

Review

Computational Methods Summarizing Mutational Patterns in Cancer: Promise and Limitations for Clinical Applications

Andrew Patterson ^{1,2} , Abdurrahman Elbasir ², Bin Tian ² and Noam Auslander ^{2,3,*}

¹ Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

² The Wistar Institute, Philadelphia, PA 19104, USA

³ Department of Cancer Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

* Correspondence: nauslander@wistar.org

Simple Summary: Cancer is a complex disease that develops over time through accumulated mutations in DNA that transform normal cells into a cancerous state. To fully capture the complexity of the cancer genome, computational methods have been developed to summarize the mutational patterns of cancer, distinguish causal oncogenic mutations, and determine clinically useful mutational patterns. In this review, we survey different computational approaches with an emphasis on important clinical roles and provide insights into better integration of computational methods for clinical use.

Abstract: Since the rise of next-generation sequencing technologies, the catalogue of mutations in cancer has been continuously expanding. To address the complexity of the cancer-genomic landscape and extract meaningful insights, numerous computational approaches have been developed over the last two decades. In this review, we survey the current leading computational methods to derive intricate mutational patterns in the context of clinical relevance. We begin with mutation signatures, explaining first how mutation signatures were developed and then examining the utility of studies using mutation signatures to correlate environmental effects on the cancer genome. Next, we examine current clinical research that employs mutation signatures and discuss the potential use cases and challenges of mutation signatures in clinical decision-making. We then examine computational studies developing tools to investigate complex patterns of mutations beyond the context of mutational signatures. We survey methods to identify cancer-driver genes, from single-driver studies to pathway and network analyses. In addition, we review methods inferring complex combinations of mutations for clinical tasks and using mutations integrated with multi-omics data to better predict cancer phenotypes. We examine the use of these tools for either discovery or prediction, including prediction of tumor origin, treatment outcomes, prognosis, and cancer typing. We further discuss the main limitations preventing widespread clinical integration of computational tools for the diagnosis and treatment of cancer. We end by proposing solutions to address these challenges using recent advances in machine learning.

Keywords: cancer genomics; mutation signatures; machine learning; bioinformatics; clinical predictors; cancer drivers



Citation: Patterson, A.; Elbasir, A.; Tian, B.; Auslander, N. Computational Methods Summarizing Mutational Patterns in Cancer: Promise and Limitations for Clinical Applications. *Cancers* **2023**, *15*, 1958. <https://doi.org/10.3390/cancers15071958>

Academic Editors: Teresa M. Przytycka, Damian Wójtcowicz and Roded Sharan

Received: 27 October 2022

Revised: 24 February 2023

Accepted: 9 March 2023

Published: 24 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Cancer has historically been studied using genetic techniques, with the goal to identify gene-driver mutations that confer selective advantage and drive cells into a cancerous state. Driver mutations are distinguished from passenger mutations, which accumulate in the genome due to the changes undergone in the cancer cell as it becomes cancerous [1,2]. These approaches have led to several landmark discoveries and treatment successes, in particular, targeted therapies (Box 1) [3]. A prominent example is BRCA1/2 mutations in breast and ovarian cancers [4,5], which allowed for revolutionary treatment success for

patients harboring the mutations by exploiting synthetic lethality (Box 1) through PARP inhibitors [6]. Identification of genes in the MAPK pathway, including BRAF and KRAS, has allowed for potent anticancer treatments in melanoma [7,8] and non-small-cell lung cancers [9,10]. IDH1 and IDH2 genes are inhibited in the treatment of AML [11] and gliomas [12], and ALK genes are targeted in lung cancers [13–15]. Furthermore, drugs targeting HER2 are a major treatment strategy for HER2-positive breast cancers [16–21]. However, most cancers are not driven purely by single-gene mutations; different genes or combinations of genes may confer a similar cancer phenotype. An understanding of how changes in multiple mutations or in the entire genome affect different cancers, and unraveling the biological sources of cancer mutations, has been a burgeoning field over the last decade [1,22].

Box 1. Definition of select terms.

Targeted therapies	Therapies targeting a specific protein associated with a disease
Synthetic lethality	A type of interaction wherein a single event is tolerable but co-occurrence of two or more events is lethal
Driver mutation	A mutation that provides a selective advantage to a cell and transforms a cell into a cancerous state
Passenger mutation	A mutation that is a result, but not a direct cause, of a cell becoming cancerous
Mutagenic process	Anything that causes damage to DNA or induces mutations in DNA, such as UV light, radiation, or alkylating agents
Non-negative matrix factorization (in progress)	Unsupervised mathematical method wherein a single large nonnegative matrix is decomposed into two or more smaller matrices
COSMIC	Catalogue of Somatic Mutations in Cancer: https://cancer.sanger.ac.uk/cosmic , accessed on 12 March 2023
Signature 1	Mutation signature associated with age
Signature 2	Mutation signature associated with the mutagenic effects of APOBEC activity
Signature 4	Mutation signature associated with tobacco smoke
Signature 7	Mutation signature associated with UV exposure
Signature 10	Mutation signature associated with POLE proofreading errors
Signature 16	Mutation signature associated with alcohol consumption
Signature 18	Mutation signature associated with the mutagenic effects of the MUTYH gene
DDR	DNA damage repair, a network of processes that repairs damaged DNA
MMR	Mismatch repair, a DDR pathway involved in detecting and repairing DNA mismatches
BER	Base-excision repair, a pathway that repairs typically small-scale mutations by first removing only the base and leaving an abasic site, which is later removed and replaced with other nucleotides
NER	Nucleotide-excision repair, a pathway that repairs mutations by entirely removing mutated sections of DNA
HR	Homologous recombination, a pathway repairing double-strand DNA damage that uses another strand of DNA as a template for repair
NHEJ	Non-homologous end joining, a pathway repairing double-strand DNA damage that involves attaching two strands of broken DNA together.
Logistic regression	A regression model for supervised classification
LASSO logistic regression	A regression model that uses L1 regularization
Random forest	An ensemble machine-learning model that combines decision trees produced by bagging
ICI	Immune-checkpoint inhibitors, a class of cancer drugs that suppresses pro-tumor immune-system regulatory effects
Supervised learning	Machine-learning strategies wherein the classes of outcomes are known
Unsupervised learning	Machine-learning strategies wherein the task of the model is to cluster the data into previously unidentified classes or discover the underlying classes
Neural network	A machine-learning model that connects the input data to a desired output classification, where nodes connected by edges apply non-linear transformations to the data passed through the network
Deep learning	Machine-learning models that are composed of multiple layers of neural networks stacked over one another (giving rise to the term “deep”)
Overfitting	Fitting a particular data point too well and therefore failing to predict on other data
Underfitting	Not fitting the data well enough and inferring simplified decision rules that may not be optimized for any dataset
Graph convolutional networks	Neural-network architectures that represent graph data for learning tasks

2. Mutation-Signatures Background

One of the earliest and most important computational tools to uncover patterns of mutations arising through different mutagenic processes (Box 1) is mutation signatures, which triggered this revolution in understanding the holistic cancer genome (Figure 1). Mutation signatures were first developed by extracting patterns of nucleotide transitions within the mutations in whole-genome-sequencing data in a small cohort of breast cancers in 2012 [23]. In 2013, this principle was confirmed and expanded upon into a much larger dataset across different cancers [24]. These mutation signatures were a paradigm shift in understanding changes in the human genome in the context of cancer, as they allow patterns of historical mutations to be associated across the entire genome with biological, environmental, cancer-specific, and even cancer-treatment-specific effects (Figure 1A).

