Annu. Rev. Genom. Hum. Genet. 2019. 20:X–X
https://doi.org/10.1146/annurev-genom-083117-021712

Measuring Clonal Evolution in Cancer with Genomics

Marc J. Williams[1], Andrea Sottoriva[2], Trevor A. Graham[1]

[1] Evolution and Cancer Laboratory, Barts Cancer Institute, Queen Mary University of London, London EC1M 6BQ, United Kingdom;

[2] Evolutionary Genomics and Modelling Lab, Centre for Evolution and Cancer, The Institute of Cancer Research, London SM2 5NG, United Kingdom

email: marc.williams@qmul.ac.uk, t.graham@qmul.ac.uk

Keywords

Cancer, Evolution, Genomics, Population Genetics

Abstracts

Cancers originate from somatic cells in the human body that have accumulated genetic alterations. These mutations modify the phenotype of the cells, allowing them to escape the homeostatic regulation that maintains normal cell number. Viewed through the lens of evolutionary biology, the transformation of normal cells into malignant cells is evolution in action. Evolution continues throughout cancer growth, progression, treatment resistance, and disease relapse, driven by adaptation to changes in the cancer's environment, and intratumor heterogeneity is an inevitable consequence of this evolutionary process. Genomics provides a powerful means to characterize tumor evolution, enabling quantitative measurement of evolving clones across space and time. In this review, we discuss concepts and approaches to quantify and measure this evolutionary process in cancer using genomics.

CLONAL EVOLUTION IN CANCER

Abnormal cell growth and survival are the root proximal cause of cancer. Each time a cell divides, the DNA in the cell is copied and passed on to daughter cells. While this process is relatively accurate, the fidelity is not perfect. Because somatic cells reproduce asexually, any mutations that arise during DNA replication and are not repaired are passed on to the daughter cells and all subsequent descendants. Given a conservative estimate of the base-pair mutation rate of 10^{-9} and a human genome consisting of approximately 3×10^9 base pairs, it is likely that every cell division will introduce new mutations into the daughter cells. Furthermore, tumors are often subject to genomic instability, either through increased point mutation rates due to defective DNA repair processes (17) or through chromosomal instability (8). Billions of cell divisions coupled with imperfect DNA copying makes intratumor heterogeneity inevitable.

The advent of high-throughput genomics and its application to cancer has validated cancer progression as an evolutionary process, that is likely to be driven by (sometimes complex) genetic alterations (103), and intratumor heterogeneity is pervasive (74). More specifically, tumors can harbor many thousands of point mutations (62) and in many cases can be highly aneuploid (44), and intratumor heterogeneity is present for both types of mutations, meaning that every cell in a tumor is likely to be genetically distinct from the others. This incredible diversity and genomic complexity are now beginning to be understood in terms of evolutionary principles.

In this review we discuss approaches to quantify evolutionary dynamics in human cancers. It is our hypothesis that such quantitative measurement will enable more accurate prediction of the future course of the disease, and concomitantly facilitate the design of improved treatment strategies. Such an approach has already shown promise in some cancer types, Maley et al (68a) have shown that diversity is a prognostic biomarker in the pre-malignant disease Barrett's oesophagus. This has been validated in subsequent studies (71a) and a pan cancer analysis along similar lines also showed that diversity of the clonal composition of tumors was prognostic (6a). In non-small-cell lung cancer, heterogeneity of copy number alterations (but not single nucleotide variants) has also been demonstrated to correlate with patient outcome (49). Similarly in prostate cancers, the percentage of the genome altered has also been shown to be clinically predictive (35a). These results make sense from an evolutionary perspective, diversity can be thought of as measuring the "evolvability", i.e. the capacity of the population to adapt to its environment, and thus the more "evolvable" lesions will have worse prognosis. Quantitative

measurements of the fitness of subpopulations of cells within a tumor has also been shown to correlate with the efficacy of immunotherapy (65b). These studies all demonstrate that applying evolutionary principles to improve patient stratification is promising. Our belief is that precise quantification of the different components of the evolutionary system will provide greater predictive power and accuracy.

COMPONENTS OF THE EVOLUTIONARY SYSTEM

Like every evolutionary system, clonal evolution in cancer is shaped by the fundamental evolutionary forces: (stochastic) mutation, (stochastic) genetic drift, and (arguably deterministic) selection (65a). Mutation is a stochastic process that introduces new variation into the population. Genetic drift describes the stochastic changes in clone size due to random effects that lead to cancer cell growth or death. Selection, on the other hand, is in principle deterministic: A combination of genotype and phenotype that is adapted to a particular environment will always be expected to grow (110). We acknowledge that stochastic effects also affect the growth of a new lineage when it is small, but once the population of the lineage is large enough to overcome the genetic drift barrier, the expansion of the lineage becomes predictable (the switch from stochastic to deterministic behavior for a positively selected lineage occurs when the population size N of the lineage is greater than the inverse of the selection benefit s that the lineage experiences, e.g., $1/s$) (33), as also shown experimentally in yeast (65).

The growing field of cancer evolution interrogates the relative and combined contributions of these evolutionary components. Large sequencing studies such as the Cancer Genome Atlas have uncovered many recurrent so-called driver mutations (mutations that lead to a positively selected phenotype and therefore expansion of the clone of cells carrying the driver mutation) across cancer types (3). These types of analysis particularly highlight the importance of clonal selection in cancer development. The mutation rate itself has also received considerable attention. The mutation burden varies considerably across cancers (63), suggesting large differences in the underlying mutation rate between individual tumors and tumor types. The realization that different mutational processes (a combined term for the interrelated processes of mutagenesis and defective DNA repair), such as damage from UV light or defective mismatch repair, each leave distinctive (i.e. nonrandom) patterns of mutation across the genome has been instructive in

mapping genetic mutations to underlying biological process (3). On the other hand, the role of stochastic drift in shaping tumor evolution has been largely neglected.

The field of population genetics provides a quantitative framework with which to study evolution and has proved useful for understanding organismal evolution (66). It is perhaps the area of biology that has made the most use of mathematical theories, and for good reason. The difficulty in conducting experiments over evolutionary timescales necessitates theoretical approaches to explain patterns of genetic diversity within populations. Population genetics models provide a framework to explore the distinct evolutionary components of mutation, drift, and selection in a principled way. Since cancers also evolve, population genetics theory can also be applied to cancer, albeit with some adaptations necessary to represent the differences between organismal and somatic cell evolution (47). We discuss applications of population genetics theory in cancer in this review.

Classifying Tumor Evolution

There have been many efforts to categorize cancer evolution into one of several models (or modes), with the important aim of finding similarities in the evolutionary trajectories of different cancers so that experience in the treatment of one cancer that evolves according to a particular mode can guide treatment of similar cancers (68). Frequently discussed models of cancer evolution include neutral, punctuated, branched, and linear (25). Below, we discuss each of these terms.

Linear evolution describes the process where successively fitter mutants arise and sweep to fixation, replacing less fit lineages. While such selective sweeps do occur in cancers, the inevitable ongoing mutations mean that cancers always contains genetic heterogeneity. Thus, a sweep of a driver mutation should not be conflated with causing genetic homogeneity, and indeed, all cancer evolution is therefore in some sense branched. Nevertheless, the term branched evolution is typically used to describe the scenario where multiple subclones, each with selective growth advantages and effectively a distinct phenotype, co-occur within the tumor (39). Neutral evolution is the converse situation, wherein all subclonal lineages have equal fitness, and so is a description of what happens in the absence of selection. Neutral evolution and branched evolution are thus distinct —one describing the scenario where there are no differences in fitness between lineages (selective coefficient of each subclone $s = 0$), and the other describing differential fitness between some lineages ($s > 0$ for at least one subclone). The term effectively-

neutral can also be useful, denoting the situation where selection is weak and does not cause a detectable deviation from neutral evolutionary dynamics (104). It is important to appreciate however, that even in the case where evolutionary dynamics within a tumor are entirely neutral, the prior accumulation of driver events that led to tumor initiation must be driven by selection and adaptation (101).

