

Published in final edited form as:

Curr Opin Genet Dev. 2014 February ; 0: 52–60. doi:10.1016/j.gde.2013.11.014.

Mutational Signatures: The Patterns of Somatic Mutations Hidden in Cancer Genomes

Ludmil B. Alexandrov¹ and Michael R. Stratton¹

¹Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA

Abstract

All cancers originate from a single cell that starts to behave abnormally due to the acquired somatic mutations in its genome. Until recently, the knowledge of the mutational processes that cause these somatic mutations has been very limited. Recent advances in sequencing technologies and the development of novel mathematical approaches have allowed deciphering the patterns of somatic mutations caused by different mutational processes. Here, we summarize our current understanding of mutational patterns and mutational signatures in light of both the somatic cell paradigm of cancer research and the recent developments in the field of cancer genomics.

INTRODUCTION

Long before the discovery of the double helix [1], it was well established that ultraviolet light (UV) can cause tumours of the skin [2]. While the mechanism was unclear at this time, it was hypothesized that successive doses of UV radiation result in accelerating the relative rate of cell proliferation [3]. The paradigm shifting discovery that the genetic material is contained within a deoxyribonucleic acid led to many studies in the late 1950s and throughout the 1960s examining how organisms protect their DNA from endogenous and exogenous mutations, and a focus was given to ultraviolet induced mutations (reviewed in ref [4]). It was established that exposure to UV light can lead to the formation of dimers of any two adjacent pyrimidine bases on the same DNA strand with a preference for thymine-thymine dimers [4]. It was further shown that UV irradiation damage predominantly results in cytosine to thymine or cytosine-cytosine to thymine-thymine changes, preferentially occurring at these pyrimidine dimers (*i.e.*, C>T or CC>TT DNA mutations at dipyrimidine sites) [5,6]. This was the first detailed characterization of the pattern of DNA changes occurring due to the activity of an exogenous mutagen and, as such, the very first description of a signature of a mutational process.

While these early studies established the mutational signature of UV light, it was unclear whether UV induced mutations are present and involved in the neoplastic expansion of human cancers. The development of the DNA sequencing technique with chain-terminating inhibitors by Fred Sanger [7] allowed rapid examination of the genetic material contained in

Address for correspondence: Ludmil B. Alexandrov, Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, United Kingdom, la2@sanger.ac.uk.

cancer cells. In the early 1990s, two studies sequenced exons of the gene *TP53* [8,9] from several patients and provided experimental evidence that aflatoxin and UV light leave distinct patterns (consistent with the ones observed in experimental systems) of DNA mutations respectively in hepatocellular and squamous-cell carcinomas. These studies confirmed that the mutational signatures of carcinogens are left as “evidence” in the genomes of cancer cells [10] thus spawning research which first examined the mutations across *TP53* and later across multiple genes and even whole cancer genomes in order to provide a better understanding of the mutational processes involved in human carcinogenesis.

MUTATIONAL PATTERNS OF *TP53*

Multiple independent studies used Sanger sequencing of some (or all) exons of a cancer gene to provide clues to the aetiology of both endogenous and exogenous factors of human carcinogenesis. *TP53* was usually selected for this analysis due to its high prevalence of somatic mutations in almost all tumour classes [11]. Commonly, each of these studies involved multiple samples of a cancer type that were examined for somatic mutations in *TP53* (studies reviewed in refs [11-13]). The *TP53* somatic mutations were aggregated, their spectrum was reported as specific for the given cancer type, and this spectrum was then compared to mutations generated experimentally in *in vitro* or *in vivo* systems [11,13]. It should be noted that the mutational spectra of other genes, albeit rarely, were also used for such analysis [14].

These early studies revealed a significant heterogeneity of the *TP53* spectra across different cancer types, which allowed associating some patterns of mutation to known carcinogens. Here, we provide a brief summary of some of the more important findings while details could be found in refs [11-13]. The *TP53* spectrum of skin carcinomas exhibited C>T and CC>TT mutations at dipyrimidines (all substitutions and dinucleotide substitutions are referred to by the pyrimidine(s) of the mutated Watson-Crick base pair). This was consistent with the *in vitro* described mutational signature of UV light. The *TP53* mutational spectrum derived from lung cancers in tobacco smokers was overwhelmed by C>A substitutions, which coincided with the class of mutation produced experimentally as a result of bulky adduct formation by tobacco carcinogens on guanine [15]. In other tobacco associated cancers, such as oesophageal and head and neck tumours, C>A mutations (while still ubiquitous) were less common while there was a significant increase of T>C mutations. Interestingly, in both smokers and non-smokers, C>T and C>G mutations at non-CpG sites were elevated when compared to all other cancer types, with bladder tumours harbouring the most C>G mutations [11]. Additionally, it was demonstrated that C>A transversions were common in hepatocellular cancers and these mutations were believed to be associated with aflatoxin, a known carcinogen commonly found in food from southern Africa and Asia [16]. Lastly, all cancer types harboured at least some C>T mutations at CpG dinucleotides (mutated base underlined), a process attributed to the normal cellular event of deamination of 5-methylcytosine [11].

The analyses of *TP53* spectra were the first attempts to bridge the gap between molecular cancer genetics and epidemiology [17]. The large number of studies examining *TP53* spectra

required a computational resource to facilitate and retrieve the already identified somatic mutations. At first these data were managed by the researchers that were generating it but in 1994 the International Agency for Research on Cancer (IARC) started to maintain a database while providing a free access to it [17]. The first release of the IARC *TP53* database contained ~3,000 somatic mutations [18] while the most recent version (R16) released in November of 2012, which can be found at <http://p53.iarc.fr/>, contains almost 30,000 somatic mutations in *TP53*.

Though extremely informative, the data gathered from single gene studies have significant limitations. In these studies, the spectrum of a cancer type is reported by aggregating mutations from multiple samples. This may be adequate when a single mutational process generates the majority of mutations in the particular cancer (*e.g.*, UV light is the predominant mutational process in melanoma [19]). However, usually multiple mutational processes are operative in a single cancer sample, and combining their mutations generates a mixed composition of the patterns of somatic mutations. In most cases, reporting this jumbled spectrum is uninformative for the diversity of mutational processes operative in a single cancer type or in a single cancer sample [20]. Moreover, the examined *TP53* exons are both under selection and also have a specific nucleotide sequence. This affects the opportunity for observing a somatic mutation and as such the reported spectrum can be a reflection of the processes of selection and/or the nucleotide architecture of the *TP53* gene in addition to the processes of mutation [21,22].

