Syst. Biol. 64(1):e1–e25, 2015

DOI:10.1093/sysbio/syu081
Advance Access publication October 7, 2014

Cancer Evolution: Mathematical Models and Computational Inference

NIKO BEERENWINKEL[1,2,*], ROLAND F. SCHWARZ[3], MORITZ GERSTUNG[4], AND FLORIAN MARKOWETZ[5]

[1]*Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland;* [2]*SIB Swiss Institute of Bioinformatics, 4058 Basel, Switzerland;* [3]*European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom;* [4]*Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom;* [5]*Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, CB20RE, United Kingdom*

**Correspondence to be sent to: Niko Beerenwinkel, Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 26, 4058 Basel, Switzerland; E-mail: niko.beerenwinkel@bsse.ethz.ch.*

Received 27 September 2013; reviews returned 8 January 2014; accepted 26 September 2014
Associate Editor: Olivier Gascuel

Abstract.—Cancer is a somatic evolutionary process characterized by the accumulation of mutations, which contribute to tumor growth, clinical progression, immune escape, and drug resistance development. Evolutionary theory can be used to analyze the dynamics of tumor cell populations and to make inference about the evolutionary history of a tumor from molecular data. We review recent approaches to modeling the evolution of cancer, including population dynamics models of tumor initiation and progression, phylogenetic methods to model the evolutionary relationship between tumor subclones, and probabilistic graphical models to describe dependencies among mutations. Evolutionary modeling helps to understand how tumors arise and will also play an increasingly important prognostic role in predicting disease progression and the outcome of medical interventions, such as targeted therapy. [Cancer; cancer progression; evolution; population genetics; probabilistic graphical models]

Cancer is a very heterogeneous disease. Genetic differences between people lead to differences in susceptibility (Pharoah et al. 2004), tumors develop in different organs and tissues of the body (Weinberg 2013), and cancers deriving from the same tissue can be stratified into disease subtypes based on differences in genomic measurements (Curtis et al. 2012). The genetic heterogeneity among cancer cells and the cellular heterogeneity of the tumor tissue underlie this phenotypic heterogeneity of the disease. The cancer cells in a tumor are not all identical, but form different clones, defined as sets of cancer cells that share a common genotype. Genetic and epigenetic heterogeneity poses a problem for the diagnosis and therapy of cancer. For example, it can lead to incorrect treatment decisions if a biopsy sample is not representative of other parts of the tumor (Merlo et al. 2006). In addition, tumor cells are part of the so-called tumor microenvironment, a heterogeneous tissue containing not only cancer cells, but also stromal and immune cells (Albini and Sporn 2007).

Cancer is an evolutionary process.—Despite the heterogeneity of tumors, some functional organizing principles exist, often summarized as the hallmarks of cancer, which include evasion of apoptosis and immune response, unstable DNA, and the ability to metastasize (Hanahan and Weinberg 2000; 2011). A particular successful guide for understanding and modeling cancer progression has been evolutionary theory, which has a long tradition in cancer research. Already 40 years ago, seminal work established an evolutionary view of cancer (Nowell 1976; Dexter et al. 1978; Fidler 1978), in which carcinogenesis is regarded as an evolutionary process driven by stepwise somatic mutations and clonal expansions (Fig. 2a). Within each tumor, clones can evolve that harbor selectively advantageous mutations (called drivers), neutral mutations (called passengers), and deleterious mutations. The frequencies of passenger mutations can rise in a population by chance, often because they are linked to a driver mutation and hitchhike on the expanding clone. Some mutations increase the rate of other genetic changes and microenvironmental changes can also alter the fitness effects of mutations (Greaves and Maley 2012; Barcellos-Hoff et al. 2013). Moreover, there is evidence of competition, predation, parasitism, and mutualism between co-evolving clones in and around a tumor (Merlo et al. 2006). Most of these concepts were known for decades when advanced genomic technologies renewed interest in cancer evolution in the early 2000's. In the last few years, next-generation sequencing (NGS) of cancer genomes (Stratton et al. 2009) brought additional vigor to the field and has made tumor evolution a central topic in cancer research (Greaves and Maley 2012).

Features of cancer evolution.—Evolutionary theory is well developed and it provides an extensive toolkit for any evolutionary process. Modeling the somatic evolution of cancer also benefits greatly from this body of work. However, cancer evolution has several specific features, including the following four. (i) Cancer genomes harbor complex alterations. Many tumors display genetic instability, which results in abnormal numbers of chromosomes, that is aneuploidy (Weinberg 2013), elevated mutation rates, and altered distributions of mutational patterns (Garraway and Lander 2013). In addition, some genomic alterations can be extremely complex and rearrange entire chromosomes. These alterations enable large mutational jumps and they render comparisons between cancer

genomes challenging. (ii) There are many selectively advantageous mutations. Recent cancer genome sequencing studies have identified hundreds of driver mutations (Vogelstein et al. 2013). Many of them disrupt cellular signaling pathways that are essential for multicellular organisms and may have occurred early in the evolution of multicellularity. These pathways tightly control and orchestrate cellular behavior in a tissue. In general, there are many ways to perturb a signaling pathway and hence, many possible mutations exist that increase the somatic fitness of cancer cells. In this sense, cancer evolution can be regarded as the evolution of defection (Nowak 2006a). (iii) Tumor cells are organized in specific structures. Population structure can result from interactions with the environment and from the spatial organization or the differentiation hierarchy of the tissue of origin. These structures affect the fate of mutations that occur in individual cells of the population, and hence the dynamics of tumor progression (Nowak et al. 2003). (iv) Tumorigenesis is a reproducible evolutionary process. Each tumor of the same type, or subtype, can be regarded as an independent realization of the same evolutionary process, even if some confounding factors will remain, for example, genetic background or microenvironment. Repeated observations provide an opportunity to enhance statistical inference about the evolution of tumors, which may eventually make cancer evolution more predictable (Orr 2005).

The evolutionary theory of cancer has survived 40 years of empirical observation and testing, and its components are well understood. However, at the same time, central questions remain controversial, for example, the argument of gradualism versus punctuated equilibrium, which is a long-standing debate also in species evolution (Gould and Eldredge 1993). The cancer community currently actively discusses whether tumors evolve gradually through a sequence of genetic alterations and clonal expansions that accumulate genomic lesions, or through a few punctuated changes driven by complex rearrangements (Baca et al. 2013; Shen 2013) or individual catastrophic events that shatter entire chromosomes (Stephens et al. 2011). In addition, large territories of cancer evolution still remain unexplored. For example, the evolutionary dynamics of cancer are still incompletely understood, because many parameters of this process have not yet been or can generally not be assessed experimentally, including fitness effects of mutations, generation times, population structure, the frequency of selective sweeps, and the selective effects of therapies (Merlo et al. 2006).

Controlling cancer evolution.—Research into cancer evolution not only addresses basic biological questions of tumor development and progression, but is also of clinical significance. For example, drug resistance is a major clinical problem resulting in therapeutic failure and uncontrolled disease progression. Even when patients initially respond well to cancer treatment, they often die because their tumors develop resistance to all available therapeutic avenues (Garraway and Jänne 2012). In his (1976) landmark article, Nowell wrote that "more research should be directed towards understanding and controlling the evolutionary process in tumors before it reaches the late stage seen in clinical cancer." This statement is as true now as it was 40 years ago. A recent literature survey found that even though relapse and therapeutic resistance are inherently evolutionary processes, evolutionary concepts have not yet permeated cancer research (Aktipis et al. 2011), emphasizing the great need for evolutionary approaches to cancer biology and treatment.

Tumor heterogeneity and evolution as well as its clinical implications have been reviewed recently and extensively (de Bruin et al. 2013; Bedard et al. 2013; Klein 2013; Junttila and de Sauvage 2013; Burrell et al. 2013; Meacham and Morrison 2013; Almendro et al. 2013; Aparicio and Caldas 2013; Greaves and Maley 2012; Marusyk et al. 2012; Caldas 2012; Podlaha et al. 2012; Lambert et al. 2011; Michor and Polyak 2010; Salk et al. 2010; Bowtell 2010). In contrast to these biological and medical reviews, we focus here on mathematical and statistical methodology for modeling tumorigenesis. After reviewing the types of data available for modeling the evolution of cancer, we discuss several computational models, including population dynamics models of cancer cells, phylogenetic methods of tumor subclones, and probabilistic graphical models of tumor progression. Cancer-specific terminology used in this article is explained in Table 1.

Molecular Cancer Data

The amount and the breadth of tumor molecular profiling has increased tremendously in recent years, mainly due to the advent of cost-effective high-throughput seqeuncing technologies. Genomic data on cancer stems from a variety of different sources, including (i) cell lines cultivated in laboratories, (ii) xenografts derived from patient tumors and engrafted into model organisms like mice, and (iii) clinical patient samples from biopsies. Experimentally, cell lines have several advantages over clinical samples: Initially they are genetically homogeneous, they can be kept under constant environmental conditions, and they show no contamination with normal cells. However, freed from its natural cellular context, this evolutionary process can have little in common with disease progression in patients, where, for example, the tissue microenvironment affects tumor evolution (Bissell and Hines 2011). Recent work on sequencing HeLa cells has shown that cancer cell lines can evolve to be very divergent and might be poor model systems for the disease they were derived from (Landry et al. 2013; Adey et al. 2013). Xenografts, although more closely related to a real tumor, are affected by different immune response and microenvironment in the host, but they can model tumor progression in a living organism over time with little sampling restrictions (Hidalgo et al. 2014).

TABLE 1. A table of common terms and acronyms used in cancer genomics

Term	Description
aCGH	Array CGH; microarray-based high-resolution CGH method
APC	Adenomatous polyposis coli; a human tumor suppressor gene (TSG)
Allele frequency	Fraction of cells (in NGS data, of reads) carrying a mutation
Aneuploidy	Abnormal number of (parts of) chromosomes (Fig. 1)
BAF	B allele frequency; ratio of the number of B alleles over the total ($A+B$) DNA content
BRAF	Gene coding for the B-Raf protein, a signal transduction kinase that can be involved in cancer if mutated
Carcinogenesis	The process of cancer development
Cellularity	The proportion of tumor cells in a sample
CGH	Comparative genome hybridization; cytogenetic method for analyzing copy number variations (CNVs)
Chromothripsis	The shattering of the genome in one catastrophic event (Fig. 1)
Chromoplexy	Chained rearrangement across several chromosomes (Fig. 1)
Clone	A set of tumor cells descending from the same ancestor and hence sharing its mutations
Clonal frequency	Percentage of tumor cells carrying an allele
CNV	Germline (normal) copy number variation in normal cells of the tissue and in tumor cells
CNA	Somatic copy number aberration (or alteration) in the cancer genome of tumor cells (Fig. 1)
Driver mutation	Mutation that confers a selective advantage
EGFR	Epidermal growth factor receptor; a cell-surface receptor whose altered expression is involved in cancer
Kataegis	Local hypermutation; many SNVs clustered on a short genomic segment (Fig. 1)
LOH	Loss of heterozygosity; Loss of one parental allele with or without a copy number change.
logR	Logarithmic intensity ratio of tumor and control DNA in an array CGH experiment.
NGS	Next-generation sequencing; High-throughput sequencing technologies based on massively parallel DNA amplification and sequencing
Oncogene	Gene that confers a selective advantage if hit by a gain-of-function mutation
Oncogenetic model	A model of the dependencies of events (CNAs, SNVs) in cancer development (Fig. 6)
Passenger mutation	Selectively neutral mutation
Phasing	The assignment of genomic alterations to specific haplotypes (Fig. 5)
Phylogenetic model	A model of the evolutionary relationship between different tumor samples from the same patient or between clones from the same tumor (Fig. 6)
Segmentation	The process of calling integer copy numbers from noisy logR values.
Somatic evolution	Evolution within an organism
SNP	Single nucleotide polymorphism; single base variant existing in the human population (found in normal tissue and tumor)
SNV	Single nucleotide variant; single base change that occurred in the tumor (Fig. 1)
TSG	Tumor suppressor gene; gene that confers a selective advantage if hit by a loss-of-function mutation
Tumorigenesis	The process of tumor development
Vascularization	The process of establishing blood vessels

Clinical samples are more direct reflections of the disease but incur logistic and scientific problems (Basik et al. 2013). For example, collecting multiple samples from patients can introduce biases, because patients with either very good or very poor response often contribute only few samples. Indeed, if the patient is cured after surgery or chemotherapy, no follow-up samples will be available, whereas in patients progressing very quickly, surgery is sometimes not attempted or the patient may succumb to the disease in short time. In addition, ethical and technical restrictions hinder broad collection of samples. A biopsy might be highly interesting from a scientific perspective but medically not necessary. Finally, clinical samples often suffer heavily from normal cell contamination (low cellularity) and infiltration of immune cells such as lymphocytes (Yuan et al. 2012).

The Complexity of Cancer Genomes and the Search for Driver Genes

The normal human point mutation rate is 10^{-10} per base pair per cell division (Kunkel and Bebenek 2000), and this rate is elevated in many cancers, a phenomenon termed mutator phenotype (Loeb 2001; 2011). Hence, almost every cell division introduces a mutation, which inevitably leads to genetic diversity in every proliferating cell population. Cancer genomes are characterized by complex aberrations and rearrangements, ranging from small-scale point mutations (single-nucleotide variants; SNVs), often numbering in thousands per cancer cell, to large-scale chromosomal rearrangements resulting in complex patterns of genomic architecture and copy number aberrations (CNAs) (Greenman et al. 2012; Garraway and Lander 2013). Figure 1 gives an overview of genomic changes widespread in cancer. In addition, cancer genomes show epigenetic alterations, such as changes to DNA methylation (Hong et al. 2010; Sottoriva et al. 2013b). Some aberrations, like amplifications, deletions, and point mutations, are also common to many other evolutionary processes outside of cancer. Others, such as chromosomal deletions where one copy of a chromosome region is lost, are central to explanations of cancer evolution like the two-hit hypothesis (Nordling 1953).

In the following, we highlight complex patterns of aberrations that have recently been discovered in cancer genomes and whose evolutionary role is currently being discussed. *Kataegis* refers to a pattern of localized hypermutation, that is regional clustering of substitution

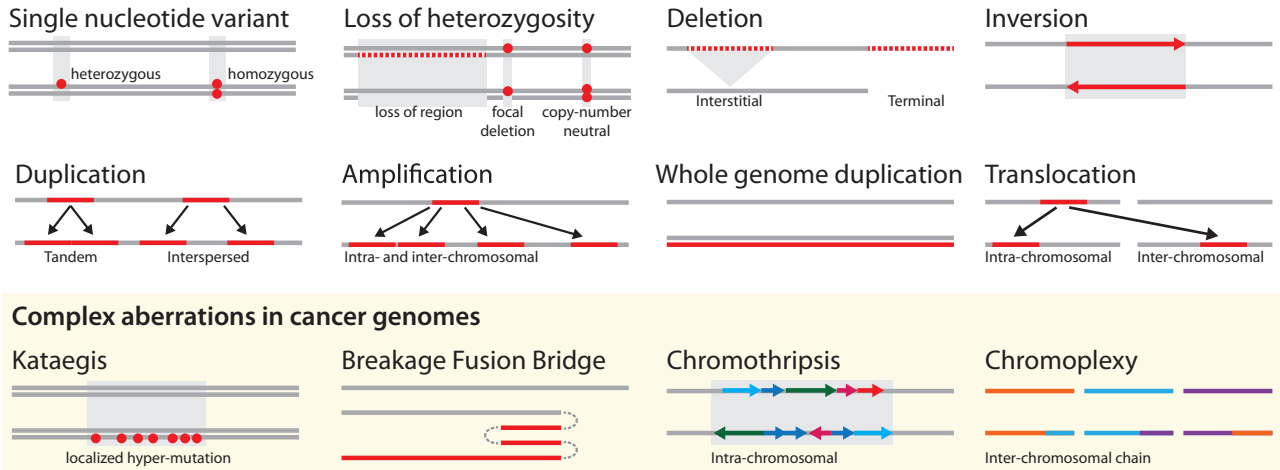

FIGURE 1. Common aberrations in cancer genomes. These events lead to the abnormal chromosome numbers (aneuploidy) and chromosome structures of a cancer genome. Lines indicate the genome with germline genome on top and cancer genome with somatic aberrations below. Double lines are used when differentiating heterozygous and homozygous changes is useful. Dots represent single nucleotide changes, whereas lines and arrows represent structural changes.

mutations, observed in breast cancer genomes (Nik-Zainal et al. 2012a). *Breakage-fusion-bridge cycles* lead to palindromic genomic patterns, which can be an early step in DNA amplification (Guenthoer et al. 2012). *Chromothripsis* (chromosome shattering) refers to a single catastrophic event in which tens to hundreds of genomic rearrangements occur at the same time (Stephens et al. 2011). Although its exact cause is unclear, it is thought to be provoked by radiation exposure at a critical time point during cell cycle when chromosomes are condensed for mitosis. Cells that survive the catastrophe can have a selective advantage due to increased tumor cell growth, and their genomes often exhibit CNA patterns oscillating between one and two copies in the chromothriptic region. *Chromoplexy* is a process similar to chromothripsis in that it involves multiple genomic rearrangement events (Baca et al. 2013). The events often occur in a chain-like fashion connecting spatially distant areas of the genome that can affect multiple drivers from the same pathway at the same time despite their location on different chromosomes. Both chromothripsis and chromoplexy show random breakage and fusion of genomic segments, but several features set them apart: Chromothripsis displays hundreds of breakpoints clustered within a single chromosome, whereas rearrangements in chromoplexy are unclustered, usually number in the tens, and include multiple chromosomes (Shen 2013). Chromothripsis appears to be a single catastrophic event early in tumor progression, whereas chromoplexy can occur multiple times during cancer evolution and has been detected at the clonal and subclonal level (Baca et al. 2013).

