Published in final edited form as:

Cancer Lett. 2013 November 1; 340(2): . doi:10.1016/j.canlet.2012.12.028.

Advances for Studying Clonal Evolution in Cancer

Li Ding^{1,2,3,4}, Benjamin J. Raphael^{5,6}, Feng Chen^{7,8}, and Michael C. Wendt^{2,4,9}

¹Department of Medicine, Division of Oncology, Washington University School of Medicine, St. Louis, MO 63110, USA

²The Genome Institute, Washington University School of Medicine, St. Louis, MO 63108, USA

³Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO 63110, USA

⁴Department of Genetics, Washington University School of Medicine, St. Louis, MO 63108, USA

⁵Center for Computational Molecular Biology, Brown University, Box 1910, Providence, RI 02912, USA

⁶Department of Computer Science, Brown University, 115 Waterman St. Providence, RI 02912, USA

⁷Department of Medicine, Renal Division, Washington University School of Medicine, St. Louis, MO 63110, USA

⁸Department of Cell Biology and Physiology, Washington University School of Medicine, St. Louis, MO 63110, USA

⁹Department of Mathematics, Washington University, St. Louis, MO 63130, USA

Abstract

The “clonal evolution” model of cancer emerged and “evolved” amid ongoing advances in technology, especially in recent years during which next generation sequencing instruments have provided ever higher resolution pictures of the genetic changes in cancer cells and heterogeneity in tumors. It has become increasingly clear that clonal evolution is not a single sequential process, but instead frequently involves simultaneous evolution of multiple subclones that co-exist because they are of similar fitness or are spatially separated. Co-evolution of subclones also occurs when they complement each other’s survival advantages. Recent studies have also shown that clonal evolution is highly heterogeneous: different individual tumors of the same type may undergo very different paths of clonal evolution. New methodological advancements, including deep digital sequencing of a mixed tumor population, single cell sequencing, and the development of more sophisticated computational tools, will continue to shape and reshape the models of clonal evolution. In turn, these will provide both an improved framework for the understanding of cancer progression and a guide for treatment strategies aimed at the elimination of all, rather than just some, of the cancer cells within a patient.

© 2013 Elsevier Ireland Ltd. All rights reserved.

Correspondence to: Li Ding.

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

1. Introduction

Paradigm-changing discoveries tend to result from the culmination of evolutionary progress and its convergence with key advances in methodology. Even the occasional revolutions brought about by exceptionally forward thinking minds have frequently benefited from the enabling technologies of their times. Cancer progression is influenced by a myriad of intrinsic and extrinsic factors, possibly ruling out any unified model of clonal evolution for all cancer types. Even individual tumors of the same type may undertake different paths of progression. In this review, we will focus on the recent methodological advances in studying cancer clonal evolution and describe some of the seminal findings that help refine the models of clonal evolution. In addition, we will discuss some of the emerging technologies and approaches in the study of clonal evolution.

1.1. History of studying clonal evolution

Cellular heterogeneity within individual tumors was observed in the 1800s by the great pathologist Rudolf Virchow and others [1] using the compound microscope. Numerous advances in experimental methodology were subsequently made, including in the areas of pathohistology, cytogenetics, and imaging. These advances in turn resulted in straightforward, qualitative theories describing how tumors “progress” and become more morphologically and clinically “malignant”. The concept of cancer being driven by sequential genetic mutations began to gain wider acceptance [2; 3]. In 1976, Peter Nowell summarized these developments in the Clonal Evolution (CE) theory that describes tumor progression as a process paralleling that of Darwinian evolution with individual tumor cells analogous to individuals of an evolving species undergoing diversification and selection [4].

The CE model proposed that a normally-functioning cell incurs an induced or spontaneous genetic change, thereby undergoing “neoplastic proliferation”. Subsequently, random genetic alterations within these neoplastic cells create new mutant cells with variable fitness, and the cellular population undergoes selection. Specifically, most of the genetic variants are detrimental and the associated cells are either eliminated by competition for resources or destroyed by host immune system cells, the latter acting like predators in the natural selection for species. Occasionally, a genetic change provides selective *advantage* to a tumor cell, which may give rise to a dominant subpopulation. The sequential rounds of diversification and selection drive the tumor progression to higher levels of malignancy[4]. The analogy to Darwinian evolution is obvious. Over the nearly 4 decades since Nowell’s landmark paper, the basic action of clonal evolution has been found in many types of tumors by the use of ever improving cytogenetic tools and molecular genetics approaches. The latest addition to the diagnostic arsenal is DNA sequencing technology that provides increasingly detailed pictures of the mutations carried by the tumor cells.

1.2. Current understanding of tumor evolution

Heterogeneity within tumors is the basis for the selection of the fittest clones, a key step in clonal evolution. However, this heterogeneity presents a major challenge in deciphering the phylogenetic structure of subclones within individual tumors. Early studies measured genetic alterations in a limited number of genes or markers from tumor samples collected at different stages and from different individuals. While these studies were necessarily incomplete, such inter-tumor comparisons of a few factors largely skirted the issues of intra-tumor heterogeneity and provided clues to the key steps along the clonal evolution pathway for a number of tumor types[5; 6]. The advances in genomic tools, from microarrays to large scale DNA sequencing, have brought about the “genome era”, in which hundreds of thousands of genes, transcripts, and proteins can be surveyed at once in cancers and other biological systems.

The rapid developments in next-generation DNA sequencing technologies now make it possible to interrogate the entire exomes, genomes, or transcriptomes of cancer samples including clinical samples [7; 8]. These advances allow the comparison of the genetic variations at single-nucleotide resolution [9]. Parallel to the advances in DNA and RNA sequencing technologies, we have witnessed an equally spectacular rise in computational power and the development of critical bioinformatics tools for the analyses of astronomical amounts of data generated [10; 11; 12; 13]. These bioinformatics tools have been critical in deducing the clonal history of individual tumors using sequencing data from tumor samples with a mixture of subclones having different genetic compositions.