Punctuation can be thought of as a catastrophic event that induces a radical change in phenotype, typically followed by strong selection for that phenotype (7, 23). The punctuated event can be the result of gradual evolution of a single isolated lineage that suddenly emerges following the evolution of the radical phenotype; this is the analogue of punctuated equilibrium, proposed by Stephen Jay Gould in the context of species evolution. In cancer genomes, there can also be large sudden changes in the genome in a single catastrophic event. These changes include chromothripsis (the shattering and reassembly of a chromosome arm) (56), chromoplexy (the interleaving of genetic material from multiple chromosomes, potentially in a single event) (100), genome doubling (13), and kataegis (localized hypermutation resulting in many single-base-pair changes) (81).

We argue that classifying cancer evolution into these apparently distinct models or modes is an illusion in many respects. After all, evolution itself is fundamentally a single process: Natural selection of competing lineages does not intermittently stop and start, but is rather always ongoing. How the evolutionary process appears at a single time point depends on how and when the tumor is sampled as well as the resolution of the assay. For example, if a tumor is sampled right after a clone has swept, then the evolution would appear to be linear, but if it is sampled just before the fixation event, then the evolution would appear to be branched (**Figure 1**). Furthermore, how the samples are taken in space could also lead to the appearance of linear (if only the sweeping clone is sampled), branched (if the sweeping clone and residual tumor population is sampled), or neutral evolution (if only the clone, or residual population, is sampled) (**Figure 1**). Spatially biased sampling and limited genetic resolution can also mean that some clones are missed and others overrepresented in the samples.

<COMP: PLEASE INSERT FIGURE 1 HERE>

Figure 1 Many modes of evolution can be seen through the history of an individual tumor. This tumor goes through distinct phases that can be characterized by many different modes.

Instead, we suggest that a useful approach is to think about the relative contributions of the different evolutionary components (mutation, drift, and selection) to the observed evolutionary dynamics. Specifically, this would mean measuring the mutation rate (separately for different types of mutations) and the distribution of fitness effects of these mutations (potentially also taking into the account the current microenvironment) and elucidating the relative importance of stochastic effects (e.g., the prevalence of drift). Together, these measures could provide an evolutionary index for cancer development (68). As a prelude to a discussion of approaches to quantify these aspects of cancer evolution, we first discuss the different evolutionary forces and their peculiarities in cancer.

Selection

The population genetics definition of selection is the increase in frequency of a particular genotype in the population due to the increased fitness of that genotype. Classically, fitness is defined as more surviving offspring per capita per generation. In tumors, fitness can be intuitively understood as the net growth rate of lineages relative to other lineages. Despite this simple framing, from a mechanistic point of view, the causes of selection can vary widely and are likely to be variable across time and space.

Positive selection, when subclones within a tumor grow more rapidly than others, is the dominant mode of selection during tumor initiation. Many of the so-called hallmarks of cancer result in increased proliferation, or the ability of cells to evade the homeostatic regulation of physiologically normal tissues (45). Sequencing of large cohorts has revealed recurrent mutations in certain genes (63), often suggesting large fitness effects (105). Other driver mutations are rarer, perhaps due to weaker selection, mutational unlikeliness, or a cancer requiring many small-effect drivers rather than a few (recurrent) strong drivers (20). Cancer-associated mutations also induce clonal expansions in what appear to be physiologically normal tissues (69, 71).

Treatment radically changes the selective pressures imposed by the tumor microenvironment and can have profound consequences. Many alterations that confer resistance likely exist in the tumor at the time treatment is applied. A concept that may be relevant to such dynamics is that of soft selective sweeps (76, 85). These sweeps occur when previously neutral mutations become adaptive following a change in the environment, resulting in multiple mutations with different genetic backgrounds simultaneously rising in frequency in the population. They are thought to be

an important adaptive mechanism in many evolutionary systems (96). By contrast, hard selective sweeps occur when mutations with large fitness advantages rapidly rise in frequency and fix in the population, resulting in a loss of diversity. Dynamics that fit the soft sweeps paradigm have been observed in cancer under treatment. For example, in colorectal cancer, *KRAS* mutations that confer treatment resistance to the targeted therapy cetuximab can rapidly rise in frequency after treatment is applied (28, 78). Treatment can therefore radically alter the fitness landscape of cancers. The ecological context within which mutations arise and how it changes over time are thus undoubtedly important (99), but more challenging to quantitatively measure than genetics (i.e. measurement of mutant allele frequencies).

An alternative mode of selection in cancer is negative or purifying selection. Negative selection arises when subclones that have reduced fitness are more likely to be lost or remain at low frequencies within the population. This type of selection is particularly important in the context of how the immune system interacts with cancer cells and is critical for predicting the efficacy of immunotherapy (97). Mutations that induce neoantigens on the cell surface can elicit an immune response that may purge lineages carrying neoantigens from the population. To counteract this, tumors employ various mechanisms of immune escape. Immunotherapies work by reactivating a dormant immune response, and their effectiveness correlates with the burden of neoantigens (64). The degree of negative selection experienced by neoantigens remains an open question. A fuller understanding of the mechanisms that immune escape tumors employ is required to realize the full potential of immunotherapies. Relatedly, there is also some speculation that the average fitness of cancer cells may decrease over time due to the constant accrual of slightly deleterious mutations (72).

Another curious property of selection in cancer is that, because tumors are growing populations, the effects of selection are less apparent (57, 101). Subclones that have fitness advantages may therefore never reach a high enough frequency within an overall rapidly growing tumor to affect the bulk clonal makeup of the tumor. In this case, the dominant clone is what is important for defining the biology of the tumor, as selection may not be strong enough (given the short timescales) to allow a new subclone to replace the dominant one.

Mutation

Mutation is crucial for evolution, as the diversity generated by mutational processes provides the substrate on which selection can act. As is the case for selection, mutation comes in different

flavors. The typical cancer genome is modified in radically different ways. The most straightforward to identify are point mutations, which are single-base-pair changes that can alter the protein-coding region and render it nonfunctional (e.g., tumor-suppressive mutations) or alter its function (e.g., oncogenic mutations). Slightly larger changes, such as insertions and deletions of base pairs (collectively called indels), can induce similar effects. Larger structural variations across the genome, including whole-genome doubling, chromosomal loss or gains, and translocations, are also common in different cancers (10).

Genome doubling is a common feature of cancer evolution (~30% of cases) and is thought to be a driver of copy number instability (13). Copy number aberrations may act synergistically with point mutations to increase or decrease the dosage of particularly important genes (12). Furthermore, some tumors are hypermutated due to inactivating mutations in DNA repair pathways (17). Current gaps in our understanding include the mutation rates of different types of mutations (passengers, drivers, structural variation, and copy number alterations) and how the various kinds of genomic instability modify the baseline rate. It appears likely that the point mutation rate is elevated in somatic tissue compared with germline tissue (77a), but the degree to which it is elevated in humans is unknown. The baseline rate of copy number alterations and structural variants in particular has proven hard to measure.

Neutral Drift

Another important aspect is the evolutionary dynamics in the absence of selection, i.e., neutral drift processes. Understanding these processes is useful to quantify the degree of diversity that we would expect to see in a tumor if all cells had the same fitness. Furthermore, neutrality provides the natural null model for molecular evolution (116). This is crucial because it enables selection to be distinguished from any variation that would be expected when there is no selection (55).

The peculiarities of cancer growth complicate how drift is manifested in cancer. Given that tumors are growing populations, modifications to classical models of drift that account for this are being developed (22). Furthermore, the expected frequencies of new mutations entering the population are also affected by the growth of tumors. For example, mutations that appear early during tumor growth will be present at a higher frequency than those that appear late, when the population is large (101). Drift will also likely be stronger during premalignant stages, when the population size of tumor cells is small, compared with when the tumor is clinically detectable

and contains billions of cells (19). High death rates can also lead to greater variability (16), but the death rate of cells—particularly after transformation, when the tumor is small (where drift effects are likely to dominate)—is unknown.