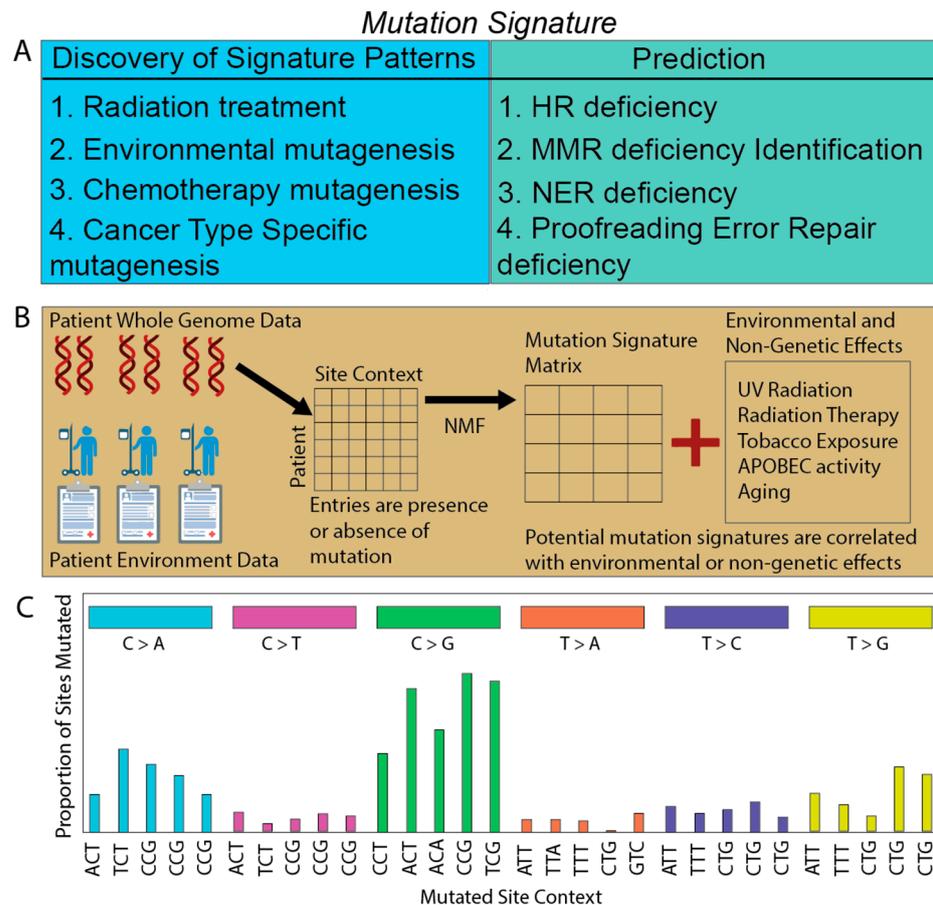


Figure 1. Mutation-signatures overview. (A) Mutation signatures have been used for discovery of genomic patterns reflecting the effects stressors have on the cancer genome and for phenotype prediction. (B) Simplified illustration of the construction of mutation signatures. Whole-genome-sequencing (WGS) data are collected and combined into a matrix. The matrix is decomposed using non-negative matrix factorization (NMF) or a similar method, and the resulting mutation-signature matrix is then correlated with environmental, patient-specific, or cancer-specific effects. (C) Simplified example of a potential mutation signature. The x-axis is site-specific nucleotide contexts. The colored boxes indicate groupings of the same nucleotide transition. The y-axis is the proportion of those context-specific sites that are mutated according to the specified transition. Only 30 of the 96 total potential sites are shown here for clarity.

2.1. Deriving Signatures of Mutations

Mutation signatures mathematically model certain types of mutations that cluster together based on co-occurrence in tumors [24–26] (Figure 1B). The original types of mutation considered were based on nucleotide triplets [24]. Mutations were classified

according to the transition from one base pair to another as defined from the pyrimidine of the Watson Crick base pair (6 potential transitions total, corresponding to C > A, C > G, C > T, T > A, T > C, T > G), as well as the nucleotide context of the surrounding two base pairs, yielding 96 total mutation types [24,26,27] (Figure 1C). The repertoire of mutation types considered has been subsequently expanded, including indels and double mutations, increasing the complexity and potential ability of the signatures to capture biological complexity across the genome [25,28].

Computationally, the first mutation-signature methods relied on the mathematical principle of non-negative matrix factorization (NMF) (Box 1), where a single large non-negative matrix is decomposed into two or more smaller matrices [29,30]. Multiplying these smaller matrices together should approximate the original input matrix. One of the decomposed matrices is the signature matrix representing the mutation signatures, which are, in a separate step, associated with outside environmental, cancer, or biological causes (see [26,27] or the supplemental information of [25] for a comprehensive mathematical explanation of mutation-signature generation, and in the first two, a comparison between different derivation methods). Subsequent checks are used to determine the optimal number of signature matrices. These include biological checks by investigating whether the cluster of mutations makes sense in the context of potential biological drivers and algorithmic checks, such as k-means clustering [24–27] (Figure 1B).

Many developments and refinements of the methods to generate mutation signatures have been suggested. Several rely on variations in NMF [24,25,31,32], but others use different methods to generate these signatures, resulting in potentially different signatures [26,27]. These methods include the NMF-based SigProfiler [25], which is a newer version from the original mutation-signature paper, updated with more data [24], as well as the NMF-based MutSpec [31] and MutSignatures [32]. Additional methods are Bayesian NMF methods such as BayesNMF [33,34] and signeR [35]; probabilistic modeling, such as pmsignature [36] and EMu [37]; PCA-based methods, such as SomaticSignatures [38] and Helmsman [39]; and basic machine learning methods, such as deconstructSigs [40]. A recent comparison that evaluated the strengths and limitations of different methods for real and simulated data indicated that probabilistic models may perform better based on simulated data [27]. Others have developed methods to assess the reproducibility of the decomposition method itself [41], but comprehensive benchmarking is still needed. These signatures can be found in the Catalogue of Somatic Mutations in Cancer (COSMIC) (Box 1) [25], and other tools have been developed to allow for data analysis of mutation signatures [42,43].

2.2. Associating Mutation Signatures with Carcinogenic Processes

Once derived, the mutation signatures are then associated with potential biological, environmental, or cancer-related phenomena, and mutations that occur in these signatures may be extracted to investigate potential clinical relevance (Figure 1B). The landmark study by Alexandrov et al. (2013) established canonical mutation signatures that were used in numerous studies across the field and have been continuously expanded on by multiple laboratories. In the original study, age was associated with mutation-signature 1, later discovered by additional data to be two similar signatures labeled signatures 1A and 1B 1 (Box 1) and correlated with a C > T transition [24]. Age was associated with these signatures because the rate of mutation did not change across different ages and was consistent across cancers, indicating a steady baseline rate of mutation [24,25,28]. Subsequent work expanded upon using mutation signatures to track rates of mutations, which found that several signatures had clock-like processes associated with the passage of time but potentially varied across different tissues [28]. Signature 2 (Box 1) was associated with a family of cytidine-deaminase enzyme (APOBEC) activity, using previous work as a guide for the expected activity of APOBEC proteins [24,25,44,45]. Further work, seeking to investigate how mutation processes act in real time on live cells, confirmed signature 2 as being associated with APOBEC activity, and also found that APOBEC activity was sporadic, a finding

that may have clinical opportunities and challenges when targeting mutagenic processes for treatment [46–48]. Another study investigating the cause of esophageal squamous-cell carcinoma found signatures associated with APOBEC activity, indicating activation of APOBEC was a driver in the formation of this cancer [49]. Individual genes may also be associated with certain mutation signatures. For example, germline mutations in the base excision-repair gene MUTYH left distinct mutation signatures corresponding to COSMIC signature 18 (Box 1) in colorectal cancers and adrenocortical carcinomas [50]. Mutation signatures have also been linked to known environmental carcinogens. Signature 4 (Box 1) mutations, which primarily involve C > A transitions on the transcribed strand, have been observed in lung, head and neck, and liver cancers and are associated with tobacco-smoke mutagens [24,25]. Studies confirming this association provided further evidence of smoking driving cancer by inducing genome-wide mutagenesis [51]. Another environmental association was found in the C > T transitions of signature 7 (Box 1), which was highly prevalent in melanoma, and indicated association with UV exposure [24,25]. Further incorporating indel mutations, multiple mutation signatures have been linked to diverse mutagenesis processes. These include substitution and indel-mutation signatures that correlated with mismatch repair and microsatellite instability in a subset of cancers [25,52]. Ionizing-radiation-mutation signatures, corresponding with single-nucleotide variations and indels, were identified in new cancer events of patients treated with radiation therapy [53]. Ionizing radiation can also interact with germline mutations to induce distinct mutation signatures, as demonstrated in TP53-deficient mice that were exposed to ionizing radiation [54]. Other environmental effects associated with mutation signatures include exposure to carcinogenic chemicals, including cobalt, vinylidene, and 1,2,3-trichloropropane. These associated effects were confirmed in both experimental mouse tumors and, in the case of 1,2,3-trichloropropane, human tumors caused by contaminated drinking water [55]. Therefore, the analysis of thousands of cancer genomes allowed the delineation of various mutational signatures and some of these signatures to be linked to endogenous and exogenous mutagenic processes. Yet, the etiology of some of these signatures remains to be discovered.

3. Clinical Applications of Mutation Signatures: Promises and Challenges

Concurrent with the development of mutation signatures was the recognition that these signatures may potentially be used in a clinical context for prognoses and treatment outcomes [23,24]. With their inherent ability to summarize genome-wide mutation patterns, mutation signatures are particularly useful when genome-wide mutagenesis is clinically relevant, or when genomic mechanisms modulating treatment outcomes are unknown (Figure 2).