The traditional CE theory describes iterative rounds of diversification and selection with the fittest subclone dominating each round [4]. As DNA sequencing and mutational detection tools have become increasingly efficient and sensitive, intra-tumor comparisons of subclones are now being conducted, including studies of spatial distribution of subclones [14] (Table 1). Such studies have led to a number of modified CE theories, including models in which multiple clones of similar fitness evolve simultaneously and models in which coevolution of subclones having potentially cooperative relations occur [15]. Nearly all tumors are highly heterogeneous and undergo dynamic changes. It is unlikely that any two cancer cells are identical at the nucleotide level. Thus, the recent application of single cell sequencing opens the door for the study of cancer clonal evolution at the highest level of resolution [16; 17; 18; 19; 20] (Table 1). In addition, cancers live within a host environment to which their changes are bilaterally coupled and the co-evolution of the “cancer-environment system” is gradually being recognized [21; 22; 23; 24].

Unlike clonal evolution models, cancer stem cell models emphasize the separation of tumor cells into tumorigenic stem cells capable of self-renewal and their more differentiated derivatives that are non-tumorigenic. The applicability of these models is still a matter of debate because of questions regarding markers used to identify/purify these cancer stem cells and the systems in testing tumorigenic potential [25; 26]. It is conceivable that even if cancers have subpopulations with innately different tumorigenic and self-renewal abilities, these cells, including the stem cells, themselves, can undergo genetic changes and selections mirroring the process of clonal evolution.

2. Computational and mathematical approaches for studying tumor evolution

Computational and mathematical work in cancer started almost 6 decades ago. Despite the complete lack of modern genetic data, early studies were able to statistically investigate germline predispositions and somatic changes that initiate cancer and promote its clonal expansion. In particular, several landmark papers established what is now loosely known as the “Knudson hypothesis”, i.e. the theory that a small number of accumulated somatic events, sometimes as few as two, are sufficient to cause cancer [2; 3; 27; 28]. The original concept was constructed indirectly according to the observation that a specific cancer frequency, f , correlates well with age, a , raised to a certain power, x . That is, $\ln(f) = x \cdot \ln(a) + C$, where C is a fitting constant.

2.1. Identifying genetic alterations for studying tumor clonality

One of the most important pre-processing (upstream) steps in the analysis of cancer sequence data is the identification of somatic variants from the raw sequence reads (Fig. 1). Significant progress has been made during the roughly 5 years since cancer genomic sequencing emerged on a significant scale. For example, there are now a number of established software systems that can reliably detect *clonal* single-nucleotide variants

(SNVs) in high-purity samples sequenced with high (>30X) coverage [29; 30]. Subclonal variants – those having variant allele frequencies (VAFs) below 10% - are much more difficult to detect reliably. Importantly, these subclonal variants are precisely the events that are indicative of the emergence or existence of tumor subclones. Recently, Nik-Zainal et al. [31] demonstrated that subclonal variants could be readily detected with ultra-high coverage (188X) genome sequencing. Further algorithmic developments in variant prediction coupled with the availability of deeper coverage data will make it possible to identify subclonal mutations directly from sequencing data. In addition, further advances in sample preparation – e.g. laser capture microdissection – will provide higher purity samples for sequencing. The isolation and sequencing of single tumor cells will directly measure subclonal mutations. We discuss these further below.

In addition to SNVs, the detection of other types of somatic mutations is important, including insertions and deletions (indels) varying in size from a few base pairs up to larger copy number aberrations involving significant portions of (or even entire) chromosome arms. In addition, structural variants such as inversions, translocations, or other rearrangements are also common in tumors. Progress in sequencing these mutations and assessing their heterogeneity has generally been slower than for SNVs, primarily because these mutations require specialized techniques for detection and analysis. These include “split reads”, read depth, statistical analysis of anomalous read-pair insert lengths, and read assembly to assess break points [32]. In theory, there are designs to detect indels at a constant false-positive error rate, say 1%, over a wide range of sizes, e.g. 10bp to 10,000bp, by judicious use of different insert types, but the required sequence coverage is still higher than what is currently economically feasible [33]. As with SNVs, algorithms for detecting these more general somatic event types are under active research.

2.2. Methods for discerning clonalities

Given the crucial relevance of clonality, an important problem is the *clonal census problem*: determining how many clones actually comprise a tumor. While the dynamics of clone birth and proliferation are governed primarily by somatic driver events, the total set of events, drivers plus passengers, provides the strongest signature, i.e. the most information, to estimate the clonal census. It will eventually become routine to sample a tumor spatially (Fig. 1), but most specimen procurements are still non-specific, so the tumor sample is assumed to be a uniform sample from the population of tumor cells. Although in some cases clones are indeed known to be spatially segregated [7], all current census methods make the uniform sampling assumption. Consequently, the numbers of reads from each clone are expected to be proportional to their clonal fractions, though also subject to the scatter that is incurred from artefactual noise, such as sequencing and alignment errors and read sampling biases. For example, a clone that represents 10% of the tumor mass should exhibit clone-specific heterozygous SNVs clustered around a variant allele frequency (VAF) of about 5%.