Two studies tried to overcome some of the single gene limitations by leveraging a targeted capillary sequencing approach of large number of genes. A survey of the 518 protein kinase genes in 25 human breast cancer samples revealed 92 somatic mutations (90 substitutions and 2 indels) in which C>T transitions and C>G transversions preceded by thymine (*i.e.*, C>T and C>G at TpC, mutated base is underlined) occurred with a higher than expected frequency [23]. This survey was later expanded to 210 cancer samples and it revealed more than 1,000 somatic mutations with significant variations in their patterns across the examined twelve cancer types [24]. Only a small fraction of the mutations reported in these screens are likely to be affected by selection [25], thus indicating that the observed mutational patterns reflect the operative mutational processes in the analyzed samples and not the processes of negative or positive selection.

MUTATIONAL PATTERNS IDENTIFIED IN NEXT GENERATION SEQUENCING DATA

The development of second-generation sequencing technologies allowed examination of cancer exomes (*i.e.*, the combined protein coding exons) and even whole cancer genomes. Sequencing cancer exomes has been generally preferred as the majority of known cancer-causing driver somatic substitutions, indels, and copy number changes (although generally not rearrangements) [21] are located in protein coding genes. As the nucleotide sequence of protein coding genes is ~1% of the whole genome, analysis of exomes is considered an advantageous and cost effective methodology for discovering the genes involved in neoplastic development. As a result, many studies have focused predominantly on the generation and analysis of exome sequences [26].

Early next generation sequencing studies started revealing patterns of somatic substitutions in different cancer types. In 2010, two back-to-back studies in *Nature* reported the patterns of somatic mutations in a malignant melanoma [27] and small cell lung carcinoma [28]. As expected, a strong signature of tobacco carcinogens was found in the genome of the lung cancer, while the mutational signature of ultraviolet light overwhelmed the melanoma genome. These studies demonstrated the value of whole genome sequencing for evaluating signatures of mutational processes by providing greater resolution and mechanistic insight into mutational signatures due to known carcinogens, for example through the identification of a lower prevalence of mutations over the footprints of genes.

Multiple independent studies and international consortiums started sequencing large numbers of samples from both cancer genomes and exomes [26]. An integrated genomic characterization was reported for many different cancer types including: acute lymphoblast leukemia [29-31], acute myeloid leukemia [32], breast cancer [33-35], chronic lymphocytic leukemia [36,37], colorectal cancer [38,39], esophageal cancer [40], glioblastoma [41], cancers of the head and neck [42,43], kidney cancer [44-46], liver cancer [47,48], lung cancer [49-54], lymphomas [55,56], melanoma [57-60], multiple myeloma [61], ovarian cancer [62], pancreatic cancer [63,64], prostate cancer [65-68], stomach cancer [69-71], uterine cancer [72], and several different types of pediatric tumors [73-79]. While these studies focused on the identification of novel cancer genes, mutational spectra were usually reported for each of the examined samples and some studies even tried to associate certain types of somatic mutations with the activity of mutagens or the failure of DNA repair mechanisms. A brief summary of the mutational patterns identified in these cancer genomics studies is provided in the next paragraph.

In lung cancer, comparison between tobacco smokers and non-smokers revealed that smokers have on average 10-fold increase in the burden of somatic mutations in their cancer genomes [50,51]. Consistent with the experimental evidence for tobacco carcinogens, this elevation is mainly due to the increase of the number of C>A transversions [15]. Examination of the cancer genomes of melanomas confirmed that the majority of mutations are C>T and CC>TT at dipyrimidines in the ultraviolet-associated tumors, while acral melanomas exhibit predominantly C>T transitions at CpG sites [59,60]. In glioblastoma multiforme, it was demonstrated that treatment with an alkylating agent, such as temozolomide, significantly elevates the numbers of somatic mutations and results in a distinct mutational pattern of C>T transitions [41]. In chronic lymphocytic leukemia, it was observed that samples with mutations in the immunoglobulin genes have a higher proportion of T>G transversions [36]. This mutational pattern and its immediate sequencing context are consistent with the activity of the error-prone polymerase η during somatic hypermutation [36,80]. In endometrial and colorectal tumors, a set of ultra-hypermutators with increased mutational frequency of transversions was associated with somatic mutations in polymerase ϵ [44,72]. Microsatellite unstable gastric cancer were observed to have a higher mutation prevalence of both C>T transitions and C>A transversions [71]. Examining the cancer exomes of patients with urothelial carcinoma (of the upper urinary tract) revealed a large number of somatic mutations with an unique pattern of T>A transversions predominately located at CpTpG sites and possessing a very strong transcription strand bias [81]. This

pattern of mutations was associated with exposure to aristolochic acid. In esophageal cancer, a high prevalence of T>G transversions was observed [40] while certain breast cancer genomes were found to be overwhelmed with C>T and C>G mutations at TpC sites [35].

These next generation sequencing studies provided an unbiased look into the patterns of DNA changes across cancer genomes. While they resolved some of the previous limitations from *TP53* studies (mostly by examining large portions of the human genome which are usually not under selection and which have a nucleotide context that is representative of the whole human genome) they still did not address the important issue of examining mixtures of mutations generated by different mutational processes.

MUTATIONAL SIGNATURES DERIVED FROM PATTERNS OF SOMATIC MUTATIONS

The somatic mutations in a cancer genome are the cumulative result of the mutational processes that have been operative since the very first division of the fertilized egg, from which the cancer cell was derived [21,22]. Each of these mutations was caused by the activity of endogenous and/or exogenous mutational processes with different strengths (Figure 1). Some of these processes have been active throughout the whole lifetime of the cancer patient while others have been sporadically triggered, for example, due to lifestyle choices (Figure 1). While examining patterns of somatic mutations can provide an indication of the aetiology of the operative mutational processes, it doesn't allow deciphering the individual mutational signatures that are operative in each sample as usually the pattern of a sequenced cancer genome does not resemble any of the operative mutational processes (Figure 1).