The complexity of cancer genomes and the presence of mutator phenotypes make it challenging to separate driver from passenger mutations. To identify genes under positive somatic selection, one can detect an excess of nonsynonymous somatic mutations, that is, a high dN/dS ratio, in cancer genome sequences. The same genes are often under purifying selection in intergenerational terms leading to a depletion of nonsynonymous polymorphisms in the human population. Based on the idea of a high somatic dN/dS, (Greenman et al. (2006)) formulated a hypothesis test in a Poisson regression framework for discovering cancer driver genes, which was applied to identify 120 driver genes among 518 protein kinases in a cohort of 210 cancer samples (Greenman et al. 2007). More recent methods incorporate additional covariates, such as replication timing and gene expression data to refine estimates of the local mutation rate (Lawrence et al. 2013). Gonzalez-Perez et al. (2013) also accounted for the functional impact of mutations, as predicted, for example, by SIFT (Kumar et al. 2009) and PolyPhen2 (Adzhubei et al. 2010). In addition, they used evolutionary sequence conservation and clustering of mutations within each gene to identify driver genes. Recently, Lawrence et al. (2014) analyzed 4,742 cancers to present a list of 219 recurrently mutated cancer genes. As the authors suggest, this list may grow further in the future, as many driver genes are only infrequently mutated.

Intratumor Heterogeneity and the Detection of Subclonal Alterations

It has long been known that tumors are composed of multiple cellular subpopulations with different genotypes (Nowell 1976), and modern genomic techniques have refined this observation (Burrell et al. 2013). Analyzing single cells is the most informative approach to assess the heterogeneity within a tumor. Cell sorting can be used to detect cellular phenotypic heterogeneity in blood cancers (Amir et al. 2013) and immunofluorescence *in situ* hybridization to highlight the genetic diversity of individual loci (Almendro et al.

2014). Progress in single-cell genomics (Shapiro et al. 2013) allows sequencing genomes of individual cells taken from a tumor (Navin et al. 2011; Hou et al. 2012; Xu et al. 2012; Potter et al. 2013). However, in most studies, the samples used are a mixture of cancer cells and stromal cells. In the following, we discuss how to analyze clonal architecture from genomic profiles of mixed samples.

Genomic data is typically obtained by NGS or by DNA microarrays. Sequencing has the advantage of being able to detect somatic SNVs as well as local tumor copy numbers by read depth analysis. By contrast, SNP arrays generally do not allow *de novo* discovery of SNVs, but the SNP probes allow for allele-specific copy number inference by considering bi-allelic frequency, that is, the ratio of the frequencies of the two parental alleles. The main objective when calling SNVs is to distinguish sequencing errors from true variants and separating germline from somatic changes. Algorithms solving this problem either employ frequentist statistical methods for modeling the distribution of variants per site in the genome, such as deepSNV (Gerstung et al. 2012), Varscan (Koboldt et al. 2012), and LoFreq (Wilm et al. 2012), or employ a Bayesian classifier framework, such as MuTect (Cibulskis et al. 2013).

Classical methods for CNA calling, called segmentation, are often ill-suited for cancer samples, because they do not take differences in cellularity nor changes in tumor ploidy into account. For those methods that do, segmentation broadly follows the same principles in most implementations: Normalized array intensities or normalized read counts are defined as the log ratio, *logR*, between the local DNA copy numbers in a mixture of cancer and normal cells and the average copy numbers in the mixture,

$$logR = \frac{\alpha(n_i^A + n_i^B) + 2(1-\alpha)}{\alpha P + 2(1-\alpha)} \quad (1)$$

Here, α is the cellularity, P is the average tumor ploidy to which array intensities and read counts are normalized, and n_i^A and n_i^B are the integer copy numbers of the two parental alleles at locus i. The B allele frequency (*BAF*)

$$BAF = \frac{n_i^B}{n_i^A + n_i^B} \quad (2)$$

is the ratio between the *B* allele and the total allele count, which can be obtained from both sequencing and SNP arrays. It provides additional information about the copy number state at a certain genomic locus, as often only one allele is amplified or deleted.

Traditionally, cellularity was assessed by visual analysis of tumor cells, either manually by a pathologist or via image analysis (Yuan et al. 2012). Laser capture microdissection (Emmert-Buck et al. 1996) can be used to select more homogeneous areas from mixed tissue sections (Navin et al. 2010), but the procedure is time-consuming and generally only used in small studies. Most current methods for mixed samples estimate both cellularity and average tumor ploidy during segmentation, including PICNIC (Greenman et al. 2010), ABSOLUTE (Carter et al. 2012), and ASCAT (Van Loo et al. 2010). ASCAT calls copy numbers specifically for each allele, which results in two integer vectors of copy numbers, one for each parental allele. Because typically no linkage information between adjacent loci is available, at each site the larger of the two copy numbers, by convention, defines the major and the smaller the minor allele.

Unlike SNVs, where allelic frequencies can be directly derived from read counts, CNA calling is difficult in populations with subclonal structure, because the mixture of subclones leads to noninteger copy numbers which introduce deviations from the expected log-ratio (Equation 1) in array data or from the expected read counts in sequencing data.

To address this issue, Oesper et al. (2013) proposed THETA, an algorithm that infers the most likely collection of genomes and their proportions from NGS data, in the case where CNAs distinguish subpopulations. Nik-Zainal et al. (2012b) introduced the Battenberg algorithm (named after a checkered genomic pattern resembling Battenberg cake), which first assigns all SNPs to known haplotypes, a task known as phasing. Then it tests within haplotype blocks for small deviations of the *BAF* values (Equation 2) from those expected at normal diploid loci to assess whether there is sufficient evidence for a subclonal copy number change.

Population Dynamics of Cancer Cells

Population genetics provides a well-developed mathematical theory of evolution (Durrett 2002; Ewens 2004), and many of these models and techniques have been applied to cancer. The most basic models assume no interactions among tumor cells and ignore any structure of the population. These strong simplifications result in mathematically tractable models that allow for calculating quantities of interest such as intratumor genetic diversity, the probability of fixation of a new mutant, or the age of the tumor. For this reason, models of well-mixed populations with constant selection are widely and successfully applied to the evolution of cancer, and they provide the starting point for more complex models. The population size, N, is an important parameter affecting not only the evolutionary dynamics of the population, but also the appropriate choice of the mathematical model. In small populations, allele sampling effects are more pronounced requiring stochastic modeling, whereas large populations often behave almost deterministically allowing for models based on differential equations.

Historically, multistage theory was the first approach to model tumor progression based only on cancer incidence data. Later, with the availability of protein and DNA sequence data, classical population genetics models of asexual populations have been applied to tumorigenesis, and several new models have been

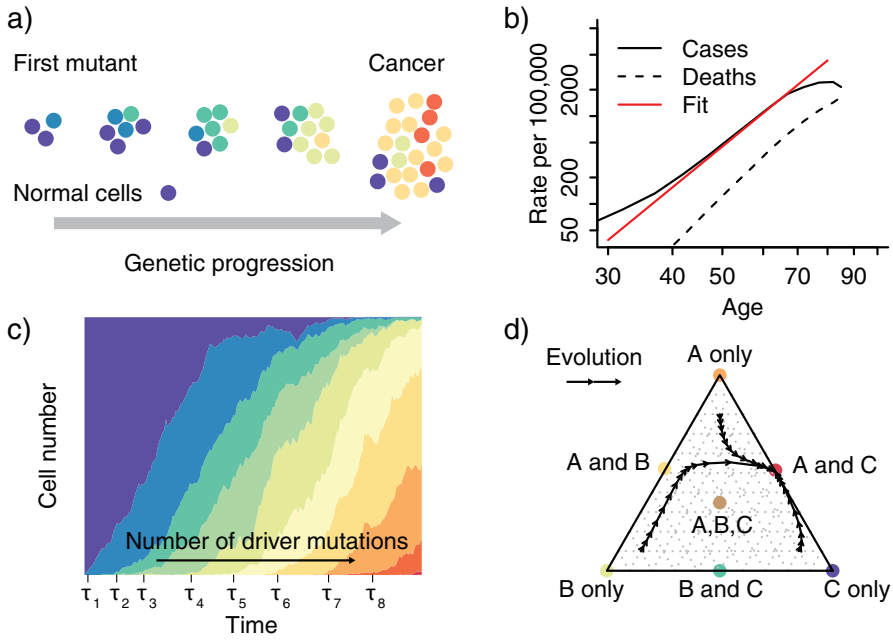

FIGURE 2. Modeling the population dynamics of cancer cells. a) Schematic illustration of the genetic progression from initially healthy cells (normal cells) to an invasive cancer by accumulating driver mutations. b) Age-incidence curves rise sharply above the age of 50 and are informative about the dynamics of tumor progression. The straight line shows a fit with power 4.8. The log-log-linear dependency of incidence on age is used in multistage theory to estimate the number of rate-limiting steps in cancer progression from incidence data. c) Population genetics models such as the Wright–Fisher model can be used to model the accumulation of driver mutations through multiple clonal expansions and to derive the average waiting times τ_k for a given number of alterations k. d) Dynamics of a three-strategy game corresponding to cell types A, B, and C. While simple additive fitness models always lead to the survival of the fittest, evolutionary game theory accounts for cellular interactions and allows for more complex dynamics, such as stable coexistence of cell types. Indicated here is a stable equilibrium with strategy A and C, but not B, which is reached from all three starting conditions via the indicated evolutionary paths.

developed to address specific aspects of the somatic evolution of cancer. In this section, we first review classical multistage theory. Then, models are discussed that are either stochastic or deterministic and assume either a well-mixed or a structured population, followed by hybrid models. Lastly, we address modeling of cellular interactions using evolutionary games.

Multistage Theory

Multistage theory models the probability of developing cancer as a function of age. The kinetics of tumor initiation and progression are usually unobserved, but different models can be tested and parameters can be inferred by fitting epidemiological age-incidence curves. In 1953, Nordling observed that cancer incidence rises sharply with higher ages and that it can be approximated by a monomial in age of degree six (Fig. 2b). Based on this observation, he postulated the existence of six independent rate-limiting steps during carcinogenesis. The underlying rationale of his inference was the following: If a single transforming step towards cancer occurs stochastically at a constant and small rate u, then the probability to observe this transition after time t is approximately ut. The cumulative probability of k successive steps is therefore proportional to t^k. Although the transforming steps occur stochastically in each patient, they will manifest on the population level, which allows the number of rate-limiting steps to be related to the observed age-incidence curves. Armitage and Doll (1954) repeated Nordling's analysis one year later and estimated that the exponent of the age-incidence curves ranged between 4.97 and 6.48 across a variety of cancer types. Using a similar reasoning, Knudson (1971) attributed the differences in the incidence of sporadic and hereditary retinoblastoma, a childhood cancer of the eye, to the existence of two independent rate-limiting mutations.

These early findings motivated the development of the so-called multistage theory of carcinogenesis (Frank 2007), which predicts that cancer incidence, I, defined as the first derivative of the cumulative number of cases, depends on age t as

$$I(t) \propto u_1 \cdots u_k t^{k-1} \tag{3}$$

where k is the number of stages and u_i the transition rate from stage $i-1$ to i. Multistage theory has been extended to explicitly model the different kinetics of tumor initiation, namely the acquisition of the first transforming mutation in a renewing tissue, and the subsequent progression phases in the expanding tumor (Armitage and Doll 1957; Luebeck and Moolgavkar 2002; Calabrese et al. 2004; Meza et al. 2008; Luebeck et al. 2013). These stochastic models usually require fewer rate-limiting steps, as the rise in incidence is partly explained

by exponential growth. Recently, multistage models have been used to optimize cancer screening strategies on the population level (Jeon et al. 2008; Dewanji et al. 2011).

The existence of up to six rate-limiting steps in cancer development has been widely accepted and appears to resemble the prevalence of driver mutations found in comprehensive cancer sequencing studies, where estimates range between two and eight drivers per tumor (Vogelstein et al. 2013).

Stochastic Models of Well-mixed Populations

The assumption of a well-mixed cancer cell population, although questionable especially for solid tumors, is frequently made, and the mathematical approaches are best developed for this case. The most basic models assume constant population size.

Moran process and Wright–Fisher process.—The Moran process and the Wright–Fisher process are the standard models for finite populations of constant size N. We assume two different cell types, referred to as normal and mutant (e.g., healthy and cancer cells, or two tumor cell types with different mutational patterns), with fitness f_1 and f_2, respectively. The Moran process defines a discrete-time Markov chain that keeps track of the number of mutants, $X(t)$, in the population in generation t (Moran 1958). In each step of the process, a cell is randomly selected with probability proportional to its fitness to divide and produce one offspring, and another cell is selected for death uniformly at random. Each birth–death event leaves the total population size, N, unchanged and can alter the number of mutants, $X(t)$, by at most one. The mutant subpopulation increases if a mutant is selected for reproduction and a normal cell for death. The probabilities of these two events are proportional to $f_2X(t)$ and $N-X(t)$, respectively. The Moran process is the birth–death process defined by these transition probabilities. It has two absorbing states, namely extinction ($X=0$) and fixation ($X=N$) of the mutant.

In case of neutral evolution ($f_1=f_2$), the Moran process describes neutral drift of the population, that is, the fluctuation in allele frequencies due only to random offspring sampling Kimura 1983. In this setting, the probability of X mutants to reach fixation is $x=X/N$. If mutants have a selective advantage ($f_1<f_2$), then their fixation probability is

$$\rho_X=\frac{1-(f_2/f_1)^{-X}}{1-(f_2/f_1)^{-N}} \tag{4}$$

and the mean time to fixation is $(-2N)^2[x\log x+(1-x)\log(1-x)]$ (Nowak 2006a). If u denotes the mutation rate, then the total mutation rate in a tumor of N cells is Nu and the rate of evolution, that is, the rate at which the entire population shifts from normal to mutant cells is $Nu\rho_1$. In the neutral case, the rate of evolution is equal to the mutation rate, because then $\rho_1=1/N$ (Kimura 1983).

The Wright–Fisher process is very similar to the Moran process, but in each generation, the entire population is drawn at random from the previous generation (Fisher 1930; Wright 1931). For two cell types and no selection, it is defined by the binomial sampling $[X(t+1)\,|\,X(t)]\sim\text{Binom}(N,\,x(t))$. Thus, cells are assumed to be synchronized and one generation in the Wright–Fisher process corresponds to N generations in the Moran process. With this rescaling, both processes have the same fixation probability, (essentially) the same fixation time, and the same diffusion limit, that is, they agree in the limit of large population size $N\to\infty$ (Ewens 2004). For computer simulations, the Wright–Fisher process generally allows drawing samples more efficiently.

When applied to cancer evolution, the Wright–Fisher process has been generalized to multiple cell types, representing genetically different tumor subclones, using multinomial sampling, and it has been extended to account for additional evolutionary forces, including mutation, selection, and genetic instability (Beerenwinkel et al. 2007a; Datta et al. 2013).

The coalescent.—The coalescent is a stochastic process based on the Wright–Fisher process. It establishes a connection to observed sequence data by making inference about population parameters from a contemporary finite sample. In the coalescent, genealogies are generated by tracing coalescent events between lineages backwards in time (Kingman 1982). In the diffusion limit, $N\to\infty$, two lineages coalesce at rate $\binom{j}{2}$, when there are j individuals left (Rosenberg and Nordborg 2002). The coalescent accounts for mutations by imposing them on the random genealogy. It has originally been developed for neutral evolution, but has later been extended to account for selection (Neuhauser and Krone 1997).

The coalescent allows for inferring characteristic evolutionary parameters such as population size, mutation rate, or the time to the most recent common ancestor from sampled sequences. Nicolas et al. (2007) apply the coalescent to DNA methylation patterns of differentiated cells to estimate the number of stem cells in a human colonic crypt, that is, the number of cells at risk of initiating colon cancer. Statistical inference in the coalescent can be computationally demanding and several approximations to ML or Bayesian estimation have been proposed, most notably Approximate Bayesian Computation, which avoids evaluation of the likelihood function (Marjoram et al. 2003; Haccou et al. 2005).

Branching processes.—Branching processes are well-studied stochastic models for populations of finite, but fluctuating, size (Athreya and Ney 1972; Kimmel and Axelrod 2002; Haccou et al. 2005). The basic assumption is that individuals produce a random number of offspring, each giving rise to an independent lineage that behaves identically to its parent, that is, after a certain lifetime, it will again produce a random number

of offspring drawn from the same distribution. Most branching processes are inherently unstable: Eventually, the population will either die out or grow indefinitely. On the other hand, if a branching process with constant lifetime (i.e., a Galton–Watson process) and Poisson offspring distribution is conditioned on constant population size, one obtains the Wright–Fisher process (Haccou et al. 2005).

The probability generating function of the offspring distribution is the main mathematical tool for analyzing branching processes. It allows for computing several quantities of interest, including the extinction probability and time, and the probability of a mutant to arise and to establish its lineage in the population. For example, using a branching process with three different cell types, Danesh et al. (2012) have modeled ovarian cancer growth and progression, with the goal of identifying a window of opportunity for screening, that is, a time period during which tumor diagnosis is feasible (tumor large enough), but treatment still possible (tumor not progressed too far). Different tumor subclones can be modeled by multitype branching processes, where fitness advantages translate into altered offspring distributions. Because fitness parameters are generally unknown for cancer cells, Durrett et al. (2010) modeled them as random variables. They studied the effect of bounded versus unbounded fitness distributions on genetic tumor diversity and found that it depends only on the maximum attainable fitness advance, but not on the specific form of the fitness distribution (Durrett et al. 2011).

The development of resistance to targeted cancer treatment has been modeled by density-dependent branching processes. Here, tumor cell growth (and hence offspring distribution) is limited by tumor size due to geometric and metabolic constraints (Bozic et al. 2012). This modification removes the instability of the branching process and introduces a steady state in which the population size fluctuates around a constant value. The probability of tumor eradication under therapy can be computed in this model by considering the generation of resistance mutations during initial expansion, steady state, and treatment.

