Looking beyond drivers and passengers in cancer genome sequencing data

S. De* & S. Ganesan

Center for Cancer Systems and Computational Biology, Rutgers Cancer Institute of New Jersey, New Brunswick, USA

*Correspondence to: Dr Subhajyoti De, Center for Cancer Systems and Computational Biology, Rutgers Cancer Institute of New Jersey, 195 Little Albany Street, Room No. 3550, New Brunswick, NJ 08901, USA. Tel: +1-732-235-8558; E-mail: subhajyoti.de@rutgers.edu

Cancer arises as a result of acquired changes in the DNA sequence of the genome of somatic cells. A subset of the genetic changes, dubbed driver mutations, propels tumor growth, and the remaining changes are passengers, apparently inconsequential for neoplastic transformation. Massive genome sequencing of thousands of tumors from all major cancer types has enabled cataloging of the so-called driver and passenger mutations, and facilitated molecular classification of cancer, guiding precision medicine approach for the patients. Nonetheless, innovative analyses of cancer genomics data has led to novel, sometimes serendipitous findings that have aided to our understanding of other aspects of the biology of the disease and opened up new frontiers. For instance, emerging findings show that mutational patterns in cancer genomes can help detect signatures of known and novel DNA damage and repair processes, provide a likely chronological account of genomic changes in cancer genomes, and allow revisiting the models of cancer evolution. These findings have stimulated original approaches to identify disease etiology, stratify patients, target the disease, and monitor patient responses, complementing driver-mutation centric approaches. In this review, we discuss these emerging approaches and unexpected breakthroughs, and their implications for basic cancer research and clinical practices.

Key words: genomics, sequencing, mutation signatures, cancer evolution, heterogeneity, precision medicine

Introduction

In 2015, the Cancer Genome Atlas (TCGA) [1], an NIH funded Herculean project to characterize genomes, transcriptomes, and epigenomes of tumors from almost all major cancer types, came to an end after genetic profiling of ~10 000 cases [2]. Completion of the TCGA, which has been complemented by initiatives such as the International Cancer Genome Consortium (ICGC) [3], marks the end of the beginning of an era and a paradigm shift in the approach to study this set of diseases that still remains a leading cause of morbidity and mortality globally.

It is now evident that cancer is not a single disease, but a group of diseases characterized by abnormal genomes, which carry acquired genetic changes such as point mutations, amplifications, deletions, translocations, complex rearrangements, ploidy changes, or a combination of those. During malignant transformation and tumor growth, a small subset of mutations known as ‘driver’ mutations are thought to be critical for driving malignant transformation and supporting the hallmarks of neoplastic growth, while the remaining alterations, dubbed ‘passengers’, are presumed to be benign or neutral i.e. they do not contribute to changes in cellular fitness [4, 5]. The dominant viewpoint, until recently, was that passenger mutations have little implications for understanding the biology and treatment of cancer, and a hunt for ‘drivers’ and drugs to target them primarily propelled cancer research; it is beginning to change.

The TCGA [1], ICGC [3], and other cancer genomics projects have greatly facilitated discovery of major oncogenic ‘drivers’ and candidate drug targets. But perhaps a major reward of genomic characterization of cancer was discovery of novel oncogenic ‘drivers’ and candidate drug targets. But perhaps a major reward of genomic characterization of cancer was discovery of novel oncogenic ‘drivers’ and candidate drug targets. But perhaps a major reward of genomic characterization of cancer was discovery of novel oncogenic ‘drivers’ and candidate drug targets. But perhaps a major reward of genomic characterization of cancer was discovery of novel oncogenic ‘drivers’ and candidate drug targets. But perhaps a major reward of genomic characterization of cancer was discovery of novel oncogenic ‘drivers’ and candidate drug targets. But perhaps a major reward of genomic characterization of cancer was discovery of novel oncogenic ‘drivers’ and candidate drug targets.
In 2007, a pioneering study sequenced exonic regions in 11 breast and 11 colorectal cancer samples, showcasing for the first time, that mutational landscapes of these cancers are characterized by the presence of a few, frequently mutated genes (‘mountains’) and an larger number of genes (‘hills’) that are mutated at low frequencies [6]. Although initial efforts were only able to robustly re-identify known cancer driver genes, it showed feasibility of cancer genome sequencing, and subsequent efforts by international consortia and investigator-initiated projects led to discovery of novel cancer genes including therapeutic targets. A classic example is IDH1, recurrent mutations in which were first discovered in glioblastoma [7], quickly followed by additional findings in other cancer types. Currently, the Cancer Gene Census [8] catalogs 595 driver genes that have recurrent oncogenic point mutation, copy number alteration, or fusions in one or more cancer types. In addition, there are hundreds of additional ‘potential driver’ genes cataloged, a number that continues to grow. This has prompted the question how many cancer driver genes we expect to identify. Analyzing mutation data from ~5000 human cancer samples across 21 major cancer types, and, based on a saturation analysis, it was predicted that nearly all of the key driver genes could be detected by sequencing 600–5000 samples per tumor type, depending on the background mutation frequency in respective cancer types [9]. Nevertheless, the data suggest that only a few cancer genes are mutated at a high frequency in any given tumor-type (>20%), while most others are mutated at low to intermediate frequencies (2–20%)—reiterating the decade-old observation that mutation landscape of cancer genomes typically has a few mountains and many more hills and mounds [6].