Recently, a theoretical model and computational framework that allows decomposing distinct patterns of somatic mutations from a set of cancer samples was developed [20]. The mathematical model was an extension of the well-known blind source separation problem, in which original signals need to be separated from a set of mixed signals [82], and the algorithm was based on a method used in face recognition software that allows meaningfully learning distinct parts of objects [83]. The algorithm deciphers the minimal set of mutational signatures that optimally explains the proportion of each mutation type found in each cancer sample and then the method estimates the contribution of each signature to each cancer sample (see ref [20] for more details about this method, including a discussion of its limitations).

Initial application of this approach was performed on the somatic substitutions derived from the whole genomes of 21 breast cancer patients [33]. In order to increase the resolution of the derived mutational signatures, substitutions were examined using their immediate sequencing context. This included the base immediately 5' before to the somatic mutation and the base immediately 3' after the somatic mutation; thus resulting in 96 mutation types - 16 different trinucleotides for each of the 6 types of somatic substitutions. For example, C>T mutations were extended to include C>T with (5' adenine): ApCpA, ApCpC, ApCpG, ApCpT; (5' cytosine): CpCpA, CpCpC, CpCpG, CpCpT; (5' guanine): GpCpA, GpCpC, GpCpG, GpCpT; and (5' thymine): TpCpA, TpCpC, TpCpG, TpCpT. Including the

immediate sequence context allows better differentiation between different mutational processes; for example, distinguishing between C>T mutations due to the formation UV-light induced photodimers (i.e., C>T mutations at dipyrimidine sites such as TpCpC or CpCpC) from C>T mutations due to deamination of 5-methylcytosine (i.e., C>T mutations at CpG sites). The mutational catalogues of the 21 breast cancer genomes were generated, including each of the 96 mutation types, and applying the newly developed method to these catalogues revealed multiple distinct mutational signatures of substitutions. As expected, a mutational signature with features of C>T mutations at CpG sites was identified in most samples, thus reflecting the activity of normal endogenous cellular processes. Further, a mutational signature with C>X mutations at TpC sites was identified and based on similarity between its mutational pattern and *in vivo* experimental data, it was proposed that this process is due to the activity of the APOBEC family of deaminases and more specifically APOBEC1, APOBEC3A, and/or APOBEC3B [84,85]. Additionally, a rather uniform mutational signature (no prominent features across trinucleotides) was also identified and, interestingly, the activity of this mutational signature in each of the 21 samples allowed separation (by unsupervised hierarchical clustering) of *BRCA1* and *BRCA2* wild-type breast tumours from *BRCA1* and *BRCA2* germline mutants. Another mutational signature with unknown aetiology and mutations predominately at C>G at TpC was also identified. In addition to these genome-wide signatures, a localized hypermutation (termed *kataegis*) was observed in some of the breast cancer samples. This localized hypermutation was predominantly constituted of C>T and C>G substitutions at TpC trinucleotides and it was speculated that it is also due to the activity of the APOBEC enzymes. Lastly, deciphering the independent mutational signatures operative in these breast cancer samples provided the means for timing their activity across different cancer sub-clones [86].

This initial analysis of the mutational signatures operative in the 21 breast cancer genomes revealed several intriguing mutational processes but its focus was predominantly on substitutions. However, the newly developed approach for deciphering mutational signatures also allows extending mutational signature analysis over an arbitrary selected set of biologically meaningful mutation types [20]. To demonstrate its applicability, the mutational catalogues of the 21 breast cancer genomes were extended to include double nucleotide substitutions, indels at microhomologies, indels at mono/polynucleotide repeats, and even a complex mutation type such as *kataegis*. Reanalysing these mutational catalogues demonstrated that *kataegis* separates as its own mutational process. Further, double nucleotide substitutions and indels at microhomologies associated predominantly with the activity of the previously identified uniform mutational process. Lastly, indels at mono/polynucleotide repeats did not strongly associate with any of the previously described mutational processes [20].

Extending the previously defined mutational catalogues illustrated the possibility of incorporating additional mutation types and it revealed some associations between substitutions and indels thus providing more biological insight into the identified mutational processes [20]. Further biological insight was derived by analysing mutational catalogues that incorporate the transcriptional strand on which a substitution resides in the footprints of a gene. Thus, the previously defined 96 substitution types were extended to 192 mutation

types. For example, the number of C>T mutations at TpCpA were split into two categories: the number of C>T mutations at TpCpA occurring on the untranscribed strand of a gene and the number of C>T mutations at TpCpA occurring on the transcribed strand. In general, one would expect that these two numbers are approximately the same unless the mutational processes are influenced by activity of the transcriptional machinery. This could happen, for example, due to recruitment of the transcription-coupled component of nucleotide excision repair (NER) [87]. If a mutational process has a higher number of C>A substitutions on the transcribed strand compared to the C>A substitutions on the untranscribed strand (*i.e.*, note that C>A mutations on the untranscribed strand is the same as G>T mutations on the transcribed strand), this could indicate that the mutations caused by this process are being repaired by NER. As such, this analysis provides a further insight into the operative mutational processes and their interaction with cellular repair processes. A known example of such strand bias due to interplay between a mutational process and a repair mechanism is the formation of photodimers due to UV-light exposure that are repaired by NER and result in a higher number of C>T mutations on the untranscribed strand [87].

Analysing the transcriptional strand bias of the mutational signatures operative in the 21 breast genomes revealed a weak strand bias of C>A mutations with unknown aetiology [20]. Interestingly, deciphering mutational signatures from 100 breast cancer exomes revealed exactly the same trinucleotide mutational signatures but with a different strand bias. Specifically, there was an elevation of C>X mutations at TpCpT on the transcribed strand of exomes, which was absent in the complete gene footprints derived from the 21 whole genome sequences [20]. This transcriptional strand bias could be indicative of exon-specific repair processes that are active in the cell.

The extensive mutational signature analysis performed on the 21 breast cancer genomes was recently expanded and mutational signatures (including substitutions, indels, dinucleotide substitutions, *kataegis*, and strand bias) were deciphered from 30 different types of human cancer [19]. The previously developed computational framework was applied to almost 5 million somatic mutations identified in 7,042 cancer samples (507 from whole genome and 6,535 from whole exome sequences). This included both previously published samples and newly sequenced whole genomes. The analysis revealed 27 distinct mutational signatures [19]. 22 of these 27 mutational signatures were validated (*i.e.*, confirmed by orthogonal technologies or other approaches), 3 were associated with technology-specific sequencing artefacts, and 2 of the mutational signatures remain un-validated due to the lack of access to the relevant biological samples.