Diffusion approximation.—Stochastic evolutionary models can be approximated by differential equation models to obtain simpler models that are more tractable and easier to interpret. The diffusion approximation is based on the assumption that the population size, N, is large and that the change per generation is small. It is given by the master equation, also called Kolmogorov forward, or Fokker–Planck equation (Fisher 1922; Kolmogorov 1931; Wright 1945; Feller 1951), for the probability density $\psi(x,t)$ of the relative allele frequency x at time t in an evolutionary Markov process $X(t)$, as

$$\frac{\partial \psi(x,t)}{\partial t} = -\frac{\partial}{\partial x} M(x)\psi(x,t) + \frac{1}{2}\frac{\partial^2}{\partial x^2} V(x)\psi(x,t) \qquad (5)$$

The second-order differential operator depends only on the mean $M(x)$ and the variance $V(x)$ of the Markov process $X(t)$. This framework allows for analyzing or constructing evolutionary models with specific directional (M) and undirectional (V) forces. For example, at equilibrium of the diffusion limit ($\partial\psi/\partial t = 0$), the Wright–Fisher process becomes

$$\psi(x) \propto x^{\theta-1}(1-x)^{\theta-1}e^{\sigma x} \qquad (6)$$

where $\theta = 2Nu$ and $\sigma = 2Ns$ are the scaled mutation and selection parameters, respectively, and $s = f_2 - f_1$ is the selective advantage of the fitter allele. This distribution reveals the strong impact of selection on large populations and of random genetic drift on small populations.

Tomasetti et al. (2013) used the diffusion approximation of the Moran process to calculate the expected number of passenger mutations that accumulate in the precancer phase to be BuT, where B is the total number of DNA bases sequenced, and T the number of times the tissue has self-renewed before tumor initiation. Using DNA sequencing data, they found that at least half of all somatic mutations in tumors of self-renewing tissues occur before the onset of neoplasia.

Diffusion theory is also useful for computing fixation probabilities and mean fixation times by considering the conditional density $\psi(x,t|x(0))$ (Ewens 2004). For example, in the Wright–Fisher process, the fixation probability of a new mutant in a haploid population is approximately $2s$. The average fixation time is approximately $2Ns$ in the limit of weak selection, $Ns \ll 1$, and $2\log(N-1)/s$ for strong selection (Otto and Whitlock 2001).

Tumor initiation and progression models.—Specific evolutionary models have been proposed to describe the dynamics of tumor initiation and progression. The two basic genetic events driving carcinogenesis are gain-of-function mutations in oncogenes and loss-of-function mutations in tumor suppressor genes (TSGs). Oncogenes are activated by specific point mutations, gene amplification, or chromosomal fusion. In a well-mixed cell population of size N with constant mutation rate u, to activate the oncogene the fixation probability by time t is

$$P(t) = 1 - \exp(-Nu\rho_1 t) \qquad (7)$$

where $Nu\rho_1$ is the rate of evolution (Michor et al. 2004). The equation shows that the accumulation of oncogene-activating mutations is faster in large than in small compartments. Thus, the organization of self-renewing tissues into many small compartments, such as, for example, the stem cell pools in colonic crypts, from which the tissue is derived, protects against cancer initiation.

The dynamics of TSG inactivation are more complex, because here two alleles need to be hit, either by two point mutations, or by one point mutation and loss of heterozygosity (LOH). For simplicity, we assume that the

first hit does not confer any selective advantage. Then, in small populations and for short time spans t, the TSG fixation probability $P(t)$ is quadratic in t, indicating two rate-limiting steps as in Knudson's two-hit theory. In intermediate populations, however, the second mutation may arise before the first one has reached fixation, a phenomenon termed stochastic tunneling. As a result, in intermediate populations, the TSG fixation probability is linear in t. Finally, there is no rate-limiting step in large populations (Nowak 2006a; Komarova 2007). Genetic instability can result in elevated mutation rates, which overall may or may not accelerate TSG inactivation, depending on the rate at which genetic instability is acquired and on the possible fitness costs it incurs.

The initial tumor cell lives in a hostile environment and generally has several possibilities to accumulate driver mutations to survive and expand. Using multitype branching processes, the evolutionary escape dynamics have been derived for arbitrary fitness landscapes defined on any genotype network, that is, any set of genotypes connected by mutation (Iwasa et al. 2003; Iwasa et al. 2004). Assuming that the initial tumor cell is too unfit to survive in the long run, the risk of escape is the probability of the population developing, before extinction, additional mutations that allow for escaping the selective pressure. For a given genotype network, the risk of escape is proportional to the per-site mutation rates and to the polynomial

$$\sum_{i_0 \to i_1 \to \cdots \to i_k} f_{i_1} \cdots f_{i_{k-1}} \tag{8}$$

in the fitness values (Beerenwinkel et al. 2006). Here, the sum runs over all possible mutational pathways in the network from the initial and unfit genotype i_0 to the escape type i_k, where a mutational pathway is a sequence of viable mutants. The risk polynomial (Equation 8) captures the impact of the topology of genotype space on evolutionary escape. The larger the number of alternative escape pathways, the higher is the risk of escape. Studies on protein evolution have shown that only few mutational paths can lead to fitter proteins (Weinreich et al. 2006). Hence, understanding these mutational constraints for the evolutionary escape of cancer cells may make tumorigenesis more predictable.

During tumor progression, selectively advantageous mutations give rise to clonal expansions, which drive tumor growth. The series of selective sweeps is often described as a sequence of traveling mutant waves (Fig. 2c). To understand this process, mathematical models have been devised that address questions about the speed of evolution, the waiting time distribution, and the clone size distribution (Park et al. 2010). They are typically based on the Moran or the Wright–Fisher process, and different approximations have been proposed to arrive at the quantities of interest.

For example, assuming Wright–Fisher dynamics, the average waiting time for the first cell with k mutations to appear in a population of size N has been approximated as

$$\tau_k = \frac{k}{2s \log N} \left(\log \frac{s}{ud} \right)^2 \tag{9}$$

where d is the number of driver genes, each of which confers the selective advantage s (Beerenwinkel et al. 2007a). Thus, the waiting time is approximately linear in the number of mutations and tumor progression is driven mainly by selection. Schöllnberger et al. (2010) have identified the successive clonal expansions predicted by this model with the rate-limiting steps in multistage theory.

The rough approximation (Equation 9) has been refined and generalized in several ways. Using a branching process, Bozic et al. (2010) showed that in a growing population, the acquisition of subsequent driver mutations becomes increasingly faster. They estimated the average selective advantage per driver mutation from experimental data to be as small as $s \approx 0.004$. In addition, the stochasticity of the branching process provides an explanation for the huge heterogeneity in tumor sizes and progression times observed clinically. This model also allows for relating the expected number of driver mutations, k, to passenger mutations, n,

$$n = \frac{\nu}{2s} \log \frac{4ks^2}{u^2} \log k \tag{10}$$

where ν is the rate of acquisition of neutral mutations, and u the driver mutation rate.

Durrett et al. (2009) considered a branching process approximation of the Moran process to arrive, in a rigorous analysis, at waiting times generalizing Equation 9 and the results for TSG inactivation (Durrett and Mayberry 2011). Furthermore, mutation accumulation has also been studied in exponentially growing cancer cell populations (Iwasa et al. 2006; Haeno et al. 2007; Durrett and Moseley 2010). McFarland et al. (2013) used a stochastic birth–death process to study the effect of moderately deleterious, rather than neutral, mutations on tumor progression and to explore cancer treatments that exploit their genetic load. In an attempt to calculate waiting times to cancer under mutational order constraints, Gerstung and Beerenwinkel (2010) considered conditionally independent Poisson waiting times for each mutation and derived the waiting time, τ_k, accounting for all mutational pathways in genotype space.

Models for the evolution of drug resistance.—The evolution of drug resistance is frequently observed during cancer treatment. For example, *BRAF* mutant melanomas treated by a targeted inhibitor often acquire resistance through mutations in other genes after initial response (Nazarian et al. 2010). A classical model for testing whether resistance is acquired during treatment or caused by pre-existing subclones is the model of Luria and Delbrück (1943). In their pioneering work on resistance of bacteria to infection with a bacteriophage, they calculated the relation between the mean and

the variance of the number of drug resistant colonies emerging in an exponentially growing population. If resistance is acquired at the time of infection and cell-specific, then the number of resistant clones follows Poisson statistics, whereas the variance is larger if pre-existing resistant clones are selected, because resistant clones are not independent. Recently, Diaz Jr et al. (2012) used this model to conclude that in colorectal carcinomas, resistance to EGFR inhibitors, which is frequent and occurs after a rather constant time of treatment, is caused by pre-existing subclones. The analysis of Goldie and Coldman (1979) demonstrated that resistant subclones exist in a tumor with probability proportional to tumor size and mutation rate. Bozic et al. (2013) used branching processes to compute the probability of mono- and combination therapy success under the assumption that single mutations confer resistance to an individual drug and showed that combination therapy is much more likely to succeed.

Stochastic Models of Structured Populations

The evolutionary dynamics of structured populations can differ from those in well-mixed populations. Population structure may result from cell differentiation into functional groups or from separation of cells into spatial compartments. For example, the simplest form of a regular grid is the n-dimensional lattice, where the number of interaction partners is identical for all cells. Komarova (2006, 2007) has shown that the fixation probability in the one-dimensional lattice is smaller than in a mixed population, because the number of interaction partners is constrained to the surrounding neighbors, which suppresses the effect of selection.

Linear systems.—A simple one-dimensional system is motivated by the colonic crypt, which is organized as a vertical cylinder with a small number of stem cells at the bottom and the colonic epithelium at the top end. Stem cells divide slowly and produce differentiating cells that move up the cylinder until they reach the epithelium where they eventually shed off. Because of its radial symmetry, this process can be idealized by a linear chain of length N with a stem cell on one end and the epithelium on the other end (Nowak et al. 2003). As all but the stem cells are only transiently present in this system, only mutations in the stem cell will reach fixation, thereby reducing the number of cells at risk from N to just one. The process of shedding is regulated by the *APC* TSG, whose deactivation results in the accumulation of cells and their outgrowth as polyps. To inactivate both alleles of the tumor suppressor *APC*, the stem cell needs to be hit at least once followed by a second hit either in the stem cell or in a differentiating cell, as a double mutation in a differentiating cell is rather unlikely due to their short lifespan. The linear architecture of the colonic crypt can also be modeled by consecutive compartments of stem cells, differentiated cells, and transit cells, in which tumor growth is caused by imbalances between the rates of exchange between these (Johnston et al. 2007). Recently Zhao and Michor (2013) extended the linear process model to include spatially different growth kinetics along the crypt and found that the experimentally observed kinetics with an increased proliferation rate closer to the stem cell at the base of the crypt best suppressed the evolution of TSGs.

Cellular automata.—The replication dynamics of cells on a discrete lattice in discrete time defines a cellular automaton (Deutsch and Moreira 2002). Each cell is represented by a separate object that stores the position, movement, and cell identity. The movement and replication dynamics are modeled by a set of rules accounting for the cellular state and its response to neighboring cells and microenvironmental stimuli. Cellular automata have the advantage that the fate of every cell is explicitly modeled. This imposes, however, a large computational cost when large ensembles of cells are simulated, and, in general, analytical results are infeasible.

Thalhauser et al. (2010) analyzed 1D and 2D spatial generalizations of the Moran process by including cell migration. They found that migration has a positive effect on the ability of a single mutant cell to invade a pre-existing colony and that large-scale cell death selects for the migratory phenotype. This finding may explain how chemotherapy provides a selection mechanism for highly invasive cancer cells.

Perfahl et al. (2011) proposed a complex 3D multiscale model of tumorigenesis that accounts for vascularization by coupling several processes, including blood flow, angiogenesis, nutrient and growth factor transport, cell movement, and interactions between normal and tumor cells. The agent-based model is analytically intractable, but in forward simulations, the spatio-temporal evolution of a vascular tumor can be investigated, including its response to therapy.

Deterministic Models of Well-mixed Populations

In large well-mixed populations, stochastic effects can be negligible such that deterministic models of evolution based on dynamical systems can be applied that describe the mean behavior of the evolutionary system. Denoting by x_i the frequency of genotype i and by $\dot{x}_i$ its derivative with respect to time, the replicator equation (Schuster and Sigmund 1983) is

$$\dot{x}_i = x_i[f_i(x) - \phi(x)], \qquad i = 1, \ldots, n \tag{11}$$

where $f_i(x) = f_i(x_1, \ldots, x_n)$ denotes the fitness which, in general, depends on the frequencies of all other genotypes. The term $\phi(x) = \sum_j f_j(x) x_j$ is the average fitness of the population. In the special case of constant fitness, $f_i(x) = f_i$, the replicator equation is termed selection equation. In this case, the genotype with highest fitness reaches fixation, whereas all others go extinct; this is commonly referred to as survival of the fittest (Nowak 2006a). The selection equation can be

extended to account for mutation with probability q_{ij} from type i to type j to obtain the so-called quasispecies equation $\dot{x}_i = \sum_j x_j f_j q_{ji} - \phi(x) x_i$. This model predicts an error threshold, that is, a critical mutation rate beyond which the population cannot be maintained due to loss of vital genetic information. Using this approach, Solé and Deisboeck (2004) explored the effect of genetic instability on tumor progression and found that an error threshold exists in mutator phenotype cancer cell populations, whereas Brumer et al. (2006) compared the effect of microsatellite and chromosomal instability on tumors.

Deterministic Models of Structured Populations

At a macroscopic scale, the number of cancer cells in a given volume may be approximated by a continuous density. The continuum approximation of the dynamics on regular grids is given by partial differential equations (PDEs). A random spatial movement is then approximated as diffusion and a directed movement by physical drift (not to be confused with random genetic drift). The solutions of PDEs can be efficiently computed, and in certain limit cases, there exist analytical solutions or solutions that can be approximated by analytically tractable models. The latter approaches have the advantage that one can directly assess the influence of certain model parameters (Murray 2002). PDEs have been used to model the dynamics of tumor cell density and its dependency on other diffusive factors such as nutrients and growth factors. Many authors have studied avascular tumor growth to elucidate the role of nutrient flux and necrotic signaling during early tumor growth (Roose et al. 2007). Avascular tumor growth is characterized by an early exponential spherical expansion up to a diameter of about 1mm, above which cells on the inside become necrotic and growth saturates.

Population structure due to differentiation exists, for example, in the hierarchically organized hematopoietic system. It consists of a few thousand slowly replicating hematopoietic stem cells which differentiate in multiple steps into different mature blood cells. The evolutionary dynamics of the average number of cells in a given stage can be modeled by a system of ordinary differential equations. Such models show that the hierarchical organization suppresses the accumulation of mutations and correctly predict the clonal diversity in childhood acute lymphoblastic leukaemia (Werner et al. 2013). The different rates of cellular proliferation in stem cells and differentiating cells can also explain the biphasic response to imatinib treatment in chronic myeloid leukemia as well as the rates of relapse due to resistance mutations (Michor et al. 2005).

Hybrid Models

Hybrid approaches of PDEs and cellular automata have been used to model the discrete dynamics of cell fates coupled to diffusive signals and nutrients fluxes (Anderson and Quaranta 2008). Dormann and Deutsch (2002) used a hybrid cellular automaton to model the avascular growth of tumor cells. Vascularization is a hallmark of cancer, and vascular endothelial growth factor (VEGF) signaling can be therapeutically targeted. Alarcón et al. (2005) used a multiscale model for the vascular tumor growth accounting for tumor-normal cell interactions and vascularization through VEGF signaling. Owen et al. (2009) used a similar model to understand the dynamics of vascularization under different drug pressures to optimize drug efficacy. A precise understanding of cancer growth kinetics may help optimize surgery and therapy duration, and eventually, quantitative models measuring the effect of drugs could be used for finding optimal drug dosage.

Modeling Cellular Interactions using Evolutionary Games

Except for the replicator equation (Equation 11), all models discussed above assume constant fitness. However, the somatic fitness of cancer cells is likely to be density-dependent, because tumor cells interact with each other and with stromal cells (Poste et al. 1981). Evolutionary game theory provides a mathematical framework for modeling such interactions (Maynard Smith 1982). Here, fitness is the expected outcome of a game, which is defined by a payoff matrix. The dynamics of an evolutionary game can be either stochastic or deterministic and populations may be structured or not.

When two different cell types, or strategies, i and j meet and interact, then M_{ij} denotes the payoff, that is, the benefit or harm, that cell type i receives from the interaction. In a well-mixed population, the expected payoff is a linear function of the population frequencies x_j, such that the fitness of type i becomes

$$f_i(x) = \sum_j M_{ij} x_j \tag{12}$$

For infinite population size, the dynamics can be modeled by the replicator equation (Equation 11). The resulting dynamical behavior is much richer than in the situation with constant fitness, as it allows for multiple stable equilibria with coexistence of different cell types and for oscillatory patterns (Fig. 2d). In situations with multiple equilibria, the fixed point reached may also depend on the initial composition of the population. In the stochastic finite population size case, solutions of the corresponding Moran process have been derived in the limits of weak and strong selection (Taylor et al. 2004; Fudenberg et al. 2006).

A prototype two-strategy game offering insights into the evolution of cooperation and defection is the Prisoner's Dilemma (Axelrod and Hamilton 1981). In this game, cooperators pay a cost for others to achieve a benefit. Defectors, however, do not pay this cost, but nevertheless receive the benefit when playing against a cooperator. This leads to a situation where

in the presence of co-operating cells a defector has a higher expected payoff leading to a selective advantage. The defecting strategy is the only evolutionarily stable strategy as it is a Nash equilibrium of the game. However, it does not maximize the overall population fitness as the payoff between two defectors is zero. The Prisoner's Dilemma game illustrates the somatic evolutionary advantage of defecting tumor cells in tissues of cooperative cells, which evolution generated in multicellular organisms (Nowak 2006a). It also raises the question under which conditions co-operativity can evolve. These conditions include different levels of reciprocity and group selection (Nowak 2006b).