3.1. DNA-Damage-Repair Footprints and Clinical Applications of Mutation Signatures

DNA damage repair (DDR) is a complex network comprising multiple DNA-repair pathways, damage-tolerance processes, and cell-cycle checkpoints, with multiple interacting components assessing and maintaining genomic integrity [22,56,57]. Impairment of DDR components leads to genomic instability, a central characteristic of almost all human cancers [58,59]. Several forms of genomic instability have been found in tumors and associated with different DDR pathways [59]. Single-strand DDR pathways include mismatch repair (MMR) (Box 1), base-excision repair (BER) (Box 1), and nucleotide-excision repair (NER) (Box 1). Impairments of these mechanisms lead to genome-wide accumulation of base-pair mutations, involving base substitutions, deletions, or insertions of a few nucleotides, as well as local copy-number amplifications and deletions [56]. Homologous recombination (HR) (Box 1) and non-homologous end joining (NHEJ) are double-strand DDR pathways correcting DNA double-strand breaks (DSBs), which can lead to genomic imbalances and translocations [57,60,61].

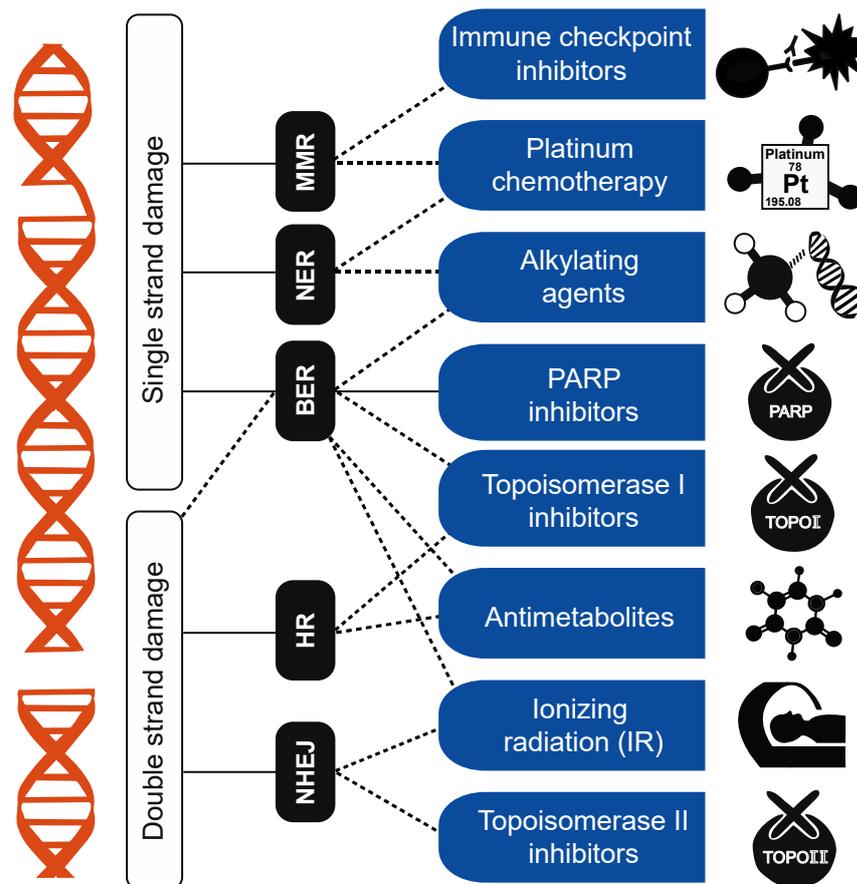


Figure 2. Potential clinical utility of genome-wide DNA-damage signatures: approved cancer drugs that induce DNA damage and associations with specific damage-repair pathways. DNA-damage-inducing drugs (lefthand blue boxes) activate DDR pathways (middle black boxes), directly or indirectly (solid and dashed lines, respectively). DDR pathways repair single- or double-strand damage, and impairment in those pathways leads to whole-genome signatures with potential clinical utility for DNA-damage-inducing drugs.

Disruption in DDR pathways induces genome-wide mutagenesis, and some DDR pathways are linked to responses to specific treatments, including chemoradiation and targeted therapies. Mutation signatures become useful in such cases, as they can examine patterns of DDR deficiencies throughout the genome.

This concept has been most clearly shown in applications to HR-deficient cancers. Loss of HR results in increased sensitivity to inhibition of the BER gene PARP1. The absence of PARP allows for unrepaired single-strand breaks to accumulate, and these breaks collide with replication forks and induce cytotoxic double-strand breaks. When HR deficient, cells are unable to repair those breaks, leading to genomic instability and cell death [62,63]. Therefore, strategies to infer HR deficiency in tumors are particularly useful for treatments targeting HR-deficient cells. One important tool developed to identify HR deficiencies in breast cancer is HRDetect, which is based on a LASSO logistic-regression model (Box 1) that uses mutation signatures associated with substitutions, indels, and rearrangements as feature inputs to the model [64]. Subsequent analysis showed that this tool was able to identify HR repair-deficient patients (HRD) irrespective of their HRD germline, genetic, or epigenetic status [65,66]. HRDetect was also shown to potentially be able to identify patients that would respond to platinum treatments [67]. The benefit of HRDetect and similar tools is the identification of patients that are sensitive to PARP inhibitors or platinum treatment but that could be missed in the traditional HR-deficiency screen [33,64,67–69]. HRDetect was used in a secondary endpoint of a phase II clinical trial examining PARP inhibitors

for triple-negative breast-cancer patients, with success in identifying HR-deficient tumors that could be missed using current clinical practice [69]. Recently, other tools have also been developed to detect HR deficiencies using mutation signatures, including CHORD and SigMA, which use a random-forest (Box 1) and likelihood-based approach (Box 1) to classification, respectively [68,70].

Other treatments targeting HRD cancers are currently in clinical trials, where mutation signatures may become useful. These treatments target different proteins involved in the HR pathway, for example, ATR inhibitors [71]. ATR inhibitors (ATRi) may selectively kill HRD cells [72]. ATR-induced cell death has also been shown in PARP-resistant cancers, indicating the complementarity of this approach with PARP [73,74]. ATRi for treatment of HRD cancers is currently in clinical trials [75]. Therefore, models using mutation signatures could also provide a way to identify patients that would benefit from ATRi therapy.

Mutation signatures can also infer MMR deficiencies (MMRd). Importantly, MMRd is an approved biomarker for immune-checkpoint inhibitors (ICI) (Box 1) [76], and similar to HR deficiencies, MMRd leaves distinct mutational footprints on the genome. MMRDetect is a tool developed to infer mutation signatures descriptive of MMRd using a logistic-regression model (Box 1) incorporating mutation signatures associated with MMRd (Table 1) [77]. Although direct sequencing of potential causal genes (such as *MSH2*, *MSH6*, *PMS2*, and *MLH1*) are clinically available for MMR [78,79], research has shown that these genes may potentially be epigenetically regulated rather than genetically mutated [80,81], posing a challenge for MMRd detection through genomic screening. Analyzing the effects of MMR across the genome using mutation signatures could complement identification of cancers deficient in MMR that may be susceptible to certain treatments. These treatments primarily involve immune-checkpoint-inhibitor therapy, but recent work demonstrated that inhibiting Werner helicases in MMRd tumors may induce synthetic lethality and potentially allow for additional treatment options [78,82,83] Further supporting this notion, studies carried out in pancreatic cancer found associations between MMR signatures and antitumor immune activation, even when canonical HR or MMR genes were not germline mutated in the tumors (Table 1) [84].

Table 1. Clinical applications of mutation signatures.

Category	Descriptive Mutational Process	Clinical Use
Clinically relevant DDR pathways	Homologous recombination (HR)	Biomarker for PARP-inhibitor sensitivity [64–66]
		Biomarker for platinum-treatment sensitivity [67]
		Biomarker for ATRi-inhibitor sensitivity [71,73–75]
	Mismatch repair (MMR)	Immune-checkpoint-inhibitor biomarker [77]
		Identification of Werner-helicase-sensitive patients [78,82,83]
		Potential biomarker for antitumor immune activation [84]
	Nucleotide excision repair (NER)	Biomarker for platinum-treatment sensitivity [34,85]
Biomarker of ERCC2 deficiency [34,85]		
	Proofreading errors	Biomarker of POLE deficiency [86,87]
Characterization of clinically relevant phenomena	Radiation treatment	Identification of radiation-driver tumors [53]
		Identification of genes with potential contra-indications of radiation therapy [54,88]
	Chemotherapy	Tumorigenic effects of 5-FU [88,89]
		Tumorigenic effects of platinum and capecitabine treatments
	Environmental	Screening for aristolochic-acid damage [90–92]
		Alcohol-consumption signatures across cancers [93–96]
Cancer-type specific mutagenesis	Identification of different subtypes of esophageal cancer [97]	
	Identification of secondary tumors of unknown origin [98]	

Other associations between cancer treatments and distinct DDR pathways include ERCC2 helicase in the NER pathway (Table 1). Mutated ERCC2 produces a distinct mutational signature that serves as a marker for disruption in the NER pathway [34]. Mutation signatures corresponding to NER patterns similar to ERCC2 disruption could provide a biomarker for cisplatin or similar platinum treatment [34,85].

