 134].

The distribution of variant allele frequencies critically depends on the underlying mode of tumor growth [21, 135]. For example, in colorectal tumors characterized by Big Bang growth dynamics, after transformation, the tumor expands in the absence of stringent selection, compatible with effectively neutral evolution [21]. Hence, in this model, the timing of a mutation is the fundamental determinant of its frequency and later arising subclonal mutations will be present in increasingly diminutive fractions of the tumor cell population (Figure 1A). As such, genetic diversity will be vastly underestimated [21], as has been predicted theoretically [127]. In a model of effectively neutral evolution, subclonal variants are not functionally relevant for primary tumor growth, but they nonetheless provide a potentially rich substrate for subsequent selection in the context of therapy. In the alternate

scenarios of continuous strong positive selection (Figure 1B) or branched evolution (co-occurrence of multiple selectively advantageous clones, Figure 1C) after transformation, subclonal variants that confer a selective growth advantage may attain high frequency (depending on various population parameters). In this case, they may be more readily detected, but in the absence of multiple sampling data, it is not always feasible to determine whether they are high frequency subclonal (private) SSNVs or public alterations due to the inherent noise associated with NGS (Figure 2).

Numerous methods have been developed to perform somatic variant calling from tumor-normal pairs, including *deepSNV* [136] and *MuTect* [137], which aim to achieve greater sensitivity for subclonal variants. However in a typical NGS study, the limits of detection corresponds to a variant allele frequency (VAF) of ~10% [137]. Critically, subclonal variants *alone* are informative in chronicling the dynamics of growth after transformation, as variants that were accrued prior to transformation will have either been lost or else present in the founding tumor and hence clonal in the final tumor. On the other hand, it is possible to over-estimate ITH when calling variants from multiple tumor samples in the same patient that have non-uniform coverage or purity. Purity differences can be especially notable in comparisons of paired primaries and metastases [34] or paired pre versus post-treatment samples [138, 139] and hence are important to account for. The increasing availability of MRS also motivates methods such as *MultiSNV* [140] that aim to jointly call SSNVs from multiple samples. By exploiting pseudo-repeated measurements, such an approach can potentially aid the detection of false positives and false negatives. Deep targeted sequencing or digital PCR are commonly used for orthogonal validation of candidate SSNVs discovered through unbiased surveys. However, PCR based approaches (including NGS library preparation) can suffer from “jackpot effects” due to stochastic fluctuations and errors introduced during early PCR steps that are amplified exponentially, resulting in differences in biased allele frequency estimates [141–143]. Such artifacts can be mitigated to varying extents by modifying library preparation techniques to incorporate unique molecular identifiers and via circle [144] or duplex sequencing [145], although these are not currently scalable to whole exome/genome analyses.

Once SSNVs have been called, it is common to adjust their VAF values for tumor purity and copy number to obtain an estimate of their cellular prevalence or cancer cell fraction (CCF). However, this is a non-trivial task since multiple parameters are unknown (purity, ploidy, number of subclones) and often need to be estimated simultaneously. Moreover, these estimates depend on the relative ordering of CNAs, SSNVs, and genome doubling events during tumor evolution.

The accurate identification of subclonal CNAs from NGS is a challenging task given that the distribution of short reads along the genome is non-uniform, with high variability amongst adjacent loci at standard sequencing depths. As a result, only CNAs with relatively high frequency can be reliably detected from current NGS data, and this depends on the segment length and copy number states. Methods to predict the subclonal CNAs from tumor sequencing data include *TITAN* [129], *TheTA* [146], *CloneHD* [147], and *CloneCNA* [148]. Tools designed to infer CNAs from multi-region tumor sequencing data are currently

lacking, but have the potential to improve the detection of subclonal CNAs, as well as the inference of tumor phylogenies, as discussed below.

Various tools have been developed to estimate the number of subclonal clusters from tumor sequencing data by grouping mutations on the basis of their allele frequencies, [133, 147, 149, 150]. The majority of these approaches aim to define ‘clonal’ relationships based on the assumption that mutations with similar frequencies descended from the same ancestral cell and that a limited number of dominant subclones with distinct mutational profiles have undergone clonal expansion. However, a single bulk sample may fail to delineate clonal alterations versus subclonal alterations that attained high frequency to selection or drift (Figure 2). More generally, the inference of the number of subclones, their proportions and genotypes from single sample bulk tumor sequencing is challenging with the solution underdetermined under most conditions [81, 147, 151, 152]. MRS data should facilitate these inferences and have recently been used to reconstruct tumor phylogenies, as discussed below. Single cell sequencing enables the direct measurement of clonal genotypes, and hence avoids the need for deconvolution, making it a powerful approach [153, 154]. However, this is complicated by noisy and incomplete measurements and remains impractical on a large scale. Ultimately, a combination of bulk and single cell sequencing may help to further resolve tumor subclonal architecture.

3.2 Phylogenetic approaches to interpret ITH

Phylogenetic trees are the canonical structure for describing the genealogical relationship among tumor cells. They offer a potentially powerful and theoretically grounded approach to study heterogeneity within and between lesions [155, 156], drawing from the rich field of molecular phylogenetics, the principles and methods of which are reviewed elsewhere [157]. Here, we briefly outline considerations for the analysis of cancer genome sequencing data. When modeling NGS data, it is important to account for systematic errors and to increase robustness to model violations. In the case of cancer genomes, not only is this complicated by uncertainty in SSNV and CNA estimates (as discussed above), but by extensive genetic diversity due to mixtures of cells for which both the number of clones and their relative proportions are unknown. The issue of sampling bias in solid tumors further necessitates that multiple regions of temporally separated lesions be assayed to robustly reconstruct their ancestral relationship, although this has seldom been done in practice.

To date, various approaches have been employed to reconstruct sample or subclone phylogenies from tumor sequencing data. Several studies have employed standard phylogenetic approaches, including parsimony [158], maximum-likelihood [159] and distance-based methods such as the neighbor-joining algorithm [160] to delineate the relationship between tumor samples based on SSNVs (often from CNA-neutral regions) or CNA breakpoints. For example, distance methods were used to cluster SSNVs across multiple bulk samples in renal cell carcinoma [13] and CNAs in single breast cancer cells [32]. As distance-based methods take as input summary statistics, they cannot account for complex character (SSNV, CNA) states. Likewise, the exclusion of SSNVs in regions of CNA to avoid the complex issue of phasing ignores potentially valuable information, although for tumors that are largely diploid or have sufficient events in non-diploid regions,

this simplifying approach may be reasonable. CNAs also represent a powerful feature for phylogenetic inference, but require custom methods due to the horizontal dependencies between adjacent loci and overlapping alterations. An early method, termed *TuMult* [161] sought to overcome the issue of overlapping aberrations by employing copy number breakpoints (which should be persistent). More recently, *MEDICC* was developed to jointly phase parental alleles and perform tree reconstruction [162]. This approach takes as input integer copy number profiles and B-allele frequencies and has been used to reconstruct tumor phylogenies from multi-region copy number profiles [21, 163].

Other studies have sought to perform the challenging task of subclonal deconvolution so that subclones can be modeled as species or taxa. For example, by performing *in silico* phasing of subclonal SSNVs and exploiting the pigeon-hole principle, a probable phylogeny was inferred from high coverage bulk WGS data of a breast tumor [164]. However, phylogeny reconstruction from a single bulk tumor sample will generally lead to an incomplete and potentially erroneous tree [37, 165]. Numerous methods have been described to automate deconvolution, including those that can take as input multiple samples [132, 166, 167]. However, even with multiple samples, there is no guarantee that low VAF SSNVs can be accurately phased. More recently, several computational methods have been developed to jointly model SSNVs and CNAs [81, 131, 133] towards more accurate reconstruction of clonal trees. This represents an important methodological advance and importantly suggest that many possible trees may be consistent with the data [81]. However, these methods often couple the task of mutation clustering and phylogeny reconstruction and require prior clustering of CCF values derived from WGS/WES to reduce the feature space to a limited number of subclones due to the computationally intensive nature of enumerating a large possible tree-space. Hence, this too can result in a loss of information. However, this new class of methods is in its relative infancy and is likely to remain an area of active research. Likewise, several recent methods have been described to reconstruct phylogenies from single cell data [168, 169], where phase is known, but the data are generally sparse and noisy due to allelic dropout. Single cell WGS has the potential to directly resolve clonal genotypes, however this will require improved genome coverage and mutation sensitivity since errors can distort lineage reconstruction [170].

Given the absence of ground truth phylogenies for actual tumors, simulation studies provide the primary means to evaluate a methods performance. Hence, it is important that simulations used for benchmarking capture relevant aspects of tumor growth, which can be influenced by numerous factors (as discussed in *Section 4*). It is also essential to consider the underlying model assumptions and robustness to violations. For example, an infinite-site model is commonly assumed such that each site can mutate only once and persists. However, in practice, convergent evolution could result in the acquisition of the same mutation more than once and mutations could be lost due to loss of heterozygosity (LOH), thereby violating this assumption. Some methods also require assumptions about the underlying models of molecular evolution, which may require refinement in cancer.

Despite the power of phylogenetic approaches to delineate tumor architecture and interpret ITH, there are limitations. In particular, phylogenies alone do not report on the underlying dynamics [130]. However, genealogy-based statistical inference methods, stemming from

the development of coalescent theory [171] and the advent of routine large scale sequencing efforts have transformed modern computational population genetics approaches [172]. Techniques such as Approximate Bayesian Computing (ABC) have been used extensively in population genetics to infer posterior parameter distributions from molecular data when using stochastic models for which likelihoods cannot be calculated [172, 173]. Related approaches based on coalescent modeling and Bayesian inference have been used to characterize stem cell dynamics in human colon crypts [174]. More recently, ABC has been integrated within 3-dimensional agent-based computational models of tumor growth that represent cancer as an evolutionary process to enable the inference of patient-specific tumor dynamics such as the mutation rate and cancer stem cell fraction [37], as well as subclone fitness differences and the mutational timeline [21].

3.3 Ecological and population genetic measures of ITH

The relationship between genetic diversity in cancer and population evolvability may provide prognostic and therapeutic information [26]. Importantly, it is the total diversity (including both functional and selectively neutral diversity) that likely determines the evolvability of cancer cells since in the absence of diversity natural selection is not operative. This is in-line with the Dykhuizen-Hartl effect proposed by Kimura, in which neutral mutations may later become adaptive in an altered environment [175]. Hence, ITH may be a proxy for the underlying rate of evolutionary processes operative in a tumor. While it is not clear how to best sample a tumor or to quantify ITH, several measures have been borrowed from ecology and population genetics for this task and others have been defined based on recent cancer genomic data.

One widely used measure of genetic diversity within a population is heterozygosity [176], or the probability that two sequences chosen at random from the population have different allelic types, which is equivalent to Simpson's Diversity Index in ecology [177]:

$$D=1 - \sum_i f_i^2 \quad \text{Eq.(1)}$$

For asexually reproducing populations such as cancer, the heterozygosity is the probability that two cells chosen at random have different sequences. For example, Ling et al. [127] derived that the expected heterozygosity at time t (denoted by H_t) during tumor growth can be expressed by:

$$H_t=1 - \frac{e^{-2u}}{N_t - 1} - \sum_{j=2}^t \left(\frac{e^{-2uj}}{N_{t+1-j} - 1} \prod_{i=0}^{j-2} \left(1 - \frac{1}{N_{t-i} - 1} \right) \right) \quad \text{Eq.(2)}$$

where N_t is tumor size at time t starting with a single transformed cell at time 0 and assuming a well-mixed neutrally growing population of synchronous cell generations. Here, u is the mutation rate per cell division for the 30Mb exonic coding regions, namely $u=\mu \times 3 \times 10^7$ (e.g. corresponding to $u=0.03$ when the per-base pair mutation rate is $\mu = 10^{-9}$). A cancer genome can diversify rapidly during tumor growth as illustrated in Figure 3 for a

logistic growth model. In particular, $H > 0.8$ when the tumor is barely detectable at 1 million cells. Thus, tumor growth may accompany rapid genetic diversification and a small tumor can already harbor extensive diversity. This prediction is consistent with recent findings that early tumors display high-levels of genomic diversity [33, 46, 178].