This largest cancer genomics analysis to date provided the first global roadmap describing the signatures of mutational processes in human cancer. Each of the cancer types had at least two mutational signatures operative in it, while some (*e.g.*, cancers of the liver and uterus) had up to 6 distinct mutational processes. Remarkably, most of the cancer samples had at least two mutational signatures active in them. Aetiology was proposed for 11 of the 22 validated mutational signatures. Two of the mutational signatures were associated with age of patient at cancer diagnosis and these signatures were present in 26 of the 30 cancer types and more than 70% of the samples. These two processes exhibit clear features of C>T at CpG sites and most likely reflect mutations due to normal cellular processes (*e.g.*,

deamination of 5-methylcytosine, errors due to DNA replication, *etc.*) and probably account for the majority of somatic mutations prior to neoplastic development.

Based on similarity with *in vivo* experimental data, two mutational processes (termed Signature 2 and 13) were associated with the activity of the APOBEC family of deaminases. These two signatures exhibit predominantly C>T and C>G mutations at TpC sites and were observed in 16 of the 30 cancer types (~17% of all examined cancer samples) [19]. As such, the activity of these mutational signatures (and respectively the APOBEC enzymes) is one of the most significant human carcinogens with prevalence superseding that of tobacco smoking and exposure to UV light. Recently, further evidence was provided for the involvement of APOBEC3B in human cancers, as its expression was elevated in tumours compared to their matched normal samples [88,89].

By comparing the substitution patterns of all signatures with experimental data, one of the mutational signatures was associated with exposure to ultraviolet light while another with benzo[a]pyrene, a known tobacco carcinogen. The signature associated with UV-light exhibited a higher presence of CC>TT dinucleotide substitutions as well as a strand bias indicative of the formation of photodimers, which further confirmed the association. In contrast, a mutational signature associated in lung cancer exhibited predominantly C>A mutations with a transcriptional strand bias suggesting the formation of bulky adducts on guanine. Interestingly, this mutational signature was also associated with CC>AA dinucleotide substitutions with a strong strand bias. Statistical tests comparing smokers with non-smokers in two cancer types (*viz.*, lung adenocarcinoma and tumours of the head and neck) confirmed a highly significant elevation of this “tobacco smoking signature” in smokers indicating that it was due to tobacco mutagens.

Further statistical analysis was performed to associate mutations in genes with the presence of mutational signatures. Distinct mutational signatures were associated with: 1) mutations in BRCA1/2 in breast and pancreatic cancers; 2) failure of the DNA mismatch repair pathway (*e.g.*, due to methylation of the MLH1 promoter) in colorectal cancers; 3) hypermutation of the immunoglobulin gene in CLL; 4) recurring polymerase ϵ mutations in uterine and colorectal cancers. Interestingly, the mutational signature associated with failure of DNA mismatch repair was observed in 9 different cancer types. While this process was operative in ~20% of colorectal cancers and ~15% of uterine cancers, it was also found in at least 1% of cancer samples in another 7 cancer types. Another interesting observation was that while almost all BRCA1/2 mutants exhibit a specific mutational signature, there were also BRCA1/2 wild-type samples with high number of mutations due to this mutational process. Thus, it is possible that some BRCA1/2 wild-type samples might harbour somatic mutations or other abnormalities that result in a failure of homologous repair and activation of this mutational process.

Chemotherapy treatment could cause its own set of somatic mutations[24]. Examining the pre-treatment history of all 7,042 cancer samples revealed that melanomas and glioblastomas pre-treated with an alkylating agent exhibit a distinct mutational signature.

The performed global analysis was able to propose an association for 11 of the 22 validated mutation signatures, while the origins and aetiology of the other 11 mutational signatures remains unknown. Lastly, this study also examined the presence of loci of *kataegis* across human cancer and it revealed that *kataegis* is not confined only to breast cancer but it is also present in at least another 7 cancer types including pancreas, lung, liver, medulloblastoma, CLL, B-cell lymphomas, and ALL.

CONCLUSIONS AND FUTURE PROMISES

In the past five decades, analysis of mutational patterns has evolved from *in vitro* observation of DNA changes caused by ultraviolet light, to examination of the mutational spectra generated by sequencing single cancer genes in multiple samples, to performing targeted capillary sequencing screens of multiple genes across hundreds of samples, and more recently to large-scale analysis of the genomes of thousands of cancer patients revealing the signatures of the mutational processes involved in the development of their tumours. In the next decade, thousands of new whole cancer genomes across the majority of cancer types [26] will be generated, which will allow the creation of a final and comprehensive map of mutational signatures. The generation of such a mutagenesis map will most likely require the refinement of existing mathematical methods to accurately examine all known types of somatic mutations: substitutions, indels, copy number variations, structural rearrangements, and potentially even epigenetic changes. These analyses of next generation sequencing data must be complemented with experimental work revealing the aetiology of the identified mutational processes.

REFERENCES

1. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*. 1953; 171(4356):737–738. [PubMed: 13054692]
2. Rusch HP, Baumann CA. Tumor production in mice with ultraviolet irradiation. *Am J Cancer*. 1939; 35:55–62.
3. Blum HF. On the mechanism of cancer induction by ultraviolet radiation. *Journal of the National Cancer Institute*. 1950; 11(3):463–495. [PubMed: 14824912]
4. Witkin EM. Ultraviolet-induced mutation and DNA repair. *Annual review of microbiology*. 1969; 23:487–514. [* This review summarizes the very early efforts to understand UV-light induced DNA mutations.]
5. Howard BD, Tessman I. Identification of the altered bases in mutated single-stranded DNA. Ii. *In vivo* mutagenesis by 5-bromodeoxyuridine and 2-aminopurine. *Journal of molecular biology*. 1964; 9:364–371. [PubMed: 14202272]
6. Setlow RB, Carrier WL. Pyrimidine dimers in ultraviolet-irradiated dna's. *Journal of molecular biology*. 1966; 17(1):237–254. [PubMed: 4289765]
7. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chainterminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*. 1977; 74(12):5463–5467. [PubMed: 271968]
8. Brash DE, Rudolph JA, Simon JA, Lin A, McKenna GJ, Baden HP, Halperin AJ, Ponten J. A role for sunlight in skin cancer: Uv-induced p53 mutations in squamous cell carcinoma. *Proceedings of the National Academy of Sciences of the United States of America*. 1991; 88(22):10124–10128. [PubMed: 1946433] [* This article provides one of the early evidences that UV light leaves a specific mutational imprint the DNA of a cancer cell, viz., C>T substitutions at dipyrimidines. This imprint is consistent with the one observed in previous *in vitro* experiments.]