As with much of what has been discussed here, the clonal census problem is still open and remains an active area of research. Current studies have examined the distribution of variant allele frequencies (VAFs) for detected SNVs [34]. Assuming that the noise, or scatter, itself is Gaussian, the simplest test for the presence of more than one clone is whether the distribution of VAFs fits a normal distribution. If the distribution of VAFs is not Gaussian, then the task is to estimate the number of clones that are present. A number of methods are currently used to separate the VAF distribution into a mixture of individual clonal components, including “peak counting” based on kernel density estimation[35] and various clustering procedures[36]. Besides sampling uncertainties and sequencing errors, there are still many confounders that must be controlled for in the clonal census problem, including the presence of copy number alterations and loss of heterozygosity (LOH). Future

improvements in census methods will doubtless involve integrating other mutation types such as structural variants.

2.3. Clonality and phylogenetic relationships

As discussed above, a general feature of tumor biology is clonal expansion[4], and this is now being addressed from a number of different perspectives. Complete sequencing of a patient is quickly coming to mean sequencing not only the whole genome of the primary tumor, but all metastases and even auxiliaries like xenografts, if they are created [37]. Given that these are evolving cell populations and given their evolutionary duality with populations of individuals (mentioned above), investigators have begun to apply phylogenetic analysis tools to discern relationships among tumors[5; 38]. As with traditional phylogenetic trees, topology, placement, and branch length are indicative of both lineage and timing. The task is especially facilitated by applications like PHYLP that can use reference and variant read counts as estimations of allele frequencies[39]. There is general anticipation that the cost of DNA sequencing will continue to decrease, suggesting that multiple–time–point (MTP) sequencing will eventually become commonplace[37]. Here, resequencing will be done either at various intervals or clinical milestones, such as relapse, posing the proposition of reconstructing more comprehensive tumor histories. Such information will yield much improved empirical knowledge on the rates of tumor evolution, expansion, and tendencies to relapse, metastasize, and re–metastasize. Thus far, phylogenetic tools have been used only in a single capacity: to assess the relationships between a primary tumor and its metastases at a single snapshot in time. However, advances in instrumentation are continuing to lower the cost of cancer genome sequencing and we anticipate that phylogenetic analysis will expand commensurately with the availability of larger datasets. There have already been projects where genomes from several time points have been sequenced [40]. Such information will be useful in evaluating how relationships evolve over time, and in showing how any new metastases fit into the phylogeny. Moreover, the known time intervals between sequences will help to calibrate models of tumor evolution and clonal expansion.

Cancer sequence data sets have traditionally been collected at depths of roughly 30X [1], but the improving economies of sequencing are also now fostering deeper data sets. For example, it is becoming routine practice to generate 60X to 90X (or even higher) depth for tumor sequences. These depths are necessary for capturing low variant–allele–frequency events associated with minor clones. Consider that if a minor clone comprises 3% of the mass of a pure tumor, it would take an average of about 67X depth until even a single read representing a heterozygous somatic event would appear. Deeper data will improve cancer phylogenetics not only by furnishing more complete sets of events for the analyses, but also in estimating variant allele frequencies more accurately. The combination of very deep and multiple time point sequencing will eventually enable a sort of direct observation of the clonal development process itself within a tumor. That is, we expect to be able to create a time–progression of phylogenies showing the relationships, including births and deaths, of all the clones within a tumor. In the longer term, single–cell sequencing (discussed further below) could enable the highest phylogenetic resolution, revealing essentially the complete genealogy of the cellular complement.

2.4. Mathematical considerations for clonal evolution

Since the advent of next–generation sequencing of cancer genomes, mathematical investigations of cancer dynamics have expanded dramatically. Such work is now concentrating on the big frontiers in cancer, including driver and clonal expansion mechanisms and relationships, time–evolution and scales, and latency periods, as well as focusing on the ways in which any or all of these aspects may be exploited to improve

diagnosis, prognosis, and treatment. Most of these approaches are still statistical, meaning that they try to discern largely unspecific relationships *post hoc* from data alone.

Before discussing specific efforts, let us note two caveats. First, the statistical approach implies the problem of having sufficient power to detect associations. Though there are a few instances where effects are quite strong[41], experience suggests most will be much more subtle, whereby hundreds, or even thousands of samples will be required to obtain reasonable chances of success. Studies of such size are now becoming feasible with continuing advancements in sequencing instrumentation. Second, all higher-level statistical analysis and model testing depend strongly on the lower-level data analysis, i.e. sequence determination, mapping and assembly, somatic and germline detection, etc. Strictly speaking, none of these problems is completely solved, the consequence being that investigators generally have imperfect lists of variants with which to work, lists that include non-existent variants (false-positives) and exclude real variants (false negatives). False-positives are generally less serious because they often can be resolved by validation[29]. False-negatives arise from low signal in the sequencing data, but are often the very events one requires to determine clonality. Importantly, it is not (yet) possible to know whether all events have been captured for a given amount of sequence. Indeed, the optimal depth of sequence for cancer samples is still a matter of some debate[42; 43].

2.5. Physical Modeling and Future Work

Clonal evolution models are all still highly idealized, not only in the significant simplifying assumptions they make, but also in neglecting the spatial-temporal nature of tumors and numerous intracellular and extracellular process like chaperoning of heat shock proteins and the influence of the microenvironment, respectively. Physical modeling (Fig. 1) has made some progress in these areas, for example in putting a mathematical description[44] to models that were strictly qualitative[45]. Generally speaking, physical models are based on sets of differential equations, often of the Lotka-Volterra type[46], that can describe the diffusion-reaction dynamics of various biological factors. The associated literature has expanded dramatically, with papers reporting on numerous spatial-temporal aspects, including chemotaxis[47], implications of hormone therapies[48], integration with medical imaging data[49], and the defeat of micro-environmental proliferation barriers[50].