Further non-Darwinian variability may arise from spatial phenomena such as gene surfing, where mutations acquired on the expanding front of a population rise in frequency (in cancer, the expanding front could be thought of as the invasive edge), as has been demonstrated in bacteria and human populations (36, 84). Such processes likely occur in solid tumors as well (58), although a clear understanding of the mechanisms of the spatial growth of solid tumors is still lacking, and current data may lack power to uncover these phenomena (1).

QUANTIFYING CANCER EVOLUTION

Despite the many unknowns and challenges, quantifying the evolutionary components of cancer is becoming more realistic. Recent advances in high-throughput assays enable precise measurements of biological parameters in patient samples, and parallel developments in experimental systems allow experimentation within an evolving cancer model. Taken together, these advances provide exciting opportunities to produce quantitative measurements of cancer evolution both in model systems and *in vivo* across space and time. Mathematical theories are also necessary to extract maximal information from these data (5). Models from population genetics have been adapted to consider some of the peculiarities of cancer evolution (30). Models have also been developed to investigate diverse aspects of cancer evolution, such as the rate at which driver and passenger mutations accumulate (15), the likelihood of acquiring mutations that confer resistance (48), and the waiting time to cancer (9). A large body of theory is thus available or can be adapted to interpret cancer genomic studies.

One difficulty that theoreticians face is how to fit theoretical models to what are often noisy data. Generative Bayesian modelling is well suited to such tasks, as aspects of the data-generating procedure can be incorporated directly. Other Bayesian methods, such as approximate Bayesian computation, allow researchers to fit simulation-based models to data in a principled way (106). For example, sequencing introduces various sources of noise that can be modeled by an appropriate choice of distribution (e.g., beta-binomial for overdispersed coverage) or by generating synthetic data within a simulation-based framework to compare against real data.

Broadly, approaches to measure cancer evolution and its consequences for patient outcomes can be split into two groups. First, similarities and differences between cancer genomes across large cohorts hold valuable information. Recurrent patterns may show evidence of convergent evolution on particular phenotypes, while the absence of particular features can reveal signals of negative selection. Second, intratumor heterogeneity gives insight into the evolutionary dynamics of individual malignancies.

LEVERAGING COHORT-LEVEL INFORMATION TO STUDY CANCER EVOLUTION

Tens of thousands of cancers have been sequenced using next-generation sequencing technology. These data provide an unparalleled resource to study recurrent patterns that drive the progression of the disease and uncover signatures that may correlate with patient outcome.

Driver Mutations

Data on driver mutations are perhaps best exemplified by large cancer sequencing studies such as the Cancer Genome Atlas and the International Cancer Genome Consortium, which have focused on discovering these mutations. These studies have demonstrated that there are very few highly recurrent mutations and a long tail of rare driver mutations. Only a handful of mutations occur at appreciable frequencies across all cancer types; for example, only mutations in *TP53* and *PIK3CA* occur at a frequency of greater than 10% across cancer types (51). These results demonstrate a large degree of heterogeneity in cancer drivers.

Genomic changes other than point mutations also undoubtedly drive disease progression. Indeed, the first driver mutation to be identified was the *BCR-ABL* translocation in chronic myeloid leukemia (95). In general, however, identifying structural variation that drives the disease is more technically challenging because many cancers display genomic instability. Distinguishing changes that may modify fitness from mutations that arise due to the instability remains challenging due to the lack of a background structural mutation rate model.

Mutational Signatures

Another aspect that has also become prominent in recent years is the genomic context in which mutations arise. In a seminal study, Alexandrov et al. (3) found that mutations occurring in specific sequence contexts could be assigned to distinct signatures. Some of these signatures

were found across cancer types and are thought to be age related, whereas others are cancer specific. Examples include tobacco-specific mutational signatures in lung cancers (2) and mutations due to defects in DNA repair pathways. For example, C>A mutations, particularly when flanked by a C and A base (i.e., CCA>CAA), are enriched in tobacco smokers. Many signatures remain of unknown biological origin, however, suggesting the presence of unknown mutagens. Identifying the signatures also poses challenges, as several putatively distinct signatures appear similar and so are difficult to distinguish with high confidence.

Nonetheless, mutational signatures potentially provide a window into past exposures that may be useful for designing preventative strategies or elucidating unknown mutagens (4). There is also potential to shed light on the relative contributions of mutation and selection. Certain mutational processes may predispose people to certain driver mutations, while other driver mutations may be less mutationally likely but occur frequently because of the greater increase in fitness conferred by the mutation. Indeed, a mathematical treatise that explored these relationships showed that the *BRAF* V600E mutation, a common driver in many cancer types, is unlikely to occur given its mutational context and the typical mutational signatures associated with *BRAF*-positive cancers, but it is highly selected (105). Similar approaches from other groups have shown complementary results (18, 86). An experimental approach using CRISPR/Cas9 editing on cancer cell lines was also able to deconvolve the effects of selection and mutations and identify the most functionally active domains in *TP53* (42). A similar approach was applied to screen for the effects of all possible *BRCA* mutations (35). Such experimental approaches together with larger data sets may enable characterization of the fitness effects of single point mutations across the whole genome.

The concept of mutational signatures has recently been extended to the analysis of copy number alterations (67). Particular signatures in the genomes of ovarian cancer, which is known to be driven by chromosomal abnormalities, correlate with patient outcome (67, 111). Extensions of this approach to other cancer types will shed further light on chromosomal mutational processes.

dN/dS

dN/dS—the ratio of nonsynonymous mutations to synonymous mutations normalized by the nonsynonymous and synonymous mutation likelihoods, respectively—is another method that can be used to infer selection. Originally developed for comparative genomics in species evolution,

dN/dS quantifies whether there are more mutations that alter amino acid sequences (which are potentially adaptive) than would be expected by chance. Synonymous mutations are assumed to be neutral and provide a measure of the baseline mutation rate. An excess of nonsynonymous mutations (dN/dS > 1) indicates positive selection, a deficit of nonsynonymous mutations (dN/dS < 1) indicates negative (stabilizing) selection, and equal rates of the two mutation types (dN/dS = 1) implies neutral evolution. Interestingly, in cancer, studies have reported an exome-wide dN/dS close to 1 (70, 112), suggesting the absence of negative selection and that most mutations are neutral. Driver genes, such as *TP53* and *NOTCH1*, can have dN/dS values far in excess of 1 (70). Zapata et al. (119) were able to find strong signals of negative selection by restricting the analysis to neoantigens—mutations that cause cell surface markers that can be recognized by the immune system. There may be complex relationships between the clonality of neoantigens and their recognition by the immune system (38, 65b, 73).

In practical terms, robustly measuring dN/dS in cancer genomes is challenging, as the null distribution for mutations in cancer genomes is complex and varies across the genome (109), and calculating dN/dS therefore also requires the simultaneous inference of the local mutation rate (70, 112). Moreover, the theoretical basis of dN/dS in species evolution relies on many assumptions that are largely violated in cancer. Principal among these are the assumptions that evolution has occurred over timescales long enough that selection has had time to act and, relatedly, that dN/dS quantifies the relative rate at which mutations fix in the population. In cancers, intratumor heterogeneity is pervasive, so many mutations are subclonal (not fixed), indicating that selection has not yet had time to entirely sort the fittest clones. Moreover, traditional dN/dS models assume populations of a fixed size, which of course is also violated in cancer. Theoretical studies have shown that when some of these assumptions are violated, interpreting dN/dS becomes difficult (59). The full linkage of mutations in cancer genomes (e.g., lack of recombination) means that neutral or deleterious mutations commonly hitchhike on the within positively selected clones. This can lead to the masking of signals of negative and positive selection, as the evolutionary signal is integrated over these distinct processes. In theory, a dN/dS value of 1 could be the result of a combination of positively and negatively selected clones that combine to produce a dN/dS of 1, rather than strict neutrality (21). Furthermore, because dN/dS measurements in cancer are calculated by pooling together many different patients, the dN/dS value of the cohort can be driven by outliers, with few patients having many positively selected

driver mutations driving the whole signal (46, 104a). Nevertheless, dN/dS methods adapted to cancer, such as those recently proposed (70, 112, 119) remain powerful tools to detect selection in clonal evolution.