Other than canonical DDR pathways, proofreading errors also induce distinct mutation signatures, potentially allowing for the development of similar methods to MMR and HR mutation-signature tools. For example, POLE proofreading errors are associated with Signature 10 (Box 1), which could be associated with immune-checkpoint-inhibitor therapy sensitivity (Table 1) [86,87]. Overall, the link between specific DDR pathways and mechanisms or sensitivity of distinct cancer treatments warrants more work exploring this association through mutation signatures.

3.2. Mutation Signatures as Clinical-Discovery Tools

Due to their ability to elucidate associations between exogenous or endogenous mutagenesis and cancer, mutation signatures are useful for studying clinical phenomena when the underlying mechanisms and genetic markers are unknown. Therefore, these signatures may be useful for clinical development and discovery (Figure 2).

Radiation therapy has long been recognized as a potential driver of new cancers [99,100], but markers distinguishing radiation-induced tumors are unknown. Mutation signatures have been used to differentiate cancers driven by radiation therapy as opposed to cancer relapse or recurrence (Table 1) [53]. Another study applied mutation signatures to identify an association between TP53 deficiency and radiation-induced secondary cancers in mice (Table 1) [54]. Similarly, a potential association with radiation and mutation signatures was found in mutation-signature ID12, with higher mutation-signature activity in HRD tumors compared to non-HRD tumors (Table 1) [88]. Therefore, mutation signatures have been useful for identifying patterns linked with a distinct mutation that in turn may be used as a marker for patients that should not be treated with radiation therapy.

Mutation signatures are being used to investigate the effects of other cancer treatments on the genome, allowing both a better understanding of the mechanism of the treatments and potential indications or contra-indications of the treatment. For example, using mutation signatures, 5-FU was found to induce numerous T > G substitutions throughout the genome, indicating a potential tumorigenic effect of this chemotherapy drug (Table 1) [89]. Further work has also shown mutation-signature associations with platinum therapies and capecitabine and confirmed 5-FU associations, with increasing time and doses of drugs producing higher mutation-signature signal (Table 1) [88].

Mutation signatures have also driven discovery of clinically relevant environmental carcinogens through patterns of mutations in the genome. Aristolochic acid (AA) is a chemical found in plants used in herbal remedies. In different cancers, and in bladder cancers in particular, the presence of AA-associated signatures provided evidence that AA has a mutagenic effect on the genome, demonstrating the potential of mutation signatures as a screening tool (Table 1) [90–92]. Evidence from several studies on esophageal squamous-cell carcinoma also found associations between alcohol consumption and several mutation signatures [93,94]. Specifically, mutation signature 16 (Box 1), associated with alcohol consumption, was also present in liver cancers [95]. Similarly, a study across many different cancers found a distinct mutation signature associated with alcohol consumption in HNSC, ESCA, and LIHC and proposed a mechanism of mutation involving acetaldehyde (Table 1) [96]. These and similar signatures summarizing cancer-risk factors may inform patients and possibly be developed into screening practices.

Another promising use of mutation signatures is as a biomarker for different cancer types or cell types. Mutation signatures were used to distinguish different cell types within esophageal adenocarcinoma, with the potential to directly target these different subtypes for different therapy treatments (Table 1) [97]. Recent work has also shown that distinct patterns of mutation signatures combined with additional tumor information can be used

with machine learning to identify secondary tumors of unknown primary, which can greatly facilitate targeted treatment of the cancer (Table 1) [98].

The clinical potential of mutation signatures in other contexts has been mentioned in multiple studies, for example, for predicting immunotherapy response [1,86,87]. In practice, however, mutation signatures have so far demonstrated clinical utility as a biomarker only when whole-genome changes reflect the outcome of interest or as a tool for clinical discovery when underlying mutagenic processes are unknown. In clinical practice, summarizing a mutagenic process to a defined set of genes or markers is both more interpretable to clinicians and requires sequencing fewer genomic regions. Therefore, mutation signatures are useful in the path to defining mutagenic processes and finding associated markers to be used in the clinic.

4. Beyond Mutation Signatures: Computational Approaches to Infer Clinically Relevant Patterns of Mutations

In addition to mutation signatures, other methods have been developed to discover patterns of cancer mutations that drive cancer development and underlie clinical outcomes (Figure 3). The majority of these methods derive patterns of mutations using supervised- or unsupervised-learning strategies (Box 1), which can then be directly correlated with a clinical outcome of interest (Figure 3A). A fundamental goal of these emerging techniques is the identification of cancer drivers. Discovering mutated genes that are drivers of tumorigenesis and distinct from genes that are merely passengers is essential to understanding cancer development and finding the causal players that may be clinically targeted [101]. Therefore, a comprehensive catalogue of driver mutations can improve diagnosis and prognosis and provide for new drug targets [102,103]. In recent years, as sequencing data has become increasingly available, several methods have been developed that use machine-learning techniques to distinguish potential driver mutations from passenger mutations (Figure 3B). These methods have steadily advanced to incorporate different aspects of the genome. Early work in this field involved developing methods analyzing the frequency of mutations in genes within cancers to separate out potential driver genes from passengers, such as MutSigCV [104], inVex [105], and MuSiC [106]. Later approaches incorporated functional impact by predicting the changes to the amino acids linked to a mutation and predicting the impact of a mutation to the function of a gene. Such tools include the random-forest-based CHASM [107–109], polyphen2 [110], e-Driver [111], and SIFT [112,113], which were adapted to cancer mutations. Taking this functional concept further, other algorithms use the structure of the protein itself to predict relevance to cancer. These include MSEA [114], which combines mutation frequency and protein-domain structure to predict driver genes, and iPAC [115] or GraphPAC [116], which use tertiary structure to predict driver mutations. More specialized methods such as ActiveDriver [117] have focused on mutations in phosphorylation or similar post-translational regulation sites (Table 2).

Methods have also shifted from focusing on features of single genes to accounting for more complex patterns, such as gene networks and pathways (Box 1) (Figure 3C). These approaches seek to leverage the knowledge that genes do not operate in isolation but act as part of a larger whole, where mutations in similar pathways or network locations may produce similar effects. For example, HotNet2 uses a heat-diffusion model to identify mutated subnetworks, providing more information about the mutational landscape than mutation data alone [118]. This work allowed for the identification of rare driver mutations in the TCGA compared to previous studies focusing on purely mutation-based analysis. Other network approaches include MUFFIN, which used the mutation data in network neighbors to discover cancer drivers, even with a subset of the data [119], and Paradigm, which used curated pathways with a gene-factor graph-modeling approach to discover cancer drivers (Table 2) [120]. Newer methods have expanded on this network-based analysis to discover modules of tumor–gene interactions with potential diagnostic and therapeutic significance [121] and have also incorporated non-coding mutations, path-

ways, and network analysis [122]. Beyond network or pathway analysis, a recent study developed a deep-learning model (Box 1) for the background mutation rates to identify patterns of positive selection and find driver mutations in coding and non-coding regions [123]. Another method, boostDM, combined mutational data across cancers with gradient-boosting tree algorithms (Box 1) to produce a series of interpretable models for the identification of cancer drivers, and it has even been reported that this method outperforms experimental large-scale saturation-mutagenesis experiments (Table 2) [124]. A recent benchmarking and comparison of these methods found that four methods were most effective at predicting drivers [125], namely, the random-forest-based CHASM [107–109] and DEOGEN2 [126], the PCA-based CTAT-cancer [127], and the deep residual neural-network-based PrimateAI [128] (Table 3).