9. Ozturk M. P53 mutation in hepatocellular carcinoma after aflatoxin exposure. *Lancet*. 1991; 338(8779):1356–1359. [PubMed: 1682737] [* This article provides one of the early evidences that aflatoxin leaves a specific mutational imprint on the DNA of a cancer cell, viz., C>A substitutions. This imprint is consistent with the one observed in previous experiments.]
10. Vogelstein B, Kinzler KW. Carcinogens leave fingerprints. *Nature*. 1992; 355(6357):209–210. [PubMed: 1309937]
11. Greenblatt MS, Bennett WP, Hollstein M, Harris CC. Mutations in the p53 tumor suppressor gene: Clues to cancer etiology and molecular pathogenesis. *Cancer research*. 1994; 54(18):4855–4878. [PubMed: 8069852] [** This very detailed review summarizes the studies analyzing *TP53* mutational spectra and their relationship with epidemiology.]
12. Hollstein M, Sidransky D, Vogelstein B, Harris CC. P53 mutations in human cancers. *Science*. 1991; 253(5015):49–53. [PubMed: 1905840]
13. Hollstein M, Hergenbahn M, Yang Q, Bartsch H, Wang ZQ, Hainaut P. New approaches to understanding p53 gene tumor mutation spectra. *Mutation research*. 1999; 431(2):199–209. [PubMed: 10635987]
14. Capella G, Cronauer-Mitra S, Pienado MA, Perucho M. Frequency and spectrum of mutations at codons 12 and 13 of the c-k-ras gene in human tumors. *Environmental health perspectives*. 1991; 93:125–131. [PubMed: 1685441]
15. Rodin SN, Rodin AS. Origins and selection of p53 mutations in lung carcinogenesis. *Seminars in cancer biology*. 2005; 15(2):103–112. [PubMed: 15652455]
16. Wogan GN. Aflatoxins as risk factors for hepatocellular carcinoma in humans. *Cancer research*. 1992; 52(7 Suppl):2114s–2118s. [PubMed: 1311989]
17. Hainaut P, Olivier M, Pfeifer GP. *Tp53* mutation spectrum in lung cancers and mutagenic signature of components of tobacco smoke: Lessons from the iarc tp53 mutation database. *Mutagenesis*. 2001; 16(6):551–553. author reply 555–556. [PubMed: 11682648]
18. Cariello NF, Cui L, Beroud C, Soussi T. Database and software for the analysis of mutations in the human p53 gene. *Cancer research*. 1994; 54(16):4454–4460. [PubMed: 8044795]
19. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, Boyault S, et al. Signatures of mutational processes in human cancer. *Nature*. 2013; 500(7463):415–421. [PubMed: 23945592] [** This study provides the first comprehensive map of the signatures of mutational processes across 30 different types of human cancer derived from 7,042 patients.]
20. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. *Cell reports*. 2013; 3(1):246–259. [PubMed: 23318258] [** This study provides a mathematical approach and computational framework that allow identifying mutational signatures from next generation sequencing data.]
21. Stratton MR. Exploring the genomes of cancer cells: Progress and promise. *Science*. 2011; 331(6024):1553–1558. [PubMed: 21436442]
22. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009; 458(7239):719–724. [PubMed: 19360079]
23. Stephens P, Edkins S, Davies H, Greenman C, Cox C, Hunter C, Bignell G, Teague J, Smith R, Stevens C, O’Meara S, et al. A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer. *Nature genetics*. 2005; 37(6):590–592. [PubMed: 15908952]
24. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, et al. Patterns of somatic mutation in human cancer genomes. *Nature*. 2007; 446(7132):153–158. [PubMed: 17344846]
25. Rubin AF, Green P. Mutation patterns in cancer genomes. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106(51):21766–21770. [PubMed: 19995982]
26. International Cancer Genome C. Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabe RR, Bhan MK, Calvo F, Eerola I, Gerhard DS, et al. International network of cancer genome projects. *Nature*. 2010; 464(7291):993–998. [PubMed: 20393554]
27. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, Varela I, Lin ML, Ordenez GR, Bignell GR, Ye K, et al. A comprehensive catalogue of somatic mutations

- from a human cancer genome. *Nature*. 2010; 463(7278):191–196. [PubMed: 20016485] [* The Pleasance *et al.* back-to-back articles in *Nature* demonstrate the value of whole genome sequencing for evaluating signatures of mutational processes by providing greater resolution and mechanistic insight.]
28. Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, Lin ML, Beare D, Lau KW, Greenman C, Varela I, et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature*. 2010; 463(7278):184–190. [PubMed: 20016488] [* The Pleasance *et al.* back-to-back articles in *Nature* demonstrate the value of whole genome sequencing for evaluating signatures of mutational processes by providing greater resolution and mechanistic insight.]
 29. De Keersmaecker K, Atak ZK, Li N, Vicente C, Patchett S, Girardi T, Gianfelici V, Geerdens E, Clappier E, Porcu M, Lahortiga I, et al. Exome sequencing identifies mutation in *cnot3* and ribosomal genes *rpl5* and *rpl10* in t-cell acute lymphoblastic leukemia. *Nature genetics*. 2013; 45(2):186–190. [PubMed: 23263491]
 30. Holmfeldt L, Wei L, Diaz-Flores E, Walsh M, Zhang J, Ding L, Payne-Turner D, Churchman M, Andersson A, Chen SC, McCastlain K, et al. The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nature genetics*. 2013; 45(3):242–252. [PubMed: 23334668]
 31. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, Easton J, Chen X, Wang J, Rusch M, Lu C, et al. The genetic basis of early t-cell precursor acute lymphoblastic leukaemia. *Nature*. 2012; 481(7380):157–163. [PubMed: 22237106]
 32. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, Ritchey JK, Young MA, Lamprecht T, McLellan MD, McMichael JF, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*. 2012; 481(7382):506–510. [PubMed: 22237025]
 33. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, Menzies A, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell*. 2012; 149(5):979–993. [PubMed: 22608084] [*** This study demonstrates that multiple mutational processes are operative in 21 breast cancer sample. It also demonstrates the existence of localized somatic hypermutation, termed *kataegis*, for the first time.]
 34. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, Turashvili G, Ding J, Tse K, Haffari G, Bashashati A, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*. 2012; 486(7403):395–399. [PubMed: 22495314]
 35. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, Bignell GR, Yates LR, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature*. 2012; 486(7403):400–404. [PubMed: 22722201]
 36. Puente XS, Pinyol M, Quesada V, Conde L, Ordonez GR, Villamor N, Escaramis G, Jares P, Bea S, Gonzalez-Diaz M, Bassaganyas L, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2011; 475(7354):101–105. [PubMed: 21642962]
 37. Quesada V, Conde L, Villamor N, Ordonez GR, Jares P, Bassaganyas L, Ramsay AJ, Bea S, Pinyol M, Martinez-Trillos A, Lopez-Guerra M, et al. Exome sequencing identifies recurrent mutations of the splicing factor *sf3b1* gene in chronic lymphocytic leukemia. *Nature genetics*. 2012; 44(1):47–52. [PubMed: 22158541]
 38. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012; 487(7407):330–337. [PubMed: 22810696]
 39. Seshagiri S, Stawiski EW, Durinck S, Modrusan Z, Storm EE, Conboy CB, Chaudhuri S, Guan Y, Janakiraman V, Jaiswal BS, Guillory J, et al. Recurrent r-spondin fusions in colon cancer. *Nature*. 2012; 488(7413):660–664. [PubMed: 22895193]
 40. Dulak AM, Stojanov P, Peng S, Lawrence MS, Fox C, Stewart C, Bandla S, Imamura Y, Schumacher SE, Shefler E, McKenna A, et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nature genetics*. 2013; 45(5):478–486. [PubMed: 23525077]
 41. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, Olivi A, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science*. 2008; 321(5897):1807–1812. [PubMed: 18772396]

42. Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, Fakhry C, Xie TX, Zhang J, Wang J, Zhang N, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in notch1. *Science*. 2011; 333(6046):1154–1157. [PubMed: 21798897]
43. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, Kryukov GV, Lawrence MS, Sougnez C, McKenna A, Shefler E, et al. The mutational landscape of head and neck squamous cell carcinoma. *Science*. 2011; 333(6046):1157–1160. [PubMed: 21798893]
44. Cancer Genome Atlas Research N. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*. 2013; 499(7456):43–49. [PubMed: 23792563]
45. Guo G, Gui Y, Gao S, Tang A, Hu X, Huang Y, Jia W, Li Z, He M, Sun L, Song P, et al. Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma. *Nature genetics*. 2012; 44(1):17–19. [PubMed: 22138691]
46. Pena-Llopis S, Vega-Rubin-de-Celis S, Liao A, Leng N, Pavia-Jimenez A, Wang S, Yamasaki T, Zhrebker L, Sivanand S, Spence P, Kinch L, et al. Bap1 loss defines a new class of renal cell carcinoma. *Nature genetics*. 2012; 44(7):751–759. [PubMed: 22683710]
47. Kan Z, Zheng H, Liu X, Li S, Barber TD, Gong Z, Gao H, Hao K, Willard MD, Xu J, Hauptschein R, et al. Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome research*. 2013
48. Fujimoto A, Totoki Y, Abe T, Boroevich KA, Hosoda F, Nguyen HH, Aoki M, Hosono N, Kubo M, Miya F, Arai Y, et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nature genetics*. 2012; 44(7):760–764. [PubMed: 22634756]
49. Seo JS, Ju YS, Lee WC, Shin JY, Lee JK, Bleazard T, Lee J, Jung YJ, Kim JO, Shin JY, Yu SB, et al. The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome research*. 2012; 22(11):2109–2119. [PubMed: 22975805]
50. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti M, Sivachenko A, Sougnez C, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*. 2012; 150(6):1107–1120. [PubMed: 22980975]
51. Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, Maher CA, Fulton R, Fulton L, Wallis J, Chen K, et al. Genomic landscape of nonsmall cell lung cancer in smokers and never-smokers. *Cell*. 2012; 150(6):1121–1134. [PubMed: 22980976]
52. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, Sougnez C, Greulich H, Muzny DM, Morgan MB, Fulton L, et al. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature*. 2008; 455(7216):1069–1075. [PubMed: 18948947]
53. Peifer M, Fernandez-Cuesta L, Sos ML, George J, Seidel D, Kasper LH, Plenker D, Leenders F, Sun R, Zander T, Menon R, et al. Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nature genetics*. 2012; 44(10):1104–1110. [PubMed: 22941188]
54. Rudin CM, Durinck S, Stawiski EW, Poirier JT, Modrusan Z, Shames DS, Bergbower EA, Guan Y, Shin J, Guillory J, Rivers CS, et al. Comprehensive genomic analysis identifies *sox2* as a frequently amplified gene in small-cell lung cancer. *Nature genetics*. 2012; 44(10):1111–1116. [PubMed: 22941189]
55. Love C, Sun Z, Jima D, Li G, Zhang J, Miles R, Richards KL, Dunphy CH, Choi WW, Srivastava G, Lugar PL, et al. The genetic landscape of mutations in burkitt lymphoma. *Nature genetics*. 2012; 44(12):1321–1325. [PubMed: 23143597]
56. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, Johnson NA, Severson TM, Chiu R, Field M, Jackman S, et al. Frequent mutation of histone-modifying genes in non-hodgkin lymphoma. *Nature*. 2011; 476(7360):298–303. [PubMed: 21796119]
57. Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. Highly recurrent *tert* promoter mutations in human melanoma. *Science*. 2013; 339(6122):957–959. [PubMed: 23348506]
58. Stark MS, Woods SL, Gartside MG, Bonazzi VF, Dutton-Regester K, Aoude LG, Chow D, Sereduk C, Niemi NM, Tang N, Ellis JJ, et al. Frequent somatic mutations in *map3k5* and *map3k9* in metastatic melanoma identified by exome sequencing. *Nature genetics*. 2012; 44(2):165–169. [PubMed: 22197930]