Genetic heterogeneity within and between regional samples has also been quantified using the Shannon index [179]:

$$H = - \sum_i f_i \times \ln(f_i) \quad \text{Eq.(3)}$$

where f_i is the frequency of the allelic type or cell type, i . For example, the Shannon Index was used to measure microsatellite and copy number heterogeneity based on FISH in Barrett's esophagus where high diversity was predictive of progression to esophageal cancer [26]. It has similarly been used to and to quantify neutral methylation tag and copy number diversity in colorectal tumors [21, 37], where high ITH was observed at different genomic levels within and between single glands. The Shannon index was also used to quantify genetic diversity based on immunoFISH of paired primary, lymph node and distant metastatic breast tissue [113], where diversity was found to be highest in distant metastases. Comparisons of pre versus post neoadjuvant treated breast tumors suggests that lower pre-treatment diversity was associated with pathologic complete response, whereas levels did not change in tumors that failed to respond to treatment [20]. Importantly, the extent of ITH was comparable irrespective of the chromosomal region analyzed. A survey of various measures of clonal diversity was reported by Merlo et al. [180].

More recently, the MATH index which corresponds to the ratio of the median absolute deviation (MAD) and the median of the VAF values was described [181] and used to quantify ITH in head and neck squamous cell carcinoma (HNSCC), where it was shown to correlate with worse outcome [27].

$$MATH = 100 \times \frac{MAD(VAF)}{Median(VAF)} \quad \text{Eq.(4)}$$

Other approaches to quantify ITH include determination of the number of subclones and their frequency based on shared CCF values [149, 150], although this requires additional estimates and assumptions, as outlined in the preceding section. Zhang et al. found that lung adenocarcinomas with higher subclonal mutational burden ($n=3/11$ cases) were more likely to relapse [29]. However, this was again based on a limited number of cases. Additionally, in post-chemotherapy treated pediatric kidney cancer, tumors with microdiversity exhibited worse disease-specific survival [182]. Other studies have reported an association between survival and various measures of ITH independent of other clinical and molecular variables in diverse tumor types [183, 184]. However, these data suggest a non-linear relationship between the number of subclones and outcome, thereby complicating the interpretation.

In summary, several studies have suggested that ITH is associated with disease progression [26] and poor prognosis [27, 28], raising the possibility that ITH may represent a general biomarker. However, lack of association is seldom reported and further studies are needed. Moreover, as ITH is already present in early lesions, it is of interest to determine whether its predictive utility depends on disease stage. Finally, it will be important to understand which genomic features and measures of ITH are most informative, and this may depend on the underlying modes of tumor evolution, which are as of yet poorly understood.

4. Evolutionary forces determining the extent of genetic heterogeneity

Mutation, genetic drift, migration and selection are the fundamental forces that collectively shape genetic diversity within a population [185]. Clonal evolution, the process of mutation accumulation and adaptation in somatic cells, likely involves each of these forces, although their collective impact on ITH is seldom considered. Here we discuss these factors in the context of classical population genetic theory and recent cancer genomic data.

4.1 Mutation

Mutation is the primary source of genetic diversity, the rate of which is a key parameter in population genetics studies [185, 186]. Although cancer genome sequencing efforts have provided a catalogue of somatic alterations across different tumor types [83, 84], the per-cell division mutation rate is not known and likely varies during the course of disease progression. Mutational burden can also vary by orders of magnitude within and between tumor types [83, 84]. Although the somatic mutation rate is generally assumed to be elevated relative to the germline [186, 187], even germline mutation rates can vary considerably within families [188]. Additionally, the number of germ-cell divisions differs in males ($\sim 10^2$ divisions) and females (20–30 divisions) since ovum undergo two meiotic and 22 mitotic cell divisions and reach maturity prior to birth, whereas in males the number of germ cell divisions increases with age [189]. Nonetheless, the mutation rate for single nucleotide variants (SNVs) in the human germline has been estimated to be $\sim 10^{-8}$ per base pair per generation [188] or $\sim 10^{-10}$ per base pair per cell division, whereas human somatic cells have been reported to accumulate 4 to 25 times more mutations than germline cells, corresponding to an estimated mutation rate of $\sim 10^{-9}$ per cell division per base pair [186, 187]. However, there is no consensus regarding the mutation rate in human tumors. The finding that some tumors exhibit high genomic instability led to the mutator hypothesis, which posits that an elevated mutation rate is a prerequisite of tumor development [96, 190, 191]. While mutator phenotypes likely accelerate tumor growth and are most effective if acquired early, they do not appear to be universally required for tumorigenesis [192, 193].

More generally, the extreme variation in cancer risk across tissues remains poorly understood. While it has been proposed that this is due to differences in the lifetime number of adult stem cell divisions, during which DNA replication errors occur [194], this simple explanation has been debated [195–197]. Recent studies have sought to characterize mutation rates in diverse mouse tissues [170] and human adult stem cells [198] by sequencing of normal clonal organoid cultures derived from primary multipotent cells. Intriguingly, the initial mouse study reported differences in the mutation rate across tissue,

as well as distinct mutational processes, whereas the subsequent study similarly suggest tissue-specific activity of mutational processes but a steady and consistent accumulation of mutations (approximately 40 de novo mutations per year) in the human small intestine, colon and liver. Hence, further investigation is needed, but this approach suggests new ways to quantify human somatic mutation rates.

4.2 Genetic drift

Genetic drift describes random fluctuations in the frequency of genetic variants in a population and is a major driving force of genetic diversity especially for small populations [176, 185]. The neutral evolution theory posits that most fixed nucleotide substitutions in a population are driven by random genetic drift rather natural selection [199]. The neutrality hypothesis has been foundational for statistical tests of neutrality and computational analyses of natural selection from population sequencing data. However, the role of genetic drift in cancer is often overlooked as selection is generally evoked to explain most aspects of tumor evolution. For instance, it is usually assumed that detectable subclones are under strong selection [13, 49, 200]. However, theory predicts that early neutral mutations during tumor growth may also be present at high frequency at the time of diagnosis since random drift predominates when tumor size is small. Much of the theory concerning genetic drift derives from the Wright-Fisher or related models, which assume constant population sizes (N genome copies) [185]. A well-known prediction is that the probability of fixation in the population for a new neutral mutation is $\rho = 1/N$. For a growing population, the probability of fixation of a new mutation depends on both the time of occurrence and the birth/death rate. Assuming a branching process, Bozic et al. [201] derived that the probability of fixation of k -th surviving lineage is given by:

$$\rho_k = (u/u - \log(\delta))^k \quad \text{Eq. (5)}$$

where u is the neutral exonic mutation rate per cell division and δ the ratio of cell death and birth rates. Although, there is limited experimental data, the ratio of death and birth rates in cancer has been estimated to range from $\delta=0.72$ in aggressive CRC metastases [202] to $\delta=0.99$ in early tumors [203]. It can be seen from this formula that early mutations can drift to fixation (clonality) with high probability during the early stages of clonal expansion when the cell death rate is high such that: $\rho_1 = 0.22$, $\rho_2 = 0.05$ and $\rho_3 = 0.01$ for $u=0.03$ ($\mu=10^{-9}$) and $\delta=0.9$. Neutral mutations may therefore be present at high frequency in a tumor and detectable by NGS. Thus, while genetic drift is a negative force that decreases whole-population diversity, it simultaneously increases the diversity of high frequency mutations due to random fluctuations (Figure 4) and hence is an important, but often overlooked, feature of tumor sequencing data.

4.3 Spatial structure and migration

Spatial structure and migration can dramatically alter the genetic diversity of a population [204, 205]. In a spatially structured population, genetic drift and selection are restricted in localized regions thus giving rise to decreased genetic diversity within a local subpopulation while increasing genetic divergence between spatially segregated subpopulations. As shown

in Figure 5A, spatial partitioning within a tumor naturally gives rise to genetic divergence between samples from distant regions. This genetic segregation can be enhanced by local stochastic extinctions due to large demographic fluctuations at the expansion front and, hence explains spatial heterogeneity in tumors [206]. These patterns are especially relevant for interpreting NGS data from solid tumors characterized by significant spatial constraints since biopsies and even resections are often obtained from a localized region of the tumor. Such local sampling is inherently limited in its ability to capture much of the (detectable) genetic diversity in a tumor (Figure 5). Recent multi-region sequencing (MRS) studies have revealed extensive genetic diversity between different samples from the same tumor suggesting that spatial constraints are significant in many tumor types [13, 16, 18, 21, 23, 29, 30, 127, 207]. Hence, tumor evolution may be driven by adaptations to distinct microenvironmental niches amongst spatially partitioned populations.

Spatial structure thwarts selection and promotes the co-existence of multiple selective clones when selection is pervasive. Martens et al. [208] modeled crypt-structured tissues to study the effect of spatial structure on colorectal tumor initiation where selection is expected to be stringent. They report that spatial structure significantly increases the waiting time to transformation due to elevated clonal interference between selectively advantageous clones compared with that in well-mixed populations such as leukemia. More generally, they proposed that in somatic cells, the damaging effect of deleterious mutations can be mitigated by differentiation and apoptosis and that tissue structure minimizes the risk of cancer [208]. However, the contributions of spatial structure to these dynamics are complex [209] and warrant further investigation. For example, spatial structure does not always impede selection. In contrast, it can accelerate complex adaption, a scenario where adaptation requires the accumulation of multiple mutations (e.g. a double hit to inactivate a tumor suppression gene), depending on the epistatic fitness landscape [210].

Cancer stem cell (CSC) organization within tumors promotes spatial structure and increases phenotypic heterogeneity [211, 212]. Due to the limited proliferative potential of non-stem cells, only mutations in stem cells can result in clonal expansions within the niche, regionally in the tumor, unless the mutations promote dedifferentiation of non-stem cells to stem cells [213]. CSC-driven tumor growth likely impedes mutational adaptation due to the decreased effective population size. Microenvironmental heterogeneity can also promote the emergence of spatial structure in tumors [214] and the plasticity of cellular phenotypes that are often regulated by epigenetic factors. Importantly, the efficiency of selection resulting from genetic mutations would be dampened under high levels of phenotypic plasticity [215].

Given the spatial constraints within solid tumors, spatial information is important for resolving the evolutionary history of a tumor and the timing of mutation occurrence. For instance, without knowledge of sampling locations, subclonal composition can be explained by differences in tumor growth dynamics, stringent spatial constraint (Figure 5A) or early subclone mixing (Figure 5B). However, knowledge of spatial location can help to constrain these scenarios. For example, if samples R1 and R2 are spatially separated, early clonal mixing is more likely than a stringent spatial model. Therefore, the integration of spatial sampling information within a computational modeling framework (Figure 5C) can aid the inference of tumor evolutionary dynamics.

Cell migration has a dual role in the generation of tumor genetic diversity. On the one hand, within-tumor migration relaxes spatial constraints, thereby augmenting the efficiency of selection, giving rise to lower genetic diversity within a tumor. On the other hand, out-of-tumor migration and metastasis will increase the genetic divergence between tumors (Figure 5D). By simulating spatial tumor growth with cell migration, Waclaw et al. [216] showed that out-of-tumor migration can dramatically accelerate tumor growth and increase tumor diversity. Recently, comparative genomic sequencing studies revealed genomic divergence between paired primaries and metastasis [34, 80, 217] and multi-focal tumors [23, 218], suggesting that migration drives tumor diversification.

4.4 Selection

4.4.1 Positive selection—Cancer is traditionally viewed to result from the stepwise accumulation of mutations, where cells transition from a normal health state to pre-malignant, malignant and migratory phenotypes. The process begins when a single somatic cell acquires a heritable mutation that confers a fitness advantage, resulting in an increased proliferation or survival phenotype that enables it to outcompete less fit cells *within* the population. Natural selection operates on this phenotypic diversity, leading to sequential waves of clonal expansions and heterogeneity amongst subclones [7]. Within the current view, ongoing strong selection is assumed to be the primary force and to govern all stages of tumor progression from initiation to subsequent primary tumor growth, as well as metastasis where the acquisition of additional ‘drivers’ results in continual selective sweeps [8, 36, 84]. In contrast, mutation genetic drift, population structure, and migration are often overlooked, although they represent key evolutionary forces that influence the expansion, fixation and extinction of subclones, as outlined in the preceding sections. The population dynamics and distribution of passenger, advantageous and deleterious mutations can be understood through population genetics models, where the relative importance depends on key parameters such as population size. For example, genetic drift within a small population can render selection less efficient, by allowing for the fixation of deleterious mutations and loss of advantageous mutations.