The concept of ‘driver’ and the expanding catalog of driver mutations are not without debate. Some of the early drivers were detected based on the frequency of mutations in high-throughput studies (e.g. 20/20 rule [10], which states that genes with ≥20% truncating mutations are tumor suppressors, whereas genes with >20% of missense mutations in recurrent positions are oncogenes). While frequency-based approaches are useful for prioritizing candidates for functional assessment, they are not definitive classifiers themselves. For instance, 20/20 rule correctly classifies most of the well-known cancer genes, but fails to identify other cases mostly the lowly recurrent ones [11]. Frequency-based classification of drivers without considering appropriate context has led to occasional false positives [12]. Nonetheless, some cancer mutations might be important for tumor initiation (e.g. APC in colon cancer [13]), while others could play a role in tumor growth (e.g. VEGF [14]) or metastasis. Function of some of the cancer genes such as TP53 could be more complex. Furthermore, evidence for extensive epistasis among cancer mutations [15] mean that their functional consequences could be context-dependent. For example, tumor cells with BRAF mutations are sensitive to PARP inhibition, leading to the clinical use of PARP inhibitors in ovarian cancer patients [16]. In addition, some of the somatic genomic alterations and classic cancer gene mutations could be detected in pathologically benign white blood cells (e.g. DNMT3A) and other normal tissues [17–20], suggesting that defining such genes as cancer drivers could be misleading, as their overall effect may depend on the presence of other cooperating genomic alterations. Taken together, driver-passerenger classification might be an over-simplification, and it would be more relevant to describe the cancer-associated genes and mutations in the light of the tumor hallmarks (e.g. apoptosis inhibitor, angiogenesis promoter, etc.).

Nonetheless, cancer gene-based approaches have facilitated molecular classification of cancer. Historically clinical management of cancer has been guided by the organ system classification, but the approach is beginning to change. Pan-cancer analyses [21] indicate that two tumors with same pathological classification and identical morphologic features can have different mutation profiles; yet cancers arising from two different organs can have highly similar set of driver events and mutational profiles, and that they might respond similar to targeted therapies—emphasizing the significance of molecular classification of cancer [22]. This paradigm accelerates our understanding of the biology of cancer transcending the boundary of tissue of origin, but also opens up possibilities that drugs approved for one cancer-type might also benefit patients with other types of cancer. This has propelled the concept of basket trials [23], where patients from different cancer types having the same oncogenic mutations are treated with the same drug in a clinical trial setting. An example of a large basket trial currently enrolling is the NCI Molecular Analysis for Therapy Choice (MATCH) trial [24], in which, >200 actionable genomic alterations are assessed in pretreated solid tumors and lymphomas of different histology and the patients are then paired to investigational drugs that target the mutation of interest. Although different cells may have the same driver mutation, signaling circuits present in different cell types may require different approaches to effectively target this driver mutation in different cancers. For example, feedback activation of EGFR signaling limits activity of BRAF inhibitors in BRAF V600E mutant colon cancer, and thus combined targeting of both BRAF and EGFR is required to effectively target BRAF V600E in colon cancer [25]. Cataloging of major drivers and molecular classification of cancer, which were the cornerstones of the cancer genome projects, are unsurprisingly transforming the way patients are stratified and treated, as discussed in the Precision Medicine section.