A Computational Approaches to Discovering Cancer Drivers

Discovery of Cancer Drivers	Prediction
1. By mutation frequency	1. Predicting response to treatment
2. By pathway/network analysis	2. Prediction cancer subtypes
3. Multiomics Integration	3. Prediction of cancerous state

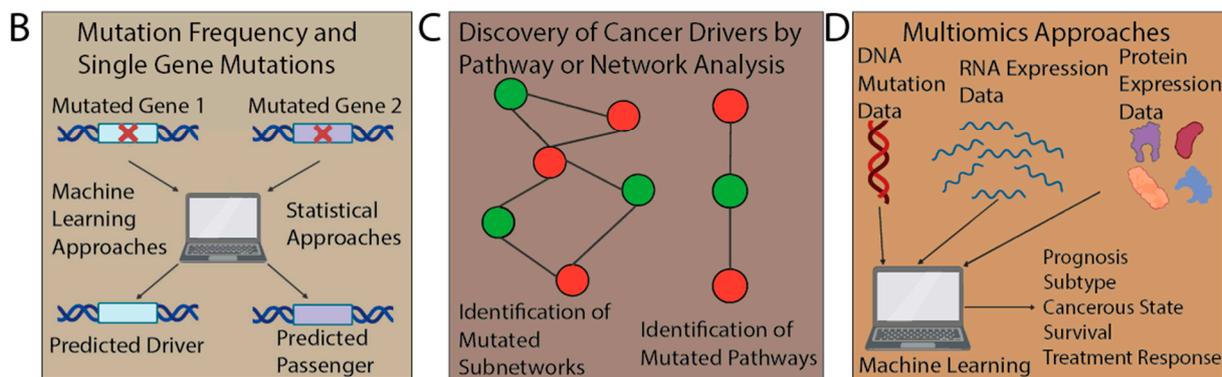


Figure 3. Overview of computational approaches for using mutational patterns beyond mutation signatures. (A) Computational approaches are used for discovery of cancer-driver mutations and for the prediction of cancer phenotypes using mutational patterns. (B) Single gene for distinction of cancer-driver mutations. (C) Network- and pathway-based methods to predict driver mutations and use mutational data for cancer-phenotype prediction. (D) Multi-omics approaches integrate mutations with different data types to improve discovery of cancer drivers and prediction of cancer phenotypes.

Other than discovery of driver mutations, methods have used pathway and network information to identify patterns of mutations to predict treatment outcomes, allowing for more biologically interpretable models (Table 2) [129,130]. An early representative study used network-based stratification to combine mutation data and gene networks to predict patient responses, tumor types, and histology [131]. A method to de-novo identify significantly mutated subnetworks has revealed known and new mutated pathways in cancer. Mutation data aggregated into biological processes were used as input to different machine-learning classifiers to predict immunotherapy response in melanoma and to understand biologically what occurs in immunotherapy response and resistance [132]. Pathway-based methods have also been developed for scoring responses to different cancer treatments, showing applications in both drug discovery and clinical selection of drugs [133]. Pathways and mutation data were also used to identify cancer subtypes and prognostic indications

of several of those subtypes [134]. In another study, mutated pathways were correlated with different DNA-damage-response mechanisms to detect tumors mainly associated with aneuploidy and those with defective DNA repair or microsatellite instability, thus identifying groups of mutated genes that predict patients' outcomes [135]. Recent work using deep learning has used pathway information, mutations, and copy-number variation to predict patient response to immunotherapy in melanoma [136]. An important benefit of these pathway-based approaches is an emphasis on biological interpretation of predictions, which are often considered more important than model performance (Table 2) [137].

Mutations in a single gene or within a specific pathway may not be sufficient for characterizing cancer development or clinical outcomes. More complex patterns and interactions between mutations confer more information for clinical-prediction tasks. Methods to identify combinations of mutations were used to distinguish tumors from healthy tissues [138], to find patterns of mutually exclusive mutations [139,140] and epistasis [140,141], and to predict patient survival and immunotherapy benefit (Table 2) [142]. Somatic mutations were analyzed by unsupervised NMF and supervised machine-learning methods to predict breast-cancer subtypes, with potential therapeutic significance [143]. Combinations of passenger mutations were recently used in a deep-learning neural network to classify metastatic tumors of unknown origin [144], and found that passengers conferred more information for predicting the tissue of origin. Some computational methods identified mutation patterns to infer the order of mutations in tumor evolution [145–149] or used timing of mutations [150], clonality [151,152], and machine-learning models [153] to predict clinical outcomes (Tables 2 and 3).

Some methods have incorporated tumor mutations with other types of data to predict response to cancer therapies (Tables 2 and 3, Figure 3D) [154]. For instance, in breast cancer, patient response or resistance to paclitaxel or gemcitabine was predicted using SVM models applied to gene mutations, copy number, and expression [155]. This study found that the mutation data alone were not sufficiently informative, likely due to sparsity. Studies have also incorporated genomic and transcriptomic information to predict ICI response and extract clinically relevant targets using a logistic-regression model [156]. Mutation data were incorporated with gene-expression-based diagnostic models to correlate clinically relevant mutations with gene-expression patterns in HCC, allowing for the identification of HCC cells compared to normal liver cells [157]. Other work has used multiomics integration of mutations and other data types with interaction and pathway information to predict ovarian-cancer outcomes [158]. A multiomics approach incorporated mutations, transcription information, epigenetics, and drug targets in a deep-learning framework to predict drug repurposing for cancer treatment [159]. Mutations in specific driver genes were also included in a multiomics integration through deep learning to predict survival in liver-cancer cases [160]. Multiomics integration has also been used to predict TMB in lung-cancer patients, which may potentially be clinically relevant for predicting response to immunotherapy in many cancers (Table 2) [161]. However, the clinical utility of multiomics integration has not been fully demonstrated, where limited amount of complex data is a serious bottleneck for development of computational methods to infer clinically relevant multiomics patterns [162].

Table 2. Methods inferring clinically relevant mutation patterns beyond mutation signatures.

Task Category	Sub-Category	Clinical Relevance	Example Methods	
Identifying cancer drivers	Cancer drivers by mutation frequency	Cancer-driver discovery; Obtaining cancer drivers for prognosis, cancer identification, and treatment	Methods based on mutation frequencies: MutSigCV [104], Invec [105], Music [106]	
			Amino-acid and functional-impact changes: Chasm [107–109], polyphen2 [110], SIFT [112,113]	
			Protein structure: MSEA [114], iPACT [115], GraphPac [116]	
			Phosphorylation-site mutation: ActiveDriver [117]	
			Heat diffusion: HotNet2 [118]	
	Cancer drivers by pathway		Mutated neighbors: MUFFIN [119]	
			Curated pathways and gene-factor modeling: Paradigm [120]	
			Network-based modules [121]	
			Network-based coding and non-coding modules [122]	
			Deep-learning cancer-driver analysis [123]	
Exploring mutated pathways	Predicting outcomes using pathways	Patient-prognosis prediction	Identification of genes associated with DNA-damage response and clinical outcomes [135]	
		Patient response to immunotherapy	Machine learning on clinical mutation data to predict patient response to ICI in melanoma and other cancers [132]	
		Deep learning on pathway information, mutations, and copy-number variation to predict melanoma outcomes [136]		
	Pathways of cancer subtypes	Detecting drug targets through pathways	Drug discovery and tailored treatments [133]	
		Cancer-subtype identification	Cancer-subtype identification and prognosis [134]	
		Prediction of patient response, tumor type, and histology	Gene-network-based stratification using mutation data for prediction [131]	
		Identifying complex patterns of multiple mutations	Inferring interactions between mutations	Interactions conferring sensitivity
Clustering samples	Cancer-type identification		Epistatic effects of genes [140,141]	
			Unsupervised NMF and supervised ML to identify cancer subtypes [143]	
Inferring order of mutations	Identifying of tumors vs. healthy tissues		Inferring timing of mutations	Applying deep-learning neural network to passenger mutations to classify metastatic cancers of unknown origin [144]
				Gene-combination analysis [138]
	Determining timing for predicting clinical outcomes			Mutation patterns to infer order of mutation events [145–149]
		Mutation timing to predict clinical outcome [150]		
Multiomics approach: integrating mutations with other data types	Multiomics outcome prediction	Chemotherapy response or resistance	Using SVM on mutations, copy number, and expression for chemotherapy prediction [155]	
		ICI response or resistance	Genomic and transcriptomic information for response or resistance to ICI [156]	
		Prediction of patient outcomes	Mutation, interaction, and pathway information to identify ovarian-cancer outcomes [158]	
		Mutation-burden prediction for ICI therapy	Lung-cancer mutation-burden prediction using a multiomics approach [160]	
	Cancer classification	Identification of cancerous vs. non-cancerous cells	Identification of HCC cells from normal cells through mutation and expression information [157]	
	Identification of drug targets	Drug repositioning	Mutations, expression, epigenetics, drug targets, and deep learning for drug repositioning [159]	

5. Major Challenges for Clinical Utility of Complex and Data-Driven Mutational Patterns

Despite substantial efforts to identify clinically relevant cancer mutations and patterns, complex patterns beyond single-gene mutations have not been integrated into the clinic. There are several challenges for computational approaches that have prohibited the clinical success of data-driven mutational patterns, which are outlined below, along with potential ways forward to overcome these challenges (Figure 4).