3. New perspectives and implications

3.1. Single cell sequencing

A promising new approach for cancer heterogeneity studies is the sequencing and analysis of single cancer cells from a tumor. Molecular cytogenetic techniques such as fluorescent in situ hybridization (FISH) and spectral karyotyping (SKY) have been in use for decades to measure single tumor cells and have shown the variability between the chromosomal contents of individual tumor cells. Ideally, whole-genome sequencing of a sufficient number of such cells at sufficient quality would directly quantify the genomic heterogeneity within a tumor. However, there are two challenges that presently impede such studies: (1) technical limitations in sequencing technologies that demand DNA amplification prior to DNA sequencing, and (2) selection and sampling strategies of single cells to appropriately measure the sequence diversity within a tumor.

Current DNA sequencing technologies require a quantity of input DNA far exceeding the DNA of a single human/cancer cell. Thus, single cell sequencing studies first must use whole-genome amplification (WGA) techniques. Various WGA techniques have been employed for cancer studies include degenerate oligonucleotide primed (DOP) WGA[18] and multiple displacement amplification (MDA)[19; 20]. The key problem with MDA and other genome amplification techniques is that amplification biases results in the form of

unequal coverage of the genome. These coverage biases obviously make it difficult to identify somatic mutations of all types, including single-nucleotide variants, copy number aberrations, and structural aberrations.

The sensitivity to detect single nucleotide variants (or small indels) is most affected by allele dropout (ADO), i.e. the failure of amplification of one of two alleles in a heterozygote. Using MDA[20], Hou et al. reported ~8 and 15% ADO from single-cell whole-genome sequencing of lymphoblastoid cell lines, although others have reported much higher ADO rates of ~26-40%[51]. The sensitivity and specificity for detecting copy number aberrations is affected by unequal coverage of the genome. Building on earlier work in assessing copy number aberrations in multiple tumor subpopulations[17], Wigler and colleagues introduced Single nucleus sequencing (SNS)[18]. SNS performs low-coverage sequencing using WGA of DNA extracted from flow sorted nuclei. Using flow cytometry, tumor subpopulations can be separated by ploidy, separately analyzing diploid, hyperdiploid, and aneuploid subpopulations. Low coverage of the genome (~6%) is achieved, but this is sufficient to analyze large copy number aberrations by examining read counts in variable-sized bins. Recently, this technique has been further optimized for copy number aberration studies using lower coverage sequencing[52].

The unequal coverage characteristic of SNS makes the analysis of smaller copy number variants and structural variants difficult. Without even coverage, genome assembly is extremely difficult. In bacterial genomes, which are smaller and generally less repetitive than human genomes, there has been some recent success in *de novo* assembly from single cell sequencing data[53]. Success in overcoming technical hurdles in obtaining single-cell sequencing data has been coupled with advances in *de novo* assembly algorithms that handle uneven coverages from single-cell data [54]. It remains to be seen if these successes will scale to human genomes, and more importantly to cancer genomes where extensive aneuploidy might present additional challenges.

3.2. Mutational significance and driver events

The focus of early statistical methods that used next generation sequencing data was to identify driver mutations from specific gene lists, but this quickly matured to examination of exomes[55] and whole genomes[56]. Assessment of the genewise mutational significance was sometimes rudimentary, but a larger problem was the realization of enormous heterogeneities in cancers of the same type[57; 58]. This led to an expanded focus on pathways and other biologically-relevant groupings of genes (Fig. 1) under the hypothesis that the relevant dynamics localize at the level of the group[59]. Improved tests are now available that exploit the details of such configurations, e.g. the distribution of gene sizes within the group and the distribution of mutations among the samples[60].

One of the limitations of such mutation enrichment analysis is that gene groups are typically predefined by a curated database, but these sources represent a minuscule fraction of the total number of potentially-relevant gene combinations that exist in the human genome. For instance, there are more than 10^{82} possible sets of 25 or fewer genes. To better address this enormous genomic design space, new and sophisticated algorithms[61; 62] have been developed to identify mutated subnetworks from large-scale protein-protein interaction networks [63]. Consider an algorithm such as HotNet[64], which addresses the combinatorial aspect of interactions. For example, two proteins associated with mutated genes that interact with one another directly, or interact perhaps through a third protein, which itself engages only a few other proteins would tend to be more meaningful than if the intermediary were highly connected elsewhere too. However, this aspect must be weighed against the nature of the interactions, the degree of separation, etc. Hotnet quantifies this problem by inferring an “influence metric” based on an analog of heat diffusion. That is,

diffusion in a candidate network is modeled by invoking a node source, q , and calculating how a resultant scalar T distributes itself according to a diffusion equation of the form $\frac{\partial T}{\partial t} = \nabla^2 T + q$, where t is time. The resulting T distribution indicates influence between any pair of nodes, thereby providing a mechanism for distinguishing incidental sub-networks from those that are actually meaningful. While such networks do not yet contain the full human interactome, they do allow the identification of combinations of mutations that in a way that is less biased than existing annotations of pathways. More recently, the Dendrix algorithm [64] identifies combinations of mutations without any prior knowledge of gene interactions, instead examining statistical patterns of in recurrence of these combinations across patients.

The general topic of assessing mutational significance is yet another active area and efforts in the near term will continue to focus on software improvements, e.g. combining network algorithms with sophisticated significance tests, on extending the testing scope, e.g. for loci that are in physical proximity[57], and on integrating other data types that are now amenable to high-throughput processing, e.g. methylation and RNA-seq. In general, methodological development will likely continue to follow well-established concepts from mathematical statistics, for example in searching for correlations between somatic events and changes in expression.