Cancer genomes typically have 10^2–10^6 somatic point mutations (10^-1–10^3 per Mb) [12], and up to ~10% of genome could be affected by focal genomic alterations [26], depending on the cancer histology. While a vast majority of these are passengers i.e. are thought to be bystander events that do not contribute to tumor genesis, associated mutational signatures can help detect patterns and extents of the underlying mutational process. These mutational signatures can identify different types of DNA damage and/or DNA repair defects, providing clues to disease etiology, and also timeline of cancer development in some cases (Figure 1).

The proof of principle of this approach was demonstrated in two seminal papers in 2010, which characterized mutation patterns of a malignant melanoma [27] and a small cell lung carcinoma [28] cell lines using whole genome sequencing, and assessed the mutation patterns in cancer genomes and compared them to the known patterns of mutations induced by carcinogenic exposures such as ultraviolet radiation or tobacco. Unsurprisingly, mutational signature
of ultraviolet radiation dominated the melanoma genome, while mutational signatures in the genome of lung cancer were consistent with effects of tobacco carcinogens. Genome sequences of melanoma typically have a massive burden of C > T and CC > TT substitutions at dipyrimidine context due to UV-related DNA damage, while acral melanomas have an excess of C > T transitions at CpG sites [27, 29, 30]. On the other hand, lung cancers arising in smokers tend to have an order of magnitude increase in mutation burden compared with that of non-smokers. This increased mutation burden is primarily due to an increase in G > T transversions, a result of tobacco-related carcinogens [28, 31–34]. Tobacco related signature was also predominant in head and neck cancer, a subset of which also carried signatures of human papilloma virus (HPV) infection [35]. Treatment with temozolomide in glioblastoma leads to an increase in somatic mutations, primarily C > T transitions. Exposure to aristolochic acid contributed to an excess of T > A transversions at CpTpG context and had a strong transcription strand bias in urothelial carcinoma [36]. These observations based on mutational signatures can corroborate epidemiological studies and independently help establish the impact of external mutagens at the molecular level. For instance, in the settings of case–control or cohort studies relative burden of specific types of mutation signatures in cancer patients could indicate effects of exposure to known or novel carcinogens arising from lifestyle, environmental or occupational exposure, or consequences of gene–environment interactions and guide strategies for evaluation of cancer risk, patient stratification, and also prevention.

Mutation burdens seen in cancer can arise not just from exposure to exogenous mutagens, but also can be the result of acquired defects in DNA repair or replication fidelity. Somatic mutation signatures from cancer genome sequencing can also help identify defects in the DNA replication and repair pathways intrinsic to the somatic cells [37]. High somatic mutation burden is a feature of impaired functions of DNA polymerase and co-factors involved in fidelity of replication and DNA repair. Recent studies have shown that germ line and somatic mutations in the proofreading exonuclease domains of the replicative DNA polymerases POLD and POLE are associated with several cancers. POLE mutant lung, endometrial, and colorectal cancers have defective proofreading function during DNA replication leading to an ultra-mutator phenotype and excess of transversions [38, 39]. An excessive mutation burden is also seen in cancer harboring defects in mismatch repair; these cancers have microsatellite instability and their mutational landscapes are dominated by C > T and C > A substitutions, and also InDels in the settings of extended repeats [40, 41]. Chronic lymphocytic leukemia and malignant B-cell lymphomas, which have somatic immunoglobulin gene hyper-mutated phenotype associated with Activation Induced Cytidine Deaminase (AICDA), tend to have T > G transversions at ApTpN and TpTpN context, and this signature is suspected due to the error-prone POLN [42]. Activity of APOBEC family enzymes can lead to a pattern of localized hypermutation, known as kataegis, which is often co-localized with regions of somatic genome rearrangements [43, 44]. APOBEC mutation signatures were serendipitous discoveries from genomic characterization of cancer genomes. Whole genome sequencing has revealed that certain breast cancer genomes are burdened with C > T and C > G mutations at CpG sites [45]. Furthermore, strand-bias in substitution patterns in coding regions identified transcription-coupled repair as an important source of somatic mutations [27, 37].

