Chapter 3
Population Genetics of Neoplasms

Andrea Sottoriva and Simon Tavaré

Abstract Cancer is a complex disease of the genome that arises from the interplay of numerous underlying biological processes occurring within and between cells. Cancer population genetics aims at investigating malignant dynamics by studying the distribution of somatic alterations in cancer cell populations. Such aberrant DNA modifications lead to the development of cellular malignant traits like cancer invasion, metastasis and therapy resistance. The study of population genetics in neoplasms integrates mathematical modeling of evolving populations with molecular and epidemiological cancer data. The goal is to infer fundamental properties of tumors and predict the progression of the disease. With the ever-growing amount of data produced by genomic techniques, cancer population genetics represents a quantitative tool to begin making sense to this massive amount of information.

Keywords Mathematical modelling • Population genetics • Cancer evolution • Statistical inference

3.1 The Importance of Modeling the Population Genetics of Tumors

Mathematical population genetics, pioneered by Sewall Wright [1, 2], Ronald Fisher [3] and Patrick Moran [4], originated from the study of gene frequencies in populations of individuals to address questions about the evolution of species and their genotypes. Neoplasms can be viewed as populations of cancer cells that

A. Sottoriva (✉)
Department of Preventive Medicine, Keck School of Medicine of the University of Southern California, Los Angeles, CA 90089, USA

Norris Comprehensive Cancer Center, Keck School of Medicine of the University of Southern California, Los Angeles, CA 90089, USA
e-mail: Andrea.Sottoriva@icr.ac.uk

S. Tavaré
Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK

C. Maley, M. Greaves (eds.), *Frontiers in Cancer Research*,
DOI 10.1007/978-1-4939-6460-4_3

Table 3.1 Terminology: standard population genetics vs. population genetics of neoplasms

Standard population genetics	Population genetics of neoplasms
Individual	Cancer cell
Species	Clone
Reproduction	Mitosis
Fertilization, recombination	Somatic mutation, mitotic recombination
Speciation	Appearance of a new clone
Species evolution	Tumor progression
Natural environment	Tumor microenvironment
Selective pressure from environment, predators, etc.	Selective pressure from normal cells, other cancer clones, immune system, therapy, etc.

undergo reproduction and death, and that are subject to evolutionary forces such as selection and drift [5–7]. In the cancer scenario however, no sexual reproduction occurs. Instead, changes in the genetic heritage of cells depend only on errors occurring during DNA replication, known as somatic mutations. Hence, due to the analogy between the evolution of species and the evolution of cancer (see Table 3.1), approaches from mathematical population genetics are also employed in the study of neoplasms, for instance to reconstruct the phylogenetic history of tumors [8].

Evolutionary forces drive the accumulation of somatic mutations in cancer cells that lead to the transformation of normal tissue into a tumor, a highly dynamical process referred to as *tumorigenesis*. Genetically, tumorigenesis is disseminated by events that change the genome of cells, such as point mutations, copy number alterations, chromosomal translocations and aneuploidy [9]. The ultimate product of those events is the development of malignant traits, such as tissue invasion and metastasis [10]. Those complex cancer processes can be addressed using population genetics approaches with the aim of bringing order to apparent chaos. This is important given the observed complexity of cancer mutations and their related, affected pathways [9]. Even more crucial, if we consider the uniqueness of each single cancer identified by multiple large-cohort studies [11–14], this calls for a shift towards personalized cancer medicine.