59. Berger MF, Hodis E, Heffernan TP, Deribe YL, Lawrence MS, Protopopov A, Ivanova E, Watson IR, Nickerson E, Ghosh P, Zhang H, et al. Melanoma genome sequencing reveals frequent prex2 mutations. *Nature*. 2012; 485(7399):502–506. [PubMed: 22622578]
60. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat JP, Nickerson E, Auclair D, Li L, Place C, Dicara D, et al. A landscape of driver mutations in melanoma. *Cell*. 2012; 150(2): 251–263. [PubMed: 22817889]
61. Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, Schinzel AC, Harview CL, Brunet JP, Ahmann GJ, Adli M, Anderson KC, et al. Initial genome sequencing and analysis of multiple myeloma. *Nature*. 2011; 471(7339):467–472. [PubMed: 21430775]
62. Jones S, Wang TL, Shih Ie M, Mao TL, Nakayama K, Roden R, Glas R, Slamon D, Diaz LA Jr, Vogelstein B, Kinzler KW, et al. Frequent mutations of chromatin remodeling gene *arid1a* in ovarian clear cell carcinoma. *Science*. 2010; 330(6001):228–231. [PubMed: 20826764]
63. Jiao Y, Shi C, Edil BH, de Wilde RF, Klimstra DS, Maitra A, Schulick RD, Tang LH, Wolfgang CL, Choti MA, Velculescu VE, et al. *Daxx/atrx*, *men1*, and *mtor* pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science*. 2011; 331(6021):1199–1203. [PubMed: 21252315]
64. Wu J, Jiao Y, Dal Molin M, Maitra A, de Wilde RF, Wood LD, Eshleman JR, Goggins MG, Wolfgang CL, Canto MI, Schulick RD, et al. Whole-exome sequencing of neoplastic cysts of the pancreas reveals recurrent mutations in components of ubiquitin-dependent pathways. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108(52):21188–21193. [PubMed: 22158988]
65. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, Park K, Kitabayashi N, MacDonald TY, Ghandi M, Van Allen E, et al. Punctuated evolution of prostate cancer genomes. *Cell*. 2013; 153(3):666–677. [PubMed: 23622249]
66. Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, Quist MJ, Jing X, Lonigro RJ, Brenner JC, Asangani IA, et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature*. 2012; 487(7406):239–243. [PubMed: 22722839]
67. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat JP, White TA, Stojanov P, Van Allen E, Stransky N, Nickerson E, et al. Exome sequencing identifies recurrent *spop*, *foxa1* and *med12* mutations in prostate cancer. *Nature genetics*. 2012; 44(6):685–689. [PubMed: 22610119]
68. Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, Sboner A, Esgueva R, Pflueger D, Sougnez C, Onofrio R, et al. The genomic complexity of primary human prostate cancer. *Nature*. 2011; 470(7333):214–220. [PubMed: 21307934]
69. Zang ZJ, Cutcutache I, Poon SL, Zhang SL, McPherson JR, Tao J, Rajasegaran V, Heng HL, Deng N, Gan A, Lim KH, et al. Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nature genetics*. 2012; 44(5): 570–574. [PubMed: 22484628]
70. Wang K, Kan J, Yuen ST, Shi ST, Chu KM, Law S, Chan TL, Kan Z, Chan AS, Tsui WY, Lee SP, et al. Exome sequencing identifies frequent mutation of *arid1a* in molecular subtypes of gastric cancer. *Nature genetics*. 2011; 43(12):1219–1223. [PubMed: 22037554]
71. Nagarajan N, Bertrand D, Hillmer AM, Zang ZJ, Yao F, Jacques PE, Teo AS, Cutcutache I, Zhang Z, Lee WH, Sia YY, et al. Whole-genome reconstruction and mutational signatures in gastric cancer. *Genome biology*. 2012; 13(12):R115. [PubMed: 23237666]
72. Cancer Genome Atlas Research N. Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R, Benz CC, et al. Integrated genomic characterization of endometrial carcinoma. *Nature*. 2013; 497(7447):67–73. [PubMed: 23636398]
73. Robinson G, Parker M, Kranenburg TA, Lu C, Chen X, Ding L, Phoenix TN, Hedlund E, Wei L, Zhu X, Chalhoub N, et al. Novel mutations target distinct subgroups of medulloblastoma. *Nature*. 2012; 488(7409):43–48. [PubMed: 22722829]
74. Rausch T, Jones DT, Zapatka M, Stutz AM, Zichner T, Weischenfeldt J, Jager N, Remke M, Shih D, Northcott PA, Pfaff E, et al. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with *tp53* mutations. *Cell*. 2012; 148(1-2):59–71. [PubMed: 22265402]

75. Pugh TJ, Weeraratne SD, Archer TC, Pomeranz Krummel DA, Auclair D, Bochicchio J, Carneiro MO, Carter SL, Cibulskis K, Erlich RL, Greulich H, et al. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature*. 2012; 488(7409):106–110. [PubMed: 22820256]
76. Jones DT, Jager N, Kool M, Zichner T, Hutter B, Sultan M, Cho YJ, Pugh TJ, Hovestadt V, Stutz AM, Rausch T, et al. Dissecting the genomic complexity underlying medulloblastoma. *Nature*. 2012; 488(7409):100–105. [PubMed: 22832583]
77. Pugh TJ, Morozova O, Attiyeh EF, Asgharzadeh S, Wei JS, Auclair D, Carter SL, Cibulskis K, Hanna M, Kiezun A, Kim J, et al. The genetic landscape of high-risk neuroblastoma. *Nature genetics*. 2013; 45(3):279–284. [PubMed: 23334666]
78. Sausen M, Leary RJ, Jones S, Wu J, Reynolds CP, Liu X, Blackford A, Parmigiani G, Diaz LA Jr, Papadopoulos N, Vogelstein B, et al. Integrated genomic analyses identify *arid1a* and *arid1b* alterations in the childhood cancer neuroblastoma. *Nature genetics*. 2013; 45(1):12–17. [PubMed: 23202128]
79. Zhang J, Wu G, Miller CP, Tatevossian RG, Dalton JD, Tang B, Orisme W, Punchihewa C, Parker M, Qaddoumi I, Boop FA, et al. Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. *Nature genetics*. 2013; 45(6):602–612. [PubMed: 23583981]
80. Spencer J, Dunn-Walters DK. Hypermutation at a-t base pairs: The a nucleotide replacement spectrum is affected by adjacent nucleotides and there is no reverse complementarity of sequences flanking mutated a and t nucleotides. *Journal of immunology*. 2005; 175(8):5170–5177.
81. Hoang ML, Chen CH, Sidorenko VS, He J, Dickman KG, Yun BH, Moriya M, Niknafs N, Douville C, Karchin R, Turesky RJ, et al. Mutational signature of aristolochic acid exposure as revealed by whole-exome sequencing. *Science translational medicine*. 2013; 5(197):197ra102.
82. Comon, P. Handbook of blind source separation : Independent component analysis and blind deconvolution. Elsevier; Boston, MA: 2010.
83. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature*. 1999; 401(6755):788–791. [PubMed: 10548103]
84. Taylor BJ, Nik-Zainal S, Wu YL, Stebbings LA, Raine K, Campbell PJ, Rada C, Stratton MR, Neuberger MS. DNA deaminases induce break-associated mutation showers with implication of *apobec3b* and *3a* in breast cancer kataegis. *eLife*. 2013; 2:e00534. [PubMed: 23599896]
85. Harris RS, Petersen-Mahrt SK, Neuberger MS. Rna editing enzyme *apobec1* and some of its homologs can act as DNA mutators. *Molecular cell*. 2002; 10(5):1247–1253. [PubMed: 12453430]
86. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, Raine K, Jones D, Marshall J, Ramakrishna M, Shlien A, et al. The life history of 21 breast cancers. *Cell*. 2012; 149(5):994–1007. [PubMed: 22608083]
87. van Zeeland AA, Vreeswijk MP, de Gruijl FR, van Kranen HJ, Vrieling H, Mullenders LF. Transcription-coupled repair: Impact on uv-induced mutagenesis in cultured rodent cells and mouse skin tumors. *Mutation research*. 2005; 577(1-2):170–178. [PubMed: 15949822] [* This article explains the interplay between transcription-coupled repair and mutations induced by UV-light.]
88. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, Kiezun A, Kryukov GV, Carter SL, Saksena G, Harris S, et al. An *apobec* cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nature genetics*. 2013
89. Burns MB, Temiz NA, Harris RS. Evidence for *apobec3b* mutagenesis in multiple human cancers. *Nature genetics*. 2013