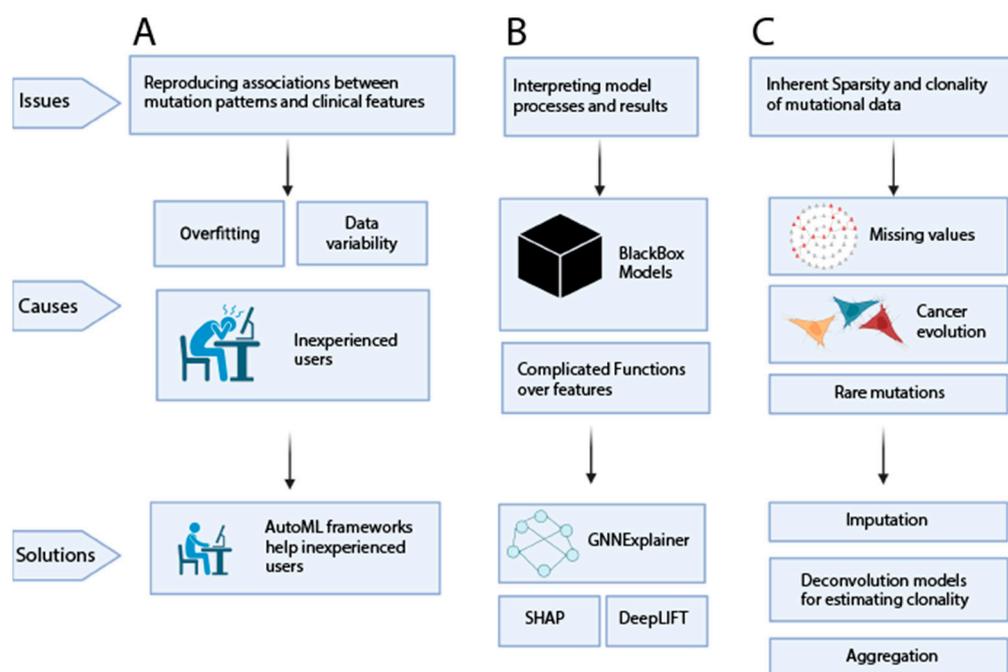


Figure 4. Key challenges for clinical integration of computationally derived mutational patterns and machine-learning methods that address these issues, including (A) reproducibility, (B) interpretability, and (C) inherent sparsity and clonality of the mutational data.

A major and fundamental challenge to overcome is the difficulty of recapitulating associations between mutational patterns and clinical features across multiple studies (Figure 4A) [163,164]. This issue of reproducibility is especially pertinent in the context of clinical significance [165]. Reproducibility issues can result from model underfitting or overfitting (Box 1) due to biological or clinical confounders, small sample settings, data sparsity, or noisy and variable data [166,167]. Both under and overfitting result in failure to generalize findings to other studies, and failure to establish clinically useful biomarkers. Other factors that can lead to unreproducible results are errors and poor documentation of code and data processing [168] and lack of availability of the software and methods used [169]. With multiple parameters and intricate biological datasets, even in well-documented studies, it can be very difficult to fully reproduce results [164]. More complex model and mutation patterns may improve the performance but also risk overfitting. It is therefore important to follow guidelines and tools for reproducible computational work [170,171]. To ensure reproducibility with an eye to clinical integration, correct training, validation, and testing practices in machine learning should be followed, along with standardized methods, automation, transparency, and good coding practices [172–174]. Studies should also ensure generalization across different, biologically independent datasets [175–177].

Tools are also being developed to assist non-specialists with ML applications (Table 3). One example for such tools are automated machine-learning (AutoML) pipelines, which handle the required tasks of applying machine learning to user-provided datasets. In recent years, several frameworks that handle hyperparameter optimization and model selection have become available [178–185]. Such frameworks can also be adapted by non-expert machine-learning users in biomedicine, which can help support reproducibility for machine-learning applications. Beyond the model itself, failure to reproduce results can also be caused by poor laboratory or data-handling practices, human error such as mislabeling, or contaminants, among other sources of variability [186,187].

Another important challenge to overcome in the path to clinical integration is the issue of biologically interpretable results (Figure 4B) [113,137,188]. An interpretable model allows for an understanding of the data that go into the model, the processes applied by the model, and of how the model arrives at the results [189–191]. This is important because clinicians

and biologists typically favor biological interpretability over black-box models [192,193], even at the expense of the predictive capability of the model. An interpretable model can also provide for follow-up biological discoveries and a better understanding of unexpected results [194]. More complex models or patterns that may demonstrate better performance are likely to be less interpretable. For example, cancer-driver identification is complex, and increasingly more sophisticated models have been developed to address this complexity, but even more complex models have not necessarily expanded on the drivers being discovered.

To address this complexity, many interpretation approaches have been proposed to provide explanations for the trained models' predictions and the features driving the model to make a specific prediction (Table 3). LIME is a popular interpretation tool that learns a new interpretable model that can better explain a less interpretable model. Numerous studies have successfully applied LIME to provide interpretation of complex models, including in biomedicine [195,196]. Another popular interpretation method is DeepLIFT [197], which calculates the contribution of neurons in a trained neural network by evaluating the difference in activation from a chosen representative reference. DeepLIFT has also been useful for interpreting model prediction in genomic datasets [198–203]. Another interpretive model is SHapley Additive exPlanations (SHAP) [204], which is based on the Shapley value from game theory. This method generates contribution values called SHAP values for each feature, which represents the differences between the actual prediction and the expected prediction of a trained model. SHAP values not only provide insight into how much each feature contributes to the prediction but also to the direction of the contribution, either towards the positive class or the negative class. Multiple biomedical studies have used SHAP to provide clear explanations of features driving predictions [124,205–209].

Another form of explanatory methods is through biological-network explanations (Table 3) [210,211]. Biological networks have been used to build network-based predictive models based on graph convolutional networks (GNN) (Box 1) [212–214]. An interpretability challenge for a GNN learning biological networks is understanding the network structure and how sub-networks contribute to the prediction. GNNExplainer [215] provides explanations of GNN-based prediction by identifying a dense sub-network structure along with a small subset of node features that play an important role in the GNN-based prediction. GNNExplainer can be used to understand the contributions of sub-networks' nodes and their roles in determining predictions, allowing for biological interpretability. Interpretation models can help bridge the gap between model developers and clinicians, potentially allowing for clinical utility of more complex model-based mutational patterns [190,191,216–218].

Another challenge for uncovering complex patterns of mutations is linked to the sparse nature of mutation data themselves (Figure 4C). Mutations, even in cancer, are generally infrequent when the entire genome or exome is taken into consideration [219,220]. This sparsity extends to other sources of biological data [220–225]. Most machine-learning models have difficulty learning and picking up patterns for prediction based on sparse data, which can lead to overfitting [219,226–230]. This results in poor reproducibility [227–230]. Feeding into this issue is the fact that cancer is highly heterogeneous, and rare events do not preclude clinical relevance [231,232]. Including the methods discussed above, aggregation of mutations can potentially mitigate this sparsity. Although aggregation may reduce sparsity, care must be taken to ensure results are biologically interpretable [233,234].

Another factor that can lead to sparsity is missing data. Missing data can result from experimental design or different types of human errors [235,236]. In addition to sparseness, missing data can also lead to biased datasets and results [237]. Several techniques have been developed to handle missing data, such as imputing missed instances with estimated values [238–240]. Machine-learning methods can also be used to perform data imputation, such as regression- and ensemble-based models [241,242]. Furthermore, several methods have been developed recently to improve the quality of data imputation [243–247].

Another challenge is linked to how mutations are accumulated in cancer evolution. As cancer develops, mutations arise in certain cell lineages, and tumor mutations are therefore clonal and not homogenous [248–250]. Different cells within the same tumor have different

clonal lineages and therefore different patterns of mutations [248,251]. Within the same patient, lineages can be very different. This complicates typical data analysis because the data being analyzed are subjected to specific clonal lineages where some mutations may be misrepresented. As a result, in bulk datasets the actual clinically relevant mutational players may be obscured [252] and the clonal composition of the tumor may change over time, especially in response to treatment. Several methods have been developed to address issues surrounding clonality [251,253–257], but more work is needed to address clonality in the context of computational tools and modeling.