3.2 Background and Previous Work

3.2.1 Modeling Tumorigenesis

Cancer initiation and progression can be represented as a branching process driven by cell division, cell death and mutation. Within normal tissues, a mutation can be introduced at cell division and can be advantageous, disadvantageous or neutral for the newborn cell. Mutations that yield a selective advantage tend to spread and become fixated within a population. Aberrations in crucial cancer genes [15], such as TP53, RAS or PTEN, provide selective advantage [16] through the increase of the net proliferation rate of cells and therefore tend to promote the outgrowth of

malignant traits and consequently drive tumor progression following patterns of clonal evolution [5–7]. These mutations are often referred to as *drivers*, in contrast to mutations that do not alter the cell behavior, called *passengers*. Driver genes are categorized into two subgroups: oncogenes that can become oncogenic upon mutation [17], and tumor-suppressor genes that protect cells from becoming malignant and must be *inactivated* by a mutation to lead to a cancer phenotype [18]. A third category is referred to as *mutator* genes, or genes that, upon a certain genomic alteration, in turn alter the frequency of generation of new mutations in other genes. For instance, if a mutation affects a gene involved in the repair of DNA alterations, such as BRCA1 [19], it will not have a direct effect on the phenotype of the cell, but it will induce the cell to accumulate further mutations faster, thus speeding up the process of malignant transformation. This phenomenon is often referred to as the *muator* phenotype [20].

The investigation of normal tissues from which the cancer originates and the different pre-malignant stages is extremely important in cancer research to understand how cancer originates and advances through the accumulation of more and more aberrant genes. In some cancers, such as colorectal carcinoma, tumorigenesis has been shown to be a step-wise process that transforms a normal colon epithelium into an adenoma and finally into a carcinoma, following the so-called model of multistage carcinogenesis proposed by Vogelstein and co-workers [21]. Each stage corresponds to the acquisition of a mutation in a cancer gene, until the right combination of multiple alterations induces a fully blown colorectal carcinoma (Fig. 3.1). Both oncogenes and tumor suppressor genes take part in the step-wise process of malignant transformation.

Importantly, this sequential progression model can be matched to colon cancer incidence data using population genetics approaches, as demonstrated by Luebeck and colleagues [22]. Moreover, analyses by Nowak and co-authors [23–26] have modeled the mechanics behind the processes of fixation of mutations in tumors, with focus on the different dynamics of oncogenes and tumor suppressor genes. The recent advent of next-generation sequencing [27] has shed new light on the landscape of mutations in tumors. Based on the mutational data by Sjöblom et al. [14], Beerenwinkel and colleagues developed a mathematical model to explore the parameters of tumorigenesis and calculate the expected waiting time to cancer, reporting an average selective advantage of driver mutations in the order of 1 % under a normal mutation rate [28]. Furthermore, Bozic and co-authors [29] introduced a model of tumor progression to predict the number of passengers in relation to drivers reported by sequencing studies and concluded that very small selective advantages, on the order of 10^{-3}, are sufficient to drive tumorigenesis.

Once the primary malignancy is established, the cancer is believed to advance to later stages in which it becomes capable of metastasis by colonizing distant sites. Nonetheless, it is not clear if metastatic cancer cells arise early or late during cancer formation [30, 31]. Using population genetics applied to pancreatic cancer sequencing data, Yachida and co-workers [32] predicted a decade of time between the first cancer cell and the rising of the metastatic clone, thus suggesting the need of further mutational steps after the seeding of the malignancy in this type of tumor.

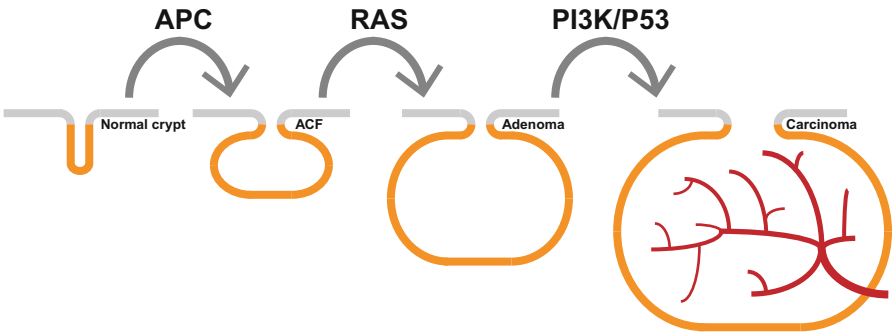

Fig. 3.1 The step-wise model of tumorigenesis in the colon. In the colon, the formation of cancers is often driven by a step-wise accumulation of multiple malignant alterations. In this case a normal crypt is transformed into an ACF (Aberrant Crypt Foci) through the silencing of tumor suppressor gene APC. In turn, an ACF becomes an adenoma via an oncogenic mutation of the RAS oncogene. Finally a carcinoma arises from further alterations in the P53 or PI3K pathways