Genomic context of not only point mutations but also copy number alterations have been highly informative. Analysis of patterns of LOH events have led to development of a ‘homologous recombination defect’ or HRD score that may identify cancers that may be sensitive to PARP inhibitors or platinum agents [46]. Computational investigations of genomic alteration end-points
have enabled quantifying prevalence of different known DNA double-strand break repair signatures [47]. Assessment of break-point repair signatures identified in whole genome sequencing data from for 560 breast cancer samples has identified distinct and novel differences in signatures of DNA DSB repair defects for BRCA1 and BRCA2 mutant cancer samples [45]. Much of this analysis was performed using short-read sequencing which are not optimal for detecting complex genomic alterations, a limitation which could be mitigated using long-read sequencing. Nonetheless, it should be possible to extend the approach to elucidate functions of other DNA repair genes and detect novel DNA DSB repair signatures.

Other works have demonstrated that somatic mutations are non-randomly distributed in the genome and that features such as chromatin, replication timing, genomic context, and nuclear localization modulate mutational landscapes of cancer genomes (reviewed in [48, 49]). Mutation analyses of pathologically benign tissues have indicated that some of the classic cancer gene mutations, as well as signatures of external mutagens and DNA repair defects could also be found at the single clone level or in clonally expanded cell populations in apparently benign tissues [17, 19, 20, 50]—indicating that some of the cancer mutations and mutational signatures are not exclusive to cancer genomes.

**Molecular clock-like signatures and tumor age**

Mutations in cancer genomes serve as the records of the journey traveled by the clonal population during the course of normal development and neoplastic transformation, and also the intrinsic and extrinsic mutagenic processes they experienced during this period. It is now inferred that perhaps half or more of the somatic mutations in certain tumors occur before the onset of neoplasia [51]. Using mutation data from ~10 000 cancer genomes across 36 cancer types, it was also found that not only the absolute number of somatic mutations but also two mutational signatures have clock-like patterns (Figure 1), i.e. there was correlation between numbers of mutations carrying the signature and age of diagnosis [52]. The molecular clocks corresponding to both the signatures showed substantial variation between cancer types, and the two signatures were not strongly correlated across cancer types indicating that perhaps they indicate distinct biological attributes. It is suspected that one of the signatures might reflect the rate of cell division [52]. A similar clock-like signature for different tissue types was observed for DNA methylation as well [53]. It is unclear whether the genetic and epigenetic clocks are mechanistically linked. A major limitation in the field is lack of ways to validate or measure rate of cell division, cell death, and mutation rate per generation in individual cancers during the course of neoplastic growth. Nonetheless, if confirmed by independent studies, estimation of chronological age and the rate of cell division from mutational signatures based on genome sequencing data for a given tumor could offer important clues into the biology of the particular tumor, including the intrinsic and extrinsic mutagenic influences that prevailed during the period of tumor initiation and clonal evolution.

**Tumor heterogeneity and clonal architecture**

A major benefit of cancer genome sequencing was gaining molecular-level insights into intra-tumor heterogeneity and clonal architecture. At the time of detection, tumor typically has clonal populations of tumor cells that are not genetically identical, even though they trace their ancestry back to the tumor-initiating cell. In the last few years, the genomic approaches (reviewed in [54, 55]) have matured sufficiently to infer tumor purity, burden of clonal and subclonal mutations, and also predict evolutionary relationship between different sub-clones within the clonal architecture in a tumor (Figure 1). This has helped identify early (clonal) and late (subclonal) occurring cancer mutations. Furthermore, integrating cancer genomic and transcriptomic data, it is possible to estimate the abundance of stromal and immune cells and their activity, which in turn have implications for cancer immunotherapy [56]. Sequencing multiple different regions in tumors has helped investigate spatial heterogeneity in cancer. While these approaches are primarily applicable to bulk tumor genomic data, single cell sequencing, discussed elsewhere [57], has complemented these approaches, providing cell-by-cell assessment of mutational landscape and clonal architecture in tumors. We think integrative analysis of single-cell sequencing and multi-region sequencing would be powerful to understand patterns of micro- and macro-heterogeneity in tumor.

Preliminary results suggest that tumor clonal architecture may not conform to a single pattern [58]. In some tumors, one or two major clones dominate [59], while others have multiple co-dominant clones [60]. Furthermore, genetically distinct clones could be spatially segregated [60], or mixed [61] indicating higher mobility of the tumor cells. The extent and geographic distribution of immune cell infiltration also vary between cancers. It might be possible that some of the genetic heterogeneity is functionally inconsequential and hence effectively fitness neutral [58]. Nevertheless, patterns of clonal mixing and heterogeneity could indirectly indicate the extent of genomic instability, evolvability, and migratory potential of the clones, and hence would have clinical implications. Indeed, reports indicate that the patterns of intra-tumor heterogeneity could be an early indicator of clinical outcome [61–63].