Various computational methods have been developed to identify ‘driver’ mutations that confer a selective growth advantage (reviewed in [219]). The majority of approaches evaluate whether a gene is altered more frequently than expected by chance based on patterns of recurrence across tumors, where it is critical to account for replication timing and variable mutation rates throughout the genome (or in hypermutated tumors) that can otherwise bias these patterns [83, 220]. In contrast to these cohort-based approaches to define putative driver genes, formal methods to detect selection in an individual tumor have been described in only a few instances [21, 221]. For example, Sottoriva et al. [21] employed a spatial computational model of tumor growth and statistical inference framework to infer subclone fitness differences. Martincorena et al. [221] evaluated average mutant clone sizes and the dN/dS ratio, a measure of the relative abundance of non-synonymous to synonymous mutations, and report evidence for strong positive selection for mutations in normal skin, although the interpretation has been debated [222, 223]. More generally, it has been assumed that the presence of subclonal known or putative drivers implies selection [13, 29, 30, 164]. For example, in a breast cancer subject to high-depth WGS high frequency subclonal

mutational clusters were observed [164], compatible with the clonal expansion of a driver mutation accompanied by numerous hitchhiking passengers. Although this is a reasonable assumption, the presence of subclonal ‘driver’ mutations alone does not necessarily guarantee that the mutation conferred a fitness advantage. Fitness estimates for specific mutations are unknown in human tumors, but mutations can be context dependent and epistatic [224, 225] with differential effects during tumor initiation versus subsequent expansion, potentially confounding the interpretation of such patterns. Moreover, many putative ‘drivers’ have yet to be functionally and phenotypically characterized. Another strategy is to evaluate evidence for convergent evolution, as has been done in ccRCC [13, 18, 226], GBM [57] and CLL [58], providing evidence for selection.

4.4.2. Negative selection—Negative selection by deleterious mutations is pervasive in the evolution of natural species and can significantly decrease population diversity through purifying selection [227, 228]. Classic molecular evolution theory posits that most newly arising mutations are slightly deleterious and purged from the population quickly [185, 199]. However, the extent to which this applies to somatic cell evolution is unclear since tumors can harbor high levels of genomic instability in which alterations might be deleterious [229]. Indeed, as discussed in Section 2.2, chromosomal aberrations can impact cellular fitness [89, 90, 97].

Given the complete genomic linkage of asexually reproducing populations (due to the lack of recombination) such as tumor cells, interference between positive and negative selection (known as the Hill-Robertson effect) [230] may be particularly strong when accompanied by high mutation rates [231]. Intriguingly, McFarland et al. [232, 233] found that deleterious mutations significantly alter tumor progression by impeding the efficiency of positive selection. They further showed via simulation studies that some pre-malignant lesions die out due to negative selection, suggesting the possibility of novel therapeutic strategies through dramatic elevation of the mutation rate. It is tempting to speculate that the apparently paradoxical relationship between variable levels of genomic instability and cellular fitness tradeoffs is compatible with mutational meltdown and Muller’s Ratchet [234, 235], wherein deleterious alterations drive population extinction. While this may be particularly acute for small asexually reproducing populations, it is not whether this applies in highly adaptive cancer cell populations or how it is influenced by the interplay between deleterious and advantageous mutations [233].

4.5 Evidence for different modes of tumor evolution

4.5.1 Pre-tumor progression: multiple clonal expansions to transformation—It has been appreciated for nearly two decades that more than half of all somatic mutations in colorectal cancers occur prior to transformation [1, 2]. Critically, the *initiating* (epi)genetic events provide a selective fitness advantage in the target cell relative to other pre-malignant cells, resulting in clonal expansion. Initiating events are often tissue type specific as is the case for *APC* in gastrointestinal tumors [3–5] and *DNMT3* in leukemias [6], suggesting that they are beneficial only in certain cellular contexts. It has long been assumed that the cell of origin may be a self-renewing stem cell and that this is the “unit” on which selection operates in cancer [8, 236], where cancer reflects a state in which the normal balance of self-

renewal and differential is lost. This is supported by observations that higher stem cell activity is associated with poor prognosis in multiple tumor types [9]. The subsequent expansion of a subclone depends on the rates of advantageous and deleterious mutations, the fitness benefit they confer in a given background population and the effective population size.

Several studies have characterized genetic diversity in precursor or pre-invasive lesions, including the seminal study by Maley et al. [26], demonstrating that clonal diversity predicts progression from Barretts esophageal to esophageal adenocarcinoma (ESCA). More recently, WGS and targeted sequencing was conducted to delineate mutational ordering in pre-invasive esophageal lesions [237], and reported that mutations in ESCA occur very early during disease development, where several additional drivers occur during later stages of progression with both diagnostic and therapeutic implications. Pre-tumor progression in the colon has also been studied by examining clonal evolution in stem cell populations [238, 239]. In particular, Baker et al. [239] reported that the number of functional stem cells is 5–6 in both normal patients and individuals with familial adenomatous polyposis (FAP) associated with germline APC mutation. In contrast, in adenomatous crypts (APC(–/–)), an increase in both functional stem cell number and the loss/replacement rate was noted as was the rate of division for colonic crypts. Other studies have sought to experimentally derive fitness measurements of common drivers (such as *APC* and *KRAS*) during tumor initiation in the mouse small intestine and report that while they each provide strong selective advantages, mutation fixation is highly inefficient due to hierarchical tissue architecture, strong genetic drift and stochastic stem cell replacement [240]. Multiple studies have revealed the early acquisition of tumor heterogeneity in non-invasive adenomas [21, 45, 47, 241]. Breast tumors also have well defined precursor lesions, facilitating studies of tumor initiation and early growth, where genetic diversity was similarly reported to occur early [178] and reconstruction of phylogenetic trees amongst pre-invasive lesions revealed lineage heterogeneity both within and between lesions [33].

Clonal labeling techniques have also been employed to delineate mechanisms of tumor initiation and growth by genetically tracking the fate of clones [242–244]. In particular, Dreissens et al. [242] reported the identification of two populations in benign papilloma cells, corresponding to a small, but persistent stem-cell-like population that rapidly divides and a slower cycling transient population that gives rise to terminally differentiated tumor cells. In parallel, mathematical modeling of tumor initiation has sought to determine the number of drivers required to initiate a tumor [245], suggesting that only three events are required for lung and colorectal cancer. As the timing of genomic events that occur prior to transformation is obscured when studying established tumors (for which the majority of data is available), it will be of continued importance to characterize precursor lesions towards identification of the cell of origin, drivers and temporal ordering of tumor initiating events as may inform treatment and prognostication.

4.5.2 Linear versus branched (selective) evolution—As described above, cancer genome sequencing efforts have revealed extensive ITH in diverse tumors. Although relatively few primary tumors of a given tissue type have been analyzed in detail via multi-region or single cell sequencing, there have been multiple reports of branched evolution [13,

17, 18, 22, 25, 29, 30, 33] (Figure 1C). Hence, these data contrast with the early view that after transformation, subsequent tumor growth typically follows a linear clonal succession model [9] (Figure 1B), wherein a selectively advantageous mutation has time to sweep through the population prior to the acquisition of additional advantageous mutations. Rather, branched clonal evolution occurs when multiple selectively advantageous subclones are present and compete via clonal interference before fixation (Figure 1C). It has been postulated that branched evolution may be indicative of adaptation to different microenvironments or that there are several routes from a normal cell to an advanced cancer [246].

It is worth noting that the observed “branched” phylogenies do not imply branched evolution (phylogenies are branching diagrams). The finding that many tumors exhibit such patterns should not be surprising since individual cells will clearly acquire distinct mutations when both selectively advantageous and neutral (passenger) mutations are considered. Moreover, in spatially structured populations, such as solid tumors, there is a high chance of detecting genetic divergence between regionally segregated samples due to enhanced genetic drift in localized regions.

4.5.3 Evidence for effectively neutral tumor evolution after transformation—

Selection is necessary for tumor initiation as the stepwise process of mutation, selection and clonal expansion is what gives rise to a cell that harbors the constellation of events required for unchecked growth and which becomes the founding tumor cell (Figure 1). While it is generally assumed that strong selection is operative and detectable throughout all stages of tumor progression, this has not been systematically examined. Rather, recent findings have challenged the notion that strong positive selection is continuously operative during the growth of a primary tumor after transformation [21, 127, 135]. For example, Sottoriva et al. [21] described a Big Bang model of colorectal tumor growth, wherein after transformation, the tumor grows as a single terminal expansion populated by a large number of heterogeneous subclones that are not subject to stringent selection. As such, in the Big Bang model, the timing of a mutation is the primary determinant of its frequency in the final tumor and most detectable subclonal alterations (resulting in ITH) arise early during growth, whereas late-arising mutations will be largely undetectable. The predictions of the Big Bang model were tested through single gland and bulk tumor profiling as well as spatial computational modeling and statistical inference on the genomic data. Intriguingly, inference on the genomic data in a patient-specific fashion revealed signals of selection (corresponding to subclone fitness differences) in the majority of carcinomas examined. However, selection was generally not sufficiently strong to result in *hard* selective sweeps that would have dramatically altered subclonal architecture (by altering clone frequencies), compatible with *effectively* neutral evolution.

Subsequent studies have corroborated aspects of ‘Big Bang’ dynamics [46, 47, 247], suggesting that they may be relatively common in colorectal tumors. Others have considered more strict models of neutrality. For example, Ling et al. reported neutral evolution in hepatocellular carcinoma via in depth multi-region profiling [127]. Borrowing established approaches from population genetics, they analyzed the site frequency spectrum (SFS) [248] in a single liver tumor where 24 samples were subject to WES and 300 samples were deeply

genotyped and found that the observed data could not reject the expected distribution arising under a neutral exponential growth model, assuming a well-mixed population. They also reported very high levels of genetic diversity, where the heterozygosity between samples was estimated $H=0.941$. Williams et al. [135] similarly employed a neutral exponential growth model and attempted to fit somatic mutation data from bulk tumor samples in TCGA, resulting in the claim that 1/3 of tumors are compatible with neutrality.

4.5.4 Punctuated evolution—Several recent studies have reported punctuated clonal evolution [31, 32, 83], karyotypic chaos [249], and other catastrophic events in various tumor types, providing further evidence that the linear, sequential clonal evolution model does not accurately describe the patterns of ITH found in many human cancers. For example, chromothripsis, a cataclysmic event involving surges of chromosomal rearrangements that occur in a single or few cell cycles, has been reported in medulloblastoma, osteosarcoma, pancreatic cancers, and other tumors [125, 250, 251]. By generating multiple driver alterations in a single cell division, a greater fitness benefit may be achieved than would be feasible through sequential alterations [252]. Chromoplexy is another type of catastrophic event, wherein chains of translocations are formed across multiple chromosomes, potentially due to double-stranded DNA breaks [253]. In contrast, Kataegis results in localized hypermutation, preferentially at cytosine in TpC dinucleotides [254, 255]. Collectively, these studies suggest that mutation rates are not constant during the course of tumor progression and highlight the importance of understanding a tumor's evolutionary trajectory (including the distinct stages of initiation and subsequent growth) in light of multiple sources of genomic alterations, including SSNVs, CNAs and SVs.

5. Evolutionary strategies to exploit intra-tumor heterogeneity therapeutically

5.1 Understanding, predicting and preventing tumor progression

Although tumor growth cannot be directly observed, patterns of genetic diversity present in a lesion reflect the evolutionary forces that gave rise to it. It is therefore not surprising that ITH since a tumor's future behavior depends on its evolutionary course. In theory, the earlier the initial clonal expansion is detected, the less diverse and less fit the cell population (since natural selection operates on diversity), and the better the prognosis. Even relatively early tumors exhibit high genomic diversity [33, 46], which may reflect processes that were already operative for some time and likely will continue in the absence of intervention. The term “early” is indeed relative as the first clonal expansions will be invisible and the lesion must be large enough to be detected and sampled. In line with this view, we have previously proposed that some tumors may be ‘born to be bad’ [21], wherein invasive potential may be specified early. We have also observed that many private mutations originate from the first few cell divisions in human colorectal adenomas [45]. These findings suggest that targeting early clonal alterations in the relevant target cells and/or perturbing the microenvironment may alter the evolutionary trajectory of some tumors. As discussed above, numerous forces shape tumor diversity and it is important to appreciate these contributions even when they cannot be disentangled. More generally, understanding the mode of tumor evolution may help to pinpoint the drivers of this process. This will require a combination of detailed

‘omic’ analyses of clinical samples and relevant model systems, as well as computational modeling to delineate the dynamics of this process.

Recently, several studies have described the use of computational and mathematical modeling to gain insight into complex phenomena during tumor evolution [21, 256], as well as strategies to prevent unwanted clonal evolutionary processes [257–260]. For example, patterns of ITH can be analyzed using genealogy-based statistical inference methods within spatial computational models of tumor growth to infer patient-specific tumor parameters such as the mutation rate, subclone fitness differences, and mutational timeline [21]. This approach was employed to test the predictions of a Big Bang model of colorectal tumor growth. As it takes as input genomic data and requires minimal assumptions about the underlying process, it can also reveal unexpected features in the data.