Finally however, it is important to note that the step-wise model of carcinogenesis is not fully understood and, most importantly, it does not describe all the scenarios in which tumors arise. As we will discuss in the next session, it should therefore be used with caution as it may be misleading in certain contexts [33].

3.2.2 Cancer as a Complex System

Cancer is the result of billions of highly dynamical non-linear interactions between billions of cells and their surrounding microenvironment. It follows that a malignancy is more than the mere sum of its parts; in other words, it is *a complex system* [34]. To study such objects it is fundamental to employ models that are able to handle the countless interactions occurring in tumors. With recent advances, models of tumor growth and development have become more and more sophisticated, incorporating the simulation of complex malignant processes, such as invasion [35–37], cancer stem cell dynamics [38], angiogenesis [39] and therapeutic response [38, 40]. The enormous advances in digitalizing disease information, as well as the development of ever more detailed models of tumor dynamics, has led to formulation of a new computationally oriented cancer field referred to as *mathematical oncology* [41]. To date, the most challenging task for this field is the integration of population genetics approaches, and more generally mathematical and computational models, with cancer data.

At the genomic level, one of the most important characteristics of cancers that has been recently exposed is intra-tumor heterogeneity (ITH). This feature refers to the fact that neoplasms are not composed of a set of uniform cells, but rather of a

highly diverse and dynamical population with different phenotypic and genotypic traits in constant competition for space and resources. This finding is rather unsurprising as it naturally follows from the clonal evolution model of malignancies [7]. Indeed, this property was already known at the histopathological level [42], nevertheless only recently has genomic data demonstrated that this is a common features of tumors [43–46] and that it predicts branching evolutionary patterns [47–49], in contrast to the sequential progression model described above. The complexity revealed by these studies further indicates the need for the use of modeling to tackle cancer evolutionary dynamics.

3.2.3 Fitting Multi-Scale Models to Genomic Data

Whether drivers or passengers, mutations are the result of the underlying biological processes that occurred in normal and malignant tissues. Therefore, mutations in a cell record its past mitotic history. Moreover, different cellular dynamics (i.e. cellular organization, mutation rate, etc.) generate different mutational patterns that can be used to infer the characteristics of such cellular processes. The landscape of genomic alterations in cells can therefore be seen as a *molecular clock* that describes the history of cell lineages and can be used not only to reconstruct phylogenetic trees of cell populations, but also to deconvolute tissue dynamics via the mutational signatures it contains. Within the context of the normal colon crypt, which is the proliferative unit in the colon and at the origin of colorectal cancer, molecular clock data can be coupled with modeling to infer fundamental properties of human crypts, such as the number of stem cells responsible for the crypt cell renewal [50, 51]. This approach can be extended to populations of cancer cells to study the dynamics of growth and clonal expansion [52, 53].

A new *in silico* framework to model cancer systems has been proposed that integrates multi-scale spatial modeling of single cells with single-molecule genomic data, with the aim of inferring properties of malignancies [54]. The computational model represents the *unknowns* of the biological system (the tumor) as parameters to be probed using a method called Approximate Bayesian Computation (ABC) [55]. ABC is a statistical inference technique that can be used when likelihoods are incomputable, as for agent-based models [51]. This framework (Fig. 3.2) allows computation of posterior probability distributions of the parameters, given the genomic data observed. This corresponds to performing an *in vivo* indirect measurement of experimental parameters in human systems that would be otherwise impossible in wet labs. This new framework offers a computational-oriented method, complementary to biological and clinical experimentation, to study malignant and pre-malignant systems directly in humans, using molecular data directly sampled from patients.