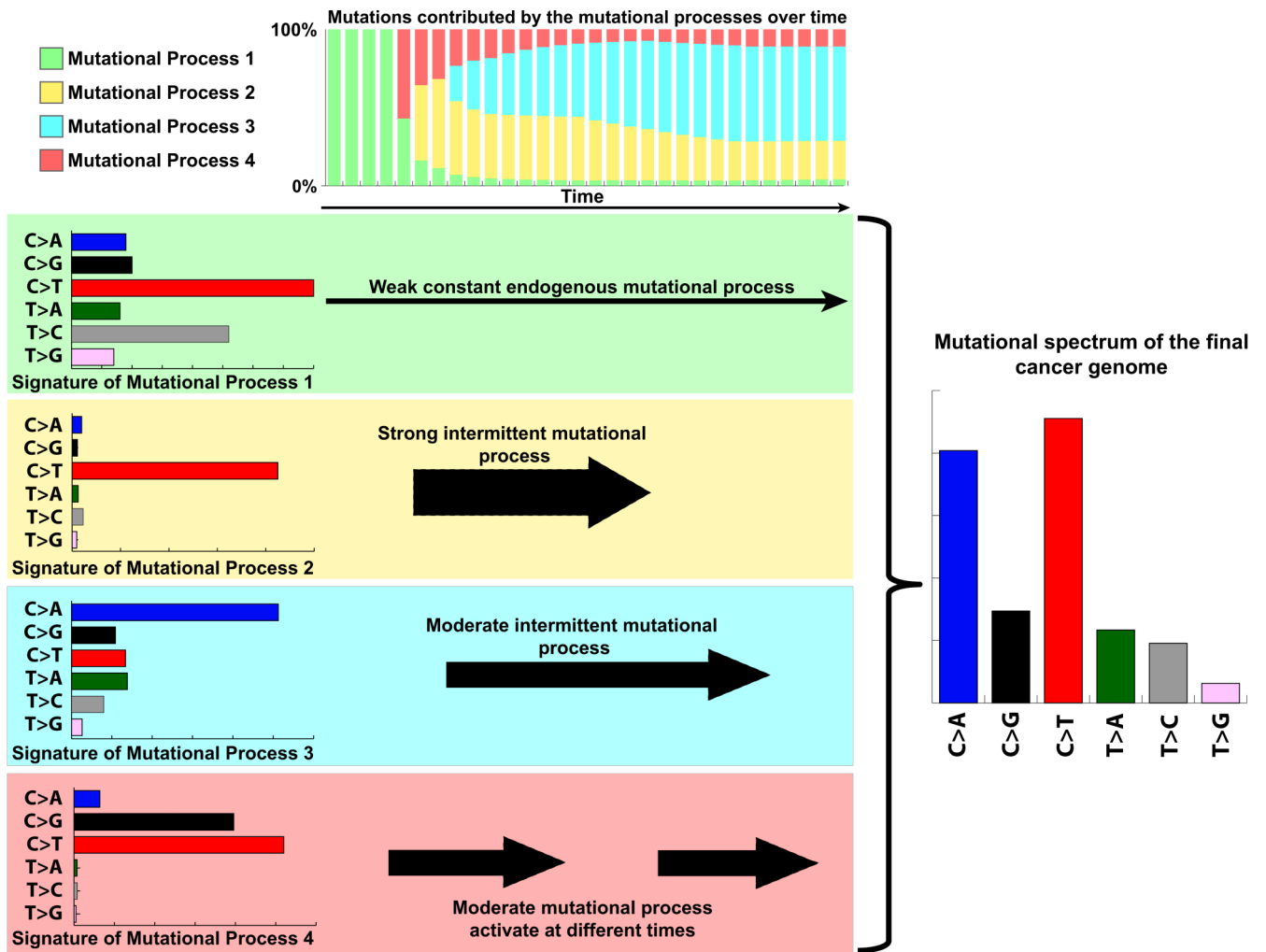


Figure 1. Mutational processes operative in a cancer

This simulated example illustrates four distinct mutational processes with variable strengths operative at different times throughout the lifetime of the patient. Each of these processes has a unique mutational signature exemplified by the six classes of somatic substitutions. At the beginning, all mutations in the cell (from which the cancer was eventually developed) were due to the activity of the endogenous mutational process 1. As time progresses, the other mutational process get activated and the spectrum of the cell continues to change. Note that the final sequenced cancer genome does not resemble any of the operative mutational signatures.