3.3. Clonal evolution and drug treatment/resistant

Clonal evolution is thought to be not only one of the most crucial aspects of cancer, but also the key to its treatment. The ability to discern new clones, especially early in their appearance, and to establish what modalities will be effective on such clones will be important in the near-term clinical environment. Consequently, the modeling of clonality and clonal expansions is also a growing trend in statistical genetics. At a high level, these efforts assume a particular population evolution process, e.g. Wright-Fisher reproduction[65], Galton-Watson branching[66], or Moran replacement[67], along with various other tractability assumptions like synchronous division and constant mutation rate. They then try to discern parameters of the resulting evolutionary model from sequence data. These models yield various testable predictions that will become more relevant with future multiple time point data sets. For example, several papers report explicit expressions for the timing of various clinical milestones such as latency, the appearance of new clones, relapse, and metastases[68]. There are numerous open issues, for example quantification of selection strength and questions regarding the functionality of passenger mutations as “cancer molecular clocks”[69] that such models may eventually help to resolve.

4. Conclusion

Cancer initiation and progression are governed by complex, non-linear mechanisms. In the shorter-term, computational and mathematical work will probably continue to be divided into narrow, question-specific areas and to be aided by the rapid accumulation of genomic data. Indeed, given the heterogeneity of the cancers, the field could diverge further into cancer-specific models, for example for those tumors having a mutator phenotype versus those without. Over the longer term, the goal will certainly be to unify the physical and statistical genomics viewpoints in order to rigorously account for all the mechanisms and scales important in cancer. This will be a much more ambitious undertaking and likely to become one of the major research programs in cancer biology.

Acknowledgments

The authors are supported by grants from the NIH (U01HG006517, U54HG003079, R01HG005690, R01DK081592, and R01DK087960). We thank Joshua McMichael for help with figure preparation and Michael D. McLellan for providing cross-study comparison.

References

1. Brown TM, Fee E. Rudolf Carl Virchow: medical scientist, social reformer, role model. *Am J Public Health*. 2006; 96:2104–2105. [PubMed: 17077410]
2. Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A*. 1971; 68:820–823. [PubMed: 5279523]
3. Nordling CO. A new theory on cancer-inducing mechanism. *Br J Cancer*. 1953; 7:68–72. [PubMed: 13051507]
4. Nowell PC. The clonal evolution of tumor cell populations. *Science*. 1976; 194:23–28. [PubMed: 959840]
5. Navin NE, Hicks J. Tracing the tumor lineage. *Mol Oncol*. 2010; 4:267–283. [PubMed: 20537601]
6. Garcia SB, Novelli M, Wright NA. The clonal origin and clonal evolution of epithelial tumours. *Int J Exp Pathol*. 2000; 81:89–116. [PubMed: 10762440]
7. Greaves M, Maley CC. Clonal evolution in cancer. *Nature*. 2012; 481:306–313. [PubMed: 22258609]
8. Guan YF, Li GR, Wang RJ, Yi YT, Yang L, Jiang D, Zhang XP, Peng Y. Application of next-generation sequencing in clinical oncology to advance personalized treatment of cancer. *Chin J Cancer*. 2012; 31:463–470. [PubMed: 22980418]
9. Ma QC, Ennis CA, Aparicio S. Opening Pandora's Box--the new biology of driver mutations and clonal evolution in cancer as revealed by next generation sequencing. *Curr Opin Genet Dev*. 2012; 22:3–9. [PubMed: 22386266]
10. Wooley JC, Ye Y. Metagenomics: Facts and Artifacts, and Computational Challenges*. *Journal of computer science and technology*. 2009; 25:71–81. [PubMed: 20648230]
11. Piro RM, Di Cunto F. Computational approaches to disease-gene prediction: rationale, classification and successes. *The FEBS journal*. 2012; 279:678–696. [PubMed: 2221742]
12. Wang Q, Xia J, Jia P, Pao W, Zhao Z. Application of next generation sequencing to human gene fusion detection: computational tools, features and perspectives. *Briefings in bioinformatics*. 2012
13. Medvedev P, Stanciu M, Brudno M. Computational methods for discovering structural variation with next-generation sequencing. *Nat Methods*. 2009; 6:S13–20. [PubMed: 19844226]
14. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, Varela I, Phillimore B, Begum S, McDonald NQ, Butler A, Jones D, Raine K, Latimer C, Santos CR, Nohadani M, Eklund AC, Spencer-Dene B, Clark G, Pickering L, Stamp G, Gore M, Szallasi Z, Downward J, Futreal PA, Swanton C. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*. 2012; 366:883–892. [PubMed: 22397650]
15. Snuderl M, Fazlollahi L, Le LP, Nitta M, Zhelyazkova BH, Davidson CJ, Akhavanfard S, Cahill DP, Aldape KD, Betensky RA, Louis DN, Iafrate AJ. Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. *Cancer Cell*. 2011; 20:810–817. [PubMed: 22137795]
16. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods*. 2009; 6:377–382. [PubMed: 19349980]
17. Navin N, Krasnitz A, Rodgers L, Cook K, Meth J, Kendall J, Riggs M, Eberling Y, Troge J, Grubor V, Levy D, Lundin P, Maner S, Zetterberg A, Hicks J, Wigler M. Inferring tumor progression from genomic heterogeneity. *Genome Res*. 2010; 20:68–80. [PubMed: 19903760]
18. 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 WR, Hicks J, Wigler M. Tumour evolution inferred by single-cell sequencing. *Nature*. 2011
19. 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. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell*. 2012; 148:886–895. [PubMed: 22385958]
20. 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. Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell*. 2012; 148:873–885. [PubMed: 22385957]
21. Lisanti MP, Martinez-Outschoorn UE, Chiavarina B, Pavlides S, Whitaker-Menezes D, Tsigos A, Witkiewicz A, Lin Z, Balliet R, Howell A, Sotgia F. Understanding the “lethal” drivers of tumor-stroma co-evolution: emerging role(s) for hypoxia, oxidative stress and autophagy/mitophagy in the tumor micro-environment. *Cancer Biol Ther*. 2010; 10:537–542. [PubMed: 20861671]
 22. Jossan S, Matsuo Y, Chung LW, Zhou HE, Wang R. Tumor-stroma co-evolution in prostate cancer progression and metastasis. *Semin Cell Dev Biol*. 2010; 21:26–32. [PubMed: 19948237]
 23. Lorusso G, Ruegg C. The tumor microenvironment and its contribution to tumor evolution toward metastasis. *Histochem Cell Biol*. 2008; 130:1091–1103. [PubMed: 18987874]
 24. Peng CW, Liu XL, Liu X, Li Y. Co-evolution of cancer microenvironment reveals distinctive patterns of gastric cancer invasion: laboratory evidence and clinical significance. *Journal of translational medicine*. 2010; 8:101. [PubMed: 20950454]
 25. Clevers H. The cancer stem cell: premises, promises and challenges. *Nat Med*. 2011; 17:313–319. [PubMed: 21386835]
 26. Magee JA, Piskounova E, Morrison SJ. Cancer stem cells: impact, heterogeneity, and uncertainty. *Cancer Cell*. 2012; 21:283–296. [PubMed: 22439924]
 27. Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br J Cancer*. 1954; 8:1–12. [PubMed: 13172380]
 28. Armitage P, Doll R. A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. *Br J Cancer*. 1957; 11:161–169. [PubMed: 13460138]
 29. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012; 22:568–576. [PubMed: 22300766]
 30. Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, Ley TJ, Mardis ER, Wilson RK, Ding L. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics*. 2012; 28:311–317. [PubMed: 22155872]
 31. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, Raine K, Jones D, Marshall J, Ramakrishna M, Shlien A, Cooke SL, Hinton J, Menzies A, Stebbings LA, Leroy C, Jia M, Rance R, Mudie LJ, Gamble SJ, Stephens PJ, McLaren S, Tarpey PS, Papaemmanuil E, Davies HR, Varela I, McBride DJ, Bignell GR, Leung K, Butler AP, Teague JW, Martin S, Jonsson G, Mariani O, Boyault S, Miron P, Fatima A, Langerod A, Aparicio SA, Tutt A, Sieuwerts AM, Borg A, Thomas G, Salomon AV, Richardson AL, Borresen-Dale AL, Futreal PA, Stratton MR, Campbell PJ. The life history of 21 breast cancers. *Cell*. 2012; 149:994–1007. [PubMed: 22608083]
 32. Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl MC, Zhang Q, Locke DP, Shi X, Fulton RS, Ley TJ, Wilson RK, Ding L, Mardis ER. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods*. 2009; 6:677–681. [PubMed: 19668202]
 33. Wendl MC, Wilson RK. Statistical aspects of discerning indel-type structural variation via DNA sequence alignment. *BMC Genomics*. 2009; 10:359. [PubMed: 19656394]
 34. Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC, Wartman LD, Lamprecht TL, Liu F, Xia J, Kandoth C, Fulton RS, McLellan MD, Dooling DJ, Wallis JW, Chen K, Harris CC, Schmidt HK, Kalicki-Veizer JM, Lu C, Zhang Q, Lin L, O’Laughlin MD, McMichael JF, Delehaanty KD, Fulton LA, Magrini VJ, McGrath SD, Demeter RT, Vickery TL, Hundal J, Cook LL, Swift GW, Reed JP, Alldredge PA, Wylie TN, Walker JR, Watson MA, Heath SE, Shannon WD, Varghese N, Nagarajan R, Payton JE, Baty JD, Kulkarni S, Kline JM, Tomasson MH, Westervelt P, Walter MJ, Graubert TA, DiPersio JF, Ding L, Mardis ER, Wilson RK. The origin and evolution of mutations in acute myeloid leukemia. *Cell*. 2012; 150:264–278. [PubMed: 22817890]
 35. Parzen E. On Estimation of a Probability Density Function and Mode. *Ann Math Statist*. 1962; 33:1065–1076.