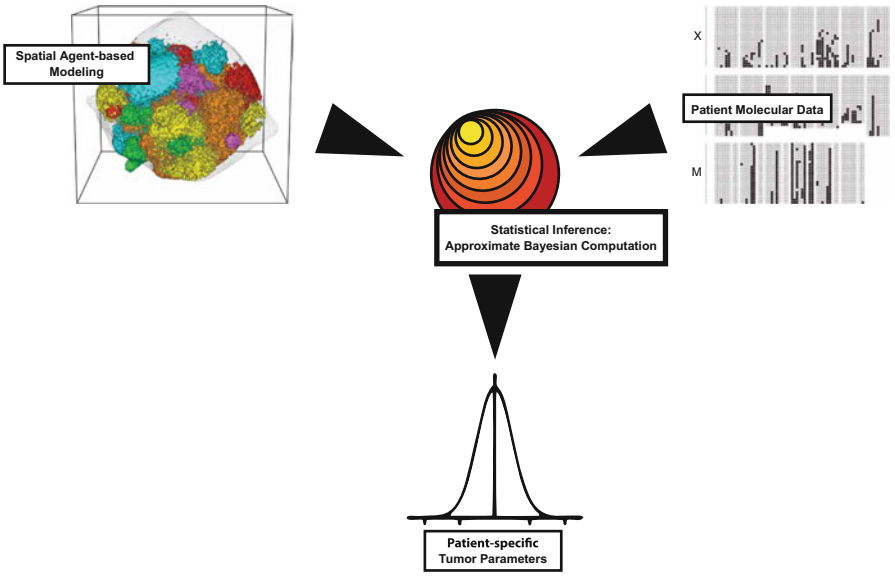

Fig. 3.2 Statistical inference framework on cancer genomic data. It is possible to integrate spatial agent-based computational models with cancer molecular data from a patient using a type of Bayesian statistical inference names Approximate Bayesian Computation (ABC). This allows estimating patient-specific tumor parameters, corresponding to an indirect measurement of cancer characteristics directly from the patient molecular profile

3.3 Short-Term Open Questions

The overwhelming amount of genomic data has shed new light on our understanding of malignant processes. At the same time this flow of information is revealing an ever more striking complexity of tumors and of biological systems in general. With the increase in efficiency, throughput and precision, we will soon start to routinely sequence single cells, thousands of times for each tumor. So far, the more we have descended into smaller and more precise scales, the more the results have been difficult to interpret. Just by observing and analyzing the data it has becomes almost impossible to make sense of the underlying biology, simply because of the complexity of the processes that generate such data. Often validation of model parameters and predictions in cancer has relied on qualitative estimates and approximate prediction from experimental observations. Improving this approach by coupling and validating models to cancer molecular information is a necessary step, yet it can be extremely challenging.

The most important task is for the modeling to keep up with the scale of these data. This is not always done by incorporating more complexity, but rather by designing models to be integrated with data in the first place. Hence a data-driven modeling perspective is necessary to solve the open questions in cancer dynamics and most importantly, to shift the field towards more quantitative methods. One of the major challenges in this task is that genomic data derive from sampling of neoplasms often made up of several billion cells. On top of that, due to intra-tumor heterogeneity,