**Emerging models of tumor evolution**

Studying tumor heterogeneity and clonal architecture is incomplete without understanding them in the light of tumor evolution. Integration of mathematical frameworks with cancer genome sequencing data has allowed revisiting the classical model of cancer evolution and proposing new models (Figure 2).

Integrating cross-sectional cancer genomic data with classical evolutionary model, it was [64] proposed that in glioblastoma, there were two common temporal sequences of events: RAS pathway activation followed by TP53 inactivation and PI3K activation and RAS activation before AKT activation. Combining tumor genome sequencing data with epidemiologic data, it was suggested that only three driver mutations could be sufficient for the formation of lung and colon cancer [65]. Vogelstein et al. [66] proposed that, on average, a driver mutation offers ~0.4% selective advantage. However, as we discussed above, not all driver mutations are equal,
Genomic data driven mathematical models have also been used to predict the timing and mode of metastasis. Based on the mathematical model and genomic data from primary and metastasis prostate cancer, it was proposed that between the occurrence of the initiating mutation and the birth of the metastasis founder cell, there might be at least a decade of waiting time, and that the founder cell might need ～5 years to acquire metastatic properties [75]. However, other studies in different cancer types have found similar drivers in paired primary and metastasis [76, 77], and even small invasive cancers often have metastatic potentials, suggesting that in other cases invasive phenotype is inseparable from the metastatic phenotype.

Going forward, the models could benefit from incorporating emerging data (e.g. single-cell sequencing) and concepts (e.g. chromothripsis). For instance, estimates of somatic mutation burden in tumor and locus-specific mutation rate from cancer genome projects [12] could benefit the models. Epistasis among the cancer gene mutations (and possibly other somatic mutations) seems common [15] and should be incorporated. Furthermore, tumor–stroma–immune cell interactions and their effects on both growth and immune surveillance can have major impact on clonal dynamics [78]. In addition, relative fitness (rather than absolute fitness) of a clone relative to other clones in a microenvironment influences its success in clonal competition in the tumor microenvironment [79]. Incorporating these features into tumor evolution models remains an active area of development. Nevertheless, while nuances of the models can be debated and some of the results are provocative, these studies offer the early foundations of genomic data-driven models of tumorigenesis, generating falsifiable predictions about disease progression and treatment outcomes in manner that was not possible in the pre-genomics era.

Implications for precision medicine

Genomics-driven discovery of novel driver mutations and molecular classification of cancer have accelerated rational design...
strategies for cancer prevention, patient stratification, development of new drugs, and treatment options in clinical settings (Figure 1). This concept is epitomized by the Precision Medicine Initiative, launched in 2015 in the United States [80] (and subsequently similar efforts were undertaken also in other countries). Many NCI-designated comprehensive cancer centers have adopted strategies to use targeted gene panel sequencing to characterize individual patient’s cancers, and recommend treatments based on their oncogenic mutation patterns. Computational methods (e.g. SomVarIUS [81]) can identify somatic mutations from targeted sequencing data from patient derived tumor samples without matched normals. Multiple CLIA-certified providers (e.g. Foundation Medicine [82]) now offer targeted sequencing of known cancer-associated genes in cancer in clinical settings with high reliability. Targeted sequencing has typically detected clinically actionable genomic alterations in three-fourth of cases [82] and can lead to implementation of targeted therapy in a significant subset of patients [83]. The efficacy of such an approach is greatly limited at present by the lack of effective treatments for most driver mutations; genomic sequencing alone cannot improve outcome without a corresponding development of new effective targeted therapeutics. Genomic classification of cancer will be crucial to guide development of clinical trials that can evaluate novel therapies.

Genomic approaches are also beginning to be used to monitor tumor progression and emergence of resistance during treatment. For instance, advances in cancer genomics has enabled identification of existing and acquired drug-resistant mutations, even when the resistant mutations are present at low allele frequencies [84, 85]. For instance, deep sequencing of target regions have been able to identify KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer as early as 10 months prior to radiographic documentation of disease progression [86]. In cases where serial samples can be obtained (e.g. chronic lymphocytic leukemia, liquid biopsy for solid tumors), tracking cancer-associated mutations during the course of treatment and relapse [87, 88] holds enormous potential to advance clinical management of the patients, significantly beyond the current standard of care. This has been extensively reviewed elsewhere.