Tumors consist of many cell types and interactions may occur at different levels. Vascularization is one hallmark of cancer, and tumor cells generally require the support of other stromal cells. Tumor cells also compete for resources, which introduces an interaction among them. Evolutionary game theory allows for modeling interactions between cancer and stromal cells and also cooperation among tumor cell types. Tomlinson (1997) has investigated the dynamics of multiple tumor cells where one cell type produces a substance cytotoxic to another and under which conditions a cytotoxic cell type can spread through the tumor. Gatenby and Vincent (2003) analyzed the evolution of cancer cells under growth control mechanisms competing for resources and found that such mechanisms can lead to multiple coexisting tumor cell types. Similarly, Axelrod et al. (2006) studied under which conditions cooperation among tumor cells can evolve. Gerstung et al. (2011b) investigated coexistence of multiple tumor types interacting with stromal cells using affine fitness functions, which include a constant fitness contribution in addition to the expected payoff. Due to the symmetry of the replicator equation, the corresponding replicator dynamics can be transformed to an equivalent game with different payoff (Stadler 1991). Other applications include the work by Dingli et al. (2009), who showed that the phenotypic variability commonly observed in multiple myeloma, a blood cancer, can be explained by an evolutionary game. Basanta et al. (2012) have modeled the dynamics of different prostate tumor strategies and their interactions with stromal tissue and have analyzed how the resulting dynamics can be influenced by therapy.

Tumor Phylogeny

The reconstruction of evolutionary trees is a classical topic in computational and evolutionary biology with a wealth of algorithms and models of sequence evolution (Felsenstein 2003), and these methods can be applied to the somatic evolution of cancer (Fig. 6a). Cancer cells divide and accumulate mutations and genomic rearrangements to form clonal subpopulations, the taxa in the phylogenetic tree. Tumor phylogenies represent the evolutionary history of its subclones and can be used to test different hypotheses about tumor evolution (Navin and Hicks 2010). However, the specific features of tumor evolution and cancer data pose challenges to the direct application of classical phylogenetic models. For example, clinical samples contain an unknown number of novel cancer genomes with admixture of normal tissue, whereas classical phylogenetic approaches assume that taxa are known a priori. Additionally, short read NGS does not reveal complete haplotypes, but only individual alterations without information about their co-occurrence.

Many different types of data and cellular properties have been used for evolutionary analyses in cancer, including microsatellites (Frumkin et al. 2008) and lentiviral barcoding (Nolan-Stevaux et al. 2013), but in the following we mostly limit the discussion to SNVs and CNAs, two widely used data types (Gerlinger et al. 2012; Carlson et al. 2012). Inference from complex events like chromothripsis might become very important in the future, but currently only the basic concepts have been described and robust inference methods to identify these events are lacking (Korbel and Campbell 2013). We end the section with a short overview of the current developments in single-cell sequencing, which will provide new opportunities for tumor phylogeny reconstruction in the future. For a summary of the software discussed in this section, see Table 2.

Phylogenetic Tree Reconstruction from SNVs

Sequencing a population of cancer cells allows to infer SNVs and their allele frequencies (Fig. 3). The allele frequencies need to be corrected for copy number alterations, LOH, and normal contamination to estimate the percentage of cancer cells carrying the SNV (Shah et al. 2012; Roth et al. 2014). Because it is unknown in which genome a given SNV occurred, prior to tree reconstruction, SNVs are clustered into sets of mutations with common frequency (Fig. 3d). This clustering is often performed using Bayesian mixture models, either finite ones (Larson and Fridley 2013) or nonparametric ones in which the number of mixture components is estimated together with their frequencies and densities (Shah et al. 2012; Nik-Zainal et al. 2012b; Roth et al. 2014).

For manual phylogenetic tree reconstruction from inferred SNV frequencies, Nik-Zainal et al. (2012b) made two assumptions: (i) no mutation occurs twice in the course of cancer evolution (infinite sites assumption), and (ii) no mutation is ever lost (no back mutations). These assumptions translate into two basic principles: The pigeonhole principle or Dirichlet's Box (Fig. 4a), which in the simplest case states that if the sum of the clonal frequencies of two SNVs is greater than 100%, at least one cell must have contained both SNVs. Because the same mutation cannot be gained twice independently by the first assumption, one clone must be the ancestor of the other. The second assumption implies that the clone with the higher clonal frequency, that is, the larger number of cells carrying the SNV, must be the ancestor (Fig. 4b). These assumptions, although naive

TABLE 2. Software tools implementing phylogenetic methods for reconstructing within-patient and within-tumor evolutionary tumor histories.

Software	Data	Model / Inference	References
PhyloSub[a]	SNV	Tree-stick-breaking process, binomial / MCMC	(Adams et al. 2010)
PyClone[b]	SNV	Dirichlet Process, beta-binomial / MCMC	(Roth et al. 2014)
SciClone[c]	SNV	Beta mixture model	Miller et al. 2014
Clomial[d]	SNV	Binomial / EM	(Zare et al. 2014)
Trap[e]	SNV	Exhaustive search under constraints	(Strino et al. 2013)
CloneHD[f]	SNV + CNA	HMM, EM, Variational Bayes	(Fischer et al. 2014)
ThetA[g]	CNA	Maximum likelihood	(Oesper et al. 2013)
cancerTiming[h]	CNA	Maximum likelihood	(Purdom et al. 2013)
GRAFT[i]	CNA	Partial maximum likelihood	(Greenman et al. 2012)
MEDICC[j]	CNA	Finite state transducer, Minimum-event distance	(Schwarz et al. 2014b)
TuMult[k]	CNA	Breakpoint distance	(Letouzé et al. 2010)
TITAN[l]	CNA	HMM / EM	(Ha et al. 2014)

Notes: SNV, single-nucleotide variant; CNA, copy number aberration; MCMC, Markov-chain monte carlo; EM, expectation maximization; HMM, Hidden Markov Model;
[a]https://github.com/morrislab/phylosub
[b]http://compbio.bccrc.ca/software/pyclone
[c]https://github.com/genome/sciclone
[d]http://www.bioconductor.org/packages/devel/bioc/html/Clomial.html
[e]http://sourceforge.net/projects/klugerlab/files/TrAp
[f]https://github.com/andrej-fischer/cloneHD
[g]https://github.com/raphael-group/THetA
[h]http://cran.r-project.org/web/packages/cancerTiming
[i]http://www.sanger.ac.uk/genetics/CGP/Software/GRAFT
[j]https://bitbucket.org/rfs/medicc
[k]http://bioserv.rpbs.univ-paris-diderot.fr/ letouze/TuMult
[l]http://compbio.bccrc.ca/software/titan/

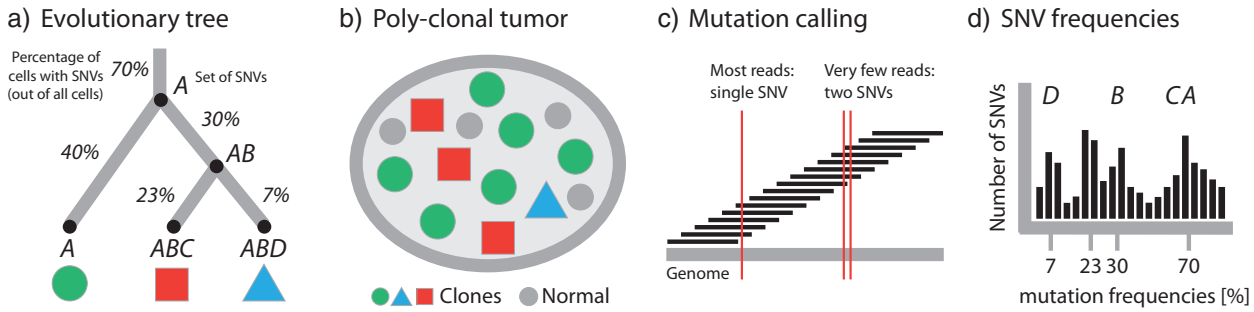

FIGURE 3. Inferring tumor phylogeny from next-generation sequencing data. a) Subclones are related to each other by an evolutionary process of acquisition of mutations. In this example, the three clones (leaf nodes) are characterized by different combinations of the four single nucleotide variant (SNV) sets *A*, *B*, *C*, and *D*. The percentages on the edges of the tree indicate the fraction of cells with this particular set of SNVs, e.g., 70% of all cells carry *A*, 40% additionally carry *B*, and only 7% carry *A*, *B*, and *D*. b) The evolutionary history of a tumor gives rise to a heterogeneous collection of normal cells (small discs) and cancer subclones (large discs, triangles, squares). Internal nodes that have been fully replaced by their descendants (like the one carrying SNV sets *A* and *B* without *C* or *D*) are no longer part of the tumor. c) Sequencing data consist of short reads covering (parts of) the cancer genome. Comparison to the germline DNA of the same patient allows to identify SNVs and other genomic aberrations. Since reads are short, most will only cover a single SNV. In few cases, pairs of SNVs are covered, which allows to assess patterns of co-occurrence and mutual exclusivity between SNVs. d) The sets of SNVs distinguishing the subclones cluster in the SNV frequency distribution. The mean of each cluster (*x*-axis) is the fraction of cells carrying this set of SNVs. The goal of tumor phylogenetics is to infer the evolutionary tree (a) from the mutations observed in the sequencing data (c) and their frequencies (d).

from a general phylogenetics perspective, appear to be justified in the cancer setting. Recent sequencing efforts showed that HeLa cells have accumulated about four million SNVs (Adey et al. 2013). Even with such a large number of mutations the probability of the same site being affected twice is about 1/1000, assuming a uniform mutation rate across the genome. Back mutations are accordingly less likely.

Computational methods have been proposed that implement these two rules. Strino et al. (2013) used a linear algebra approach that makes use of parsimony and sparsity assumptions to limit the number of possible trees. They show that their approach works for up to 25 aberrations, which makes extensive feature selection necessary, but it may be applied to the output of the mixture models discussed above by treating each

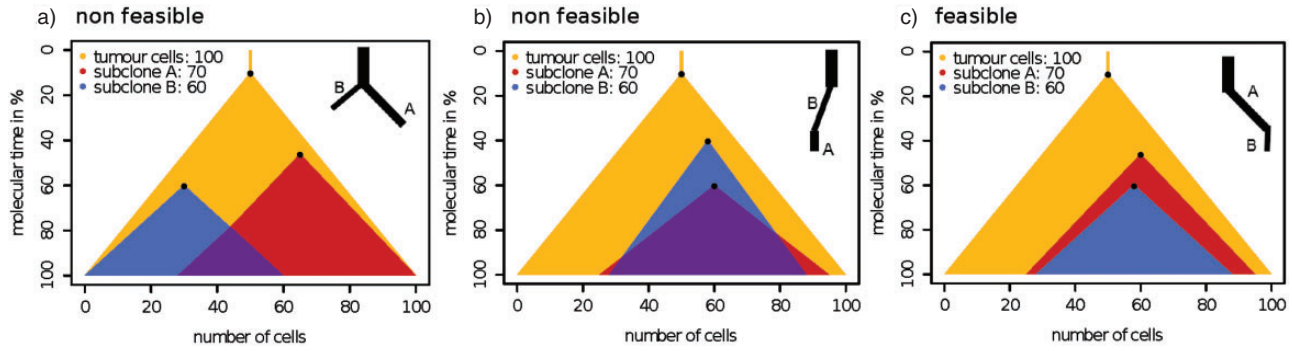

FIGURE 4. Two simple principles for tree inference from SNVs. For a given set of subclones and their respective clonal fractions, each illustrated by a triangle with a dot at the top vertex representing the clonal origin, two conditions need to be met for a potential phylogeny to be considered feasible: a) Dirichlet's box: If two SNV frequencies (small triangles inside large triangle) sum to more than 100%, then some cancer cells must contain both SNVs (overlap of the two small triangles). In a tree-like evolutionary process some cells must have acquired the same mutation independently, which in cancer, is considered highly unlikely. Hence, one of the two subclones (small triangles) is ancestral to the other. b) Larger ancestor: In this case, if one clonal fraction is larger than the other, the larger must be the ancestor; otherwise cancer cells would have lost the previously gained mutation (nonoverlapping regions between the two small triangles at the bottom), which again is considered highly unlikely. The most likely feasible solution is shown in c), where both principles are met (and the two small triangles are nested).

cluster as a single aberration. However, Nik-Zainal et al. (2012b) demonstrated the limitations of sequential SNV clustering and tree reconstruction. In their analysis, one of the clusters was spread over three different branches of the tree. Such inconsistencies may be avoided by combining clustering and tree reconstruction into a single step. Recently, Zare et al. (2014) proposed a generalization of the approach implemented by Strino et al. (2013) to multiple samples per patient. Jiao et al. (2014) present a joint approach based on interleaving two stick-breaking processes, which results in a hierarchy of clusters (Adams et al. 2010). A recent implementation of a nonparametric Bayesian clustering approach that implements those principles is the work of Roth et al. (2014), which jointly infers posterior density estimates over both the clustering structure and cellular prevalence of the clones using MCMC sampling. An alternative is the work of Miller et al. (2014) which uses a variational Bayesian mixture model for subclone identification. Finally, Fischer et al. (2014) make use of the combined information in CNAs and SNVs to perform clonal decomposition using Hidden Markov Models.

It is important to note that single allele frequencies offer only a very limited snapshot of the tumor phylogeny. In particular, many different trees can agree with the same pattern of allele frequencies and only few topological constraints exist to limit the space of solutions (Fig. 4). Jiao et al. (2014) discuss this issue and find that SNV clusters can always be ordered into a linear cascade (with the most frequent one on top and the least frequent one at the bottom) and in a fork (unless the sum of frequencies of the child nodes is larger than the frequency of the parent node). Having multiple samples can further constrain the tree topology and, for example, predict a fork if the frequencies in the different samples put the clones into contradictory linear orders (Jiao et al. 2014). In summary, single allele frequencies offer only weak evidence for tumor phylogenies. Sequencing technologies with longer reads might alleviate these limitations in the near future, because more reads will carry multiple SNVs and patterns of co-occurance or mutual exclusivity will help to refine tree structure.

Phylogenetic Tree Reconstruction from CNAs

In addition to SNVs, many cancers display a large amount of genomic rearrangements resulting in CNAs, which provide another source of data for inferring evolutionary relationships among cancer genomes. Copy number profiles are affected by the same challenges as SNVs, such as normal admixture and subclonal genomic changes. Attempts to address these issues include Sector Ploidy Profiling (Navin et al. 2010), which involves macrodissection of a physical sample into sectors followed by cell sorting according to total DNA content, which results in more genomically homogeneous cell populations. Copy number profiles were then computed by segmenting intensities derived from two-color aCGH microarrays. Trees were reconstructed using distances based on Pearson correlation between the *logR* values followed by neighbor-joining tree inference (Saitou and Nei 1987). The authors found that the breast cancers they studied could be divided into two groups, one genetically homogeneous group, called monogenomic tumors, and one genetically heterogeneous group, called polygenomic tumors.

Alternatively Oesper et al. (2013) have proposed a computational approach for subclonal decomposition from copy number profiles based on genome-wide segmented read depth information. In the same spirit Ha et al. (2014) implemented a Hidden Markov Model-based approach for identification of subclonal copy number profiles.

In any case, for tree reconstruction from copy number profiles the traditionally used Euclidean or correlation distances are ill-suited. Genomic sites affected by rearrangement events are not independent

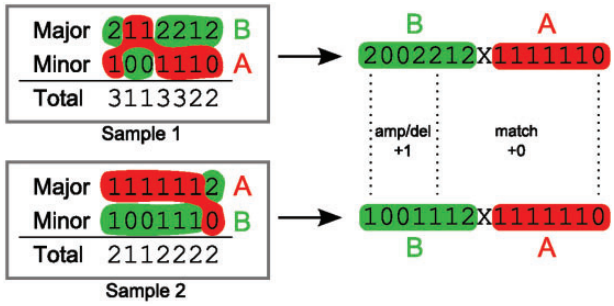

FIGURE 5. Phasing copy number profiles. While SNP arrays are capable of determining a major and minor copy number for the two parental alleles, their assignment (phasing) to the two actual physical alleles A and B is unknown. Because evolutionary events happen on the physical copies, correct phasing is essential for determining evolutionary distances. In this example, the two major copy number profiles between sample 1 and sample 2 (left) have a distance of two events (one amplification at position 1 and one amplification spanning positions 4 and 5), while the minor copy number profiles are identical, yielding a total of two events between the genomes of sample 1 and sample 2. Optimal assignment (right) to the alleles A and B reduces the evolutionary distance to a single amplification event spanning the first five genomic loci. This is also not evident from the total copy number (the sum of major and minor) which would still require two separate events.

and identically distributed. By the nature of the DNA replication process all genomic rearrangement events have a specific start and end, duplicating or deleting all bases between those two loci (Hastings et al. 2009). This mechanism results in strong dependencies between adjacent loci that are not accounted for by Euclidean or correlation distances, because they consider all loci independently, rather than the actual events.

To address this limitation, Letouzé et al. (2010) proposed the TuMult method, which works on breakpoints, that is, loci at which the copy number changes, instead of full genomic profiles. Using a novel breakpoint distance they estimate the number of genomic events between two copy number profiles, and these distances are then used for tree reconstruction. TuMult has been applied by Sottoriva et al. (2013a) to estimating phylogenies of glioblastoma. Like earlier methods, *TuMult* does not take allele-specific copy numbers into account.

Greenman et al. (2012) and Purdom et al. (2013) developed related algorithms for estimating the order of genomic rearrangement events. Although tree reconstruction is not the primary focus of these studies, ordering events involves solving similar problems as for estimating evolutionary distances. For example, for each site, the major and minor copy numbers must be assigned to one of the two physical alleles, that is, phased (Fig. 5). Greenman et al. (2012) use external linkage information that can, for example, be obtained from HapMap and a graph-theoretical approach to find the most likely assignment of each copy number to either of the two alleles. Ultimately, the method finds the most likely clonal ordering by solving this problem over all sampled genomes of a tumor. The phasing problem is solved in a similar fashion by the Battenberg algorithm (Nik-Zainal et al. 2012a).