genomic data can contain detailed information about sub-populations of cells, or even single cells. Hence there is the need to model the system in a multi-scale fashion, from the very small to the very large scale, in order to mirror the information embedded in genomic data. To date, computational models have simulated a relatively small number, on the order of 100,000 s, of cells and their interactions. This is orders of magnitude less than the number of cells in a 7 cm-diameter tumor, containing >100 billion cells. Moreover, to investigate the parameter space of malignant processes it is necessary to run a single 100-billion cell simulation for tens of millions of times, thus yielding an apparently unmanageable $10^{11} \times 10^{7} = 10^{18}$ computational problem. To tackle this, besides efficient coding, there are two crucial principles to take into consideration. The first is to identify the most important driving processes that will need to be simulated, at the same time approximating those processes that are believed to be secondary or for which no information is contained in the data. For instance, if the position of multiple samples from a tumor is recorded, that information must be modeled. However, when single cell positions within a sample are unknown, it can be neglected in the simulation as well. Thus it is important to tightly couple the level of detail of the simulation to the specific experimental settings and the type of data. As it is impossible to simulate every known malignant process, it important to identify the central mechanisms we want to investigate. Here we will illustrate a few examples of fundamental processes that need to be taken into the account:

1. Cell division. The most important process in cancer, uncontrolled proliferation ultimately leads this disease to be fatal and therefore must be modeled accurately.
2. Mutations. Cell division is accompanied by genomic alterations that are also the primary source of molecular data we currently collect on the behavior of cancer cells. The mutation process therefore needs to be simulated with high precision.
3. Cell hierarchical structure. Accumulating evidence indicate that tumors are organized into a hierarchy of cells, similarly to normal tissues, in which small groups of so-called cancer stem cells have stem-like abilities, such as the capacity of self-renew and giving rise to differentiated progeny. The bulk of the tumor is instead composed of cells with limited replicative potential [56]. The hierarchical cellular structure of malignancies is therefore another fundamental mechanism that needs to be taken into consideration -- it has remarkable effects on the evolutionary dynamics of the tumor [38, 57].
4. Intra-tumor heterogeneity. Tumors are not a homogeneous population of cells but rather a mixture of sub-clones with different phenotypic and genotypic characteristics. Such variation is an important feature of tumors and it is responsible for the failure of therapy and the development of treatment resistance [7]. This variation is complex and involves the acquisition of different traits from different groups of cancer cells, their interaction with one other and with the microenvironment. Due to the large number of variables involved in this mechanism, this remains one of the most complex to model.

Although we may have multiple samples from the same malignancy, the genomic material in most cases refers to a small sub-population of sampled cells in the original neoplasm, for instance a section or a core-punch. This means that, although a

sample should be representative of the neoplasm, only a small part of the tumor is actually examined. Thus it is not useful to simulate in full detail the behavior of billions of cells that may never be sampled [54].

Furthermore, achieving the right level of detail of a simulation must be complemented with the development of new algorithms for statistical inference to integrate with the raising complexity of mathematical models. This introduces problems related to parallelization of inference methods and the use of adaptive approaches to more efficiently explore the parameter space [58].

3.4 Long-Term Open Questions

We are in need of comprehensive computational frameworks in which we can make predictions and test hypotheses. Evolutionary cancer biology provides the underlying theoretical paradigm within which to develop the models and the mathematical tools that will allow tackling the apparent chaos reported by genomic data. As we have previously discussed, the field of population genetics of neoplasms incorporates a broad spectrum of mathematical models and statistical techniques for inference from molecular data. All these concepts have existed in separate disciplines for a long time; only recently have they been assembled together and applied to cancer research. This paradigm has already shown the potential to become an important instrument in cancer investigation. Nevertheless it needs substantial further development to be fruitfully integrated with modern genomic data produced by next generation sequencing.

It appears that in the field of cancer genomics, theory does not keep up with data. In general models of biological processes exist, but they are fragmented into compartmental boxes describing different phenomena. For example protein-protein interactions, RNA transcription and cellular signaling are all events for which a substantial amount of information is known. However, although in reality these processes are highly interconnected, a theoretical framework that links them all does not yet exist. Thus new findings do not refer to a general model of reference in which results could be easily tested, as it is for the standard model in physics. Thus, a paradigm founded on computation and mathematical representation of biological processes is a long-term goal in the field of evolutionary cancer biology.