36. H D, L P, W S. Review of Three Latent Class Cluster Analysis Packages: Latent Gold, poLCA, and mCLUST. *The American Statistician*. 2009; 63:81–91.
37. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, Harris CC, McLellan MD, Fulton RS, Fulton LL, Abbott RM, Hoog J, Dooling DJ, Koboldt DC, Schmidt H, Kalicki J, Zhang Q, Chen L, Lin L, Wendl MC, McMichael JF, Magrini VJ, Cook L, McGrath SD, Vickery TL, Appelbaum E, Deschryver K, Davies S, Guintoli T, Crowder R, Tao Y, Snider JE, Smith SM, Dukes AF, Sanderson GE, Pohl CS, Delehaanty KD, Fronick CC, Pape KA, Reed JS, Robinson JS, Hodges JS, Schierding W, Dees ND, Shen D, Locke DP, Wiechert ME, Eldred JM, Peck JB, Oberkfell BJ, Lolofie JT, Du F, Hawkins AE, O’Laughlin MD, Bernard KE, Cunningham M, Elliott G, Mason MD, Thompson DM Jr, Ivanovich JL, Goodfellow PJ, Perou CM, Weinstock GM, Aft R, Watson M, Ley TJ, Wilson RK, Mardis ER. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature*. 464:999–1005. [PubMed: 20393555]
38. Greenman CD, Pleasance ED, Newman S, Yang F, Fu B, Nik-Zainal S, Jones D, Lau KW, Carter N, Edwards PA, Futreal PA, Stratton MR, Campbell PJ. Estimation of rearrangement phylogeny for cancer genomes. *Genome Res*. 2012; 22:346–361. [PubMed: 21994251]
39. Felsenstein J. PHYLIP -- Phylogeny Inference Package (Version 3.2). *Cladistics*. 1989; 5:164–166.
40. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, Ritchey JK, Young MA, Lamprecht T, McLellan MD, McMichael JF, Wallis JW, Lu C, Shen D, Harris CC, Dooling DJ, Fulton RS, Fulton LL, Chen K, Schmidt H, Kalicki-Veizer J, Magrini VJ, Cook L, McGrath SD, Vickery TL, Wendl MC, Heath S, Watson MA, Link DC, Tomasson MH, Shannon WD, Payton JE, Kulkarni S, Westervelt P, Walter MJ, Graubert TA, Mardis ER, Wilson RK, DiPersio JF. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*. 2012; 481:506–510. [PubMed: 22237025]
41. Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, Kandoth C, Payton JE, Baty J, Welch J, Harris CC, Lichti CF, Townsend RR, Fulton RS, Dooling DJ, Koboldt DC, Schmidt H, Zhang Q, Osborne JR, Lin L, O’Laughlin M, McMichael JF, Delehaanty KD, McGrath SD, Fulton LA, Magrini VJ, Vickery TL, Hundal J, Cook LL, Conyers JJ, Swift GW, Reed JP, Alldredge PA, Wylie T, Walker J, Kalicki J, Watson MA, Heath S, Shannon WD, Varghese N, Nagarajan R, Westervelt P, Tomasson MH, Link DC, Graubert TA, DiPersio JF, Mardis ER, Wilson RK. DNMT3A mutations in acute myeloid leukemia. *N Engl J Med*. 2010; 363:2424–2433. [PubMed: 21067377]
42. Wendl MC, Wilson RK. Aspects of coverage in medical DNA sequencing. *BMC Bioinformatics*. 2008; 9:239. [PubMed: 18485222]
43. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet*. 2010; 11:685–696. [PubMed: 20847746]
44. Gatenby RA, Vincent TL. An evolutionary model of carcinogenesis. *Cancer Res*. 2003; 63:6212–6220. [PubMed: 14559806]
45. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell*. 1990; 61:759–767. [PubMed: 2188735]
46. Gatenby RA, Gawlinski ET. A reaction-diffusion model of cancer invasion. *Cancer Res*. 1996; 56:5745–5753. [PubMed: 8971186]
47. Vital-Lopez FG, Armaou A, Hutnik M, Maranas CD. Modeling the Effect of Chemotaxis on Glioblastoma Tumor Progression. *AIChE Journal*. 2011; 57:778–792.
48. Tanaka G, Hirata Y, Goldenberg SL, Bruchovsky N, Aihara K. Mathematical modelling of prostate cancer growth and its application to hormone therapy. *Philos Transact A Math Phys Eng Sci*. 2010; 368:5029–5044.
49. Konukoglu E, Clatz O, Menze BH, Stieltjes B, Weber MA, Mandonnet E, Delingette H, Ayache N. Image guided personalization of reaction-diffusion type tumor growth models using modified anisotropic eikonal equations. *IEEE Trans Med Imaging*. 2010; 29:77–95. [PubMed: 19605320]
50. Gatenby RA, Gillies RJ. A microenvironmental model of carcinogenesis. *Nat Rev Cancer*. 2008; 8:56–61. [PubMed: 18059462]
51. Gundry M, Li W, Maqbool SB, Vijg J. Direct, genome-wide assessment of DNA mutations in single cells. *Nucleic Acids Res*. 2012; 40:2032–2040. [PubMed: 22086961]