INTRATUMOR HETEROGENEITY AS A WINDOW INTO CANCER EVOLUTION

Due to the continual acquisition of genetic mutations in a growing tumor and the unavoidable intratumor genetic heterogeneity that results, a cancer genome—or, more accurately, the differences between the genomes of single cells in a cancer—contains a record of the tumor's evolutionary history. Each new mutation acquired by a single cell in a cancer will be passed on to its daughter cells, and these mutations thus record ancestral relationships between cells (**Figure 2*a***). Fundamentally, the genome of every cancer cell is an imperfect copy of another cancer cell that existed in the past. This simple observation—that heterogeneity emerges from cell divisions coupled with the occurrence of new mutations—therefore allows one to infer the past history of a cancer, or indeed any somatic tissue. Genomic analysis thus provides a window to study cancer evolution that circumvents one of the biggest issues facing the study of cancer as an evolutionary system: the inability to follow cancers *in vivo* unperturbed over time, due to clinical necessity and ethical issues. This provides a perspective that is different from those of the cohort-level approaches, as inferences can be made on a patient-by-patient basis.

<COMP: PLEASE INSERT FIGURE 2 HERE>

Figure 2 (*a*) Mutations act as labels that can track the evolutionary dynamics of cancer cell populations, here each different colored dot is a distinct mutation which accumulate during cell division (*b*) To uncover these dynamics, bulk sequencing pools all alleles together from the sample and enables measurement of the frequency of mutations in the pool. Low-frequency variants may be missed however, as the depth of sequencing at each locus is limited. Due to the alleles being pooled, information on the co-occurrence of mutations is in general lost. (*c*) Single-cell sequencing potentially provides far greater resolution of cell lineages than bulk sequencing as each cell is individually tagged and sequenced.

Lineage Tracing

Lineage tracing refers to the use of a heritable label (such as somatic mutations) to track clonal descendants over time. Coupled with theoretical models to explain the resulting clone-size distribution measurements and ancestral relationships, lineage tracing provides a powerful

method with which to study evolution and population dynamics. The concept of lineage tracing has also been used extensively in model systems, particularly in developmental biology and stem cell biology to track the progeny of particular cell types of interest (14). Advances in experimental techniques have been used in recent years to look at the evolutionary dynamics of tumors in model systems (29, 88, 110).

Both natural and artificial labels have been used for lineage tracing in cancers and experimental model systems. Labeling of cell populations with fluorescent reporters has been used in mouse models of tumor growth, identifying stem cell populations in squamous skin tumors (29) and the clonal dynamics required for the formation of skin tumors (94). The same approach has shown that stem cell dynamics in expanding colorectal neoplasms are plastic and entirely driven by the environment (64a). That somatic mutations are natural labels is at the heart of phylogenetic principles applied to cancer. Early examples of this kind of approach used X-chromosome inactivation as a lineage marker to demonstrate that cancers were of single-cell origin (34), and the now sophisticated field of cancer phylogenetics makes use of multiple types of mutations simultaneously to infer the evolutionary history of a tumor (98).

High-Throughput Experimental Lineage Tracing

Many experimental systems have been developed that rely on lineage tracing to quantify population dynamics. Fluorescent reporter constructs have long been used to identify putative stem cells and determine the statistical properties of cell fate outcomes (14, 54). Such experiments have also been used to quantify the fitness of oncogenic mutations during tumor initiation. Vermeulen et al. (110) showed that *KRAS* and *APC* mutations were two to four times as likely as other mutations to reach fixation in mouse colonic crypts**[**AU: Edits OK? (If not, please clarify)**]**. Experimental systems based on fluorescent reporters tend to suffer from a lack of resolution because only a small number of clones can be tracked over time. To circumvent these issues, high-throughput lineage-tracing protocols have been developed via the use of multiplexed DNA barcodes. These barcodes can be inserted via viral transfection into the genomes of single cells and provide a unique tag for each cell. Millions of clones can be traced simultaneously with this approach via sequencing of pools of barcoded cells (11). Barcode libraries are constructed such that each transfected cell carries a unique label that can be used to measure its size. Deep sequencing of pools of barcoded cells therefore measures the sizes of lineages within the population. As only a few base pairs of the genome need to be sequenced (the

barcodes), high-depth coverage of the barcode is employed, which results in high-resolution tracking of individual lineages.

Levy et al. (65) used this approach to measure lineage sizes in serial passages of yeast cells. Using these data together with theoretical population genetics enabled them to quantify when fitter lineages emerged and the distribution of fitness effects by identifying lineages that increased in size faster than could be expected from stochastic neutral drift. Similar experimental strategies have recently been applied to cancer model systems. For example, Rogers et al. (88, 89) measured the fitness of 11 tumor suppressor pathways by using CRISPR/Cas9 genome editing to introduce mutations followed by barcoding to measure tumor size in mouse models of lung cancer. They found that mutations in *SETD2* and *LKB1* had the largest fitness effect and resulted in the largest tumors. Lan et al. (60) used a barcoding approach in glioblastoma models in mice and showed that intratumor heterogeneity in glioblastoma was driven largely by the stochastic fate of cells in a stem cell hierarchy, while treatment-resistant clones could be identified via deviations from this model. More complex experimental strategies are likely to provide further insight; for example, barcoding potentially allows tracking of the size of competing lineages over time, which could be used to measure the relative fitness of subclones or the effects of the immune system on clonal diversity.

Deep Sequencing

Lineage size is the crucial piece of information that is revealed by deep sequencing studies and enables the population dynamics to be inferred. A useful way to summarize the information from a deep sequencing experiment is by plotting a histogram of the mutation frequencies. In cancer this is commonly referred to as the variant allele frequency distribution (or the cancer cell fraction distribution, after correction for tumor cellularity, ploidy, and relative copy number). In population genetics, this distribution is known as the site frequency spectrum, and there is a considerable body of work devoted to exploiting it to measure evolutionary dynamics (52).

Different demographic models make different predictions about the shapes of these distributions. For example, the classic Wright–Fisher model predicts that the number of mutations present at a frequency f, follows a $1/f$ dependency (33), while a model that assumes exponential growth predicts a $1/f^2$ relationship (53), a result closely related to the Luria–Delbrück distribution (121)**.** Using this exponential growth model ($1/f^2$) as a neutral null model for cancer evolution, Williams et al. (114) found that approximately 30% of cancers fitted the model well.

Subpopulations of cells that possess fitness differences can cause deviations from the expected distribution under neutrality. Mutations that hitchhike on the back of any selected lineage will be found at the same frequency and uniquely mark the subpopulations of cells (43). These mutations appear as striking clusters in the variant allele frequency distributions of deeply sequenced tumors (82). Many tools have been developed to identify these clusters (77, 90). The generative model assumed by these methods (variant allele frequency distribution explained by multiple distinct clusters, each corresponding to a distinct clonal population) is often violated because many mutations appear at similar frequencies but are present in distinct lineages, a natural consequence of population growth (46). Taking this into account, Williams et al. (115) used a theoretical model to show that information encoded in these clusters enables estimation of the relative fitness advantages of different subclones within a tumor. Deviations from the $1/f^2$ distribution may also arise due to different demographic histories; analytical results are available for power law or boundary driver growth under certain assumptions, where power law scaling for the size distribution is also expected but with different exponents (36, 80).

Sequencing depth remains a major factor that limits the resolution of the data. The typical sequencing depth is currently approximately 100× for whole-exome sequencing and approximately 40× for whole-genome sequencing. Greater depth is required to detect small subclones, which may be biologically important (e.g., in the detection of rare drug-resistant clones). Deeper sequencing of samples as well as high-fidelity sequencing approaches to overcome technical noise limitations at low frequencies can allow the probing of evolutionary dynamics at higher resolution (93).