6. Summary

With the introduction of next-generation sequencing, numerous causal and actionable mutations have been identified and used clinically as biomarkers or for new targeted therapies. Due to the increasing realization of the vast complexity underlying tumorigenesis, future clinical breakthroughs are likely to increasingly rely on computational methods to identify these clinically actionable patterns of mutations. Mutation signatures allow for exploration of intricate patterns of mutations in cancer, effectively identifying mutational patterns to describe DDR pathways and environmental effects. However, mutation signatures require extensive sequencing of cancer genomes, limiting clinical applications beyond these purposes. Other methods have been developed to uncover complex patterns of mutations for clinical use. These include methods that identify drivers of cancer, methods that predict clinical outcomes by integrating mutations with biological pathways, and methods incorporating other types of omics. However, such methods have yet to be integrated into the clinic. The major challenges for clinical integration of computationally driven mutational patterns are lack of reproducibility, the difficulty of interpreting complex models, and issues associated with intrinsic attributes of cancer-mutational data, such as sparsity and clonality. State-of-the-art computational and machine learning can be adjusted to address these issues, improving the interpretation of complex models and enhancing reproducibility. With the consistent accumulation of cancer-genomic datasets and the complexity of cancer genomes, many of the next great clinical breakthroughs in cancer research will rely on computational tools to fully understand the complicated patterns of mutations that characterize cancer.

Table 3. Summary of tools reviewed in this article, software resources, and mention of the review section where tools are referenced. All referenced tools and websites were accessed between 16 February 2023 and 19 February 2023.

Method Name	Method Description	Code/Tool	Reference	Review Section
IntOGen	A method to access the database of mutational-cancer drivers	https://www.intogen.org/search	[2]	1
SigProfiler	Framework for deciphering mutation signatures from mutational catalogues of cancer genomes	https://www.mathworks.com/matlabcentral/fileexchange/38724-sigprofiler	[24–27]	2.1
MutSpec	Somatic-mutation analysis in human and mouse	https://toolshed.g2.bx.psu.edu/	[31]	2.1
MutSignatures	Cancer-mutation-signatures analysis	https://github.com/dami82/mutSignatures	[32]	2.1
SigneR	Bayesian approach to discover mutation signatures	http://bioconductor.org/packages/release/bioc/html/signer.html	[35]	2.1

Table 3. Cont.

Method Name	Method Description	Code/Tool	Reference	Review Section
pmsignature	Probabilistic model to infer and visualize cancer-mutation signatures	https://github.com/friend1ws/pmsignature https://friend1ws.shinyapps.io/pmsignature_shiny/	[36]	2.1
SomaticSignatures	Inferring characteristics of mutation signatures	https://www.bioconductor.org/packages/release/bioc/html/SomaticSignatures.html	[38]	2.1
Helmsman	Mutation-signature analysis	https://github.com/carjed/helmsman	[39]	2.1
deconstructSigs	Mutation signature by machine learning	https://github.com/raerose01/deconstructSigs	[40]	2.1
SignatureEstimation	Discovering the existence of mutation signatures in cancer	https://www.ncbi.nlm.nih.gov/CBBresearch/Przytycka/index.cgi#signatureestimation	[41]	2.1
Signal	Mutation-signature analysis	https://github.com/Nik-Zainal-Group/signature.tools.lib	[42]	2.1
MutationalPatterns	Comprehensive analysis of mutation processes across the genome	http://bioconductor.org/packages/release/bioc/html/MutationalPatterns.html	[43]	2.1
	Identification of mutation signatures	https://github.com/team113sanger/mouse-mutation-signatures	[55]	2.2
CHORD	Classifier identifying homologous recombination deficiency across cancers	https://github.com/UMCUGenetics/CHORD	[68]	3.1
SigMA	Identification of mutation signatures	https://github.com/parklab/SigMA	[70]	3.1
mutfootprints	Identification of mutation footprint of and for cancer treatment	https://bitbucket.org/bbglab/mutfootprints/src/master/	[88]	3.2
	Identification of mutation signatures	https://github.com/UMCUGenetics/5FU	[89]	3.2
CUPLR	Classification of primary-tumor identity of metastatic tumors	https://github.com/UMCUGenetics/CUPLR	[98]	3.2
MutSigCV	Identification of mutated genes in cancer	https://software.broadinstitute.org/cancer/cga/mutsig	[104]	4
inVex	Identification of positive selection for non-silent mutations	https://software.broadinstitute.org/cancer/cga/invex	[105]	4
MuSiC	Identification of mutational relevance in cancer genome	http://gmt.genome.wustl.edu/	[106]	4
CHASM	Identification of important biological single-nucleotide mutations in cancer	http://wiki.chasmsoftware.org/index.php/Main_Page	[107–109]	4
PolyPhen-2	Classification of missense-mutation damaging effects on protein	http://genetics.bwh.harvard.edu/pph2/	[110]	4
e-Driver	Identification of protein functional regions driving cancer	https://github.com/eduardporta/e-Driver	[111]	
SIFT	Classification of amino-acid-substitution impact on proteins	https://sift.bii.a-star.edu.sg/	[112,113]	4

Table 3. Cont.

Method Name	Method Description	Code/Tool	Reference	Review Section
MSEA	Classification of cancer genes based on patterns of mutation hotspots	https://github.com/bsml320/MSEA	[114]	4
iPAC	Identification of non-random somatic mutations in proteins	http://www.bioconductor.org/packages/2.12/bioc/html/iPAC.html	[115]	4
GraphPAC	Identification of non-random somatic mutations in proteins	http://bioconductor.org/packages/release/bioc/html/GraphPAC.html	[116]	4
ActiveDriver	Effect of mutation on post-translational signaling	http://www.baderlab.org/Software/ActiveDriver	[117]	4
HotNet2	Identification of rare somatic-mutation combinations in pathways and protein complexes	http://compbio-research.cs.brown.edu/pancancer/hotnet2/#/ http://compbio.cs.brown.edu/software/	[118]	4
MUFFINN	Cancer-gene detection through network analysis of somatic mutations	http://www.inetbio.org/muffinn/	[119]	4
boostDM	Identification of driver mutations in cancer genes from observed mutations in human tumors	https://zenodo.org/record/4813082#.Y9L38dLMKV4	[124]	4
DEOGEN2/MutaFrame	Classification of single-amino-acid variant loss in human proteins	http://babylone.3bio.ulb.ac.be/MutaFrame/	[126]	4
PrimateAI	Classification of clinical impact of human mutations	https://basespace.illumina.com/s/cPgCSmecvhb4	[128]	4
	Classification of immune-checkpoint-inhibitor therapy response	https://github.com/AuslanderLab/Mutated_pathway_ICI_prediction	[132]	4
	Identification of associations between driver mutations and chromosomal aberrations	https://github.com/noamaus/INTERPLAY-TUMOR-CODES	[135]	4
KP-NET	Classification of immunotherapy response	https://github.com/0219zhang/KP-NET	[136]	4
	Causal identifications of individual instances of cancer	https://bitbucket.org/sajal000/multihit-combinations/src/master/	[138]	4
CLICnet	Identification of somatic-mutation combinations that predict cancer survival	https://github.com/gussow/clicnet	[142]	4
	Classification of primary and metastatic tumors	https://github.com/ICGC-TCGA-PanCancer/TumorType-WGS	[144]	4
SMASH	Identification of somatic-mutation associations	https://github.com/Sun-lab/SMASH	[152]	4
	Learning evolution of a tumor through mutational time series	https://github.com/noamaus/LSTM-Mutational-series	[153]	4
	Classification outcomes of checkpoint inhibition by tumor and immune-signal combination	https://zenodo.org/record/5528497#.Y9Ps1dLMKV4	[156]	4
DeepDRK	Drug response prediction	https://github.com/wangyc82/DeepDRK	[159]	4

Table 3. Cont.

Method Name	Method Description	Code/Tool	Reference	Review Section
MetAML	Prediction of metagenomics-based tasks	https://github.com/segatalab/metaml	[176]	5
	Generalization in machine learning for dataset characteristics	https://github.com/pietrobarbiero/dataset-characteristics	[177]	5
Auptimizer	Hyperparameter optimization	https://github.com/LGE-ARC-AdvancedAI/auptimizer	[178]	5
TPOT	Automated ML-tree-based optimization pipeline	https://github.com/EpistasisLab/tpot	[181,182]	5
Hyperband	Hyperparameter optimization	https://github.com/automl/pylearningcurvepredictor	[183]	5
DanQ	Classification of the function of DNA de novo mutations from sequences	http://github.com/uci-cbcl/DanQ	[188]	5
	An explainable machine learning tool of severity-level predictions of COVID-19 patients	https://github.com/freddygabbay/covid19explainableML	[196]	5
DeepLIFT	An explainable machine-learning tool	https://github.com/kundajelab/deeplift	[197]	5
SpliceRover	Classification of donor and acceptor splice site	http://bioit2.irc.ugent.be/rover/splicerover/	[199]	5
RIDDLE	Imputation technique using deep learning	https://github.com/jisungk/RIDDLE	[200]	5
P-NET	Classification of prostate cancer	https://github.com/marakeby/pnet_prostate_paper	[203]	5
SHAP	An explainable machine learning tool	https://github.com/slundberg/shap	[204]	5
devCellPy	Classification of cell types across complex annotation hierarchies	https://github.com/devCellPy-Team/devCellPy	[205]	5
BCrystal	An interpretable sequence-based protein-crystallization predictor	https://github.com/raghvendra5688/BCrystal	[206]	5
MetaNet	Metastatic-risk assessment of a primary tumor	https://github.com/WangLabHKUST/METANET-analysis	[207]	5
Ocelot	Prediction of relationships across histone modifications	https://github.com/GuanLab/Ocelot	[208]	5
DeepHF	Optimization of CRISPR guide RNA design using deep learning for two high-fidelity Cas9 variants	https://github.com/izhangcd/DeepHF http://www.deephf.com/#/home	[209]	5
MTGCN	Identification of cancer-driver genes	https://github.com/weiba/MTGCN	[213]	5
GNNExplainer	An explainable graph neural-network tool	https://github.com/RexYing/gnn-model-explainer	[215]	5
SBMClone	Identification of tumor clones in sparse single-cell-mutation data	https://github.com/raphael-group/SBMClone	[221]	5
Mix-MMM	Identification of mutation signatures from sparse mutation data	https://github.com/itaysason/Mix-MMM	[222]	5

Table 3. Cont.