52. Baslan T, Kendall J, Rodgers L, Cox H, Riggs M, Stepansky A, Troge J, Ravi K, Esposito D, Lakshmi B, Wigler M, Navin N, Hicks J. Genome-wide copy number analysis of single cells. *Nat Protoc.* 2012; 7:1024–1041. [PubMed: 22555242]
53. Chitsaz H, Yee-Greenbaum JL, Tesler G, Lombardo MJ, Dupont CL, Badger JH, Novotny M, Rusch DB, Fraser LJ, Gormley NA, Schulz-Trieglaff O, Smith GP, Evers DJ, Pevzner PA, Lasken RS. Efficient de novo assembly of single-cell bacterial genomes from short-read data sets. *Nat Biotechnol.* 2011; 29:915–921. [PubMed: 21926975]
54. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012; 19:455–477. [PubMed: 22506599]
55. Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV, Ballinger DG, Sparks AB, Hartigan J, Smith DR, Suh E, Papadopoulos N, Buckhaults P, Markowitz SD, Parmigiani G, Kinzler KW, Velculescu VE, Vogelstein B. The genomic landscapes of human breast and colorectal cancers. *Science.* 2007; 318:1108–1113. [PubMed: 17932254]
56. Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, Dooling D, Dunford-Shore BH, McGrath S, Hickenbotham M, Cook L, Abbott R, Larson DE, Koboldt DC, Pohl C, Smith S, Hawkins A, Abbott S, Locke D, Hillier LW, Miner T, Fulton L, Magrini V, Wylie T, Glasscock J, Conyers J, Sander N, Shi X, Osborne JR, Minx P, Gordon D, Chinwalla A, Zhao Y, Ries RE, Payton JE, Westervelt P, Tomasson MH, Watson M, Baty J, Ivanovich J, Heath S, Shannon WD, Nagarajan R, Walter MJ, Link DC, Graubert TA, DiPersio JF, Wilson RK. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature.* 2008; 456:66–72. [PubMed: 18987736]
57. Dees ND, Zhang Q, Kandoth C, Wendl MC, Schierding W, Koboldt DC, Mooney TB, Callaway MB, Dooling D, Mardis ER, Wilson RK, Ding L. MuSiC: Identifying mutational significance in cancer genomes. *Genome Res.* 2012; 22:1589–1598. [PubMed: 22759861]
58. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE. The consensus coding sequences of human breast and colorectal cancers. *Science.* 2006; 314:268–274. [PubMed: 16959974]
59. Lin J, Gan CM, Zhang X, Jones S, Sjoblom T, Wood LD, Parsons DW, Papadopoulos N, Kinzler KW, Vogelstein B, Parmigiani G, Velculescu VE. A multidimensional analysis of genes mutated in breast and colorectal cancers. *Genome Res.* 2007; 17:1304–1318. [PubMed: 17693572]
60. Wendl MC, Wallis JW, Lin L, Kandoth C, Mardis ER, Wilson RK, Ding L. PathScan: a tool for discerning mutational significance in groups of putative cancer genes. *Bioinformatics.* 2011; 27:1595–1602. [PubMed: 21498403]
61. Vandin F, Upfal E, Raphael BJ. Algorithms for detecting significantly mutated pathways in cancer. *J Comput Biol.* 2011; 18:507–522. [PubMed: 21385051]
62. Ciriello G, Cerami E, Sander C, Schultz N. Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res.* 2012; 22:398–406. [PubMed: 21908773]
63. Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, Surendranath V, Niranjana V, Muthusamy B, Gandhi TK, Gronborg M, Ibarrola N, Deshpande N, Shanker K, Shivashankar HN, Rashmi BP, Ramya MA, Zhao Z, Chandrika KN, Padma N, Harsha HC, Yatish AJ, Kavitha MP, Menezes M, Choudhury DR, Suresh S, Ghosh N, Saravana R, Chandran S, Krishna S, Joy M, Anand SK, Madavan V, Joseph A, Wong GW, Schiemann WP, Constantinescu SN, Huang L, Khosravi-Far R, Steen H, Tewari M, Ghaffari S, Blobe GC, Dang CV, Garcia JG, Pevsner J, Jensen ON, Roepstorff P, Deshpande KS, Chinnaiyan AM, Hamosh A, Chakravarti A, Pandey A. Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res.* 2003; 13:2363–2371. [PubMed: 14525934]
64. Vandin F, Upfal E, Raphael BJ. De novo discovery of mutated driver pathways in cancer. *Genome Res.* 2012; 22:375–385. [PubMed: 21653252]