Copy number aberrations can also be combined with point mutations to infer patterns of evolution in single bulk samples. The proportion of mutations observed on amplified chromosomes compared with one chromosome can be used to time the appearance of copy number alterations (50). This type of analysis using whole-genome sequencing in breast cancer indicated that chromosomal alterations accumulate steadily in this cancer type but are not some of the earliest events (82). Similar analysis has been conducted pan-cancer to identify the timing of important events across cancer types (41), showing that biallelic loss of tumor suppressors is often an early event, while genomic instability increases at later stages. Again, high-depth sequencing is preferred because it facilitates the greater resolution necessary to be able to accurately map point mutations to amplified or single chromosomes.

Multiregion Sequencing

A second approach to infer tumor evolution is to take multiple samples from a tumor and measure the (genetic) differences between samples. This approach has been used for many years. For example, Tsao et al. (107, 108) used genetic divergence (genetic distance between samples) as a summary statistic between spatially distinct regions in mismatch-repair-deficient tumors; in conjunction with a computational model, this approach enabled them to estimate the ages of adenomas (premalignant lesions) versus cancers. Perhaps surprisingly, the adenomas and carcinomas were of similar ages.

Multiregion sequencing studies are now commonplace. The biggest contribution of such studies thus far has been demonstrating the extensiveness of intratumor heterogeneity across cancer types. In one prominent study, Gerlinger et al. (40) used whole-exome sequencing of different tumor regions to profile clear cell renal carcinomas and found a large degree of intratumor heterogeneity. Later studies from the same group found evidence of convergent evolution with distinct putative driver mutations in *SETD2* on different branches of the phylogenetic tree (39). Further multiregion sequencing studies have shown that intratumor heterogeneity is pervasive across cancer types, including lung (27, 120), breast (117) (8a), lymphoma (83), brain (102) (102a), and colon cancers (101) (24), among others.

Phylogenetic analysis reveals the temporal order of events and has shown that driver mutations are often truncal on the phylogenetic tree—that is, found ubiquitously across all sampled regions (87). This finding suggests that most of the important driver events are acquired early relative to the time patients present with symptoms of their disease. This is particularly true in some cancer types, such as colon and lung cancer (101, 120), while kidney cancers, for example, often appear to have subclonal driver mutations (39).

Multiregion sequencing together with phylogenetic analysis has also been useful in determining how evolution is influenced by environmental factors (27) and elucidating the seeding patterns of metastasis (32, 75). Studies such as the Tracking Cancer Evolution Through Therapy (TRACERx) clinical trial are currently under way to determine the effects of intratumor heterogeneity on patient prognosis using multiregion sequencing assays in multiple cancer types (49) (108a). Machine learning approaches are also being developed to integrate information from cohorts of multiregion sequencing data to uncover repeated evolutionary trajectories (19a).

It is important to note that the patterns that emerge from phylogenetic analysis of tumor samples are subject to multiple confounding factors. First, accurately reconstructing the phylogenetic relationships between tumor sites remains challenging (98). Other issues include sampling bias, where samples may not be taken uniformly across the tumor mass and/or may inadvertently be confined to particular subclones, resulting in an unrepresentative sample of the tumor. Relatedly, the typical limited sampling (four to five samples per tumor) can result in misclassifying truncal mutations (113). Perhaps the biggest challenge is that bulk tumor samples (e.g., a biopsy consisting of hundreds of thousands to millions of cells) potentially consist of multiple subclones, and therefore the phylogenetic relationships should ideally be constructed based on the deconvolved clonal structure (6), but deconvolving bulk tumor samples into subclones remains technically challenging (see above). Additionally, clonal mixing in three-dimensional tumors makes relating phylogenies to the underlying evolutionary dynamics difficult (104).

Single-Cell Sequencing

Recent advances in single-cell sequencing resolve some of the issues of studies that rely on bulk sequencing but also introduce new challenges (79). Single-cell sequencing potentially provides unparalleled resolution of tumor genetic diversity, identifying subclonal populations that would most likely be missed by conventional bulk sequencing (**Figure 2**). Single-cell sequencing thus provides opportunities for fine-grained analysis of cancer evolution and has the benefit of avoiding the need to resolve the clonal structure through complicated clustering-based approaches, as each cell is a pure sample by its very nature. This type of approach does, however, come with its own set of problems. In particular, the degree of technical noise is higher than that of other sequencing approaches, and issues of sampling bias remain, given that only 100 cells of a tumor comprising billions of cells are typically sampled.

Technical issues in single-cell sequencing technology arise from the low quantity of DNA extracted from single cells, meaning that whole-genome amplification is generally required to generate sufficient DNA for sequencing. This additional step introduces technical artifacts such as nonuniform coverage and allele dropout (26). Single-nucleotide variants are also difficult to accurately detect due to high technical error rates (91). For this reason, copy number profiling, which is more robust, is generally preferred. Sophisticated single-cell-specific algorithms have been developed for analyzing these data (91, 92, 100a). Recent technical advances have shown

that single-cell sequencing without whole-genome amplification is possible (118) and that pooling single cells can make single-nucleotide variant calling more robust (92). Further advances in this area are likely to provide exquisitely fine-grained data for evolutionary analyses. As methods to reduce noise continue to improve and sequencing costs continue to fall, single-cell sequencing will no doubt become the preferred assay for evolutionary analysis of tumors.

Single-cell sequencing studies have already demonstrated the power and advantages of this approach over other methodologies and revealed interesting aspects of the evolutionary process. Gao et al. (37) used single-cell sequencing to look at aneuploidy in triple-negative breast cancer. Interestingly, copy number alterations appeared to be spatially and therefore temporally stable, suggesting that large-scale copy number changes are perhaps rare events during tumor evolution in breast cancer. Single-cell sequencing will no doubt be particularly illuminating as it is applied to investigate chromosomal (in)stability in other tumor types. Another interesting study that employed single-cell sequencing investigated the temporal dynamics of cancer evolution using a patient-derived xenograft model (31). Interestingly, the authors found that minor clones often come to dominate the tumor population, suggesting that some clones acquire large fitness advantages, which are likely necessary to induce such large expansions. The degree to which small subpopulations influence evolutionary trajectories in general will be interesting to observe as single-cell sequencing expands in scope.

From genotype to phenotype

Elucidating the phenotypic effects that drive adaptive change is challenging from DNA sequencing alone, as a multitude of epigenetic and environmental factors determine whether or not a mutant protein is produced, and whether or not the mutant protein can play a functional role in the cell's current context. For evolutionary analysis this is a very important consideration, as, after all, selection acts on phenotypes rather than genotypes. To explore the mapping between phenotype and genotype, genomics can be augmented with other measurement modalities (multi-omics) that measure the transcriptome, epigenome or immune cell repertoire for example. For example, integrative molecular analysis using the Cancer Genome Atlas dataset has explored the pan-cancer immune landscape (105a) and the association of the cell of origin with molecular features (46a) amongst other things. Zhang et al. (120a) used a multi-omics approach combining whole genome sequencing, transcriptomics and immune cell receptor sequencing to demonstrate

that the spatial organization of the immune microenvironment influences the clonal dynamics in ovarian cancer. A longitudinal study in Follicular Lymphoma also uncovered immune signatures that correlate with patient prognosis using a combination of genetics, T cell repertoire sequencing and transcriptomics (6b). Genetics has also been combined with machine learning based image analysis, Heindle et al. (45a) demonstrated that defects in the DNA repair machinery of cells correlate with morphological changes in ovarian cancer cells, and that diversity of morphological phenotypes correlated with survival. Multi-omics approaches are also being extended to the single cell arena through the simultaneous measurement of genomes and transcriptomes from the same cell, allowing for investigation into changes in gene dosage due to chromosomal aneuploidies and effects of gene fusions at the single cell level (66a). These few studies give some insight into how a complete picture of the evolutionary dynamics, and the molecular mechanisms than ultimately drive them, can be elucidated from multi-omic measurement. Future studies combining the population genetics based approaches using genomics outlined in this review with assays that interrogate the phenotypic changes occurring during cancer evolution will provide new perspectives on cancer evolutionary dynamics.