3.5 Current Obstacles to Progress

To predict and understand the behavior of a complex disease like cancer we need to model the interaction among its many underlying mechanisms. This implies a substantial increase in the complexity of the models we design. In this scenario we encounter the risk of the models becoming overly complicated and unmanageable

both from a computational and experimental point of view. One of the chief challenges in any model is to define a large body of reliable parameters, and eventually infer a few unknown ones. However, quantitative measurements in biological systems are often hard to obtain due to the variability of the conditions and the limitation of the technologies. Cancer cells show also a remarkably different behavior for different patients, different stages of a single lesion and even within the same tumor at a given time. On top of these issues, the measurement of parameters *in vivo* can be particularly difficult if not impossible in humans. Relying on animal models or cell cultures may represent a solution in some cases, but it is still necessary to bear in mind that those models do not necessarily reproduce the behaviors of the malignancy in humans.

Besides parameterization, the growing complexity of models could introduce important computational limitations. To quantitatively infer unknown biological parameters it is necessary to use statistical inference techniques that in turn necessitate of a large number of simulation instances to produce reliable posterior distributions. For multi-scale models for which it may take hours to run a simulation this may signify the impossibility of doing any inference with them. Another obstacle to addressing questions in cancer research is that often, molecular data have not been collected for the purpose of performing modeling. The whole research plan may assume modeling as an auxiliary tool to extract sensible information out of large-scale genomic data. This limits the modeling with additional assumptions that must be made in the design of the mathematical model and often the difficulty of collecting further information about the biological system that may be useful for modeling purposes only.

On top of these issues there is the *fragmentation* of our understanding of different biological processes. This reflects the long-standing molecular reductionist approach to biology which, although tremendously successful in tackling some complex problems, shows frustrating limitations [41]. At the same time, the models are also fragmented, as a common collaborative effort to model cancer as a whole is not yet present.

3.6 Overcoming the Obstacles to Progress

To prevent overly complex models it is necessary to integrate the large set of known mechanisms occurring in the disease, but in doing so, to simplify each single mechanism to model their global emergent behavior while neglecting unnecessary complexity. This is a general principle of any complex system model and it has begun to be applied also in mathematical oncology [31]. Pursuing the simulation of several simple interlinked processes while paying attention to maintain the dependencies among them also reduces the problems of parameterization. Integrating the simulation of multiple biological processes (e.g. different genomic alterations, cancer stem cell hierarchy, intra-tumor heterogeneity) within one coherent mathematical framework is not a simple task, yet it appears now to be a necessary step to position population genetics of neoplasms as a valid complementary research approach in cancer.

In order to overcome the issues related to the design of data collection, we propose that molecular data acquisition should be more model-oriented and hence performed while keeping an eye on the modeling requirements and limitations. So the paradigm *inverts* a modeling-aided data analysis of molecular information to experiment-aided modeling and simulation of cancer processes. This implies the need to define experiments and data acquisition in a manner that is sensible both in terms of modeling and in terms of biological and clinical significance.

Giving shape to a theoretical reference framework is of great importance. This will require a joint effort from different fields such as evolutionary biology, mathematics, computer science, genetics and oncology.

3.7 Conclusion

Population genetics applied to neoplasms is a broad research field that integrates many radically different techniques such as cancer biology, evolutionary theory, mathematical modeling and statistical inference methods. In the coming years, all these subjects should be assembled to define a coherent and powerful set of tools to investigate the progression and development of cancer. Cancer population genetics is a new yet very promising approach for the analysis of cancer data and for the understanding of cancer as a complex system. This paradigm is also heavily quantitative and focused on extracting measurements from cancer data in order to reveal the dynamics of malignant processes that are today largely unknown. Placing all these concepts, originating from different fields of science, within one coherent mathematical framework is no easy task, however the potential of this technique is large and there is huge space for improvement and for the development of new methods to perform *in silico* cancer research.