65. Beerenwinkel N, Antal T, Dingli D, Traulsen A, Kinzler KW, Velculescu VE, Vogelstein B, Nowak MA. Genetic progression and the waiting time to cancer. *PLoS Comput Biol.* 2007; 3:e225. [PubMed: 17997597]
66. Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, Chen S, Karchin R, Kinzler KW, Vogelstein B, Nowak MA. Accumulation of driver and passenger mutations during tumor progression. *Proc Natl Acad Sci U S A.* 2010; 107:18545–18550. [PubMed: 20876136]
67. Nowak MA, Michor F, Iwasa Y. The linear process of somatic evolution. *Proc Natl Acad Sci U S A.* 2003; 100:14966–14969. [PubMed: 14657359]
68. Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, Kamiyama M, Hruban RH, Eshleman JR, Nowak MA, Velculescu VE, Kinzler KW, Vogelstein B, Iacobuzio-Donahue CA. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature.* 2010; 467:1114–1117. [PubMed: 20981102]
69. Shibata D. Mutation and epigenetic molecular clocks in cancer. *Carcinogenesis.* 2011; 32:123–128. [PubMed: 21076057]

Fig. 1.
History and approaches for studying tumor evolution.

Table 1

Recent clonal evolution studies using next generation sequencing approaches.

Study	Technology	Cohort	Disease	Findings
Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. NEJM (2012)	WXS, SNP Array, mRNA array	4	metastatic renal-cell carcinoma, pre/post treatment	Intratumoral heterogeneity, metastasis, and treatment resistance emerge from distinct subclonal mutations within a single tumor
Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. Cancer Cell (2011)	SNP Array, FISH	350+206 (TC GA)	Glioblastoma (GBM)	Divergent subclones evolve from a common ancestor and accumulate independent drivers via mosaic tyrosine kinase amplifications
Inferring tumor progression from genomic heterogeneity. GR (2010)	CGH, macro dissection, flow sorting, FISH	20	Breast (primary ductal breast carcinomas)	Distinct heterogeneous clones expand from a common ancestor with accumulation of unique copy number and structural aberrations patterns
Tumour evolution inferred by single-cell sequencing. Nature (2011).	Single Cell WGS, CNV	2	Breast Cancer	Tumor growth is typified by punctuated clonal expansions accompanied by copy number changes that may lead to clonal divergence
Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell (2012)	WXS + Single Cell WXS	1	Kidney (Clear cell renal cell carcinoma - ccRCC)	Single cell sequencing identifies intratumoral heterogeneity with distinct mutation patterns but no single dominant clone.
Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. Cell (2012)	Single Cell WXS	1	Myeloproliferative Neoplasm (MPN)	Single cell sequencing establishes monoclonal origin of an essential thrombocythemia (ET).
The life history of 21 breast cancers. Cell (2012)	WGS	21	Breast Cancers, Various Subtypes	Typical breast tumors consist of one dominant clone with additional minor subclones sharing a common origin.
The origin and evolution of mutations in acute myeloid leukemia. Cell (2012)	WGS	24	AML M1/M3 primary tumors	Clonal origins of AML reflect a small number of driver mutations accompanied by of a larger number of random background mutations that accumulate in stem cells over time.
Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature (2010)	WGS	1	Breast Cancer (Basal-like – Primary, Metastasis, Xeno graft)	Patterns of clonal enrichment and divergence emerge in metastasis and xenograft from a primary tumor.
Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome	WGS	8	Acute myeloid leukemia relapse (AML)	Mutational patterns establish clonal expansion and tumor evolution in AML relapse (post chemotherapy)

Study	Technology	Cohort	Disease	Findings
sequencing. Nature 481 (2012)				