SUMMARY AND OUTLOOK

Genomics, coupled with mathematical evolutionary theory, facilitates quantitative measurements of the evolutionary dynamics that underpin cancer development. Our opinion is that this quantitative dissection of the components of cancer evolution—namely, mutation, drift, and the many facets that lead to selection—will provide an objective way to categorize cancer evolution that is more robust than current broad-scale classifications into evolutionary modes. Such quantitative information is also inherently mechanistic, describing precisely how a cancer will change over time. Thus, we suggest that quantitative measurement of cancer evolution will provide a way to accurately forecast cancer evolution (61)—for example, to predict with confidence how, and how quickly, a tumor will evolve with or without treatment. For example, a recent study on drug resistance in colorectal cancer demonstrated that longitudinal liquid biopsies and mathematical modelling of tumor evolution allowed prediction of tumor recurrence time (53a). We suggest that the ultimate goal of measuring cancer evolution is to facilitate clinical decision-making, and that quantitative knowledge of the evolutionary dynamics of cancer development is the route to accurate prediction of disease course.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

Acknowledgements

T.G. is supported by the Wellcome Trust (202778/Z/16/Z), Cancer Research UK (A19771) and the National Institute of Health (NCI U54 CA217376). A.S. is supported by the Wellcome Trust (202778/B/16/Z) and Cancer Research UK (A22909). We acknowledge funding from the National Institute of Health (NCI U54 CA217376) to A.S. This work was also supported a Wellcome Trust award to the Centre for Evolution and Cancer (105104/Z/14/Z). We are always grateful to Benjamin Werner and Christopher Barnes for discussions that helped to shape this review.

LITERATURE CITED

1. Ahmed Z, Gravel S. 2018. Intratumor heterogeneity and circulating tumor cell clusters. *Mol. Biol. Evol.* 35:2135–44
2. Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, et al. 2016. Mutational signatures associated with tobacco smoking in human cancer. *Science* 354:618–22
3. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, et al. 2013. Signatures of mutational processes in human cancer. *Nature* 500:415–21
4. Alexandrov LB, Stratton MR. 2014. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr. Opin. Genet. Dev.* 24:52–60
5. Altrock PM, Liu LL, Michor F. 2015. The mathematics of cancer: integrating quantitative models. *Nat. Rev. Cancer* 15:730–45
6. Alves JM, Prieto T, Posada D. 2017. Multiregional tumor trees are not phylogenies. *Trends Cancer* 3:546–50

6a. Andor N, Graham TA, Jansen M, Li CX, Aktipis CA, et al. 2016 Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nature Medicine* **22,** 105–113

6b. Araf S, Korfi K, Nagano A, Cummin TEC, Bentley M, et al. 2017 Longitudinal Analyses of the Genomic, Transcriptomic, and T Cell Repertoire in Diffuse Large B Cell Lymphoma Demonstrates Changes in Signaling and Immune Recognition at Relapse. *Blood* **130,** 2734

7. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, et al. 2013. Punctuated evolution of prostate cancer genomes. *Cell* 153:666–77

8. Bakhoum SF, Cantley LC. 2018. The multifaceted role of chromosomal instability in cancer and its microenvironment. *Cell* 174:1347–60

8a. Barry P, Vatsiou A, Spiteri I, Nichol D, Cresswell G, et al. 2018 The Spatiotemporal Evolution of Lymph Node Spread in Early Breast Cancer. *Clinical Cancer Research* **24,** 4763–4770

9. Beerenwinkel N, Antal T, Dingli D, Traulsen A, Kinzler KW, et al. 2007. Genetic progression and the waiting time to cancer. *PLOS Comput. Biol.* 3:e225

10. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, et al. 2010. The landscape of somatic copy-number alteration across human cancers. *Nature* 463:899–905

11. Bhang H-EC, Ruddy DA, Krishnamurthy Radhakrishna V, Caushi JX, Zhao R, et al. 2015. Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat. Med.* 21:440–48

12. Bielski CM, Donoghue MTA, Gadiya M, Hanrahan AJ, Won HH, et al. 2018. Widespread Selection for Oncogenic Mutant Allele Imbalance in Cancer. *Cancer Cell* 34:852–62.e4

13. Bielski CM, Zehir A, Penson AV, Donoghue MTA, Chatila W, et al. 2018. Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat. Genet.* 50:1189–95

14. Blanpain C, Simons BD. 2013. Unravelling stem cell dynamics by lineage tracing. *Nat. Rev. Mol. Cell Biol.* 14:489–502

15. Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, et al. 2010. Accumulation of driver and passenger mutations during tumor progression. *PNAS* 107:18545–50

16. Bozic I, Gerold JM, Nowak MA. 2016. Quantifying clonal and subclonal passenger mutations in cancer evolution. *PLOS Comput. Biol.* 12:e1004731

17. Campbell BB, Light N, Fabrizio D, Zatzman M, Fuligni F, et al. 2017. Comprehensive analysis of hypermutation in human cancer. *Cell* 171:1042–56.e10

18. Cannataro VL, Gaffney SG, Townsend JP. 2018. Effect sizes of somatic mutations in cancer. *J. Natl. Cancer Inst.* 110:1171–77

19. Cannataro VL, McKinley SA, St Mary CM. 2016. The implications of small stem cell niche sizes and the distribution of fitness effects of new mutations in aging and tumorigenesis. *Evol. Appl.* 9:565–82

19a. Caravagna G, Giarratano Y, Ramazzoti D, Tomlinson I, Graham TA, et al. 2018 Detecting repeated cancer evolution from multi- region tumor sequencing data. *Nat Methods* **15,** 1–13

20. Castro-Giner F, Ratcliffe P, Tomlinson I. 2015. The mini-driver model of polygenic cancer evolution. *Nat. Rev. Cancer* 15:680–85

21. Chen B, Shi Z, Chen Q, Shibata D, Wen H, Wu C-I. 2018. Quasi-neutral molecular evolution—when positive and negative selection cancel out. bioRxiv 330811. https://doi.org/10.1101/330811

22. Chen Y, Tong D, Wu C-I. 2017. A new formulation of random genetic drift and its application to the evolution of cell populations. *Mol. Biol. Evol.* 34:2057–64

23. Cross W, Graham TA, Wright NA. 2016. New paradigms in clonal evolution: punctuated equilibrium in cancer. *J. Pathol.* 240:126–36

24. Cross W, Kovac M, Mustonen V, Temko D, Davis H, et al. 2018. The evolutionary landscape of colorectal tumorigenesis. *Nat. Ecol. Evol.* 2:1661–72

25. Davis A, Gao R, Navin N. 2017. Tumor evolution: linear, branching, neutral or punctuated? *Biochim. Biophys. Acta Rev. Cancer* 1867:151–61

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

27. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, et al. 2014. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* 346:251–56

28. Diaz LA, Williams RT, Wu J, Kinde I, Hecht JR, et al. 2012. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* 486:537–40

29. Driessens G, Beck B, Caauwe A, Simons BD, Blanpain C. 2012. Defining the mode of tumour growth by clonal analysis. *Nature* 488:527–30

30. Durrett R. 2013. Population genetics of neutral mutations in exponentially growing cancer cell populations. *Ann. Appl. Probab.* 23:230–50

31. Eirew P, Steif A, Khattra J, Ha G, Yap D, et al. 2015. Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* 518:422–26

32. El-Kebir M, Satas G, Raphael BJ. 2018. Inferring parsimonious migration histories for metastatic cancers. *Nat. Genet.* 50:718–26

33. Ewens WJ. 2012. *Mathematical Population Genetics*, Vol. 1: *Theoretical Introduction*. New York: Springer. 2nd ed.**[**AU: OK?**]**