Another very strong asset of cancer population genetics, integrated with statistical inference, is the ability to infer biological parameters indirectly via a computational-based analysis. These results are extremely important in cancer research because they allow the investigation of cancer in humans, a task that is often difficult for wet labs for ethical reasons. We predict important new developments in this field driven by the design of new models and the huge amount of cheap, genomic and phenotypic single-cell data that will be produced within the next few years in cancer research.

References

1. Wright S (1931) Evolution in Mendelian populations. Genetics 16:97–159
2. Wright S (1949) The genetical structure of populations. Ann Eugen 15:323–354
3. Fisher RA (1930) The genetical theory of natural selection. Clarendon Press. http://archive.org/details/geneticaltheoryo031631mbp
4. Moran PAP (1958) Random processes in genetics. Math Proc Camb Philos Soc 54:60–71

5. Nowell PC (1976) The clonal evolution of tumor cell populations. Science 194:23–28
6. Merlo LM, Pepper JW, Reid BJ, Maley CC (2006) Cancer as an evolutionary and ecological process. Nat Rev Cancer 6:924–935
7. Greaves M, Maley CC (2012) Clonal evolution in cancer. Nature 481:306–313
8. Tsao JL et al (2000) Genetic reconstruction of individual colorectal tumor histories. Proc Natl Acad Sci U S A 97:1236–1241
9. Vogelstein B et al (2013) Cancer genome landscapes. Science 339:1546–1558
10. Hanahan D, Weinberg RA (2000) The hallmarks of cancer. Cell 100:57–70
11. TCGA (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455:1061–1068
12. Curtis C et al (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486:346–352
13. TCGA (2012) Comprehensive molecular characterization of human colon and rectal cancer. Nature 487:330–337
14. Sjoblom T et al (2006) The consensus coding sequences of human breast and colorectal cancers. Science 314:268–274
15. Vogelstein B, Kinzler KW (2004) Cancer genes and the pathways they control. Nat Med 10:789–799
16. Vermeulen L et al (2013) Defining stem cell dynamics in models of intestinal tumor initiation. Science 342:995–998
17. Bos JL (1989) ras oncogenes in human cancer: a review. Cancer Res 49:4682–4689
18. Weinberg RA (1991) Tumor suppressor genes. Science 254:1138–1146
19. Duncan JA, Reeves JR, Cooke TG (1998) BRCA1 and BRCA2 proteins: roles in health and disease. Mol Pathol 51:237–247
20. Loeb LA (2001) A mutator phenotype in cancer. Cancer Res 61:3230–3239
21. Fearon ER, Vogelstein B (1990) A genetic model for colorectal tumorigenesis. Cell 61:759–767
22. Luebeck EG, Moolgavkar SH (2002) Multistage carcinogenesis and the incidence of colorectal cancer. Proc Natl Acad Sci U S A 99:15095–15100
23. Michor F, Iwasa Y, Nowak MA (2004) Dynamics of cancer progression. Nat Rev Cancer 4:197–205
24. Nowak MA et al (2002) The role of chromosomal instability in tumor initiation. Proc Natl Acad Sci U S A 99:16226–16231
25. Komarova NL, Sengupta A, Nowak MA (2003) Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. J Theor Biol 223:433–450
26. Nowak MA (2006) Evolutionary dynamics : exploring the equations of life. Belknap Press of Harvard University Press
27. Shendure J, Ji H (2008) Next-generation DNA sequencing. Nat Biotechnol 26:1135–1145
28. Beerenwinkel N et al (2007) Genetic progression and the waiting time to cancer. PLoS Comput Biol 3, e225
29. Bozic I et al (2010) Accumulation of driver and passenger mutations during tumor progression. Proc Natl Acad Sci U S A 107:18545–18550
30. Bernards R, Weinberg RA (2002) A progression puzzle. Nature 418:823
31. Klein CA (2009) Parallel progression of primary tumours and metastases. Nat Rev Cancer 9:302–312
32. Yachida S et al (2010) Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature 467:1114–1117
33. Sprouffske K, Pepper JW, Maley CC (2011) Accurate reconstruction of the temporal order of mutations in neoplastic progression. Cancer Prev Res (Phila) 4:1135–1144
34. Bar-Yam Y (2003) Dynamics of complex systems. Westview Press. http://www.amazon.com/exec/obidos/ISBN=0813341213/newenglandcompleA/
35. Bearer EL et al (2009) Multiparameter computational modeling of tumor invasion. Cancer Res 69:4493–4501
36. Anderson AR, Weaver AM, Cummings PT, Quaranta V (2006) Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. Cell 127:905–915