Schwarz et al. (2014b) have recently developed MEDICC to jointly solve the problems of phasing and tree reconstruction using a minimum evolution criterion. Based on finite-state transducers (Cortes et al. 2004; Schwarz et al. 2010), MEDICC computes the minimum number of amplification and deletion events to transform one genomic profile into another. It finds allele-specific phasing, tree topology, and ancestral genomes such that the total tree length, that is, the total number of genomic events in the tree, is minimal. In an application of this approach to a large study of high-grade serous ovarian carcinoma, the authors confirmed the bi-partition of tumors into mono- and polygenomic cancers and showed that this stratification predicts resistance development (Schwarz et al. 2014a).

Single-cell Approaches

For a small number of loci, fluorescent *in situ* hybridization has been used to characterize tumor heterogeneity on a single-cell level (Almendro et al. 2014; Trinh et al. 2014) and to infer phylogenetic trees (Chowdhury et al. 2013). On a genome-wide level, recent years have brought great advances in single-cell sequencing (Shapiro et al. 2013). For example, Navin et al. (2011) used low-coverage single-nucleus sequencing to reconstruct evolutionary histories of cancer lineages based on CNAs. They employed conventional neighbor-joining using a Euclidean distance metric on the discretized integer copy number profiles for tree building. Hou et al. (2012) demonstrated clonal evolution in essential thrombocythemia tumors by single-cell whole-exome sequencing of 90 individual cells and a population-level model of evolution, whereas Xu et al. (2012) found no evidence for clonal subpopulations when sequencing individual kidney cancer cells. One of the first statistical methods for evolutionary inference from single-cell sequencing data by Kim and Simon (2014) explicitly models the high single-cell sequencing error rates and infers phylogenetic relationships as well as orderings of mutations. Single-cell sequencing is still in its infancy and brings its own dedicated challenges, such as amplification bias, but has the potential to resolve some of the limitations of current approaches, most notably the phasing problem. In the near future, we expect most clinical studies on intratumor diversity to still rely on NGS of mixed samples, because of technical challenges and cost-benefit considerations.

CANCER PROGRESSION

Genetic events that drive cancer progression generally do not occur independently of each other. Direct and indirect interactions result from nonlinear, epistatic fitness landscapes underlying the evolutionary process and introduce statistical dependencies among genetic alterations. For example, Höglund et al. (2001)

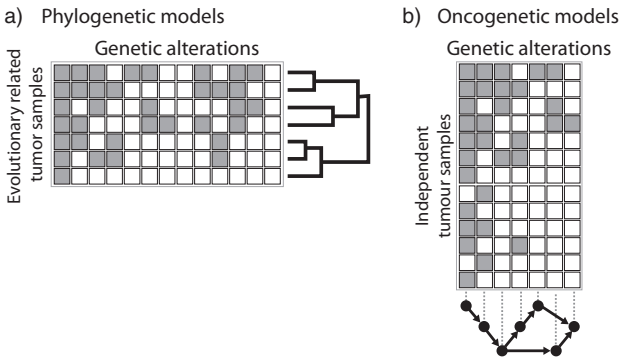

FIGURE 6. Phylogenetic versus oncogenetic models. Phylogenetic models of tumor samples (a) and oncogenetic models of cancer drivers (b) use the same type of data: genomic aberrations observed in patient tumor samples. Phylogenetic models (a) use mostly genomewide data of a small number of evolutionary-related tumor samples, either from the same patient or from different clones within the same tumor. Tumor progression models (b), on the other hand, generally concentrate on a small number of aberrations observed in a larger number of independent tumors from different patients.

analyzed cytogenetic data from 3,016 solid cancers and observed preferential combinations of alterations using principal component analysis suggesting distinct evolutionary pathways. Graphical progression models are used to estimate these dependencies (Fig. 6b). In progression models, tumor samples from different patients are regarded as independent realizations of the same stochastic evolutionary process. The data is typically cross-sectional, that is, tumors are observed at different unknown time points. Most progression network models are variations of Bayesian networks, a class of directed graphical models representing conditional independencies among random variables. Available software for modeling cancer progression is summarized in Table 3.

Trees with Unobserved Internal Vertices

Among the first approaches to estimate dependencies among cancer-driving events were distance-based phylogenetic methods. The idea is to compute distances between genetic events, rather than between tumors as discussed above, and then to compute an optimal tree, where the observed genetic alterations are represented as the leaves of the tree. Desper et al. (2000) define the distance between events X and Y as

$$-2\log P(X,Y)+\log P(X)+\log P(Y) \qquad (13)$$

which quantifies the deviation of the joint distribution $P(X,Y)$ from $P(X)P(Y)$, the one expected under the independence model. The authors subsequently apply distance-based tree reconstruction methods, such as neighbor-joining. Von Heydebreck et al. (2004) have developed an efficient maximum likelihood approach for a probabilistic Bayesian network formulation of the same tree model, in which internal vertices correspond to

TABLE 3. Software tools implementing probabilistic graphical models for estimating cancer progression.

Model	Topology	LPD	Constraints	Noise	Learning	Software	References
OT/HI	tree	discrete	monotone	no	ML	oncomodel[a]	(Desper et al. 2000; von Heydebreck et al. 2004)
OT	tree	discrete	monotone	no	MWB	oncotrees[b]	(Desper et al. 1999)
OT	tree	discrete	monotone	yes	MWB	oncotree[c]	(Szabo and Boucher 2002)
HOT	tree	discrete	monotone	yes	ML via SEM	n.a.	(Tofigh et al. 2011)
Mixture of OTs	forest	discrete	monotone	yes	ML via SEM	mtreemix[d]	(Beerenwinkel et al. 2005b; Beerenwinkel et al. 2005a)
Mixture of HOTs	forest	discrete	monotone	yes	ML via SEM	hotmix[e]	(Tofigh et al. 2011)
CBN	DAG	discrete	monotone	yes	ML	cbn[f]	(Beerenwinkel et al. 2007b)
CT-CBN	DAG	waiting time	monotone	no	ML via EM	ct-cbn[g]	(Beerenwinkel and Sullivant 2009)
Hidden CBN	DAG	waiting time	monotone	yes	ML via EM, SA	h-cbn[h]	(Gerstung et al. 2009)
Bayesian CBN	DAG	discrete	monotone	yes	MCMC	bayes-cbn[i]	(Sakoparnig and Beerenwinkel 2012)
NAM	DAG	waiting time	none	no	ML	n.a.	(Hjelm et al. 2006)
ON	DAG	discrete	(semi-)mon.	yes	MILP	diprog[j]	(Shahrabi Farahani and Lagergren 2013)
RESIC	none	RE	none	no	SM	upon request	(Attolini et al. 2010; Cheng et al. 2012)

Notes: OT, Oncogenetic tree; OT/HI, OT with hidden internal nodes; HOT, Hidden oncogenetic tree; CBN, Conjunctive Bayesian Network; CT-CBN, Continuous-time CBN; NAM, Network aberration model; ON, Oncogenetic network; LPD, local probability distribution; RESIC, Retracing the Evolutionary Steps in Cancer; DAG, directed acyclic graph; ML, maximum likelihood; EM, Expectation-Maximization algorithm; SEM, Structural EM; MCMC, Markov chain Monte Carlo; MWB, Maximum weight branching; MILP, mixed integer linear program; RE, Rate of evolution (Moran process); SA, Simulated annealing; SM, Simplex minimization by Nelder and Mead

[a]http://cran.r-project.org/web/packages/oncomodel/

[b]http://www.ncbi.nlm.nih.gov/CBBresearch/Schaffer/cgh.html

[c]http://cran.r-project.org/web/packages/Oncotree/

[d]http://mtreemix.bioinf.mpi-inf.mpg.de/

[e]https://github.com/atofigh/hotmix

[f]http://www.cbg.ethz.ch/software/cbn/

[g]http://www.cbg.ethz.ch/software/ct-cbn/

[h]http://www.cbg.ethz.ch/software/ct-cbn/

[i]http://www.cbg.ethz.ch/software/bayes-cbn/

[j]https://bitbucket.org/farahani/diprog

hidden random variables. Although there is no obvious interpretation of the internal vertices, these models capture information about co-occurrences of events, and they can help detecting preferred sequential orders of events, specifically early events in tumor progression. Indeed, if an event X occurs almost always before event Y, then the leaf representing X will be close to the path from the root of the tree to vertex representing Y.

Oncogenetic Trees

Oncogenetic tree models were introduced by Desper et al. (1999). They describe the accumulation of mutations under ordering constraints, which can be represented by a tree. The root of the tree is the wildtype without alterations. The branches at a given node describe the set of additional mutations that become possible when the node is mutated. Unlike the tree models discussed above, oncogenetic trees have no hidden internal vertices, but all vertices correspond to observed genetic alterations. The star defines the least restricted oncogenetic tree model, in which all alterations are possible at any time. The most restrictive topology is a linear chain of alterations, as in the case of sequential accumulation of *APC* → *KRAS* → *TP53* mutations during colorectal carcinogenesis, the first explicit description of cancer driver dependencies (Fearon and Vogelstein 1990). Desper et al. (1999) have proposed an efficient tree reconstruction algorithm based on an instance of maximum weight branching, a classical combinatorial optimization problem (Edmonds 1967; Karp 1971). Cancer progression models have been applied to CGH data, which records losses and gains of entire chromosome arms. For example, Jiang et al. (2000) used both distance-based and oncogenetic tree models to analyze dependencies among chromosomal aberrations in clear cell renal cell carcinoma. They found two distinct subgroups of tumors and clarified that loss of the small arm of chromosome 8 is a late event in renal cell carcinogenesis.

The original oncogenetic tree model from Desper et al. (1999) does not explicitly account for observation errors. Szabo and Boucher (2002) have made this extension to the model and the inference algorithm accounting for false positive and false negative observations. Oncogenetic trees have also been approached in the likelihood framework and extended to mixture models. Mixtures of oncogenetic trees provide a more flexible alternative to fitting a single tree, for example, in the presence of tumor subgroups, or alternative, mutually exclusive evolutionary pathways (Beerenwinkel et al. 2005a). They can be estimated by a structural EM algorithm, which, in each step, computes a maximum weight branching. To account for observation errors, a noise component with star topology is often used. Tree mixtures have been used to derive the genetic progression score, defined as the expected waiting time of the observed tumor in the model. This score improved survival predictions in glioblastoma and prostate cancers (Rahnenführer et al. 2005). Tofigh et al. (2011) have further generalized oncogenetic tree mixture models by introducing, for each event, a hidden random variable indicating a possible observation error. Their hidden-variable oncogenetic trees (HOTs) thus offer a different error model. Global structural EM algorithms for learning HOTs and mixtures of HOTs have been developed.

Progression Networks

In tumor progression networks, the assumption of a tree-like dependency structure among alterations is dropped. General Bayesian network models have been proposed for this purpose (Radmacher et al. 2001), but they are often too computationally expensive to learn from data and not straightforward to interpret as progressions in time. A common assumption is monotonicity, that is, for each event, all its predecessors in the graph are required to occur before it can happen. For example, Conjunctive Bayesian networks (CBNs) are monotone progression networks. They generalize oncogenetic trees and are defined by a partially ordered set of mutations (Beerenwinkel et al. 2007b). In continuous-time CBNs, the waiting time for each mutation is assumed to be distributed exponentially; a genotype is defined by all mutations that have accumulated before a stopping time (Beerenwinkel and Sullivant 2009). Additionally, the observed genotypes may differ from the true genotypes because of observation errors (Gerstung et al. 2009). A nested EM algorithm is used to estimate the parameters of both the error and waiting time processes in the hidden CBN (H-CBN) model. The H-CBN allows for de-noising genotypes using the maximum a posteriori (MAP) estimates based on the progression model, which were found to improve survival predictions in renal cell carcinoma. Gerstung et al. (2011a) applied the H-CBN model to genetic data from three different cancer types, and found that there is stronger evidence for temporal dependencies among signaling pathways than among individual genes, likely because there are often many different ways to hit a signaling pathway. Using the Wright–Fisher model of cancer progression (Beerenwinkel et al. 2007a), they also showed that the accumulation rates of alterations in the CBN model are approximately linearly related to the fitness advantages s. Recently, a Bayesian inference scheme has been proposed for CBNs, which allows for assessing the full posterior probability of the partial order and the parameters (Sakoparnig and Beerenwinkel 2012).

Hjelm et al. (2006) have proposed an extended network aberration model (NAM), where aberrations are grouped together. Here, events not only occur randomly in time according to intensity parameters, but also the stopping intensity of the process after each event depends on the number of aberrations grouped into the event. In addition, the strength of all pairwise dependencies are explicitly parametrized in this model. A heuristic ML method is developed owing to the increased complexity of the model. More recently,

progression network inference has been addressed using mixed integer linear programming. Shahrabi Farahani and Lagergren (2013) have devised and solved such an optimization problem for any decomposable model score, including the likelihood score and the Bayesian information criterion score. This approach works for monotone as well as for semimonotone networks, where only the presence of at least one predecessor mutation, rather than all, is required. Attolini et al. (2010) proposed a progression model in which the transition probabilities between genotypes are given by a Moran process. This method has then been extended by Cheng et al. (2012) to also account for transitions between cell states defined by altered signaling pathways.

All progression models discussed above aim at estimating the dependency structure among mutational events. Youn and Simon (2012) have proposed a statistical model for assessing the order of mutations without estimating their full dependency structure explicitly. Instead, they estimate the probability of each mutation i to occur as the k-th event in tumorigenesis directly. Although less informative about individual progression pathways, this approach may be more powerful for identifying early versus late mutations.

Sprouffske et al. (2011) used an agent-based model of a colon crypt to show that using cross-sectional data for inferring mutational order can be misleading. They emphasize the need for integrating phylogenetic methods based on intratumor samples to accurately reconstruct the evolutionary history of tumors. More generally, progression models will benefit greatly from assessing and resolving intrapatient and intratumor diversity. Integrating these two orthogonal modeling approaches (Fig. 6) is a major challenge for future work.

Applications and Perspectives

Although cancer evolution is a fascinating topic for computational and evolutionary biologists, much of the progress in the field has so far been driven by the increasing integration of research into the clinic. On the other hand, already today, evolutionary modeling has an impact on the clinical management of cancer.

Clinical Applications

Several clinical applications of evolutionary methods have been reported. Maley et al. (2006) showed that measures of clonal diversity can predict progression of the premalignant Barret's esophagus to a full-blown adenocarcinoma. In the future, this finding could enable clinical identification of high-risk patients that demand immediate treatment. Evolutionary studies can also give insight into the metastatic process and how selection pressure shapes the metastatic genotype. Khalique et al. (2009) demonstrated the clonal evolution of metastases from primary epithelial ovarian cancers using parsimony-based tree reconstruction on LOH events. In pancreatic cancer, a particularly aggressive malignancy, Campbell et al. (2010) identified genomic rearrangements that dysregulate the transition from the G1 to S phase of cell cycle and demonstrated convergent evolution among different metastases.

Complementary to studies that focus on early cancer development and the metastatic process, evolutionary studies have tried to identify sources of chemotherapy resistance. Cooke et al. (2011) showed that genetic heterogeneity indicates poor response to chemoradiotherapy in cervical cancer. In hereditary ovarian carcinomas, it was subsequently shown that secondary mutations that restore *BRCA1/2* predict chemotherapy resistance (Norquist et al. 2011). Later, an evolutionary study of high-grade serous ovarian cancers showed, for the first time, a correlation between genetic heterogeneity, patient survival, and chemotherapy resistance, and identified subclonal *NF1* deletions as potential drivers of resistant relapse (Schwarz et al. 2014a). Similarly, subclonal driver mutations have recently been identified that comprise risk factors for rapid disease progression in in chronic lymphocytic leukemia (Landau et al. 2013) and myelodysplasia (Papaemmanuil et al. 2013).

Evolutionary modeling can play an important role not only in supporting diagnostics and prognostics, but also for rationalizing treatment design (Bozic et al. 2012; Hochberg et al. 2013). For example, Chmielecki et al. (2011) use a 2-type branching process model with constraints derived from clinical data to predict dosing schedules of tyrosine kinase inhibitors against EGFR mutants. They found that the optimized dosing schemes may delay drug resistance development in lung cancer.

Outlook

Recent advancements in high-throughput molecular profiling techniques allow for the assessment of the molecular states of tumors in great detail. Cancer genome data are collected at a large scale in many clinical studies and in international consortia, such as The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC). Most of these projects initially aim at characterizing the mutational landscape of various types of cancer by sequencing the tumors of many patients at low or moderate coverage (Stratton et al., 2009; Vogelstein et al., 2013). These data will be highly informative for discovering and cataloguing common aberrations in cancer genomes (Fig. 1) and for studying inter-tumor diversity and cancer progression (Fig. 6).

However, cancer is not only a disease of the genome, but also of abnormal cellular interactions in the tumor tissue. For example, the fitness of a clone depends on its genotype and the tissue environment the cells live in. The tissue microenvironment is a complex dynamical system with multiple cellular components that can influence cancer progression and evolution (Cairns 1975; Merlo et al. 2006; Lambert et al. 2011; Greaves and Maley 2012). Cancer cells are influenced by their tissue habitat, and reciprocally, they can remodel the

tissue microenvironment to their competitive advantage (Barcellos-Hoff et al. 2013). Future studies will have to combine analyses of genetic heterogeneity with analyses of tumor tissue architecture to account for genetic variation and epigenetic variation of cancer cells as well as their interactions with surrounding cells (Marusyk et al. 2012; Barcellos-Hoff et al. 2013; Frank and Rosner 2012; Bissell and Hines 2011; Cairns 1975). Some progress is being made in this direction by jointly analyzing genomic data and pathological images (Yuan et al. 2012), but most evolutionary analyses are still performed on genetic data without any information of the tissue environment.

Once the technological hurdles of single-cell genomic profiling, such as inefficient and unbiased genome amplification, are overcome and individual cancer genomes can be identified reliably at a larger scale, tumor evolution can be studied more precisely and in greater detail. Novel and more powerful probabilistic models for these data will be required and are already being developed. They will need to take the spatial dynamics of tumors into account to allow for a systems view on cancer progression. Additionally, they will need to account for cancer-specific properties, such as generally nonhomogeneous rates of evolution. Cancer often requires deactivation of DNA repair pathways as an early driver event and in the course of clonal evolution, more of these events will follow, giving rise to mutator phenotypes. Other important questions concern cancer stem cells (Kreso and Dick 2014). Do all cancer cells have the capability of spawning new subclones and metastasize? Or is the metastatic potential limited to a small number of stem cells? The answers to these questions have important implications for the resulting tree topology, which could be fully branched, or star-like (Navin and Hicks 2010; Schwarz et al. 2014a).