34. Fialkow PJ. 1979. Clonal origin of human tumors. *Annu. Rev. Med.* 30:135–43
35. Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, et al. 2018. Accurate classification of *BRCA1* variants with saturation genome editing. *Nature* 562:217–22
35a. Fraser M, Sabelnykova VY, Yamaguchi TN, Heisler LE, Livingstone J, et al. 2017 Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* **541,** 359–364
36. Fusco D, Gralka M, Kayser J, Anderson A, Hallatschek O. 2016. Excess of mutational jackpot events in expanding populations revealed by spatial Luria-Delbrück experiments. *Nat. Commun.* 7:12760
37. Gao R, Davis A, McDonald TO, Sei E, Shi X, et al. 2016. Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat. Genet.* 48:1119–30
38. Gejman RS, Chang AY, Jones HF, DiKun K, Hakimi AA, et al. 2018. Rejection of immunogenic tumor clones is limited by clonal fraction. *eLife* 7:635
39. Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, et al. 2014. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* 46:225–33
40. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, et al. 2012. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* 366:883–92
41. Gerstung M, Jolly C, Leshchiner I, Dentro SC, Gonzalez Rosado S, et al. 2018. The evolutionary history of 2,658 cancers. bioRxiv 161562. https://doi.org/10.1101/161562
42. Giacomelli AO, Yang X, Lintner RE, McFarland JM, Duby M, et al. 2018. Mutational processes shape the landscape of *TP53* mutations in human cancer. *Nat. Genet.* 50:1381–87
43. Gillespie JH. 2000. Genetic drift in an infinite population: the pseudohitchhiking model. *Genetics* 155:909–19
44. Gordon DJ, Resio B, Pellman D. 2012. Causes and consequences of aneuploidy in cancer. *Nat. Rev. Genet.* 13:189–203
45. Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: the next generation. *Cell* 144:646–74
45a. Heindl A, Khan AM, Rodrigues DN, Eason K, Sadanandam A, et al. 2018 Microenvironmental niche divergence shapes BRCA1-dysregulated ovarian cancer morphological plasticity. *Nat Commun* 1–14

46. Heide T, Zapata L, Williams MJ, Werner B, Barnes CP, et al. 2018. Reply to 'Neutral tumor evolution?' *Nat. Genet.* 50:1633–37

46a. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, et al. 2018 Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173,** 291–304.e6

47. Hu Z, Sun R, Curtis C. 2017. A population genetics perspective on the determinants of intra-tumor heterogeneity. *Biochim. Biophys. Acta Rev. Cancer* 1867:109–26

48. Iwasa Y, Nowak MA, Michor F. 2006. Evolution of resistance during clonal expansion. *Genetics* 172:2557–66

49. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, et al. 2017. Tracking the evolution of non-small-cell lung cancer. *N. Engl. J. Med.* 376:2109–21

50. Jolly C, Van Loo P. 2018. Timing somatic events in the evolution of cancer. *Genome Biol*. 19:95

51. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, et al. 2013. Mutational landscape and significance across 12 major cancer types. *Nature* 502:333–39

52. Keinan A, Clark AG. 2012. Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* 336:740–43

53. Kessler DA, Levine H. 2014. Scaling solution in the large population limit of the general asymmetric stochastic Luria-Delbrück evolution process. *J. Stat. Phys.* 158:783–805

53a. Khan KH, Cunningham D, Werner B, Vlachogiannis G, Spiteri I, et al. 2018 Longitudinal Liquid Biopsy and Mathematical Modeling of Clonal Evolution Forecast Time to Treatment Failure in the PROSPECT-C Phase II Colorectal Cancer Clinical Trial. *Cancer Discov* 1–17

54. Klein AM, Simons BD. 2011. Universal patterns of stem cell fate in cycling adult tissues. *Development* 138:3103–11

55. Koonin EV. 2016. Splendor and misery of adaptation, or the importance of neutral null for understanding evolution. *BMC Biol.* 14:114

56. Korbel JO, Campbell PJ. 2013. Criteria for inference of chromothripsis in cancer genomes. *Cell* 152:1226–36

57. Korolev KS, Müller MJI, Karahan N, Murray AW, Hallatschek O, Nelson DR. 2012. Selective sweeps in growing microbial colonies. *Phys. Biol.* 9:026008

58. Kostadinov R, Maley CC, Kuhner MK. 2016. Bulk genotyping of biopsies can create spurious evidence for heterogeneity in mutation content. *PLOS Comput. Biol.* 12:e1004413
59. Kryazhimskiy S, Plotkin JB. 2008. The population genetics of dN/dS. *PLOS Genet.* 4:e1000304
60. Lan X, Jörg DJ, Cavalli FMG, Richards LM, Nguyen LV, et al. 2017. Fate mapping of human glioblastoma reveals an invariant stem cell hierarchy. *Nature* 549:227–32
61. Lässig M, Mustonen V, Walczak AM. 2017. Predicting evolution. *Nat. Ecol. Evol.* 1:77
62. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, et al. 2014. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505:495–501
63. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, et al. 2013. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499:214–18
64. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, et al. 2017. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* 357:409–13
64a. Lenos KJ, Miedema DM, Lodestijn SC, Nijman LE, van den Bosch T, et al. 2018 Stem cell functionality is microenvironmentally defined during tumour expansion and therapy response in colon cancer. *Nature Cell Biology* **20,** 1193–1202
65. Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS, Sherlock G. 2015. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* 519:181–86
65a. Lipinski KA, Barber LJ, Davies MN, Ashenden M, Sottoriva A, et al. 2016 Cancer Evolution and the Limits of Predictability in Precision Cancer Medicine. *Trends in Cancer* **2,** 49–63
65b. Łuksza M, Riaz N, Makarov V, Balachandran VP, Hellmann MD, et al. 2017. A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. *Nature* 27:517–20
66. Lynch M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. *PNAS* 104(Suppl. 1):8597–604
66a. Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, et al. 2015 G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat Methods* **12,** 519–522 (2015).
67. Macintyre G, Goranova TE, De Silva D, Ennis D, Piskorz AM, et al. 2018. Copy number signatures and mutational processes in ovarian carcinoma. *Nat. Genet.* 50:1262–70

68. Maley CC, Aktipis A, Graham TA, Sottoriva A, Boddy AM, et al. 2017. Classifying the evolutionary and ecological features of neoplasms. *Nat. Rev. Cancer* 17:605–19

68a. Maley CC, Galipeau PC, Finley JC, Wongsurawat VJ, Li X, et al. 2006 Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nature Genetics* **38,** 468–473

69. Martincorena I, Fowler JC, Wabik A, Lawson ARJ, Abascal F, et al. 2018. Somatic mutant clones colonize the human esophagus with age. *Science* 362:911–17

70. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, et al. 2017. Universal patterns of selection in cancer and somatic tissues. *Cell* 171:1029–41.e21

71. Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, et al. 2015. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* 348:880–86

71a. Martinez P, Timmer MR, Lau CT, Calpe S, Sancho-Serra MC et al. 2016 Dynamic clonal equilibrium and predetermined cancer risk in Barrett's oesophagus. *Nat Commun* **7,** 1–10

72. McFarland CD, Korolev KS, Kryukov GV, Sunyaev SR, Mirny LA. 2013. Impact of deleterious passenger mutations on cancer progression. *PNAS* 110:2910–15

73. McGranahan N, Furness AJS, Rosenthal R, Ramskov S, Lyngaa R, et al. 2016. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* 351:1463–69

74. McGranahan N, Swanton C. 2017. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell* 168:613–28

75. McPherson A, Roth A, Laks E, Masud T, Bashashati A, et al. 2016. Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nat. Genet.* 48:758–67

76. Messer PW, Petrov DA. 2013. Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol. Evol.* 28:659–69

77. Miller CA, White BS, Dees ND, Griffith M, Welch JS, et al. 2014. SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLOS Comput. Biol.* 10:e1003665

77a. Milholland B, Dong X, Zhang L, Hao X, Suh Y et al. 2017 Differences between germline and somatic mutation rates in humans and mice. *Nat Commun* **8,** 1–8