Other efforts have sought to outline principles for preventing cancer by exploiting host defenses against clonal evolution such as the immune response, inhibition of angiogenesis and replicative senescence [258]. For example, analysis of telomerase length regulation [261] illustrates the need for mathematical modeling in order to understand the dynamics of clones that could be restricted by replicative limits, as well as the efficacy of anti-telomerase therapies. Here, the authors demonstrate that knowing the replicative capacity of the founding tumor cell alone is insufficient as it is also essential to understand the distributions of the number of cells, the balance between cell division and death, as well as the probability of escaping replication senescence through telomerase activation. These phenomena are unlikely to be understood from experimental data alone; hence this highlights the importance of integrated experimentation and quantitative modeling.

Once tumor cells escape from natural defense mechanisms and achieve uncontrolled growth, disease progression inevitably ensues although the latency can vary considerably. Tumor reduction is often attempted via surgery, cytotoxic and/or cytostatic chemotherapy, and in some cases, molecular targeted agents. However, therapeutic resistance is virtually universal in solid tumors. Therapy itself can impose new selective pressures on the tumor cell population and it will be important to optimize treatment strategies by exploiting evolutionary principles and context-dependencies, as discussed below.

5.2 Evolutionary and ecological strategies to circumvent resistance

The fact that cancer is a clonal evolutionary process poses challenges for its effective control. Resistance can arise via several mechanisms, including the selection of pre-existing resistant subclones, *de novo* or acquired resistance, and “competitive release”, wherein elimination of treatment sensitive cells results in a substantial reduction of tumor burden and a small population where resistant clones can readily grow out without competition [262]. Hence, the interplay between cancer cell population size and their ecological niche impacts treatment response. Competitive release may be a common phenomenon given that conventional treatment strategies aim to achieve maximal tumor cell kill under the assumption that this is optimal for patient outcomes, whereas it may actually promote resistance [257]. To circumvent this, Basanta et al. [257] proposed a game theoretical framework based on an evolutionary double bind where two therapies are used sequentially such that resistance to one drug comes at a fitness cost, leaving the cells more sensitive to

the other drug. They then investigated the impact that this strategy has on cancer control. A related approach, termed adaptive therapy, administers treatment in pulses wherein removal of the drug results in a fitness penalty to the resistant cell population, thereby keeping the population size in check. Adaptive therapy was found to yield extended tumor control in preclinical breast cancer models [260] with efforts underway to evaluate this in clinical trials.

Some of the early efforts to understand resistance employed mathematical modeling to optimize drug dosing, for example in the context of EGFR-mutant NSCLC [263]. Other studies sought to explore the evolutionary dynamics of treatment response. For example, Bozic et al. demonstrated that combination targeted therapy is superior to sequential therapy and can lead to long-term disease control, but even one potential cross-resistance mutation will dampen its benefit [264]. Others have focused on understanding the impact of ITH on combinatorial treatment strategies. For example, Zhao et al. demonstrated that drugs selected as being the most beneficial single agent on the basis of clonal dominance are often suboptimal for treating heterogeneous tumors via combination therapy [265]. Rather, ITH reduces the advantage that a targeted therapy has on a specific cell population, thereby altering the efficacy of intuitive combination therapy. Zhao et al. subsequently sought to exploit temporal collateral sensitivity to circumvent resistance [259]. They report the existence of a predictable and opportunistic therapeutic window during a tumors' evolutionary trajectory and determine mechanisms of drug sensitization through a combination of genomic, phenotypic, signaling, and binding assays in conjunction with computational modeling to identify novel vulnerabilities. This integrated evolutionary and pharmacological approach highlights the potential to develop evolutionary informed treatment strategies with clinical benefit.

Collectively, these studies highlight key considerations for therapeutic strategies in light of ITH and suggest rational approaches for the design of drug combinations and dosing. Ultimately, strategies such as these that harness both evolutionary and ecological principles to target tumors may have the greatest benefit [206]. While most aspects of tumor ecology are hard to probe *in vivo* and therefore poorly understood, new technologies promise to enable fresh insights into tumor microenvironment and spatial structure, which can in turn, inform these models. It is also clear that the full armament of strategies to combat tumor progression have yet to be explored. As these efforts mature, it will be important to continue to examine model assumptions concerning tumor dynamics, as well as to generate testable predictions and to experimentally evaluate them. A combination of iterative experimentation and computational modeling should help to solidify this framework.

5.3 Implications of ITH for clinical trial design and personalized medicine

Molecularly targeted therapies have revolutionized clinical oncology and the discovery of new targets and biomarkers for patient stratification remains an active area of research. In the era of targeted therapies, the extensive evidence for ITH in diverse tumors accompanied by evidence for pre-existing resistance conferring mutations raises significant concerns about the implications for achieving a durable response [60]. In the metastatic setting, the acquisition of drug resistance can be particularly rapid. To date, multiple examples of such

pre-existing subclonal resistant mutations have been described, including KRAS mutations in the context of anti-EGFR targeted therapy in colorectal cancer [202, 266, 267], rare MEK1 mutations in BRAF mutant melanomas receiving BRAF targeted therapy [268], chemo-resistant TP53 mutations in acute myeloid leukemia [269]. These clinical examples are corroborated by *in vitro* studies using multiplexed barcode libraries to enable the detection of extremely low frequency variants that confer resistance [270], as well as by mathematical modeling, which predicts the presence of pre-existing resistant subclones [202, 264]. Complicating matters further, acquired resistance can occur in the context of treatment selective pressure, as exemplified by convergent loss of PTEN during PI(3)K α inhibition in breast cancer [271] and secondary EGFR mutations in NSCLC during anti-EGFR targeted therapy [272]. Non-genetic mechanism can also drive resistance, and have recently begun to be explored [109, 111].

Given the diversity of mechanisms a cell can employ to evade therapy, the emergence of resistance should be anticipated and actively managed. Subclonal events pose a particular challenge as they fall below the limits of most diagnostic tools, and hence are not generally considered for treatment stratification. The earlier detection of these variants could inform patient-tailored adaptive treatment strategies, particularly if this can be achieved in a minimally invasive manner via liquid biopsy. Increasing data suggest that circulating tumor DNA (ctDNA) can be used to detect the emergence of resistance [273, 274], monitor metastatic progression [275], predict recurrence and track minimal residual disease [276]. Such liquid biopsies may also help to overcome some of the challenges associated with spatial tumor heterogeneity. However, this will require detailed comparison of detailed MRS and ctDNA. In the metastatic setting it will also be important to understand whether distant metastases shed ctDNA at the same rates. For example, metastases to the brain (and primary brain tumors) are unlikely to shed significant cell free DNA into the blood due to the blood-brain barrier. Here, cerebrospinal fluid may provide an alternate, albeit less accessible, reservoir of ctDNA [277].

In practice, the clinical community is grappling with how to detect subclonal somatic alterations, whether the presence of a subclonal resistance mutation detected at low frequency should necessitate the use of an alternate therapy (should one exist) [267] and whether actionable clonal mutations in an advanced tumor are optimal targets¹⁷². As discussed above, quantitative modeling can make predictions concerning both of these scenarios [259, 265] and it will be important to pursue these approaches further. It is also of interest to evaluate clinical actionability and utility in light of VAF values and quantitative measures of ITH. The TRACERx (TRACKing non-small cell lung Cancer Evolution through therapy [Rx]) study, which aims to prospectively study cancer evolution via longitudinal and spatial sampling, coupled with analysis of ctDNA and circulating tumor cells (CTC) should help to address some of these questions [278]. A chief aim of this effort is to determine whether the clonal dominance of molecularly targetable alterations portends improved progression-free survival in advanced disease. In parallel, this study seeks to identify high-risk subclonal 'drivers' implicated in metastatic progression and to track clonal dynamics in the context of therapy.

A systematic understanding of ITH will be particularly important for the improved management of metastatic disease, which is the ultimate cause of death for most patients. To date, therapeutic advances have yet to be fully translated to the metastatic setting. Here, the analysis of genetic diversity is particularly complicated due to the challenge in obtaining clinically annotated paired primaries and metastases, the common administration of intervening therapy, and potential for differences in purity. Moreover, the extent of genetic divergence between primaries and metastases may vary according to tissue of origin with implications for the relevant targets of systemic therapy. For example, colorectal cancers routinely share canonical drivers in the primary and metastasis [66, 67, 69–71, 75–77, 279], whereas greater divergence and *de novo* putative drivers have been identified in other solid tumors [34]. To date, few studies have performed comprehensive genomic profiling of paired primaries and metastases with detailed treatment information (as therapy can confound the interpretation of the resultant genomic patterns) and the limited available data derives from small numbers of cases. Thus, additional efforts to profile metastatic disease are needed and should similarly account for the spatial sampling biases reported in primary tumors. Only then can rigorous comparisons of paired primaries and metastases be made with the potential to reveal new therapeutic vulnerabilities and to inform the dynamics of metastatic colonization [280].

It is clear that advances in precision oncology will require tight coordination between laboratory scientists, computational biologists, oncologists and clinical trial communities, as well as patient engagement. Despite the challenges that pervasive ITH and clonal dynamics pose, they beckon for evolutionary and ecological therapeutic strategies, which have only begun to be explored. Hence, there is an armament of tools that were developed to understand population dynamics for which further investigation in the context of cancer is warranted towards the goal of improving patient outcomes.

6. Conclusions

In this review, we have outlined some of the recent advances in characterizing ITH, as well as technical and methodological challenges in quantifying and interpreting these patterns. Although much focus to date has been on genotypic heterogeneity, emergent techniques may facilitate the characterization of heterogeneity *in situ* [281]. In parallel, functional heterogeneity amongst tumor cells can be assessed in patient-derived xenograft [282, 283] or organoid [284, 285] models. Moreover, clinical imaging techniques may enable non-invasive, longitudinal monitoring of ITH at a whole organ level [286]. Collectively, these approaches will help to resolve the genotype to phenotype map in cancer. We have also highlighted the power of interpreting tumor cell diversity within a population genetics framework and the need for mathematical and computational models to infer tumor dynamics. The holy grail of such an approach would be the development of predictive models to *forecast* and *steer* tumor evolution in order to achieve durable patient responses, thereby rendering cancer a chronic disease. Before this can be achieved a systematic understanding of the different modes of evolution and fitness landscapes is needed. It is also evident that the failure to accurately detect subclonal alterations will impede precision medicine. Hence, it remains of central importance to quantify both spatial and temporal genetic heterogeneity towards the goal of improved predictive and prognostic biomarkers.

Acknowledgments

This work was funded by awards from the NIH (R01CA182514), Susan G. Komen Foundation (IIR13260750), and the Breast Cancer Research Foundation (BCRF-16-032) to C.C. Z.H. is supported by an Innovative Genomics Initiative (IGI) Postdoctoral Fellowship.