Longitudinal sampling will pose both a challenge and an opportunity to phylogenetic reconstructions in cancer. Where in the traditional phylogenetics scenario all taxa are sampled at the same time point, different samples from biopsies before and after treatment might call for additional, more flexible evolutionary methodology. In this context, circulating tumor DNA has recently received a lot of attention. It has been shown that from tumor DNA found in the plasma of patients, tumor load, and genetic heterogeneity of the cancer can be inferred without the need for invasive biopsy or surgery (Forshew et al. 2012; Dawson et al. 2013; Murtaza et al. 2013). This opens many new possibilities for longitudinal sampling of patients that circumvent many of the inherent logistical and ethical complications of traditional clinical studies.

Intra-tumor genetic heterogeneity is often portrayed as a major challenge for successful targeted treatment. However, evolutionary analysis of the process leading to the observed heterogeneity could turn this perceived weakness into a strength by tailoring treatment specifically to the unique evolutionary scenario within each patient. Evolutionary models will thereby play an essential role in predicting escape mutations to treatment before they appear. Together with targeted therapy options, this approach will hopefully allow us to either ultimately eradicate the cancer or permanently restrict its growth, thus turning it into a chronic disease with low impact on quality of life.

Funding

F.M. would like to acknowledge the support of The University of Cambridge, Cancer Research UK and Hutchison Whampoa Limited. Part of this work was funded by the European Research Council (ERC Synergy Grant No. 609883 to N.B.).

References

Adams R. P., Ghahramani Z., Jordan M. I. 2010. Tree-structured stick breaking processes for hierarchical data. Adv. Neural Inf. Process. Syst. (NIPS). 23:19–27.

Adey A., Burton J.N., Kitzman J.O., Hiatt J.B., Lewis A.P., Martin B.K., Qiu R., Lee C., Shendure J. 2013. The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. Nature 500:207–211.

Adzhubei I.A., Schmidt S., Peshkin L., Ramensky V.E., Gerasimova A., Bork P., Kondrashov A.S., Sunyaev S.R. 2010. A method and server for predicting damaging missense mutations. Nat. Methods 7: 248–249.

Aktipis C.A., Kwan V.S.Y., Johnson K.A., Neuberg S.L., Maley C.C. 2011. Overlooking evolution: a systematic analysis of cancer relapse and therapeutic resistance research. PLoS ONE 6:e26100.

Alarcón T., Byrne H., Maini P. 2005. A multiple scale model for tumor growth. Multiscale Model. Simul. 3:440–475.

Albini A., Sporn M.B. 2007. The tumour microenvironment as a target for chemoprevention. Nat. Rev. Cancer 7:139–147.

Almendro V., Cheng Y.-K., Randles A., Itzkovitz S., Marusyk A., Ametller E., Gonzalez-Farre, X., Munoz M., Russnes H.G., Helland A., Rye I.H., Borresen-Dale A.-L., Maruyama R., van Oudenaarden A., Dowsett M., Jones R.L., Reis-Filho J., Gascon P., Goenen M., Michor F., Polyak K. 2014. Inference of tumor evolution during chemotherapy by computational modeling and in situ analysis of genetic and phenotypic cellular diversity. Cell Rep. 6:514–527.

Almendro V., Marusyk A., Polyak K. 2013. Cellular heterogeneity and molecular evolution in cancer. Annu. Rev. Pathol. 8:277–302.

Amir E.-a.D., Davis K.L., Tadmor M.D., Simonds E.F., Levine J.H., Bendall S.C., Shenfeld D.K., Krishnaswamy S., Nolan G.P., Pe'er D. 2013. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nat. Biotechnol. 31:545–552.

Anderson A.R.A., Quaranta V. 2008. Integrative mathematical oncology. Nat. Rev. Cancer 8:227–234.

Aparicio S., Caldas C. 2013. The implications of clonal genome evolution for cancer medicine. N. Engl. J. Med. 368:842–851.

Armitage P., Doll R. 1954. The age distribution of cancer and a multi-stage theory of carcinogenesis. Br. J. Cancer 8:1–12.

Armitage P., Doll R. 1957. A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. Br. J. Cancer 11:161–169.

Athreya K, Ney P. 1972. Branching processes. Mineola (NY): Dover.

Attolini C.S.-O., Cheng Y.K., Beroukhim R., Getz G., Abdel-Wahab O., Levine R.L., Mellinghoff I.K., Michor F. 2010. A mathematical framework to determine the temporal sequence of somatic genetic events in cancer. Proc. Natl. Acad. Sci. USA 107:17604–17609.

Axelrod R., Axelrod D.E., Pienta K.J. 2006. Evolution of cooperation among tumor cells. Proc. Natl. Acad. Sci. USA 103:13474–13479.

Axelrod R., Hamilton W.D. 1981. The evolution of cooperation. Science 211:1390–1396.

Baca S.C., Prandi D., Lawrence M.S., Mosquera J.M., Romanel A., Drier Y., Park K., Kitabayashi N., MacDonald T.Y., Ghandi M.,

Van Allen E., Kryukov G.V., Sboner A., Theurillat J.-P., Soong T.D., Nickerson E., Auclair D., Tewari A., Beltran H., Onofrio R.C., Boysen G., Guiducci C., Barbieri C.E., Cibulskis K., Sivachenko A., Carter S.L., Saksena G., Voet D., Ramos A.H., Winckler W., Cipicchio M., Ardlie K., Kantoff P.W., Berger M.F., Gabriel S.B., Golub T.R., Meyerson M., Lander E.S., Elemento O., Getz G., Demichelis F., Rubin M.A., Garraway L.A. 2013. Punctuated evolution of prostate cancer genomes. Cell 153:666–677.

Barcellos-Hoff M.H., Lyden D., Wang T.C. 2013. The evolution of the cancer niche during multistage carcinogenesis. Nat. Rev. Cancer 13:511–518.

Basanta D., Scott J.G., Fishman M.N., Ayala G., Hayward S.W., Anderson A.R.A. 2012. Investigating prostate cancer tumour-stroma interactions: Clinical and biological insights from an evolutionary game. Br. J. Cancer 106:174–181.

Basik M., Aguilar-Mahecha A., Rousseau C., Diaz Z., Tejpar S., Spatz A., Greenwood C.M.T., Batist G. 2013. Biopsies: Next-generation biospecimens for tailoring therapy. Nat. Rev. Clin. Oncol. 10: 437–450.

Bedard P.L., Hansen A.R., Ratain M.J., Siu L.L. 2013. Tumour heterogeneity in the clinic. Nature 501:355–364.

Beerenwinkel N., Antal T., Dingli D., Traulsen A., Kinzler K.W., Velculescu V.E., Vogelstein B., Nowak M.A. 2007a. Genetic progression and the waiting time to cancer. PLoS Comput. Biol. 3:e225.

Beerenwinkel N., Eriksson N., Sturmfels B. 2006. Evolution on distributive lattices. J. Theor. Biol. 242:409–420.

Beerenwinkel N., Eriksson N., Sturmfels B. 2007b. Conjunctive Bayesian networks. Bernoulli. 13:893–909.

Beerenwinkel N., Rahnenführer J., Däumer M., Hoffmann D., Kaiser R., Selbig J., Lengauer T. 2005a. Learning multiple evolutionary pathways from cross-sectional data. J. Comput. Biol. 12:584–598.

Beerenwinkel N., Rahnenführer J., Kaiser R., Hoffmann D., Selbig J., Lengauer T. 2005b. Mtreemix: a software package for learning and using mixture models of mutagenetic trees. Bioinformatics 21: 2106–2107.

Beerenwinkel N., Sullivant S. 2009. Markov models for accumulating mutations. Biometrika 96:645–661.

Bissell M.J., Hines W.C. 2011. Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. Nat. Med. 17:320–329.

Bowtell D.D.L. 2010. The genesis and evolution of high-grade serous ovarian cancer. Nat. Rev. Cancer 10:803–808.

Bozic I., Allen B., Nowak M.A. 2012. Dynamics of targeted cancer therapy. Trends Mol. Med. 18:311–316.

Bozic I., Antal T., Ohtsuki H., Carter H., Kim D., Chen S., Karchin R., Kinzler K.W., Vogelstein B., Nowak M.A. 2010. Accumulation of driver and passenger mutations during tumor progression. Proc. Natl. Acad. Sci. USA 107:18545–18550.

Bozic I., Reiter J.G., Allen B., Antal T., Chatterjee K., Shah P., Moon Y.S., Yaqubie A., Kelly N., Le D.T., Lipson E.J., Chapman P.B., Diaz Jr, L.A., Vogelstein B., Nowak M.A. 2013. Evolutionary dynamics of cancer in response to targeted combination therapy. Elife 2: e00747.

Brumer Y., Michor F., Shakhnovich E.I. 2006. Genetic instability and the quasispecies model. J. Theor. Biol. 241:216–222.

Burrell R.A., McGranahan N., Bartek J., Swanton C. 2013. The causes and consequences of genetic heterogeneity in cancer evolution. Nature 501:338–345.

Cairns J. 1975. Mutation selection and the natural history of cancer. Nature 255:197–200.

Calabrese P., Tavaré S., Shibata D. 2004. Pretumor progression: Clonal evolution of human stem cell populations. Am. J. Pathol. 164:1337–1346.

Caldas C. 2012. Cancer sequencing unravels clonal evolution. Nat. Biotechnol. 30:408–410.

Campbell P.J., Yachida S., Mudie L.J., Stephens P.J., Pleasance E.D., Stebbings L.A., Morsberger L.A., Latimer C., McLaren S., Lin M.-L., McBride D.J., Varela I., Nik-Zainal S.A., Leroy C., Jia M., Menzies A., Butler A.P., Teague J.W., Griffin C.A., Burton J., Swerdlow H., Quail M.A., Stratton M.R., Iacobuzio-Donahue C., Futreal P.A. 2010. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature 467:1109–1113.

Carlson C.A., Kas A., Kirkwood R., Hays L.E., Preston B.D., Salipante S.J., Horwitz M.S. 2012. Decoding cell lineage from acquired mutations using arbitrary deep sequencing. Nat. Methods. 9:78–80.

Carter S.L., Cibulskis K., Helman E., McKenna A., Shen H., Zack T., Laird P.W., Onofrio R.C., Winckler W., Weir B.A., Beroukhim R., Pellman D., Levine D.A., Lander E.S., Meyerson M., Getz G. 2012. Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. 30:413–421.

Cheng Y.K., Beroukhim R., Levine R.L., Mellinghoff I.K., Holland E.C., Michor F. 2012. A mathematical methodology for determining the temporal order of pathway alterations arising during gliomagenesis. PLoS Comput. Biol. 8:e1002337.

Chmielecki J., Foo J., Oxnard G.R., Hutchinson K., Ohashi K., Somwar R., Wang L., Amato K.R., Arcila M., Sos M.L., Socci N.D., Viale A., de Stanchina E., Ginsberg M.S., Thomas R.K., Kris M.G., Inoue A., Ladanyi M., Miller V.A., Michor F., Pao W. 2011. Optimization of dosing for EGFR-mutant non-small cell lung cancer with evolutionary cancer modeling. Sci. Transl. Med. 3:90ra59.

Chowdhury S.A., Shackney S.E., Heselmeyer-Haddad K., Ried T., Schäffer A.A., Schwartz R. 2013. Phylogenetic analysis of multiprobe fluorescence in situ hybridization data from tumor cell populations. Bioinformatics 29:i189–i198.

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

Cooke S.L., Temple J., Macarthur S., Zahra M.A., Tan L.T., Crawford R.A.F., Ng C.K.Y., Jimenez-Linan M., Sala E., Brenton J.D. 2011. Intra-tumour genetic heterogeneity and poor chemoradiotherapy response in cervical cancer. Br. J. Cancer 104:361–368.

Cortes C., Haffner P., Mohri M. 2004. Rational Kernels: Theory and Algorithms. J. Mach. Lern. Res. 1:1–50.

Curtis C., Shah S.P., Chin S.-F., Turashvili G., Rueda O.M., Dunning M.J., Speed D., Lynch A.G., Samarajiwa S., Yuan Y., Gräf S., Ha G., Haffari G., Bashashati A., Russell R., McKinney S., M.E.T.A.B.R.I.C.G., Langerod A., Green A., Provenzano E., Wishart G., Pinder S., Watson P., Markowetz F., Murphy L., Ellis I., Purushotham A., Borresen-Dale A.-L., Brenton J.D., Tavaré S., Caldas C., Aparicio S. 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486:346–352.

Danesh K., Durrett R., Havrilesky L.J., Myers E. 2012. A branching process model of ovarian cancer. J. Theor. Biol. 314:10–15.

Datta R.S., Gutteridge A., Swanton C., Maley C.C., Graham T.A. 2013. Modelling the evolution of genetic instability during tumour progression. Evol. Appl. 6:20–33.

Dawson S.-J., Tsui D.W.Y., Murtaza M., Biggs H., Rueda O.M., Chin S.-F., Dunning M.J., Gale D., Forshew T., Mahler-Araujo B., Rajan S., Humphray S., Becq J., Halsall D., Wallis M., Bentley D., Caldas C., Rosenfeld N. 2013. Analysis of circulating tumor DNA to monitor metastatic breast cancer. N. Engl. J. Med. 368:1199–1209.

de Bruin E.C., Taylor T.B., Swanton C. 2013. Intra-tumor heterogeneity: lessons from microbial evolution and clinical implications. Genome Med. 5:101.

Desper R., Jiang F., Kallioniemi O.P., Moch H., Papadimitriou C.H., Schäffer A.A. 1999. Inferring tree models for oncogenesis from comparative genome hybridization data. J. Comput. Biol. 6:37–51.

Desper R., Jiang F., Kallioniemi O.P., Moch H., Papadimitriou C.H., Schaffer A.A. 2000. Distance-based reconstruction of tree models for oncogenesis. J. Comput. Biol. 7:789–803.

Deutsch A., Moreira J. 2002. Cellular automaton models of tumor development: A critical review. Ad. Complex Syst. 05:247–267.

Dewanji A., Jeon J., Meza R., Luebeck E.G. 2011. Number and size distribution of colorectal adenomas under the multistage clonal expansion model of cancer. PLoS Comput. Biol. 7:e1002213.

Dexter D.L., Kowalski H.M., Blazar B.A., Fligiel Z., Vogel R., Heppner G.H. 1978. Heterogeneity of tumor cells from a single mouse mammary tumor. Cancer Res. 38:3174–3181.

Diaz Jr L.A., Williams R.T, Wu J., Kinde I.,Hecht J.R., Berlin J., Allen B., Bozic I., Reiter J.G., Nowak M.A., Kinzler K.W., Oliner K.S., Vogelstein B. 2012. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. Nature 468: 973–977.

Dingli D., Chalub F.A.C.C., Santos F.C., Van Segbroeck S., Pacheco J.M. 2009. Cancer phenotype as the outcome of an evolutionary game between normal and malignant cells. Br. J. Cancer 101:1130–1136.

Dormann S., Deutsch A. 2002. Modeling of self-organized avascular tumor growth with a hybrid cellular automaton. In Silico Biol. 2: 393–406.

Durrett R. 2002. Probability models for DNA sequence evolution. New York (NY): Springer.

Durrett R., Foo J., Leder K., Mayberry J., Michor F. 2010. Evolutionary dynamics of tumor progression with random fitness values. Theor. Popul. Biol. 78:54–66.

Durrett R., Foo J., Leder K., Mayberry J., Michor F. 2011. Intratumor heterogeneity in evolutionary models of tumor progression. Genetics. 188:461–477.

Durrett R., Mayberry J. 2011. Traveling waves of selective sweeps. Ann. Appl. Probab. 21:699–744.

Durrett R., Moseley S. 2010. Evolution of resistance and progression to disease during clonal expansion of cancer. Theor. Popul. Biol. 77:42–48.

Durrett R., Schmidt D., Schweinsberg J. 2009. A waiting time problem arising from the study of multi-stage carcinogenesis. Ann. Appl. Probab. 19:676–718.

Edmonds J. 1967. Optimum branchings.J. Res. Nat. Bur. Stand. 71B:233–240.

Emmert-Buck M.R., Bonner R.F., Smith P.D., Chuaqui R.F., Zhuang Z., Goldstein S.R., Weiss R.A., Liotta L.A. 1996. Laser capture microdissection. Science 274:998–1001.

Ewens W.J. 2004. Mathematical population genetics, 2nd ed. (*Interdisciplinary Applied Mathematics*); Vol. 27. New York (NY): Springer.

Fearon E.R., Vogelstein B. 1990. A genetic model for colorectal tumorigenesis. Cell 61:759–767.

Feller W. 1951. Diffusion processes in genetics. In: Neyman P, ed., Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. Berkeley and Los Angeles (CA): University of California Press, pp. 227–246.

Felsenstein J. 2003. Inferring phylogenies. Sunderland (MA): Sinauer Associates.

Fidler I.J. 1978. Tumor heterogeneity and the biology of cancer invasion and metastasis. Cancer Res. 38:2651–2660.

Fischer A., Vzquez-Garca I., Illingworth C.J.R., Mustonen V. 2014. High-definition reconstruction of clonal composition in cancer. Cell Rep. 7:1740–1752.

Fisher R.A. 1922. On the dominance ratio. P. Roy. Soc. Edinb. 42:321–341.

Fisher R.A. 1930. The genetical theory of natural selection. Oxford (UK): Oxford University Press.

Forshew T., Murtaza M., Parkinson C., Gale D., Tsui D.W.Y., Kaper F., Dawson S.-J., Piskorz A.M., Jimenez-Linan M., Bentley D., Hadfield J., May A.P., Caldas C., Brenton J.D., Rosenfeld N. 2012. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci. Transl. Med. 4:136ra68.