78. Misale S, Yaeger R, Hobor S, Scala E, Janakiraman M, et al. 2012. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* 486:532–36
79. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, et al. 2011. Tumour evolution inferred by single-cell sequencing. *Nature* 472:90–94
80. Nicholson MD, Antal T. 2016. Universal asymptotic clone size distribution for general population growth. *Bull. Math. Biol.* 78:2243–76
81. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, et al. 2012. Mutational processes molding the genomes of 21 breast cancers. *Cell* 149:979–93
82. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, et al. 2012. The life history of 21 breast cancers. *Cell* 149:994–1007
83. Okosun J, Bödör C, Wang J, Araf S, Yang C-Y, et al. 2014. Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. *Nat. Genet.* 46:176–81
84. Peischl S, Dupanloup I, Foucal A, Jomphe M, Bruat V, et al. 2018. Relaxed selection during a recent human expansion. *Genetics* 208:763–77
85. Pennings PS, Hermisson J. 2006. Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. *Mol. Biol. Evol.* 23:1076–84
86. Poulos RC, Wong YT, Ryan R, Pang H, Wong JWH. 2018. Analysis of 7,815 cancer exomes reveals associations between mutational processes and somatic driver mutations. *PLOS Genet*. 14:e1007779
87. Reiter JG, Makohon-Moore AP, Gerold JM, Heyde A, Attiyeh MA, et al. 2018. Minimal functional driver gene heterogeneity among untreated metastases. *Science* 361:1033–37
88. Rogers ZN, McFarland CD, Winters IP, Naranjo S, Chuang C-H, et al. 2017. A quantitative and multiplexed approach to uncover the fitness landscape of tumor suppression in vivo. *Nat. Methods* 14:737–42
89. Rogers ZN, McFarland CD, Winters IP, Seoane JA, Brady JJ, et al. 2018. Mapping the in vivo fitness landscape of lung adenocarcinoma tumor suppression in mice. *Nat. Genet.* 50:483–86
90. Roth A, Khattra J, Yap D, Wan A, Laks E, et al. 2014. PyClone: statistical inference of clonal population structure in cancer. *Nat. Methods* 11:396–98

91. Roth A, McPherson A, Laks E, Biele J, Yap D, et al. 2016. Clonal genotype and population structure inference from single-cell tumor sequencing. *Nat. Methods* 13:573–76
92. Salehi S, Steif A, Roth A, Aparicio S, Bouchard-Côté A, Shah SP. 2017. ddClone: joint statistical inference of clonal populations from single cell and bulk tumour sequencing data. *Genome Biol*. 18:44
93. Salk JJ, Schmitt MW, Loeb LA. 2018. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nat. Rev. Genet.* 19:269–85
94. Sánchez-Danés A, Hannezo E, Larsimont J-C, Liagre M, Youssef KK, et al. 2016. Defining the clonal dynamics leading to mouse skin tumour initiation. *Nature* 536:293–303
95. Sandberg A. 1980. The cytogenetics of chronic myelocytic leukemia (CML): chronic phase and blastic crisis. *Cancer Genet. Cytogenet.* 1:217–28
96. Schrider DR, Kern AD. 2017. Soft sweeps are the dominant mode of adaptation in the human genome. *Mol. Biol. Evol.* 34:1863–77
97. Schumacher TN, Schreiber RD. 2015. Neoantigens in cancer immunotherapy. *Science* 348:69–74
98. Schwartz R, Schäffer AA. 2017. The evolution of tumour phylogenetics: principles and practice. *Nat. Rev. Genet.* 18:213–29
99. Scott J, Marusyk A. 2017. Somatic clonal evolution: a selection-centric perspective. *Biochem. Biophys. Acta Rev. Cancer* 1867:139–50
100. Shen MM. 2013. Chromoplexy: a new category of complex rearrangements in the cancer genome. *Cancer Cell* 23:567–69
100a. Singer J, Kuipers J, Jahn K, Beerenwinkel N. 2018 Single-cell mutation identification via phylogenetic inference. *Nat Commun* **9,** 1–8
101. Sottoriva A, Kang H, Ma Z, Graham TA, Salomon MP, et al. 2015. A Big Bang model of human colorectal tumor growth. *Nat. Genet.* 47:209–16
102. Sottoriva A, Spiteri I, Piccirillo SGM, Touloumis A, Collins VP, et al. 2013. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *PNAS* 110:4009–14
102a. Spiteri I, Caravagna G, Cresswell GD, Vatsiou A, Nichol D, et al. 2018 Evolutionary dynamics of residual disease in human glioblastoma. *Annals of Oncology* **352,** 987–8

103. Stratton MR. 2011. Exploring the genomes of cancer cells: progress and promise. *Science* 331:1553–58

104. Sun R, Hu Z, Sottoriva A, Graham TA, Harpak A, et al. 2017. Between-region genetic divergence reflects the mode and tempo of tumor evolution. *Nat. Genet.* 49:1015–24

104a. Tarabichi M, Martincorena I, Gerstung M, Leroi AM, Markowetz F, et al. 2018 Neutral tumor evolution? *Nature Genetics* **481,** 306

105. Temko D, Tomlinson IPM, Severini S, Schuster-Böckler B, Graham TA. 2018. The effects of mutational processes and selection on driver mutations across cancer types. *Nat. Commun.* 9:1857

105a. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS et al 2018 The Immune Landscape of Cancer. *Immunity* **48,** 812–830.e14

106. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH. 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* 6:187–202

107. Tsao JL, Tavaré S, Salovaara R, Jass JR, Aaltonen LA, Shibata D. 1999. Colorectal adenoma and cancer divergence. Evidence of multilineage progression. *Am. J. Pathol.* 154:1815–24

108. Tsao JL, Yatabe Y, Salovaara R, Järvinen HJ, Mecklin JP, et al. 2000. Genetic reconstruction of individual colorectal tumor histories. *PNAS* 97:1236–41

108a. Turajlic S, Hu X, Litchfield K, Rowan A, Chambers T, et al. 2018 Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. *Cell* **173,** 1–27

109. Van den Eynden J, Larsson E. 2017. Mutational signatures are critical for proper estimation of purifying selection pressures in cancer somatic mutation data when using the dN/dS metric. *Front. Genet.* 8:415–19

110. Vermeulen L, Morrissey E, van der Heijden M, Nicholson AM, Sottoriva A, et al. 2013. Defining stem cell dynamics in models of intestinal tumor initiation. *Science* 342:995–98

111. Wang YK, Bashashati A, Anglesio MS, Cochrane DR, Grewal DS, et al. 2017. Genomic consequences of aberrant DNA repair mechanisms stratify ovarian cancer histotypes. *Nat. Genet.* 49:856–65

112. Weghorn D, Sunyaev S. 2017. Bayesian inference of negative and positive selection in human cancers. *Nat. Genet.* 49:1785–88

113. Werner B, Traulsen A, Sottoriva A, Dingli D. 2017. Detecting truly clonal alterations from multi-region profiling of tumours. *Sci. Rep.* 7:44991

114. Williams MJ, Werner B, Barnes CP, Graham TA, Sottoriva A. 2016. Identification of neutral tumor evolution across cancer types. *Nat. Genet.* 48:238–44

115. Williams MJ, Werner B, Heide T, Curtis C, Barnes CP, et al. 2018. Quantification of subclonal selection in cancer from bulk sequencing data. *Nat. Genet.* 50:895–903

116. Wu C-I, Wang H-Y, Ling S, Lu X. 2016. The ecology and evolution of cancer: the ultra-microevolutionary process. *Annu. Rev. Genet.* 50:347–69

117. Yates LR, Gerstung M, Knappskog S, Desmedt C, Gundem G, et al. 2015. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat. Med.* 21:751–59

118. Zahn H, Steif A, Laks E, Eirew P, VanInsberghe M, et al. 2017. Scalable whole-genome single-cell library preparation without preamplification. *Nat. Methods* 14:167–73

119. Zapata L, Pich O, Serrano L, Kondrashov FA, Ossowski S, Schaefer MH. 2018. Negative selection in tumor genome evolution acts on essential cellular functions and the immunopeptidome. *Genome Biol*. 19:67

120. Zhang J, Fujimoto J, Zhang J, Wedge DC, Song X, et al. 2014. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* 346:256–59

120a. Zhang AW, McPherson A, Milne K, Kroeger DR, Hamilton PT, et al. 2018 Interfaces of Malignant and Immunologic Clonal Dynamics in Ovarian Cancer. *Cell* **173,** 1–38

121. Zheng Q. 1999. Progress of a half century in the study of the Luria–Delbrück distribution. *Math. Biosci.* 162:1–32