References

1. Tsao JL, et al. Genetic reconstruction of individual colorectal tumor histories. *Proc Natl Acad Sci U S A*. 2000; 97(3):1236–41. [PubMed: 10655514]
2. Tomasetti C, Vogelstein B, Parmigiani G. Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proc Natl Acad Sci U S A*. 2013; 110(6):1999–2004. [PubMed: 23345422]
3. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell*. 1990; 61(5):759–67. [PubMed: 2188735]
4. Nakatsuru S, et al. Somatic mutation of the APC gene in gastric cancer: frequent mutations in very well differentiated adenocarcinoma and signet-ring cell carcinoma. *Hum Mol Genet*. 1992; 1(8):559–63. [PubMed: 1338691]
5. Horii A, et al. Frequent somatic mutations of the APC gene in human pancreatic cancer. *Cancer Res*. 1992; 52(23):6696–8. [PubMed: 1423316]
6. Shlush LI, et al. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature*. 2014; 506(7488):328–33. [PubMed: 24522528]
7. Nowell PC. The clonal evolution of tumor cell populations. *Science*. 1976; 194(4260):23–8. [PubMed: 959840]
8. Greaves M, Maley CC. Clonal evolution in cancer. *Nature*. 2012; 481(7381):306–13. [PubMed: 22258609]
9. Greaves M. Evolutionary determinants of cancer. *Cancer Discov*. 2015; 5(8):806–20. [PubMed: 26193902]
10. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011; 144(5):646–74. [PubMed: 21376230]
11. Heppner GH. Tumor heterogeneity. *Cancer Res*. 1984; 44(6):2259–65. [PubMed: 6372991]
12. Anderson K, et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature*. 2011; 469(7330):356–61. [PubMed: 21160474]
13. Gerlinger M, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*. 2012; 366(10):883–92. [PubMed: 22397650]
14. Shah SP, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*. 2012; 486(7403):395–9. [PubMed: 22495314]
15. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer*. 2012; 12(5):323–34. [PubMed: 22513401]
16. Sottoriva A, et al. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci U S A*. 2013; 110(10):4009–14. [PubMed: 23412337]
17. Bashashati A, et al. Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *J Pathol*. 2013; 231(1):21–34. [PubMed: 23780408]
18. Gerlinger M, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nature Genetics*. 2014; 46(3):225–33. [PubMed: 24487277]
19. Johnson BE, et al. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science*. 2014; 343(6167):189–93. [PubMed: 24336570]
20. Almendro V, et al. Inference of tumor evolution during chemotherapy by computational modeling and in situ analysis of genetic and phenotypic cellular diversity. *Cell Rep*. 2014; 6(3):514–27. [PubMed: 24462293]
21. Sottoriva A, et al. A Big Bang model of human colorectal tumor growth. *Nature Genetics*. 2015; 47(3):209–16. [PubMed: 25665006]

22. Kim H, et al. Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. *Genome Res.* 2015; 25(3):316–27. [PubMed: 25650244]
23. Boutros PC, et al. Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat Genet.* 2015; 47(7):736–45. [PubMed: 26005866]
24. Corces MR, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet.* 2016; 48(10):1193–203. [PubMed: 27526324]
25. McPherson A, et al. Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nat Genet.* 2016
26. Maley CC, et al. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat Genet.* 2006; 38(4):468–73. [PubMed: 16565718]
27. Mroz EA, et al. Intra-tumor genetic heterogeneity and mortality in head and neck cancer: analysis of data from the Cancer Genome Atlas. *PLoS Med.* 2015; 12(2):e1001786. [PubMed: 25668320]
28. Wangsa D, et al. Phylogenetic analysis of multiple FISH markers in oral tongue squamous cell carcinoma suggests that a diverse distribution of copy number changes is associated with poor prognosis. *Int J Cancer.* 2016; 138(1):98–109. [PubMed: 26175310]
29. Zhang J, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science.* 2014; 346(6206):256–9. [PubMed: 25301631]
30. de Bruin EC, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science.* 2014; 346(6206):251–6. [PubMed: 25301630]
31. Gao R, et al. Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat Genet.* 2016; 48(10):1119–30. [PubMed: 27526321]
32. Wang Y, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature.* 2014; 512(7513):155–60. [PubMed: 25079324]
33. Weng Z, et al. Cell-lineage heterogeneity and driver mutation recurrence in pre-invasive breast neoplasia. *Genome Med.* 2015; 7(1):28. [PubMed: 25918554]
34. Brastianos PK, et al. Genomic Characterization of Brain Metastases Reveals Branched Evolution and Potential Therapeutic Targets. *Cancer Discov.* 2015; 5(11):1164–77. [PubMed: 26410082]
35. Cairns J. Mutation selection and the natural history of cancer. *Nature.* 1975; 255(5505):197–200. [PubMed: 1143315]
36. Yates LR, Campbell PJ. Evolution of the cancer genome. *Nat Rev Genet.* 2012; 13(11):795–806. [PubMed: 23044827]
37. Sottoriva A, et al. Single-molecule genomic data delineate patient-specific tumor profiles and cancer stem cell organization. *Cancer Res.* 2013; 73(1):41–9. [PubMed: 23090114]
38. Cooke SL, et al. Intra-tumour genetic heterogeneity and poor chemoradiotherapy response in cervical cancer. *Br J Cancer.* 2011; 104(2):361–8. [PubMed: 21063398]
39. Mroz EA, et al. High intratumor genetic heterogeneity is related to worse outcome in patients with head and neck squamous cell carcinoma. *Cancer.* 2013; 119(16):3034–42. [PubMed: 23696076]
40. Burrell RA, et al. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature.* 2013; 501(7467):338–45. [PubMed: 24048066]
41. Swanton C. Intratumor Heterogeneity: Evolution through Space and Time. *Cancer Res.* 2012; 72(19):4875–82. [PubMed: 23002210]
42. Barber LJ, Davies MN, Gerlinger M. Dissecting cancer evolution at the macro-heterogeneity and micro-heterogeneity scale. *Curr Opin Genet Dev.* 2015; 30:1–6. [PubMed: 25555261]
43. Snuderl M, et al. Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. *Cancer Cell.* 2011; 20(6):810–7. [PubMed: 22137795]
44. Szerlip NJ, et al. Intratumoral heterogeneity of receptor tyrosine kinases EGFR and PDGFRA amplification in glioblastoma defines subpopulations with distinct growth factor response. *Proc Natl Acad Sci U S A.* 2012; 109(8):3041–6. [PubMed: 22323597]
45. Kang H, et al. Many private mutations originate from the first few divisions of a human colorectal adenoma. *Journal of Pathology.* 2015; 237(3):355–362. [PubMed: 26119426]
46. Sievers CK, et al. Subclonal diversity arises early even in small colorectal tumours and contributes to differential growth fates. *Gut.* 2016

47. Uchi R, et al. Integrated Multiregional Analysis Proposing a New Model of Colorectal Cancer Evolution. *Plos Genetics*. 2016; 12(2)
48. Suzuki Y, et al. Multiregion ultra-deep sequencing reveals early intermixing and variable levels of intratumoral heterogeneity in colorectal cancer. *Mol Oncol*. 2017; 11(2):124–139. [PubMed: 28145097]
49. Yates LR, et al. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nature Medicine*. 2015; 21(7):751–9.
50. Potter NE, et al. Single-cell mutational profiling and clonal phylogeny in cancer. *Genome Res*. 2013; 23(12):2115–25. [PubMed: 24056532]
51. Patel AP, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*. 2014; 344(6190):1396–401. [PubMed: 24925914]
52. Meyer M, et al. Single cell-derived clonal analysis of human glioblastoma links functional and genomic heterogeneity. *Proc Natl Acad Sci U S A*. 2015; 112(3):851–6. [PubMed: 25561528]
53. Zhang X, et al. Single-Cell Sequencing for Precise Cancer Research: Progress and Prospects. *Cancer Res*. 2016; 76(6):1305–12. [PubMed: 26941284]
54. Inda MM, et al. Tumor heterogeneity is an active process maintained by a mutant EGFR-induced cytokine circuit in glioblastoma. *Genes Dev*. 2010; 24(16):1731–45. [PubMed: 20713517]
55. Cleary AS, et al. Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers. *Nature*. 2014; 508(7494):113–7. [PubMed: 24695311]
56. Marusyk A, et al. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature*. 2014; 514(7520):54–8. [PubMed: 25079331]
57. Francis JM, et al. EGFR variant heterogeneity in glioblastoma resolved through single-nucleus sequencing. *Cancer Discov*. 2014; 4(8):956–71. [PubMed: 24893890]
58. Ojha J, et al. Deep sequencing identifies genetic heterogeneity and recurrent convergent evolution in chronic lymphocytic leukemia. *Blood*. 2015; 125(3):492–8. [PubMed: 25377784]
59. Holohan C, et al. Cancer drug resistance: an evolving paradigm. *Nat Rev Cancer*. 2013; 13(10):714–26. [PubMed: 24060863]
60. Schmitt MW, Loeb LA, Salk JJ. The influence of subclonal resistance mutations on targeted cancer therapy. *Nat Rev Clin Oncol*. 2016; 13(6):335–47. [PubMed: 26483300]
61. Wang J, et al. Clonal evolution of glioblastoma under therapy. *Nature Genetics*. 2016
62. Castellarin M, et al. Clonal evolution of high-grade serous ovarian carcinoma from primary to recurrent disease. *J Pathol*. 2013; 229(4):515–24. [PubMed: 22996961]
63. Ding L, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*. 2012; 481:506–510. [PubMed: 22237025]
64. Landau DA, et al. Mutations driving CLL and their evolution in progression and relapse. *Nature*. 2015; 526:525–530. [PubMed: 26466571]
65. Schramm A, et al. Mutational dynamics between primary and relapse neuroblastomas. *Nature Genetics*. 2015; 47:872–877. [PubMed: 26121086]
66. Al-Mulla F, et al. Comparative genomic hybridization analysis of primary colorectal carcinomas and their synchronous metastases. *Genes Chromosomes Cancer*. 1999; 24(4):306–14. [PubMed: 10092128]
67. Alcock HE, et al. Analysis of colorectal tumor progression by microdissection and comparative genomic hybridization. *Genes Chromosomes Cancer*. 2003; 37(4):369–80. [PubMed: 12800148]
68. Shah SP, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature*. 2009; 461(7265):809–13. [PubMed: 19812674]
69. Miranda E, et al. Genetic and epigenetic alterations in primary colorectal cancers and related lymph node and liver metastases. *Cancer*. 2013; 119(2):266–76. [PubMed: 22786759]
70. Vermaat JS, et al. Primary colorectal cancers and their subsequent hepatic metastases are genetically different: implications for selection of patients for targeted treatment. *Clin Cancer Res*. 2012; 18(3):688–99. [PubMed: 22173549]
71. Brannon A, et al. Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions. *Genome Biol*. 2014; 15(8):454. [PubMed: 25164765]

72. Ding L, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature*. 2010; 464(7291):999–1005. [PubMed: 20393555]
73. Savas P, et al. The Subclonal Architecture of Metastatic Breast Cancer: Results from a Prospective Community-Based Rapid Autopsy Program “CASCADE”. *PLoS Med*. 2016; 13(12):e1002204. [PubMed: 28027312]
74. Campbell PJ, et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature*. 2010; 467(7319):1109–13. [PubMed: 20981101]
75. Knosel T, et al. Chromosomal alterations during lymphatic and liver metastasis formation of colorectal cancer. *Neoplasia*. 2004; 6(1):23–8. [PubMed: 15068668]
76. Jiang JK, et al. Genetic changes and clonality relationship between primary colorectal cancers and their pulmonary metastases—an analysis by comparative genomic hybridization. *Genes Chromosomes Cancer*. 2005; 43(1):25–36. [PubMed: 15723340]
77. Weber JC, et al. Allelotyping analyses of synchronous primary and metastasis CIN colon cancers identified different subtypes. *Int J Cancer*. 2007; 120(3):524–32. [PubMed: 17096353]
78. Becker TE, et al. The genomic heritage of lymph node metastases: implications for clinical management of patients with breast cancer. *Ann Surg Oncol*. 2008; 15(4):1056–63. [PubMed: 18246400]
79. Haffner MC, et al. Tracking the clonal origin of lethal prostate cancer. *J Clin Invest*. 2013; 123(11):4918–22. [PubMed: 24135135]
80. Gundem G, et al. The evolutionary history of lethal metastatic prostate cancer. *Nature*. 2015; 520(7547):353–7. [PubMed: 25830880]
81. El-Kebir M, et al. Inferring the Mutational History of a Tumor Using Multi-state Perfect Phylogeny Mixtures. *Cell Syst*. 2016; 3(1):43–53. [PubMed: 27467246]
82. Kim TM, et al. Subclonal Genomic Architectures of Primary and Metastatic Colorectal Cancer Based on Intratumoral Genetic Heterogeneity. *Clin Cancer Res*. 2015; 21(19):4461–72. [PubMed: 25979483]
83. Lawrence MS, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013; 499(7457):214–8. [PubMed: 23770567]
84. Vogelstein B, et al. Cancer genome landscapes. *Science*. 2013; 339(6127):1546–58. [PubMed: 23539594]
85. Loeb LA, Springgate CF, Battula N. Errors in DNA replication as a basis of malignant changes. *Cancer Res*. 1974; 34(9):2311–21. [PubMed: 4136142]
86. Loeb LA. Human Cancers Express a Mutator Phenotype: Hypothesis, Origin, and Consequences. *Cancer Res*. 2016; 76(8):2057–9. [PubMed: 27197248]
87. Lengauer C, Kinzler KW, Vogelstein B. Genetic instability in colorectal cancers. *Nature*. 1997; 386(6625):623–7. [PubMed: 9121588]
88. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature*. 1998; 396(6712):643–9. [PubMed: 9872311]
89. Laughney AM, et al. Dynamics of Tumor Heterogeneity Derived from Clonal Karyotypic Evolution. *Cell Rep*. 2015; 12(5):809–20. [PubMed: 26212324]
90. Rowald K, et al. Negative Selection and Chromosome Instability Induced by Mad2 Overexpression Delay Breast Cancer but Facilitate Oncogene-Independent Outgrowth. *Cell Rep*. 2016; 15(12):2679–91. [PubMed: 27292643]
91. Burrell RA, et al. Replication stress links structural and numerical cancer chromosomal instability. *Nature*. 2013; 494(7438):492–6. [PubMed: 23446422]
92. Aaltonen LA, et al. Clues to the pathogenesis of familial colorectal cancer. *Science*. 1993; 260(5109):812–6. [PubMed: 8484121]
93. Kim TM, Park PJ. A genome-wide view of microsatellite instability: old stories of cancer mutations revisited with new sequencing technologies. *Cancer Res*. 2014; 74(22):6377–82. [PubMed: 25371413]
94. Lee JK, et al. Mechanisms and Consequences of Cancer Genome Instability: Lessons from Genome Sequencing Studies. *Annu Rev Pathol*. 2016; 11:283–312. [PubMed: 26907526]