37. Enderling H, Hlatky L, Hahnfeldt P (2009) Migration rules: tumours are conglomerates of self-metastases. Br J Cancer 100:1917–1925
38. Sottoriva A et al (2010) Cancer stem cell tumor model reveals invasive morphology and increased phenotypical heterogeneity. Cancer Res 70:46–56
39. Frieboes HB et al (2010) Three-dimensional multispecies nonlinear tumor growth-II: tumor invasion and angiogenesis. J Theor Biol 264:1254–1278
40. Enderling H et al (2009) Paradoxical dependencies of tumor dormancy and progression on basic cell kinetics. Cancer Res 69:8814–8821
41. Anderson AR, Quaranta V (2008) Integrative mathematical oncology. Nat Rev Cancer 8:227–234
42. Komaki K, Sano N, Tangoku A (2006) Problems in histological grading of malignancy and its clinical significance in patients with operable breast cancer. Breast Cancer 13:249–253
43. Marusyk A, Polyak K (2010) Tumor heterogeneity: causes and consequences. Biochim Biophys Acta 1805:105–117
44. Navin N et al (2010) Inferring tumor progression from genomic heterogeneity. Genome Res 20:68–80
45. Snuderl M et al (2011) Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. Cancer Cell 20:810–817
46. Szerlip NJ et al (2012) Intratumoral heterogeneity of receptor tyrosine kinases EGFR and PDGFRA amplification in glioblastoma defines subpopulations with distinct growth factor response. Proc Natl Acad Sci U S A 109:3041–3046
47. Gerlinger M et al (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 366:883–892
48. Sottoriva A et al (2013) Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. Proc Natl Acad Sci U S A 110:4009–4014
49. Nik-Zainal S et al (2012) The life history of 21 breast cancers. Cell 149:994–1007
50. Nicolas P, Kim KM, Shibata D, Tavaré S (2007) The stem cell population of the human colon crypt: analysis via methylation patterns. PLoS Comput Biol 3, e28
51. Sottoriva A, Tavaré S (2010) Integrating approximate Bayesian computation with complex agent-based models for cancer research. In: Saporta G, Lechevallier Y (eds) COMPSTAT 2010 - proceedings in computational statistics. Springer, Physica Verlag, pp 57–66
52. Siegmund KD, Marjoram P, Woo YJ, Tavaré S, Shibata D (2009) Inferring clonal expansion and cancer stem cell dynamics from DNA methylation patterns in colorectal cancers. Proc Natl Acad Sci U S A 106:4828–33
53. Siegmund KD, Marjoram P, Tavaré S, Shibata D (2009) Many colorectal cancers are 'flat' clonal expansions. Cell Cycle 8:2187–2193
54. Sottoriva A, Spiteri I, Shibata D, Curtis C, Tavaré S (2013) Single-molecule genomic data delineate patient-specific tumor profiles and cancer stem cell organization. Cancer Res 73:41–49
55. Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. Genetics 162:2025–35
56. Meacham CE, Morrison SJ (2013) Tumour heterogeneity and cancer cell plasticity. Nature 501:328–337
57. Sottoriva A, Vermeulen L, Tavaré S (2011) Modeling evolutionary dynamics of epigenetic mutations in hierarchically organized tumors. PLoS Comput Biol 7, e1001132
58. Beaumon MA, Cornuet J, Marin J, Robert CP (2009) Adaptive approximate Bayesian computation. Biometrika 1–9