The clinical benefit of cancer genome sequencing goes beyond the obvious targeting and monitoring of driver mutations. The burden of passenger mutations and their mutational signatures can be indicative of DNA repair defects, which can help identify the subset of tumors that are likely to respond to specific therapeutic strategies. Patterns of genomic changes consistent with an underlying defect in HR-mediated repair may identify cancers likely to respond to treatment with platinum agents or PARP inhibitors. Overall mutation burden may also identify patients likely to respond to immunotherapy [89]. It has been observed that tumors with excessive mutation burden, for example, those with microsatellite instability or POLE mutations tend to respond well to immunotherapy approaches such as anti-PLD-1 inhibitors and CART-based treatment [90–92].

Assessment of cancer genomic data have led to proposition of potential alternative treatment strategies that do not involve targeting the driver genes, and instead exploit other vulnerabilities in cancer genomes. Gatenby et al. [93] advocated a strategy to identify and target genes which are never mutated in cancer, and, therefore, could be essential for growth of neoplastic clone. However, some of those genes could also be crucial for normal tissue functions. Perhaps a more interesting approach is to target cancer-specific vulnerabilities that are the results of copy number losses, identifying and inhibiting synthetic lethal partner genes (CYCLOPS genes) in cancer [94].

Mutational patterns seen in normal or premalignant tissue may also reflect exposure to carcinogens, and may in the future be integrated in cancer screening. For examples, mutation signatures in cancer genomes provide mechanistic insights regarding the effects of carcinogens (e.g. UV, smoking), and when integrated with epidemiological data, such details could be informative about lifestyle, exposure, and environmental hazards contributing to cancer risk at the molecular level. The proposed Pre-Cancer Genome Atlas (PCGA) [95] makes a case for genomic characterization of pre-malignant lesions or early detection and risk stratification as well as allow for the development of novel targeted cancer interception strategies.

Outlook

Within the last decade, cancer genome sequencing has advanced both basic and clinical research in an unprecedented way, with both conventional results and serendipitous discoveries with far-reaching implications. We anticipate the trend to continue with the ICGC projects [3]. Genomics and transcriptomics reveal only one aspect of biological processes in cancer, which is a complex dynamic system. The focus of the research is now likely to shift from sequencing genomes to analyzing function of the mutations frequently detected in cancer [2]. Parallel advances in high throughput methods for studying post-translational modifications, metabolic changes, and systems-level alterations in the cellular processes [96] and also in the tumor-microenvironment will be critical for enhancing our understanding of the key aspects of the disease, and use that knowledge diagnosis, stratification, and treatment of cancer patients. At the same time, it would be important to consolidate the gains made from emerging, non-traditional analyses of mutational landscapes of cancer genomes that are shedding light in other aspects of the biology of cancer.

Acknowledgements

We thank members of the laboratories of De and Ganesan, and also the members of the Center for Systems Biology at Rutgers Cancer Institute for comments and suggestions.

Funding

The authors support from the Lung cancer Research Foundation (SD), Boettcher Foundation (SD), R01 CA142698 (SD), P30 CA072720 (SD and SG) and Rutgers Cancer Institute of New Jersey. The funders had no role in study design, data collection, and interpretation, or the decision to submit the work for publication.

Disclosure

The authors have declared no conflicts of interest.
References


72. Mardin BR, Drainas AP, Waszak SM et al. A cell-based model system

71. Zhang CZ, Spektor A, Cornils H et al. Chromothripsis from DNA dam-

70. Korbel JO, Campbell PJ. Criteria for inference of chromothripsis in can-

67. Castro-Giner F, Ratcliffe P, Tomlinson I. The mini-driver model of poly-

66. Bozic I, Antal T, Ohtsuki H et al. Accumulation of driver and passenger

64. Cheng YK, Beroukhim R, Levine RL et al. A mathematical methodology


60. Gerlinger M, Rowan AJ, Horswell S et al. Intratumor heterogeneity and


58. Barber LJ, Davies MN, Gerlinger M. Dissecting cancer evolution at the

57. Navin NE. The first five years of single-cell cancer genomics and beyond. 

and infiltrating immune cells across human cancers. Nat Med 2015; 21: 
938–945.