95. Fox EJ, Loeb LA. Lethal Mutagenesis: Targeting the Mutator Phenotype in Cancer. *Seminars in Cancer Biology*. 2010; 20(5):353–359. [PubMed: 20934515]
96. Bielas JH, et al. Human cancers express a mutator phenotype. *Proc Natl Acad Sci U S A*. 2006; 103(48):18238–42. [PubMed: 17108085]
97. Dewhurst SM, et al. Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution. *Cancer Discov*. 2014; 4(2):175–85. [PubMed: 24436049]
98. Weaver BA, et al. Aneuploidy acts both oncogenically and as a tumor suppressor. *Cancer Cell*. 2007; 11(1):25–36. [PubMed: 17189716]
99. Carter SL, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol*. 2012; 30(5):413–21. [PubMed: 22544022]
100. McGranahan N, et al. Cancer chromosomal instability: therapeutic and diagnostic challenges. *EMBO Rep*. 2012; 13(6):528–38. [PubMed: 22595889]
101. Sinicrope FA, Yang ZJ. Prognostic and predictive impact of DNA mismatch repair in the management of colorectal cancer. *Future Oncol*. 2011; 7(3):467–74. [PubMed: 21417908]
102. Burrell RA, Swanton C. The evolution of the unstable cancer genome. *Curr Opin Genet Dev*. 2014; 24:61–7. [PubMed: 24657538]
103. Zack TI, et al. Pan-cancer patterns of somatic copy number alteration. *Nat Genet*. 2013; 45(10):1134–40. [PubMed: 24071852]
104. Curtis C, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012; 486(7403):346–52. [PubMed: 22522925]
105. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. 2014; 513(7517):202–9. [PubMed: 25079317]
106. Kadoch C, Crabtree GR. Reversible disruption of mSWI/SNF (BAF) complexes by the SS18-SSX oncogenic fusion in synovial sarcoma. *Cell*. 2013; 153(1):71–85. [PubMed: 23540691]
107. Almendro V, Marusyk A, Polyak K. Cellular heterogeneity and molecular evolution in cancer. *Annu Rev Pathol*. 2013; 8:277–302. [PubMed: 23092187]
108. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. *Nature*. 2013; 501(7467):328–37. [PubMed: 24048065]
109. Sharma SV, et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell*. 2010; 141(1):69–80. [PubMed: 20371346]
110. Mazor T, et al. Intratumoral Heterogeneity of the Epigenome. *Cancer Cell*. 2016; 29(4):440–51. [PubMed: 27070699]
111. Chisholm RH, et al. Emergence of drug tolerance in cancer cell populations: an evolutionary outcome of selection, nongenetic instability, and stress-induced adaptation. *Cancer Res*. 2015; 75(6):930–9. [PubMed: 25627977]
112. Pisco AO, Huang S. Non-genetic cancer cell plasticity and therapy-induced stemness in tumour relapse: ‘What does not kill me strengthens me’. *Br J Cancer*. 2015; 112(11):1725–32. [PubMed: 25965164]
113. Almendro V, et al. Genetic and phenotypic diversity in breast tumor metastases. *Cancer Res*. 2014; 74(5):1338–48. [PubMed: 24448237]
114. Dietrich JE, Hiragi T. Stochastic patterning in the mouse pre-implantation embryo. *Development*. 2007; 134(23):4219–31. [PubMed: 17978007]
115. Chang HH, et al. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature*. 2008; 453(7194):544–7. [PubMed: 18497826]
116. Losick R, Desplan C. Stochasticity and cell fate. *Science*. 2008; 320(5872):65–8. [PubMed: 18388284]
117. Acar M, Mettetal JT, van Oudenaarden A. Stochastic switching as a survival strategy in fluctuating environments. *Nat Genet*. 2008; 40(4):471–5. [PubMed: 18362885]
118. Spencer SL, et al. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature*. 2009; 459(7245):428–32. [PubMed: 19363473]
119. Frank SA, Rosner MR. Nonheritable cellular variability accelerates the evolutionary processes of cancer. *PLoS Biol*. 2012; 10(4):e1001296. [PubMed: 22509130]

120. Gupta PB, et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell*. 2011; 146(4):633–44. [PubMed: 21854987]
121. Shlush LI, Herschkovitz D. Clonal evolution models of tumor heterogeneity. *Am Soc Clin Oncol Educ Book*. 2015:e662–5. [PubMed: 25993239]
122. Gatenby RA, et al. Cellular adaptations to hypoxia and acidosis during somatic evolution of breast cancer. *Br J Cancer*. 2007; 97(5):646–53. [PubMed: 17687336]
123. Lee HO, et al. Evolution of tumor invasiveness: the adaptive tumor microenvironment landscape model. *Cancer Res*. 2011; 71(20):6327–37. [PubMed: 21859828]
124. Lloyd MC, et al. Darwinian Dynamics of Intratumoral Heterogeneity: Not Solely Random Mutations but Also Variable Environmental Selection Forces. *Cancer Res*. 2016; 76(11):3136–44. [PubMed: 27009166]
125. Notta F, et al. A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature*. 2016; 538(7625):378–382. [PubMed: 27732578]
126. Bertos NR, Park M. Laser Capture Microdissection as a Tool to Study Tumor Stroma. *Methods Mol Biol*. 2016; 1458:13–25. [PubMed: 27581011]
127. Ling S, et al. Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. *Proc Natl Acad Sci U S A*. 2015; 112(47):E6496–505. [PubMed: 26561581]
128. Larson NB, Fridley BL. PurBayes: estimating tumor cellularity and subclonality in next-generation sequencing data. *Bioinformatics*. 2013; 29(15):1888–9. [PubMed: 23749958]
129. Ha G, et al. TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. *Genome Res*. 2014; 24(11):1881–93. [PubMed: 25060187]
130. Hu Z, Curtis C. Inferring tumor phylogenies from multi-region sequencing. *Cell Systems*. 2016; 3(1):12–4. [PubMed: 27467243]
131. Jiang Y, et al. Assessing intratumor heterogeneity and tracking longitudinal and spatial clonal evolutionary history by next-generation sequencing. *Proc Natl Acad Sci U S A*. 2016; 113(37):E5528–37. [PubMed: 27573852]
132. Niknafs N, et al. SubClonal Hierarchy Inference from Somatic Mutations: Automatic Reconstruction of Cancer Evolutionary Trees from Multi-region Next Generation Sequencing. *PLoS Comput Biol*. 2015; 11(10):e1004416. [PubMed: 26436540]
133. Deshwar AG, et al. PhyloWGS: reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. *Genome Biol*. 2015; 16:35. [PubMed: 25786235]
134. Kostadinov R, Maley CC, Kuhner MK. Bulk Genotyping of Biopsies Can Create Spurious Evidence for Heterogeneity in Mutation Content. *PLoS Comput Biol*. 2016; 12(4):e1004413. [PubMed: 27105344]
135. Williams MJ, et al. Identification of neutral tumor evolution across cancer types. *Nat Genet*. 2016
136. Gerstung M, et al. Reliable detection of subclonal single-nucleotide variants in tumour cell populations. *Nat Commun*. 2012; 3:811. [PubMed: 22549840]
137. Cibulskis K, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013; 31(3):213–9. [PubMed: 23396013]
138. Findlay JM, et al. Differential clonal evolution in oesophageal cancers in response to neo-adjuvant chemotherapy. *Nat Commun*. 2016; 7:11111. [PubMed: 27045317]
139. Swanton C, et al. Consensus on precision medicine for metastatic cancers: a report from the MAP conference. *Ann Oncol*. 2016
140. Josephidou M, Lynch AG, Tavaré S. multiSNV: a probabilistic approach for improving detection of somatic point mutations from multiple related tumour samples. *Nucleic Acids Res*. 2015; 43(9):e61. [PubMed: 25722372]
141. Smith EN, et al. Biased estimates of clonal evolution and subclonal heterogeneity can arise from PCR duplicates in deep sequencing experiments. *Genome Biol*. 2014; 15(8):420. [PubMed: 25103687]
142. Kinde I, et al. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A*. 2011; 108(23):9530–5. [PubMed: 21586637]

143. Costello M, et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res.* 2013; 41(6):e67. [PubMed: 23303777]
144. Lou DI, et al. High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing. *Proc Natl Acad Sci U S A.* 2013; 110(49):19872–7. [PubMed: 24243955]
145. Schmitt MW, et al. Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci U S A.* 2012; 109(36):14508–13. [PubMed: 22853953]
146. Oesper L, Mahmoody A, Raphael BJ. THetA: inferring intra-tumor heterogeneity from high-throughput DNA sequencing data. *Genome Biol.* 2013; 14(7):R80. [PubMed: 23895164]
147. Fischer A, et al. High-definition reconstruction of clonal composition in cancer. *Cell Rep.* 2014; 7(5):1740–52. [PubMed: 24882004]
148. Yu Z, Li A, Wang M. CloneCNA: detecting subclonal somatic copy number alterations in heterogeneous tumor samples from whole-exome sequencing data. *BMC Bioinformatics.* 2016; 17:310. [PubMed: 27538789]
149. Roth A, et al. PyClone: statistical inference of clonal population structure in cancer. *Nat Methods.* 2014; 11(4):396–8. [PubMed: 24633410]
150. Miller CA, et al. SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLoS Comput Biol.* 2014; 10(8):e1003665. [PubMed: 25102416]
151. Lonnstedt IM, et al. Deciphering clonality in aneuploid breast tumors using SNP array and sequencing data. *Genome Biol.* 2014; 15(9):470. [PubMed: 25270265]
152. El-Kebir M, et al. Multi-State Perfect Phylogeny Mixtures for Cancer Sequencing. *Cell Systems.* 2016
153. Wang Y, Navin NE. Advances and applications of single-cell sequencing technologies. *Mol Cell.* 2015; 58(4):598–609. [PubMed: 26000845]
154. Roth A, et al. Clonal genotype and population structure inference from single-cell tumor sequencing. *Nat Methods.* 2016
155. Naxerova K, Jain RK. Using tumour phylogenetics to identify the roots of metastasis in humans. *Nat Rev Clin Oncol.* 2015; 12(5):258–72. [PubMed: 25601447]
156. Zhao ZM, et al. Early and multiple origins of metastatic lineages within primary tumors. *Proc Natl Acad Sci U S A.* 2016; 113(8):2140–5. [PubMed: 26858460]
157. Yang Z, Rannala B. Molecular phylogenetics: principles and practice. *Nat Rev Genet.* 2012; 13(5):303–14. [PubMed: 22456349]
158. Yang ZH. Phylogenetic analysis using parsimony and likelihood methods. *Journal of Molecular Evolution.* 1996; 42(2):294–307. [PubMed: 8919881]
159. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol.* 1981; 17(6):368–76. [PubMed: 7288891]
160. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 1987; 4(4):406–25. [PubMed: 3447015]
161. Letouze E, et al. Analysis of the copy number profiles of several tumor samples from the same patient reveals the successive steps in tumorigenesis. *Genome Biol.* 2010; 11(7):R76. [PubMed: 20649963]
162. Schwarz RF, et al. Phylogenetic quantification of intra-tumour heterogeneity. *PLoS Comput Biol.* 2014; 10(4):e1003535. [PubMed: 24743184]
163. Schwarz RF, et al. Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. *PLoS Med.* 2015; 12(2):e1001789. [PubMed: 25710373]
164. Nik-Zainal S, et al. The life history of 21 breast cancers. *Cell.* 2012; 149(5):994–1007. [PubMed: 22608083]
165. Sprouffske K, Pepper JW, Maley CC. Accurate reconstruction of the temporal order of mutations in neoplastic progression. *Cancer Prev Res (Phila).* 2011; 4(7):1135–44. [PubMed: 21490131]
166. Malikic S, et al. Clonality inference in multiple tumor samples using phylogeny. *Bioinformatics.* 2015; 31(9):1349–56. [PubMed: 25568283]
167. Popic V, et al. Fast and scalable inference of multi-sample cancer lineages. *Genome Biol.* 2015; 16(1):91. [PubMed: 25944252]