Frank S.A. 2007. Dynamics of Cancer: Incidence, Inheritance, and Evolution. Princeton (NJ): Princeton University Press.

Frank S.A., Rosner M.R. 2012. Nonheritable cellular variability accelerates the evolutionary processes of cancer. PLoS Biol. 10:e1001296.

Frumkin D., Wasserstrom A., Itzkovitz S., Stern T., Harmelin A., Eilam R., Rechavi G., Shapiro E. 2008. Cell lineage analysis of a mouse tumor. Cancer Res. 68:5924–5931.

Fudenberg D., Nowak M.A., Taylor C., Imhof L.A. 2006. Evolutionary game dynamics in finite populations with strong selection and weak mutation. Theor. Popul. Biol. 70:352–363.

Garraway L.A., Jänne P.A. 2012. Circumventing cancer drug resistance in the era of personalized medicine. Cancer Discov. 2:214–226.

Garraway L.A., Lander E.S. 2013. Lessons from the cancer genome. Cell 153:17–37.

Gatenby R.A., Vincent T.L. 2003. An evolutionary model of carcinogenesis. Cancer Res. 63:6212–6220.

Gerlinger M., Rowan A.J., Horswell S., Larkin J., Endesfelder D., Gronroos E., Martinez P., Matthews N., Stewart A., Tarpey P., Varela I., Phillimore B., Begum S., McDonald N.Q., Butler A., Jones D., Raine K., Latimer C., Santos C.R., Nohadani M., Eklund A.C., Spencer-Dene B., Clark G., Pickering L., Stamp G., Gore M., Szallasi Z., Downward J., Futreal P.A., Swanton C. 2012. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N. Engl. J. Med. 366:883–892.

Gerstung M., Baudis M., Moch H., Beerenwinkel N. 2009. Quantifying cancer progression with conjunctive Bayesian networks. Bioinformatics 25:2809–2815.

Gerstung M., Beerenwinkel N. 2010. Waiting time models of cancer progression. Math. Pop. Stud. 17:115–135.

Gerstung M., Beisel C., Rechsteiner M., Wild P., Schraml P., Moch H., Beerenwinkel N. 2012. Reliable detection of subclonal single-nucleotide variants in tumour cell populations. Nat. Commun. 3:811.

Gerstung M., Eriksson N., Lin J., Vogelstein B., Beerenwinkel N. 2011a. The temporal order of genetic and pathway alterations in tumorigenesis. PLoS ONE 6:e27136.

Gerstung M., Nakhoul H., Beerenwinkel N. 2011b. Evolutionary games with affine fitness functions: Applications to cancer. Dynamic Games and Applications 1:370–385.

Goldie J.H., Coldman A.J. 1979. A mathematic model for relating the drug sensitivity of tumors to their spontaneous mutation rate. Cancer Treat. Rep. 63:1727–1733.

Gonzalez-Perez A., Perez-Llamas C., Deu-Pons J., Tamborero D., Schroeder M.P., Jene-Sanz A., Santos A., Lopez-Bigas N. 2013. Intogen-mutations identifies cancer drivers across tumor types. Nat. Methods 10:1081–1082.

Gould S.J., Eldredge N. 1993. Punctuated equilibrium comes of age. Nature 366:223–227.

Greaves M., Maley C.C. 2012. Clonal evolution in cancer. Nature 481:306–313.

Greenman C., Stephens P., Smith R., Dalgliesh G.L., Hunter C., Bignell G., Davies H., Teague J., Butler A., Stevens C., Edkins S., O'Meara S., Vastrik I., Schmidt E.E., Avis T., Barthorpe S., Bhamra G., Buck G., Choudhury B., Clements J., Cole J., Dicks E., Forbes S., Gray K., Halliday K., Harrison R., Hills K., Hinton J., Jenkinson A., Jones D., Menzies A., Mironenko T., Perry J., Raine K., Richardson D., Shepherd R., Small A., Tofts C., Varian J., Webb T., West S., Widaa S., Yates A., Cahill D.P., Louis D.N., Goldstraw P., Nicholson A.G., Brasseur F., Looijenga L., Weber B.L., Chiew Y.-E., DeFazio A., Greaves M.F., Green A.R., Campbell P., Birney E., Easton D.F., Chenevix-Trench G., Tan M.-H., Khoo S.K., Teh B.T., Yuen S.T., Leung S.Y., Wooster R., Futreal P.A., Stratton M.R. 2007. Patterns of somatic mutation in human cancer genomes. Nature 446:153–158.

Greenman C., Wooster R., Futreal P.A., Stratton M.R., Easton D.F. 2006. Statistical analysis of pathogenicity of somatic mutations in cancer. Genetics 173:2187–2198.

Greenman C.D., Bignell G., Butler A., Edkins S., Hinton J., Beare D., Swamy S., Santarius T., Chen L., Widaa S., Futreal P.A., Stratton M.R. 2010. PICNIC: an algorithm to predict absolute allelic copy number variation with microarray cancer data. Biostatistics 11:164–175.

Greenman C.D., Pleasance E.D., Newman S., Yang F., Fu B., Nik-Zainal S., Jones D., Lau K.W., Carter N., Edwards P.A.W., Futreal P.A., Stratton M.R., Campbell P.J. 2012. Estimation of rearrangement phylogeny for cancer genomes. Genome. Res. 22:346–361.

Guenthoer J., Diede S.J., Tanaka H., Chai X., Hsu L., Tapscott S.J., Porter P.L. 2012. Assessment of palindromes as platforms for DNA amplification in breast cancer. Genome. Res. 22:232–245.

Ha G., Roth A., Khattra J., Ho J., Yap D., Prentice L.M., Melnyk N., McPherson A., Bashashati A., Laks E., Biele J., Ding J., Le A., Rosner J., Shumansky K., Marra M.A., Gilks C.B., Huntsman D.G., McAlpine J.N., Aparicio S., Shah S.P. 2014. Titan: Inference of copy number architectures in clonal cell populations from tumor whole genome sequence data. Genome Res.

Haccou P., Jagers P., Vatutin V.A., eds. 2005. Branching processes: Variation, growth, and extinction of populations. Cambridge (UK): Cambridge University Press.

Haeno H., Iwasa Y., Michor F. 2007. The evolution of two mutations during clonal expansion. Genetics 177:2209–2221.

Hanahan D., Weinberg R.A. 2000. The hallmarks of cancer. Cell. 100:57–70.

Hanahan D., Weinberg R.A. 2011. Hallmarks of cancer: the next generation. Cell. 144:646–674.

Hastings P.J., Lupski J.R., Rosenberg S.M., Ira G. 2009. Mechanisms of change in gene copy number. Nat. Rev. Genet. 10:551–564.

Hidalgo M., Amant F., Biankin A.V., Budinsk E., Byrne A.T., Caldas C., Clarke R.B. de Jong S., Jonkers J., Mlandsmo G.M., Roman-Roman S., Seoane J., Trusolino L., Villanueva A. 2014. Patient-derived xenograft models: An emerging platform for translational cancer research. Cancer Discov. 4:998–1013.

Hjelm M., Höglund M., Lagergren J. 2006. New probabilistic network models and algorithms for oncogenesis. J. Comput. Biol. 13:853–865.

Hochberg M.E., Thomas F., Assenat E., Hibner U. 2013. Preventive evolutionary medicine of cancers. Evol. Appl. 6:134–143.

Höglund M., Gisselsson D., Mandahl N., Johansson B., Mertens F., Mitelman F., Säll T. 2001. Multivariate analyses of genomic imbalances in solid tumors reveal distinct and converging pathways of karyotypic evolution. Genes Chromosomes Cancer 31:156–171.

Hong Y.J., Marjoram P., Shibata D., Siegmund K.D. 2010. Using DNA methylation patterns to infer tumor ancestry. PLoS ONE 5:e12002.

Hou Y., Song L., Zhu P., Zhang B., Tao Y., Xu X., Li F., Wu K., Liang J., Shao D., Wu H., Ye X., Ye C., Wu R., Jian M., Chen Y., Xie W., Zhang R., Chen L., Liu X., Yao X., Zheng H., Yu C., Li Q., Gong Z., Mao M., Yang X., Yang L., Li J., Wang W., Lu Z., Gu N., Laurie G., Bolund L., Kristiansen K., Wang J., Yang H., Li Y., Zhang X., Wang J. 2012. Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. Cell 148:873–885.

Iwasa Y., Michor F., Nowak M.A. 2003. Evolutionary dynamics of escape from biomedical intervention. Proc. Biol. Sci. 270: 2573–2578.

Iwasa Y., Michor F., Nowak M.A. 2004. Evolutionary dynamics of invasion and escape. J. Theor. Biol. 226:205–214.

Iwasa Y., Nowak M.A., Michor F. 2006. Evolution of resistance during clonal expansion. Genetics 172:2557–2566.

Jeon J., Meza R., Moolgavkar S.H., Luebeck E.G. 2008. Evaluation of screening strategies for pre-malignant lesions using a biomathematical approach. Math. Biosci. 213:56–70.

Jiang F., Desper R., Papadimitriou C.H., Schffer A.A., Kallioniemi O.P., Richter J., Schraml P., Sauter G., Mihatsch M.J., Moch H. 2000. Construction of evolutionary tree models for renal cell carcinoma from comparative genomic hybridization data. Cancer Res. 60: 6503–6509.

Jiao W., Vembu S., Deshwar A.G., Stein L., Morris Q. 2014. Inferring clonal evolution of tumors from single nucleotide somatic mutations. BMC Bioinformatics 15:35.

Johnston M.D., Edwards C.M., Bodmer W.F., Maini P.K., Chapman S.J. 2007. Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer. Proc. Natl. Acad. Sci. USA 104:4008–4013.

Junttila M.R., de Sauvage F.J. 2013. Influence of tumour microenvironment heterogeneity on therapeutic response. Nature 501:346–354.

Karp R. 1971. A simple derivation of Edmonds' algorithm for optimum branching. Networks 1:265–272.

Khalique L., Ayhan A., Whittaker J.C., Singh N., Jacobs I.J., Gayther S.A., Ramus S.J. 2009. The clonal evolution of metastases from primary serous epithelial ovarian cancers. Int. J. Cancer 124: 1579–1586.

Kim K.I., Simon R. 2014. Using single cell sequencing data to model the evolutionary history of a tumor. BMC Bioinformatics 15:27.

Kimmel M., Axelrod D.E. 2002. Branching processes in biology. New York (NY): Springer.

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge (UK): Cambridge University Press.

Kingman J. 1982. The coalescent. Stoch. Proc. Appl. 13:235–248.

Klein C.A. 2013. Selection and adaptation during metastatic cancer progression. Nature 501:365–372.

Knudson A.G.J. 1971. Mutation and cancer: statistical study of retinoblastoma. Proc. Natl. Acad. Sci. USA 68:820–823.

Koboldt D.C., Zhang Q., Larson D.E., Shen D., McLellan M.D., Lin L., Miller C.A., Mardis E.R., Ding L., Wilson R.K. 2012. Varscan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 22:568–576.

Kolmogorov A.N. 1931. Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. Math. Annalen. 104:415–458.

Komarova N.L. 2006. Spatial stochastic models for cancer initiation and progression. Bull. Math. Biol. 68:1573–1599.

Komarova N.L. 2007. Loss- and gain-of-function mutations in cancer: mass-action, spatial and hierarchical models. J. Statist. Phys. 128:413–446.

Korbel J.O., Campbell P.J. 2013. Criteria for inference of chromothripsis in cancer genomes. Cell 152:1226–1236.

Kreso A., Dick J.E. 2014. Evolution of the cancer stem cell model. Cell Stem Cell 14:3.

Kumar P., Henikoff S., Ng P.C. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 4:1073–1081.

Kunkel T.A., Bebenek K. 2000. DNA replication fidelity. Annu. Rev. Biochem. 69:497–529.

Lambert G., Estvez-Salmeron L., Oh S., Liao D., Emerson B.M., Tlsty T.D., Austin R.H. 2011. An analogy between the evolution of drug resistance in bacterial communities and malignant tissues. Nat. Rev. Cancer 11:375–382.

Landau D.A., Carter S.L., Stojanov P., McKenna A., Stevenson K., Lawrence M.S., Sougnez C., Stewart C., Sivachenko A., Wang L., Wan Y., Zhang W., Shukla S.A., Vartanov A., Fernandes S.M., Saksena G., Cibulskis K., Tesar B., Gabriel S., Hacohen N., Meyerson M., Lander E.S., Neuberg D., Brown J.R., Getz G., Wu C.J. 2013. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell 152:714–726.

Landry J.J.M., Pyl P.T., Rausch T., Zichner T., Tekkedil M.M., Sttz A.M., Jauch A., Aiyar R.S., Pau G., Delhomme N., Gagneur J., Korbel J.O., Huber W., Steinmetz L.M. 2013. The genomic and transcriptomic landscape of a HeLa cell line. G3 (Bethesda) 3:1213–1224.

Larson N.B., Fridley B.L. 2013. Purbayes: estimating tumor cellularity and subclonality in next-generation sequencing data. Bioinformatics 29:1888–1889.

Lawrence M.S., Stojanov P., Mermel C.H., Robinson J.T., Garraway L.A., Golub T.R., Meyerson M., Gabriel S.B., Lander E.S., Getz G. 2014. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature. 505:495–501.

Lawrence M.S., Stojanov P., Polak P., Kryukov G.V., Cibulskis K., Sivachenko A., Carter S.L., Stewart C., Mermel C.H., Roberts S.A., Kiezun A., Hammerman P.S., McKenna A., Drier Y., Zou L., Ramos A.H., Pugh T.J., Stransky N., Helman E., Kim J., Sougnez C., Ambrogio L., Nickerson E., Shefler E., Cortes M.L., Auclair D., Saksena G., Voet D., Noble M., DiCara D., Lin P., Lichtenstein L., Heiman D.I., Fennell T., Imielinski M., Hernandez B., Hodis E., Baca S., Dulak A.M., Lohr J., Landau D.-A., Wu C.J., Melendez-Zajgla J., Hidalgo-Miranda A., Koren A., McCarroll S.A., Mora J., Lee R.S., Crompton B., Onofrio R., Parkin M., Winckler W., Ardlie K., Gabriel S.B., Roberts C.W.M., Biegel J.A., Stegmaier K., Bass A.J., Garraway L.A., Meyerson M., Golub T.R., Gordenin D.A., Sunyaev S., Lander E.S., Getz G. 2013. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 499:214–218.

Letouzé E., Allory Y., Bollet M.A., Radvanyi F., Guyon F. 2010. Analysis of the copy number profiles of several tumor samples from the same patient reveals the successive steps in tumorigenesis. Genome Biol. 11:R76.

Loeb L.A. 2001. A mutator phenotype in cancer. Cancer Res. 61: 3230–3239.

Loeb L.A. 2011. Human cancers express mutator phenotypes: origin, consequences and targeting. Nat. Rev. Cancer 11:450–457.

Luebeck E.G., Curtius K., Jeon J., Hazelton W.D. 2013. Impact of tumor progression on cancer incidence curves. Cancer Res. 73: 1086–1096.

Luebeck E.G., Moolgavkar S.H. 2002. Multistage carcinogenesis and the incidence of colorectal cancer. Proc. Natl. Acad. Sci. USA 99:15095–15100.

Luria S.E., Delbrück M. 1943. Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28:491–511.

Maley C.C., Galipeau P.C., Finley J.C., Wongsurawat V.J., Li X., Sanchez C.A., Paulson T.G., Blount P.L., Risques R.-A., Rabinovitch P.S., Reid B.J. 2006. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. Nat. Genet. 38:468–473.

Marjoram P., Molitor J., Plagnol V., Tavaré S. 2003. Markov chain Monte Carlo without likelihoods. Proc. Natl. Acad. Sci. USA 100: 15324–15328.

Marusyk A., Almendro V., Polyak K. 2012. Intra-tumour heterogeneity: a looking glass for cancer? Nat. Rev. Cancer 12:323–334.