55. Ding L, Wendl MC, McMichael JF, Raphael BJ. Expanding the computa-
tional toolbox for mining cancer genomes. Nat Rev Genet 2014; 15:
556–570.

54. Yadav VK, De S. An assessment of computational methods for estimating
purity and clonality using genomic data derived from heterogeneous 

53. Horvath S. DNA methylation age of human tissues and cell types. 
Genome Biol 2013; 14: R115.

52. Alexandrov LB, Jones PH, Wedge DC et al. Clock-like mutational proc-
tations are required for the development of lung and colorectal cancers.

51. Tomasetti C, Vogelstein B, Parmigiani G. Half or more of the somatic 
mutations in cancers of self-renewing tissues originate prior to tumor 


49. Castro-Giner F, Marchionni L, Nowak MA et al. Only three driver gene 
mutations are required for the development of lung and colorectal cancers. 
Proc Natl Acad Sci USA 2015; 112: 118–123.

mutations on cancer progression. Proc Natl Acad Sci USA 2013;
110: 2910–2915.

47. Cheng YK, Beroukhim R, Levine RL et al. A mathematical methodology
for determining the temporal order of pathway alterations arising during

and infiltrating immune cells across human cancers. Nat Med 2015; 21: 
938–945.

45. Ding L, Wendl MC, McMichael JF, Raphael BJ. Expanding the computa-
tional toolbox for mining cancer genomes. Nat Rev Genet 2014; 15:
556–570.

44. Yadav VK, De S. An assessment of computational methods for estimating
purity and clonality using genomic data derived from heterogeneous 

43. Horvath S. DNA methylation age of human tissues and cell types. 
Genome Biol 2013; 14: R115.

42. Alexandrov LB, Jones PH, Wedge DC et al. Clock-like mutational proc-
tations are required for the development of lung and colorectal cancers.

41. Tomasetti C, Vogelstein B, Parmigiani G. Half or more of the somatic 
mutations in cancers of self-renewing tissues originate prior to tumor 


mutations are required for the development of lung and colorectal cancers. 
Proc Natl Acad Sci USA 2015; 112: 118–123.

38. Bozic I, Antal T, Ohtsuki H et al. Impact of deleterious passenger
mutations on cancer progression. Proc Natl Acad Sci USA 2013;
110: 2910–2915.

37. Stephens PJ, Greenman CD, Fu B et al. Massive genomic rearrangement
acquired in a single catastrophic event during cancer development. Cell 

36. Korbel JO, Campbell PJ. Criteria for inference of chromothripsis in cancer 

35. Zhang CZ, Spektor A, Cornils H et al. Chromothripsis from DNA dam-

34. McFarland CD, Korolev KS, Kryukov GV et al. Impact of deleterious passenger
mutations on cancer progression. Proc Natl Acad Sci USA 2013;
110: 2910–2915.

33. Stephens PJ, Greenman CD, Fu B et al. Massive genomic rearrangement
acquired in a single catastrophic event during cancer development. Cell 

32. Korbel JO, Campbell PJ. Criteria for inference of chromothripsis in cancer 

31. Zhang CZ, Spektor A, Cornils H et al. Chromothripsis from DNA dam-


29. Gao R, Davis A, McDonald TO et al. Punctuated copy number evolution 
and clonal stasis in triple-negative breast cancer. Nat Genet 2016; 48:
1119–1130.


27. Yachida S, Jones S, Bozic I et al. Distant metastasis occurs late during the 

26. Brannon AR, Vakiani E, Sylvester BE et al. Comparative sequencing ana-
lysis reveals high genomic concordance between matched primary and 

25. Turalić S, Furney SJ, Lambros MB et al. Whole genome sequencing of 
matched primary and metastatic acral melanomas. Genome Res 2012;

24. Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in 


22. Smith KS, Yadav VK, Pei S et al. SomVarIUS: somatic variant identifica-

21. Frampton GM, Fichtenholtz A, Otto GA et al. Development and valid-
ation of a clinical cancer genomic profiling test based on massively paral-

20. Hirshfield KM, Tolkunov D, Zhong H et al. Clinical actionability of com-
prehensive genomic profiling for management of rare or refractory can-

19. Muller A, Cheng YK, Kulis K et al. A mathematical framework for 
visualizing tumor evolution in the presence of clonal stability. Nat 