168. Ross EM, Markowitz F. OncoNEM: inferring tumor evolution from single-cell sequencing data. *Genome Biol.* 2016; 17:69. [PubMed: 27083415]
169. Jahn K, Kuipers J, Beerenwinkel N. Tree inference for single-cell data. *Genome Biol.* 2016; 17:86. [PubMed: 27149953]
170. Behjati S, et al. Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature.* 2014; 513(7518):422–5. [PubMed: 25043003]
171. Kingman JF. The Coalescent. *Stoch Process Appl.* 1982; 13:235–248.
172. Marjoram P, Tavaré S. Modern computational approaches for analysing molecular genetic variation data. *Nat Rev Genet.* 2006; 7(10):759–70. [PubMed: 16983372]
173. Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian computation in population genetics. *Genetics.* 2002; 162(4):2025–35. [PubMed: 12524368]
174. Nicolas P, et al. The stem cell population of the human colon crypt: analysis via methylation patterns. *PLoS Comput Biol.* 2007; 3(3):e28. [PubMed: 17335343]
175. Dykhuizen D, Hartl DL. Selective neutrality of 6PGD allozymes in *E. coli* and the effects of genetic background. *Genetics.* 1980; 96(4):801–17. [PubMed: 7021316]
176. Gillespie, JH. Population genetics: a concise guide. 2nd. Baltimore, Md.: Johns Hopkins University Press; 2004. p. xivp. 214
177. Simpson EH. Measurement of diversity. *Nature.* 1949
178. Allred DC, et al. Ductal carcinoma in situ and the emergence of diversity during breast cancer evolution. *Clin Cancer Res.* 2008; 14(2):370–8. [PubMed: 18223211]
179. Shannon CE. A mathematical theory of communication. *Bell Systems Technical Journal.* 1948; 27:379–432.
180. Merlo LM, et al. A comprehensive survey of clonal diversity measures in Barrett's esophagus as biomarkers of progression to esophageal adenocarcinoma. *Cancer Prev Res (Phila).* 2010; 3(11):1388–97. [PubMed: 20947487]
181. Mroz EA, Rocco JW. MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. *Oral Oncol.* 2013; 49(3):211–5. [PubMed: 23079694]
182. Mengelbier LH, et al. Intratumoral genome diversity parallels progression and predicts outcome in pediatric cancer. *Nat Commun.* 2015; 6:6125. [PubMed: 25625758]
183. Morris LG, et al. Pan-cancer analysis of intratumor heterogeneity as a prognostic determinant of survival. *Oncotarget.* 2016; 7(9):10051–63. [PubMed: 26840267]
184. Andor N, et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat Med.* 2016; 22(1):105–13. [PubMed: 26618723]
185. Hartl, DL., Clark, AG. Principles of population genetics. 4th. Sinauer Associates; 2007.
186. Lynch M. Evolution of the mutation rate. *Trends Genet.* 2010; 26(8):345–52. [PubMed: 20594608]
187. Lynch M. Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci U S A.* 2010; 107(3):961–8. [PubMed: 20080596]
188. Conrad DF, et al. Variation in genome-wide mutation rates within and between human families. *Nat Genet.* 2011; 43(7):712–4. [PubMed: 21666693]
189. Crow JF. The origins, patterns and implications of human spontaneous mutation. *Nat Rev Genet.* 2000; 1(1):40–7. [PubMed: 11262873]
190. Loeb LA. A mutator phenotype in cancer. *Cancer Res.* 2001; 61(8):3230–9. [PubMed: 11309271]
191. Loeb LA. Mutator phenotype in cancer: origin and consequences. *Semin Cancer Biol.* 2010; 20(5):279–80. [PubMed: 20937389]
192. Tomlinson IP, Novelli MR, Bodmer WF. The mutation rate and cancer. *Proc Natl Acad Sci U S A.* 1996; 93(25):14800–3. [PubMed: 8962135]
193. Tomlinson I, Bodmer W. Selection, the mutation rate and cancer: ensuring that the tail does not wag the dog. *Nat Med.* 1999; 5(1):11–2. [PubMed: 9883827]
194. Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science.* 2015; 347(6217):78–81. [PubMed: 25554788]

195. Rozhok AI, Wahl GM, DeGregori J. A Critical Examination of the “Bad Luck” Explanation of Cancer Risk. *Cancer Prev Res (Phila)*. 2015; 8(9):762–4. [PubMed: 26122457]
196. Noble R, Kaltz O, Hochberg ME. Peto’s paradox and human cancers. *Philos Trans R Soc Lond B Biol Sci*. 2015; 370(1673)
197. Wu S, et al. Substantial contribution of extrinsic risk factors to cancer development. *Nature*. 2016; 529(7584):43–7. [PubMed: 26675728]
198. Blokzijl F, et al. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature*. 2016; 538(7624):260–264. [PubMed: 27698416]
199. Kimura, M. The neutral theory of molecular evolution. Cambridge Cambridgeshire; New York: Cambridge University Press; 1983. p. xvp. 367
200. Landau DA, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell*. 2013; 152(4):714–26. [PubMed: 23415222]
201. Bozic I, Gerold JM, Nowak MA. Quantifying Clonal and Subclonal Passenger Mutations in Cancer Evolution. *PLoS Comput Biol*. 2016; 12(2):e1004731. [PubMed: 26828429]
202. Diaz LA Jr, et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature*. 2012; 486(7404):537–40. [PubMed: 22722843]
203. Bozic I, et al. Accumulation of driver and passenger mutations during tumor progression. *Proc Natl Acad Sci U S A*. 2010; 107(43):18545–50. [PubMed: 20876136]
204. Slatkin M. Gene flow and the geographic structure of natural populations. *Science*. 1987; 236(4803):787–92. [PubMed: 3576198]
205. Excoffier L. Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Mol Ecol*. 2004; 13(4):853–64. [PubMed: 15012760]
206. Korolev KS, Xavier JB, Gore J. Turning ecology and evolution against cancer. *Nat Rev Cancer*. 2014; 14(5):371–80. [PubMed: 24739582]
207. Kim TM, et al. Regional biases in mutation screening due to intratumoural heterogeneity of prostate cancer. *J Pathol*. 2014; 233(4):425–35. [PubMed: 24870262]
208. Martens EA, et al. Spatial structure increases the waiting time for cancer. *New J Phys*. 2011; 13
209. Hindersin L, et al. Should tissue structure suppress or amplify selection to minimize cancer risk? *Biology Direct*. 2016; 11
210. Komarova NL. Spatial interactions and cooperation can change the speed of evolution of complex phenotypes. *Proc Natl Acad Sci U S A*. 2014; 111(Suppl 3):10789–95. [PubMed: 25024187]
211. Sottoriva A, et al. Cancer stem cell tumor model reveals invasive morphology and increased phenotypical heterogeneity. *Cancer Res*. 2010; 70(1):46–56. [PubMed: 20048071]
212. Poleszczuk J, Hahnfeldt P, Enderling H. Evolution and phenotypic selection of cancer stem cells. *PLoS Comput Biol*. 2015; 11(3):e1004025. [PubMed: 25742563]
213. Sprouffske K, et al. An evolutionary explanation for the presence of cancer nonstem cells in neoplasms. *Evol Appl*. 2013; 6(1):92–101. [PubMed: 23397439]
214. Carmona-Fontaine C, et al. Emergence of spatial structure in the tumor microenvironment due to the Warburg effect. *Proc Natl Acad Sci U S A*. 2013; 110(48):19402–7. [PubMed: 24218566]
215. Gallaher J, Anderson AR. Evolution of intratumoral phenotypic heterogeneity: the role of trait inheritance. *Interface Focus*. 2013; 3(4):20130016. [PubMed: 24511380]
216. Waclaw B, et al. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature*. 2015; doi: 10.1038/nature14971
217. Yachida S, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature*. 2010; 467(7319):1114–7. [PubMed: 20981102]
218. Cooper CS, et al. Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat Genet*. 2015; 47(4):367–72. [PubMed: 25730763]
219. Raphael BJ, et al. Identifying driver mutations in sequenced cancer genomes: computational approaches to enable precision medicine. *Genome Med*. 2014; 6(1):5. [PubMed: 24479672]
220. Hodis E, et al. A landscape of driver mutations in melanoma. *Cell*. 2012; 150(2):251–63. [PubMed: 22817889]

221. Martincorena I, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*. 2015; 348(6237):880–6. [PubMed: 25999502]
222. Simons BD. Deep sequencing as a probe of normal stem cell fate and preneoplasia in human epidermis. *Proceedings of the National Academy of Sciences of the United States of America*. 2016; 113(1):128–133. [PubMed: 26699486]
223. Simons BD. REPLY TO MARTINCORENA ET AL.: Evidence for constrained positive selection of cancer mutations in normal skin is lacking. *Proceedings of the National Academy of Sciences of the United States of America*. 2016; 113(9):E1130–E1131. [PubMed: 26884186]
224. Bauer B, Siebert R, Traulsen A. Cancer initiation with epistatic interactions between driver and passenger mutations. *J Theor Biol*. 2014; 358:52–60. [PubMed: 24859427]
225. Ortmann CA, et al. Effect of mutation order on myeloproliferative neoplasms. *N Engl J Med*. 2015; 372(7):601–12. [PubMed: 25671252]
226. Fisher R, et al. Development of synchronous VHL syndrome tumors reveals contingencies and constraints to tumor evolution. *Genome Biol*. 2014; 15(8):433. [PubMed: 25159823]
227. Charlesworth D, Charlesworth B, Morgan MT. The pattern of neutral molecular variation under the background selection model. *Genetics*. 1995; 141(4):1619–32. [PubMed: 8601499]
228. Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. *Genetics*. 1993; 134(4):1289–303. [PubMed: 8375663]
229. Beckman RA, Loeb LA. Negative clonal selection in tumor evolution. *Genetics*. 2005; 171(4):2123–31. [PubMed: 16143627]
230. Hill WG, Robertson A. The effect of linkage on limits to artificial selection. *Genet Res*. 1966; 8(3):269–94. [PubMed: 5980116]
231. Wu CI, et al. The Ecology and Evolution of Cancer: The Ultra-Microevolutionary Process. *Annu Rev Genet*. 2016; 50:347–369. [PubMed: 27686281]
232. McFarland CD, et al. Impact of deleterious passenger mutations on cancer progression. *Proc Natl Acad Sci U S A*. 2013; 110(8):2910–5. [PubMed: 23388632]
233. McFarland CD, Mirny LA, Korolev KS. Tug-of-war between driver and passenger mutations in cancer and other adaptive processes. *Proc Natl Acad Sci U S A*. 2014; 111(42):15138–43. [PubMed: 25277973]
234. Lynch M, et al. The mutational meltdown in asexual populations. *J Hered*. 1993; 84(5):339–44. [PubMed: 8409355]
235. Gabriel W, Lynch M, Burger R. Muller's Ratchet and Mutational Meltdowns. *Evolution*. 1993; 47(6):1744–1757. [PubMed: 28567994]
236. Kreso A, Dick JE. Evolution of the cancer stem cell model. *Cell Stem Cell*. 2014; 14(3):275–91. [PubMed: 24607403]
237. Weaver JM, et al. Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. *Nat Genet*. 2014; 46(8):837–43. [PubMed: 24952744]
238. Calabrese P, Tavaré S, Shibata D. Pretumor progression: clonal evolution of human stem cell populations. *Am J Pathol*. 2004; 164(4):1337–46. [PubMed: 15039221]
239. Baker AM, et al. Quantification of Crypt and Stem Cell Evolution in the Normal and Neoplastic Human Colon. *Cell Reports*. 2014; 8(4):940–947. [PubMed: 25127143]
240. Vermeulen L, et al. Defining stem cell dynamics in models of intestinal tumor initiation. *Science*. 2013; 342(6161):995–8. [PubMed: 24264992]
241. Sievers CK, et al. Understanding Intratumoral Heterogeneity: Lessons from the Analysis of At-Risk Tissue and Premalignant Lesions in the Colon. *Cancer Prevention Research*. 2016; 9(8):638–641. [PubMed: 27199343]
242. Driessens G, et al. Defining the mode of tumour growth by clonal analysis. *Nature*. 2012; 488(7412):527–30. [PubMed: 22854777]
243. Chen J, et al. A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature*. 2012; 488(7412):522–6. [PubMed: 22854781]
244. Schepers AG, et al. Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas. *Science*. 2012; 337(6095):730–5. [PubMed: 22855427]