Maynard Smith J. 1982. Evolution and the theory of games. Cambridge (UK): Cambridge University Press.
McFarland C.D., Korolev K.S., Kryukov G.V., Sunyaev S.R., Mirny L.A. 2013. Impact of deleterious passenger mutations on cancer progression. Proc. Natl. Acad. Sci. USA 110:2910–2915.
Meacham C.E., Morrison S.J. 2013. Tumour heterogeneity and cancer cell plasticity. Nature 501:328–337.
Merlo L.M.F., Pepper J.W., Reid B.J., Maley C.C. 2006. Cancer as an evolutionary and ecological process. Nat. Rev. Cancer 6:924–935.
Meza R., Jeon J., Moolgavkar S.H., Luebeck E.G. 2008. Age-specific incidence of cancer: Phases, transitions, and biological implications. Proc. Natl. Acad. Sci. USA 105:16284–16289.
Michor F., Hughes T.P., Iwasa Y., Branford S., Shah N.P., Sawyers C.L., Nowak M.A. 2005. Dynamics of chronic myeloid leukaemia. Nature 435:1267–1270.
Michor F., Iwasa Y., Nowak M.A. 2004. Dynamics of cancer progression. Nat. Rev. Cancer 4:197–205.
Michor F., Polyak K. 2010. The origins and implications of intratumor heterogeneity. Cancer Prev. Res. (Phila) 3:1361–1364.
Miller C.A., White B.S., Dees N.D., Griffith M., Welch J.S., Griffith O.L., Vij R., Tomasson M.H., Graubert T.A., Walter M.J., Ellis M.J., Schierding W., DiPersio J.F., Ley T.J., Mardis E.R., Wilson R.K., Ding L. 2014. Sciclone: Inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. PLoS Comput. Biol. 10:e1003665.
Moran P.A.P. 1958. Random processes in genetics. Math. Proc. Cambridge. 54:60–71.
Murray J.D. 2002. Mathematical biology, Vol. 2. 3rd ed. New York (NY): Springer.
Murtaza M., Dawson S.-J., Tsui D.W.Y., Gale D., Forshew T., Piskorz A.M., Parkinson C., Chin S.-F., Kingsbury Z., Wong A.S.C., Marass F., Humphray S., Hadfield J., Bentley D., Chin T.M., Brenton J.D., Caldas C., Rosenfeld N. 2013. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature 497:108–112.
Navin N., Kendall J., Troge J., Andrews P., Rodgers L., McIndoo J., Cook K., Stepansky A., Levy D., Esposito D., Muthuswamy L., Krasnitz A., McCombie W.R., Hicks J., Wigler M. 2011. Tumour evolution inferred by single-cell sequencing. Nature 472:90–94.
Navin N., Krasnitz A., Rodgers L., Cook K., Meth J., Kendall J., Riggs M., Eberling Y., Troge J., Grubor V., Levy D., Lundin P., Mnr S., Zetterberg A., Hicks J., Wigler M. 2010. Inferring tumor progression from genomic heterogeneity. Genome Res. 20:68–80.
Navin N.E., Hicks J. 2010. Tracing the tumor lineage. Mol. Oncol. 4:267–283.
Nazarian R., Shi H., Wang Q., Kong X., Koya R.C., Lee H., Chen Z., Lee M.-K., Attar N., Sazegar H., Chodon T., Nelson S.F., McArthur G., Sosman J.A., Ribas A., Lo R.S. 2010. Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. Nature 468:973–977.
Neuhauser C., Krone S.M. 1997. The genealogy of samples in models with selection. Genetics 145:519–534.
Nicolas P., Kim K.M., Shibata D., Tavar S. 2007. The stem cell population of the human colon crypt: analysis via methylation patterns. PLoS Comput. Biol. 3:e28.
Nik-Zainal S., Alexandrov L.B., Wedge D.C., Van Loo P., Greenman C.D., Raine K., Jones D., Hinton J., Marshall J., Stebbings L.A., Menzies A., Martin S., Leung K., Chen L., Leroy C., Ramakrishna M., Rance R., Lau K.W., Mudie L.J., Varela I., McBride D.J., Bignell G.R., Cooke S.L., Shlien A., Gamble J., Whitmore I., Maddison M., Tarpey P.S., Davies H.R., Papaemmanuil E., Stephens P.J., McLaren S., Butler A.P., Teague J.W., Jönsson G., Garber J.E., Silver D., Miron P., Fatima A., Boyault S., Langerod A., Tutt A., Martens J.W.M., Aparicio S.A.J.R., Borg A., Salomon A.V., Thomas G., Borresen-Dale A.-L., Richardson A.L., Neuberger M.S., Futreal P.A., Campbell P.J., Stratton M.R., Breast Cancer Working Group of the International Cancer Genome Consortium. 2012a. Mutational processes molding the genomes of 21 breast cancers. Cell. 149:979–993.
Nik-Zainal S., Van Loo P., Wedge D.C., Alexandrov L.B., Greenman C.D., Lau K.W., Raine K., Jones D., Marshall J., Ramakrishna M., Shlien A., Cooke S.L., Hinton J., Menzies A., Stebbings L.A., Leroy C., Jia M., Rance R., Mudie L.J., Gamble S.J., Stephens P.J., McLaren S., Tarpey P.S., Papaemmanuil E., Davies H.R., Varela I., McBride D.J., Bignell G.R., Leung K., Butler A.P., Teague J.W., Martin S., Jönsson G., Mariani O., Boyault S., Miron P., Fatima A., Langerod A., Aparicio S.A.J.R., Tutt A., Sieuwerts A.M., Borg A., Thomas G., Salomon A.V., Richardson A.L., Borresen-Dale A.-L., Futreal P.A., Stratton M.R., Campbell P.J., Breast Cancer Working Group of the International Cancer Genome Consortium. 2012b. The life history of 21 breast cancers. Cell 149:994–1007.
Nolan-Stevaux O., Tedesco D., Ragan S., Makhanov M., Chenchik A., Ruefli-Brasse A., Quon K., Kassner P.D. 2013. Measurement of cancer cell growth heterogeneity through lentiviral barcoding identifies clonal dominance as a characteristic of in vivo tumor engraftment. PLoS ONE 8:e67316.
Nordling CO. 1953. A new theory on cancer-inducing mechanism. Br. J. Cancer 7:68–72.
Norquist B., Wurz K.A., Pennil C.C., Garcia R., Gross J., Sakai W., Karlan B.Y., Taniguchi T., Swisher E.M. 2011. Secondary somatic mutations restoring BRCA1/2 predict chemotherapy resistance in hereditary ovarian carcinomas. J. Clin. Oncol. 29:3008–3015.
Nowak M.A. 2006a. Evolutionary dynamics: exploring the equations of life. Cambridge (MA): Belknap Press of Harvard University Press.
Nowak M.A. 2006b. Five rules for the evolution of cooperation. Science 314:1560–1563.
Nowak M.A., Michor F., Iwasa Y. 2003. The linear process of somatic evolution. Proc. Natl. Acad. Sci. USA 100:14966–14969.
Nowell P.C. 1976. The clonal evolution of tumor cell populations. Science 194:23–28.
Oesper L., Mahmoody A., Raphael B. 2013. Theta: Inferring intra-tumor heterogeneity from high-throughput DNA sequencing data. Genome Biol. 14:R80.
Orr H.A. 2005. The probability of parallel evolution. Evolution: Int. J. Org. Evolut. 59:216–220.
Otto S.P., Whitlock M.C. 2001. Fixation probabilities and times, John Wiley & Sons, Ltd.
Owen M.R., Alarcón T., Maini P.K., Byrne H.M. 2009. Angiogenesis and vascular remodelling in normal and cancerous tissues. J. Math. Biol. 58:689–721.
Papaemmanuil E., Gerstung M., Malcovati L., Tauro S., Gundem G., Van Loo P., Yoon C.J., Ellis P., Wedge D.C., Pellagatti A., Shlien A., Groves M.J., Forbes S.A., Raine K., Hinton J., Mudie L.J., McLaren S., Hardy C., Latimer C., Della Porta M.G., O'Meara S., Ambaglio I., Galli A., Butler A.P., Walldin G., Teague J.W. Quek L., Sternberg A., Gambacorti-Passerini C., Cross N.C.P. Green A.R., Boultwood J., Vyas P., Hellstrom-Lindberg E., Bowen D., Cazzola M., Stratton M.R., Campbell P.J., Chronic Myeloid Disorders working group of the International Cancer Genome Consortium. 2013. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood 122:3616–3627.
Park S.C., Simon D., Krug J. 2010. The speed of evolution in large asexual populations. J. Stat. Phys. 138:381–410.
Perfahl Ho., Byrne H.M., Chen T., Estrella V., Alarcon T., Lapin A., Gatenby R.A., Gillies R.J., Lloyd M.C., Maini P.K., Reuss M., Owen M.R. 2011. Multiscale modelling of vascular tumour growth in 3D: the roles of domain size and boundary conditions. PLoS ONE 6:e14790.
Pharoah P.D.P., Dunning A.M., Ponder B.A.J., Easton D.F. 2004. Association studies for finding cancer-susceptibility genetic variants. Nat. Rev. Cancer 4:850–860.
Podlaha O., Riester M., De S., Michor F. 2012. Evolution of the cancer genome. Trends Genet. 28:155–163.
Poste G., Doll J., Fidler I.J. 1981. Interactions among clonal subpopulations affect stability of the metastatic phenotype in polyclonal populations of B16 melanoma cells. Proc. Natl. Acad. Sci. USA 78:6226–6230.
Potter N.E., Ermini L., Papaemmanuil E., Cazzaniga G., Vijayaraghavan G., Titley I., Ford A., Campbell P., Kearney L., Greaves M. 2013. Single-cell mutational profiling and clonal phylogeny in cancer. Genome Res. 23:2115–2125.
Purdom E., Ho C., Grasso C.S., Quist M.J., Cho R.J., Spellman P. 2013. Methods and challenges in timing chromosomal abnormalities within cancer samples. Bioinformatics 29:3113–3120.
Radmacher M.D., Simon R., Desper R., Taetle R., Schäffer A.A., Nelson M.A. 2001. Graph models of oncogenesis with an application to melanoma. J. Theor. Biol. 212:535–548.

Rahnenführer J., Beerenwinkel N., Schulz W.A., Hartmann C., von Deimling A., Wullich B., Lengauer T. 2005. Estimating cancer survival and clinical outcome based on genetic tumor progression scores. Bioinformatics 21:2438–2446.

Roose T., Chapman S., Maini P. 2007. Mathematical models of avascular tumor growth. SIAM Review 49:179–208.

Rosenberg N.A., Nordborg M. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nat. Rev. Genet. 3:380–390.

Roth A., Khattra J., Yap D., Wan A., Justina Biele E.L., Ha G., Aparicio S., Bouchard-Cote A., Shah S.P. 2014. Pyclone: statistical inference of clonal population structure in cancer. Nature Methods. Published online 16 March 2014.

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

Sakoparnig T., Beerenwinkel N. 2012. Efficient sampling for Bayesian inference of conjunctive Bayesian networks. Bioinformatics 28:2318–2324.

Salk J.J., Fox E.J., Loeb L.A. 2010. Mutational heterogeneity in human cancers: origin and consequences. Annu. Rev. Pathol. 5:51–75.

Schöllnberger H., Beerenwinkel N., Hoogenveen R., Vineis P. 2010. Cell selection as driving force in lung and colon carcinogenesis. Cancer Res. 70:6797–6803.

Schuster P., Sigmund K. 1983. Replicator dynamics. J. Theor. Biol. 100:533–538.

Schwarz R.F., Fletcher W., Förster F., Merget B., Wolf M., Schultz J., Markowetz F. 2010. Evolutionary distances in the twilight zone–a rational kernel approach. PLoS ONE 5:e15788.

Schwarz R.F., Ng C.K., Coooke S.L., Newman S., Temple J., Piskorz A.M., Gale D., Sayal K., Murtaza M., Baldwin P.J., Rosenfeld N., Earl H.M., Sala E., Jimenez-Linan M., Parkinson C.A., Markowetz F., Brenton J.D. 2014a. Phylogenetic quantification of intra-tumor heterogeneity predicts time to relapse in high-grade serous ovarian cancer. PLoS Medicine (in revision).

Schwarz R.F., Trinh A., Sipos B., Brenton J.D., Goldman N., Markowetz F. 2014b. Phylogenetic quantification of intra-tumour heterogeneity. PLoS Comput. Biol. 10:e1003535.

Shah S.P., Roth A., Goya R., Oloumi A., Ha G., Zhao Y., Turashvili G., Ding J., Tse K., Haffari G., Bashashati A., Prentice L.M., Khattra J., Burleigh A., Yap D., Bernard V., McPherson A., Shumansky K., Crisan A., Giuliany R., Heravi-Moussavi A., Rosner J., Lai D., Birol I., Varhol R., Tam A., Dhalla N., Zeng T., Ma K., Chan S.K., Griffith M., Moradian A., Cheng S.-W.G., Morin G.B., Watson P., Gelmon K., Chia S., Chin S.-F., Curtis C., Rueda O.M., Pharoah P.D., Damaraju S., Mackey J., Hoon K., Harkins T., Tadigotla V., Sigaroudinia M., Gascard P., Tlsty T., Costello J.F., Meyer I.M., Eaves C.J., Wasserman W.W., Jones S., Huntsman D., Hirst M., Caldas C., Marra M.A., Aparicio S. 2012. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 486:395–399.

Shahrabi Farahani H., Lagergren J. 2013. Learning oncogenetic networks by reducing to MILP. PLoS ONE 8:e65773.

Shapiro E., Biezuner T., Linnarsson S. 2013. Single-cell sequencing-based technologies will revolutionize whole-organism science. Nat. Rev. Genet. 14:618–630.

Shen M.M. 2013. Chromoplexy: a new category of complex rearrangements in the cancer genome. Cancer Cell 23:567–569.

Solé R.V., Deisboeck T.S. 2004. An error catastrophe in cancer? J. Theor. Biol. 228:47–54.

Sottoriva A., Spiteri I., Piccirillo S.G.M., Touloumis A., Collins V.P., Marioni J.C., Curtis C., Watts C., Tavar S. 2013a. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. Proc. Natl. Acad. Sci. USA 110:4009–4014.

Sottoriva A., Spiteri I., Shibata D., Curtis C., Tavar S. 2013b. Single-molecule genomic data delineate patient-specific tumor profiles and cancer stem cell organization. Cancer Res. 73:41–49.

Sprouffske K., Pepper J.W., Maley C.C. 2011. Accurate reconstruction of the temporal order of mutations in neoplastic progression. Cancer Prev. Res. (Phila) 4:1135–1144.

Stadler P.F. 1991. Dynamics of autocatalytic reaction networks. IV: Inhomogeneous replicator networks. Biosystems 26:1–19.

Stephens P.J., Greenman C.D., Fu B., Yang F., Bignell G.R., Mudie L.J., Pleasance E.D., Lau K.W., Beare D., Stebbings L.A., McLaren S., Lin M.-L., McBride D.J., Varela I., Nik-Zainal S., Leroy C., Jia M., Menzies A., Butler A.P., Teague J.W., Quail M.A., Burton J., Swerdlow H., Carter N.P., Morsberger L.A., Iacobuzio-Donahue C., Follows G.A., Green A.R., Flanagan A.M., Stratton M.R., Futreal P.A., Campbell P.J. 2011. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell. 144: 27–40.

Stratton M.R., Campbell P.J., Futreal P.A. 2009. The cancer genome. Nature 458:719–724.

Strino F., Parisi F., Micsinai M., Kluger Y. 2013. Trap: a tree approach for fingerprinting subclonal tumor composition. Nucleic Acids Res. 41:e165

Szabo A., Boucher K. 2002. Estimating an oncogenetic tree when false negatives and positives are present. Math. Biosci. 176: 219–236.

Taylor C., Fudenberg D., Sasaki A., Nowak M.A. 2004. Evolutionary game dynamics in finite populations. Bull. Math. Biol. 66: 1621–1644.

Thalhauser C.J., Lowengrub J.S., Stupack D., Komarova N.L. 2010. Selection in spatial stochastic models of cancer: migration as a key modulator of fitness. Biol. Direct. 5:21.

Tofigh A., Sjölund E., Höglund M., Lagergren J. 2011. A global structural EM algorithm for a model of cancer progression. In: Shawe-Taylor J., Zemel R., Bartlett P., Pereira F., Weinberger K., editors. Advances in neural information processing systems 24. pp. 163–171. Red Hook (NY): Curran Associates, Inc.

Tomasetti C., Vogelstein B., Parmigiani G. 2013. Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. Proc. Natl. Acad. Sci. USA 110:1999–2004.

Tomlinson I.P. 1997. Game-theory models of interactions between tumour cells. Eur. J. Cancer 33:1495–1500.

Trinh A., Rye I.H., Almendro V., Helland A., Russnes H.G., Markowetz F. 2014. Goifish: a system for the quantification of single cell heterogeneity from ifish images. Genome Biol. 15:442.

Van Loo P., Nordgard S.H., Lingjærde O.C., Russnes H.G., Rye I.H., Sun W., Weigman V.J., Marynen P., Zetterberg A., Naume B., Perou C.M., Børresen-Dale A.-L., Kristensen V.N. 2010. Allele-specific copy number analysis of tumors. Proc. Natl. Acad. Sci. USA 107: 16910–16915.

Vogelstein B., Papadopoulos N., Velculescu V.E., Zhou S., Diaz L.A. Jr, Kinzler K.W. 2013. Cancer genome landscapes. Science 339: 1546–1558.

von Heydebreck A., Gunawan B., Fuzesi L. 2004. Maximum likelihood estimation of oncogenetic tree models. Biostatistics 5:545–556.

Weinberg R.A. 2013. The Biology of Cancer. 2nd ed. New York (NY): Garland Science.

Weinreich D.M., Delaney N.F., Depristo M.A., Hartl D.L. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 312:111–114.

Werner B., Dingli D., Traulsen A. 2013. A deterministic model for the occurrence and dynamics of multiple mutations in hierarchically organized tissues. J. R. Soc. Interface 10:20130349.

Wilm A., Aw P.P.K., Bertrand D., Yeo G.H.T., Ong S.H., Wong C.H., Khor C.C., Petric R., Hibberd M.L., Nagarajan N. 2012. Lofreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 40:11189–11201.

Wright S. 1931. Evolution in Mendelian populations. Genetics 16: 97–159.

Wright S. 1945. The differential equation of the distribution of gene frequencies. Proc. Natl. Acad. Sci. USA 31:382–389.

Xu X., Hou Y., Yin X., Bao L., Tang A., Song L., Li F., Tsang S., Wu K., Wu H., He W., Zeng L., Xing M., Wu R., Jiang H., Liu X., Cao D., Guo G., Hu X., Gui Y., Li Z., Xie W., Sun X., Shi M., Cai Z., Wang B., Zhong M., Li J., Lu Z., Gu N., Zhang X., Goodman L., Bolund L., Wang J., Yang H., Kristiansen K., Dean M., Li Y., Wang J. 2012. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell 148:886–895.

Youn A., Simon R. 2012. Estimating the order of mutations during tumorigenesis from tumor genome sequencing data. Bioinformatics 28:1555–1561.

Yuan Y., Failmezger H., Rueda O.M., Ali H.R., Gräf S., Chin S.-F., Schwarz R.F., Curtis C., Dunning M.J., Bardwell H., Johnson N., Doyle S., Turashvili G., Provenzano E., Aparicio S., Caldas C.,

Markowetz F. 2012. Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. Sci. Transl. Med. 4:157ra143.

Zare H., Wang J., Hu A., Weber K., Smith J., Nickerson D., Song C., Witten D., Blau C.A., Noble W.S. 2014. Inferring clonal composition from multiple sections of a breast cancer. PLoS Comput. Biol. 10:e1003703.

Zhao R., Michor F. 2013. Patterns of proliferative activity in the colonic crypt determine crypt stability and rates of somatic evolution. PLoS Comput. Biol. 9:e1003082.