245. Tomasetti C, et al. Only three driver gene mutations are required for the development of lung and colorectal cancers. *Proc Natl Acad Sci U S A*. 2015; 112(1):118–23. [PubMed: 25535351]
246. Gerlinger M, et al. Cancer: evolution within a lifetime. *Annu Rev Genet*. 2014; 48:215–36. [PubMed: 25292359]
247. Suzuki Y, et al. Multiregion ultra-deep sequencing reveals early intermixing and variable levels of intratumoral heterogeneity in colorectal cancer. *Molecular Oncology*. 2016
248. Nielsen R. Molecular signatures of natural selection. *Annu Rev Genet*. 2005; 39:197–218. [PubMed: 16285858]
249. Heng HH, et al. Stochastic cancer progression driven by non-clonal chromosome aberrations. *J Cell Physiol*. 2006; 208(2):461–72. [PubMed: 16688757]
250. Stephens PJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*. 2011; 144(1):27–40. [PubMed: 21215367]
251. Rausch T, et al. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell*. 2012; 148:1–2. 59–71.
252. Korbel JO, Campbell PJ. Criteria for inference of chromothripsis in cancer genomes. *Cell*. 2013; 152(6):1226–36. [PubMed: 23498933]
253. Baca SC, et al. Punctuated evolution of prostate cancer genomes. *Cell*. 2013; 153(3):666–77. [PubMed: 23622249]
254. Cheng C, et al. Whole-Genome Sequencing Reveals Diverse Models of Structural Variations in Esophageal Squamous Cell Carcinoma. *Am J Hum Genet*. 2016; 98(2):256–74. [PubMed: 26833333]
255. Alexandrov LB, et al. Signatures of mutational processes in human cancer. *Nature*. 2013; 500(7463):415–21. [PubMed: 23945592]
256. Waclaw B, et al. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature*. 2015; 525(7568):261–4. [PubMed: 26308893]
257. Basanta D, Gatenby RA, Anderson AR. Exploiting evolution to treat drug resistance: combination therapy and the double bind. *Mol Pharm*. 2012; 9(4):914–21. [PubMed: 22369188]
258. Rodriguez-Brenes IA, Wodarz D. Preventing clonal evolutionary processes in cancer: Insights from mathematical models. *Proc Natl Acad Sci U S A*. 2015; 112(29):8843–50. [PubMed: 26195751]
259. Zhao B, et al. Exploiting Temporal Collateral Sensitivity in Tumor Clonal Evolution. *Cell*. 2016; 165(1):234–46. [PubMed: 26924578]
260. Enriquez-Navas PM, et al. Exploiting evolutionary principles to prolong tumor control in preclinical models of breast cancer. *Sci Transl Med*. 2016; 8(327):327ra24.
261. Rodriguez-Brenes IA, Peskin CS. Quantitative theory of telomere length regulation and cellular senescence. *Proc Natl Acad Sci U S A*. 2010; 107(12):5387–92. [PubMed: 20207949]
262. Zhao B, Hemann MT, Lauffenburger DA. Modeling Tumor Clonal Evolution for Drug Combinations Design. *Trends in Cancer*. 2016; 2(3):144–158. [PubMed: 28435907]
263. Chmielecki J, et al. Optimization of dosing for EGFR-mutant non-small cell lung cancer with evolutionary cancer modeling. *Sci Transl Med*. 2011; 3(90):90ra59.
264. Bozic I, et al. Evolutionary dynamics of cancer in response to targeted combination therapy. *Elife*. 2013; 2:e00747. [PubMed: 23805382]
265. Zhao B, Hemann MT, Lauffenburger DA. Intratumor heterogeneity alters most effective drugs in designed combinations. *Proc Natl Acad Sci U S A*. 2014; 111(29):10773–8. [PubMed: 25002493]
266. Misale S, et al. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature*. 2012; 486(7404):532–6. [PubMed: 22722830]
267. Laurent-Puig P, et al. Clinical relevance of KRAS-mutated subclones detected with picodroplet digital PCR in advanced colorectal cancer treated with anti-EGFR therapy. *Clin Cancer Res*. 2015; 21(5):1087–97. [PubMed: 25248381]
268. Kemper K, et al. Intra- and inter-tumor heterogeneity in a vemurafenib-resistant melanoma patient and derived xenografts. *EMBO Mol Med*. 2015; 7(9):1104–18. [PubMed: 26105199]

269. Wong TN, et al. Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. *Nature*. 2015; 518(7540):552–5. [PubMed: 25487151]
270. Bhang HE, et al. Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat Med*. 2015; 21(5):440–8. [PubMed: 25849130]
271. Juric D, et al. Convergent loss of PTEN leads to clinical resistance to a PI(3)Kalpha inhibitor. *Nature*. 2015; 518(7538):240–4. [PubMed: 25409150]
272. Balak MN, et al. Novel D761Y and common secondary T790M mutations in epidermal growth factor receptor-mutant lung adenocarcinomas with acquired resistance to kinase inhibitors. *Clin Cancer Res*. 2006; 12(21):6494–501. [PubMed: 17085664]
273. Murtaza M, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature*. 2013; 497(7447):108–12. [PubMed: 23563269]
274. Romanel A, et al. Plasma AR and abiraterone-resistant prostate cancer. *Sci Transl Med*. 2015; 7(312):312re10.
275. Dawson SJ, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med*. 2013; 368(13):1199–209. [PubMed: 23484797]
276. Tie J, et al. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci Transl Med*. 2016; 8(346):346ra92.
277. Wang Y, et al. Detection of tumor-derived DNA in cerebrospinal fluid of patients with primary tumors of the brain and spinal cord. *Proc Natl Acad Sci U S A*. 2015; 112(31):9704–9. [PubMed: 26195750]
278. Jamal-Hanjani M, et al. Tracking genomic cancer evolution for precision medicine: the lung TRACERx study. *PLoS Biol*. 2014; 12(7):e1001906. [PubMed: 25003521]
279. Knijn N, et al. KRAS mutation analysis: a comparison between primary tumours and matched liver metastases in 305 colorectal cancer patients. *Br J Cancer*. 2011; 104(6):1020–6. [PubMed: 21364579]
280. Turajlic S, Swanton C. Metastasis as an evolutionary process. *Science*. 2016; 352(6282):169–75. [PubMed: 27124450]
281. Angelo M, et al. Multiplexed ion beam imaging of human breast tumors. *Nat Med*. 2014; 20(4):436–42. [PubMed: 24584119]
282. Bruna A, et al. A Biobank of Breast Cancer Explants with Preserved Intra-tumor Heterogeneity to Screen Anticancer Compounds. *Cell*. 2016; 167(1):260–274 e22. [PubMed: 27641504]
283. Eirew P, et al. Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature*. 2015; 518(7539):422–6. [PubMed: 25470049]
284. van de Wetering M, et al. Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell*. 2015; 161(4):933–45. [PubMed: 25957691]
285. Cantrell MA, Kuo CJ. Organoid modeling for cancer precision medicine. *Genome Med*. 2015; 7(1):32. [PubMed: 25825593]
286. Kim JY, Gatenby RA. Quantitative Clinical Imaging Methods for Monitoring Intratumoral Evolution. *Methods Mol Biol*. 2017; 1513:61–81. [PubMed: 27807831]

Highlights

1. Population genetics principles provide insight into the driving forces of intra-tumor genetic heterogeneity, including mutation, genetic drift, spatial structure, migration and selection.
2. Distinct modes of evolution (linear versus branched, selection versus effective neutrality) are operative during different stages of tumor progression and in different tumor types.
3. Spatial information obtained via multi-region sequencing can help to resolve tumor evolutionary dynamics.
4. A quantitative understanding of tumor dynamics based on patterns of intra-tumor heterogeneity may inform the prevention, detection, and treatment of cancer.

Figure 1. Schematic illustration of different phases and modes of tumor evolution

Somatic alterations (including SSNVs and CNAs) arise during cell division in normal/pre-neoplastic cells, where mutations that confer a selective growth advantage relative to the rest of the population (drivers) result in clonal expansions. Multiple such events promote transformation in a process that can occur over long time periods (indicated by hatch marks). The mutations (both drivers and passengers) that accumulate in the founding tumor cell lineage during initiation will be present in all cells in the tumor (public/clonal). After transformation, the tumor may grow in the absence of stringent selection, compatible with a

Big Bang model [Ref 21] and effectively neutral evolution (**A**) or be subject to continuous selection resulting in clonal expansions (**B and C**). In the Big Bang model, the tumor grows as a single ‘terminal’ expansion populated by numerous heterogeneous subclones that are not subject to strong selection (**A**). As such, the timing of a mutation is the primary determinant of its frequency in the final tumor and most *detectable* subclonal alterations (resulting in ITH) arise early during growth, whereas late-arising mutations will be largely undetectable. In contrast, under selection, linear evolution describes clonal successions or sweeps that occur sequentially (**B**) while branched evolution corresponds to a scenario where multiple subclonal alterations co-occur and compete during tumor growth (**C**).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 2. Schematic illustration of different modes of tumor evolution and the impact of spatial sampling on patterns of ITH

(A) After transformation, an established tumor may propagate in the absence of stringent selection (for example if the cells are already sufficiently fit), compatible with effectively neutral growth. Alternatively, the tumor may be subject to ongoing clonal selection. To illustrate these scenarios, Muller plots (<https://github.com/robjohnnoble/ggmuller>) depicting tumor subclone composition under a model of effectively neutral growth (upper panel) or strong selection (bottom panel) are shown. Descendant genotypes emerge within the parental clone, where height indicates genotype frequency (corresponding to the relative abundance of the genotype in the tumor cell population or cancer cell fraction) and the horizontal axis indicates time (generations). At the time of diagnosis or surgery, a single sample or multiple samples (denoted A, B) are sampled and subject to NGS. (B) Tumors arising under effectively neutral growth are expected to exhibit a bimodal site frequency spectrum (SFS) with a public mutational cluster centered at 0.5 and a large cluster of low VAF subclonal mutations, where θ indicates the threshold below which low frequency mutations cannot be reliably detected. This can be contrasted with the expected SFS under strong selection in which private SSNVs can attain high frequency due to regional clonal expansion of a selectively advantageous clone (purple shading).

Figure 3. Genetic diversification resulting from random neutral mutations

(A) Tumor growth via a logistic model, $N_t = KN_0 e^{rt} / (K + N_0(e^{rt} - 1))$, where $K = 10^9$, $r = 0.05$ and $N_0 = 1$ correspond to the carrying capacity, growth rate and initial tumor size, respectively. (B) The expected probability of heterozygosity, H (equivalent to Simpson's Diversity Index), during tumor growth (shown in A) where the neutral mutation rate is $u = 0.03$ per cell division in exonic coding regions. (C) The relationship between heterozygosity and tumor size.

Figure 4. Genetic drift decreases genetic diversity in the whole tumor, while increasing diversity amongst detectable mutations in NGS data

(A) The process of tumor diversification in the absence of genetic drift (and selection), in which the frequency of a mutation (or cancer cell fraction, CCF) remains constant over time, corresponding to its frequency at the time of occurrence ($1/n(t)$, where $n(t)$ is the tumor size at time t). (B) The process of tumor diversification in the presence of genetic drift in which the frequency of a mutation fluctuates stochastically during tumor growth. Early mutations whose initial frequencies are high can reach detectable frequencies, while late mutations remain non-detectable under a model of effectively neutral evolution.

Figure 5. Spatial structure and migration contribute to intra-tumor heterogeneity

(A) Schematic illustration of tumor biopsy or resection specimens derived from multiple local regions of a tumor expanding under stringent spatial constraints. Due to the spatial constraints, clones are partitioned and the genetic diversity within regions (R1, R2 or R3) is smaller than the diversity between regions (e.g. R1 vs R3). (B) While the three regions can exhibit equivalent clonal composition, one cannot distinguish whether subclone mixing occurred early during tumor growth. (C) Clonemap derived from a virtual tumor simulated within an agent-based tumor growth model with spatial constraints and subclone mixing. Each colored region represents an individual clone composed of cells that share the same mutations. While most of clones are locally restricted, the green clone is spread across distant regions of the tumor. (D) Migration and metastasis can significantly increase the divergence between tumors due to augmented genetic drift and clonal selection.