Method Name	Method Description	Code/Tool	Reference	Review Section
JDINAC	Identification of differential interaction patterns of network activation using high-dimensional sparse omics data	https://github.com/jjiadong/JDINAC	[223]	5
MoGP	Identification of patterns in amyotrophic lateral-sclerosis progression from sparse longitudinal data	https://github.com/fraenkel-lab/mogp	[225]	5
	Multi-cancer analysis of clonality in paired primary tumors and metastases	https://github.com/cancersysbio/pan-metastasis	[251]	5
CHESS	Spatial stochastic tumor-growth model to simulate multi-region sequencing data derived from spatial sampling of neoplasm	https://github.com/kchkhaidze/CHESS.cpp	[256]	5

Author Contributions: A.P., A.E. and N.A. wrote the paper, which was edited and supervised by N.A. and B.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Institutes of Health under Award Number R00CA252025.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Figures were partially created with BioRender.com accessed on 21 February 2023.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Van Hoeck, A.; Tjoonk, N.H.; van Boxtel, R.; Cuppen, E. Portrait of a Cancer: Mutational Signature Analyses for Cancer Diagnostics. *BMC Cancer* **2019**, *19*, 457. [CrossRef] [PubMed]
2. Martínez-Jiménez, F.; Muiños, F.; Sentís, I.; Deu-Pons, J.; Reyes-Salazar, I.; Arnedo-Pac, C.; Mularoni, L.; Pich, O.; Bonet, J.; Kranas, H.; et al. A Compendium of Mutational Cancer Driver Genes. *Nat. Rev. Cancer* **2020**, *20*, 555–572. [CrossRef] [PubMed]
3. Kamel, H.F.M.; Al-Amodi, H.S.A.B. Exploitation of Gene Expression and Cancer Biomarkers in Paving the Path to Era of Personalized Medicine. *Genom. Proteom. Bioinform.* **2017**, *15*, 220–235. [CrossRef]
4. Gayther, S.A.; Warren, W.; Mazoyer, S.; Russell, P.A.; Harrington, P.A.; Chiano, M.; Seal, S.; Hamoudi, R.; van Rensburg, E.J.; Dunning, A.M.; et al. Germline Mutations of the BRCA1 Gene in Breast and Ovarian Cancer Families Provide Evidence for a Genotype–Phenotype Correlation. *Nat. Genet.* **1995**, *11*, 428–433. [CrossRef]
5. Roy, R.; Chun, J.; Powell, S.N. BRCA1 and BRCA2: Different Roles in a Common Pathway of Genome Protection. *Nat. Rev. Cancer* **2012**, *12*, 68–78. [CrossRef] [PubMed]
6. Turk, A.; Wisinski, K.B. PARP Inhibition in BRCA-Mutant Breast Cancer. *Cancer* **2018**, *124*, 2498–2506. [CrossRef] [PubMed]
7. Proietti, I.; Skroza, N.; Michelini, S.; Mambrin, A.; Balduzzi, V.; Bernardini, N.; Marchesiello, A.; Tolino, E.; Volpe, S.; Maddalena, P.; et al. BRAF Inhibitors: Molecular Targeting and Immunomodulatory Actions. *Cancers* **2020**, *12*, 1823. [CrossRef]
8. Hertzman Johansson, C.; Egyhazi Brage, S. BRAF Inhibitors in Cancer Therapy. *Pharmacol. Ther.* **2014**, *142*, 176–182. [CrossRef]
9. Liu, J.; Kang, R.; Tang, D. The KRAS-G12C Inhibitor: Activity and Resistance. *Cancer Gene* **2022**, *29*, 875–878. [CrossRef]
10. Rosell, R.; Aguilar, A.; Pedraz, C.; Chaib, I. KRAS Inhibitors, Approved. *Nat. Cancer* **2021**, *2*, 1254–1256. [CrossRef]
11. Cerchione, C.; Romano, A.; Daver, N.; DiNardo, C.; Jabbour, E.J.; Konopleva, M.; Ravandi-Kashani, F.; Kadia, T.; Martelli, M.P.; Isidori, A.; et al. IDH1/IDH2 Inhibition in Acute Myeloid Leukemia. *Front. Oncol.* **2021**, *11*, 639387. [CrossRef] [PubMed]
12. Sun, X.; Turcan, S. From Laboratory Studies to Clinical Trials: Temozolomide Use in IDH-Mutant Gliomas. *Cells* **2021**, *10*, 1225. [CrossRef] [PubMed]

13. Kwak, E.L.; Bang, Y.-J.; Camidge, D.R.; Shaw, A.T.; Solomon, B.; Maki, R.G.; Ou, S.-H.I.; Dezube, B.J.; Jänne, P.A.; Costa, D.B.; et al. Anaplastic Lymphoma Kinase Inhibition in Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **2010**, *363*, 1693–1703. [[CrossRef](#)] [[PubMed](#)]
14. Hallberg, B.; Palmer, R.H. The Role of the ALK Receptor in Cancer Biology. *Ann. Oncol.* **2016**, *27*, iii4–iii15. [[CrossRef](#)] [[PubMed](#)]
15. Shaw, A.T.; Kim, D.-W.; Mehra, R.; Tan, D.S.W.; Felip, E.; Chow, L.Q.M.; Camidge, D.R.; Vansteenkiste, J.; Sharma, S.; De Pas, T.; et al. Ceritinib in ALK-Rearranged Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **2014**, *370*, 1189–1197. [[CrossRef](#)] [[PubMed](#)]
16. Slamon, D.J.; Leyland-Jones, B.; Shak, S.; Fuchs, H.; Paton, V.; Bajamonde, A.; Fleming, T.; Eiermann, W.; Wolter, J.; Pegram, M.; et al. Use of Chemotherapy plus a Monoclonal Antibody against HER2 for Metastatic Breast Cancer That Overexpresses HER2. *N. Engl. J. Med.* **2001**, *344*, 783–792. [[CrossRef](#)]
17. Swain, S.M.; Kim, S.-B.; Cortés, J.; Ro, J.; Semiglazov, V.; Campone, M.; Ciruelos, E.; Ferrero, J.-M.; Schneeweiss, A.; Knott, A.; et al. Pertuzumab, Trastuzumab, and Docetaxel for HER2-Positive Metastatic Breast Cancer (CLEOPATRA Study): Overall Survival Results from a Randomised, Double-Blind, Placebo-Controlled, Phase 3 Study. *Lancet Oncol.* **2013**, *14*, 461–471. [[CrossRef](#)]
18. Cameron, D.; Casey, M.; Press, M.; Lindquist, D.; Pienkowski, T.; Romieu, C.G.; Chan, S.; Jagiello-Gruszfeld, A.; Kaufman, B.; Crown, J.; et al. A Phase III Randomized Comparison of Lapatinib plus Capecitabine versus Capecitabine Alone in Women with Advanced Breast Cancer That Has Progressed on Trastuzumab: Updated Efficacy and Biomarker Analyses. *Breast Cancer Res. Treat.* **2008**, *112*, 533–543. [[CrossRef](#)]
19. Martin, M.; Bonnetterre, J.; Geyer, C.E.; Ito, Y.; Ro, J.; Lang, I.; Kim, S.-B.; Germa, C.; Vermette, J.; Wang, K.; et al. A Phase Two Randomised Trial of Neratinib Monotherapy versus Lapatinib plus Capecitabine Combination Therapy in Patients with HER2+ Advanced Breast Cancer. *Eur. J. Cancer* **2013**, *49*, 3763–3772. [[CrossRef](#)]
20. Johnston, S.; Pippen, J.; Pivot, X.; Lichinitser, M.; Sadeghi, S.; Dieras, V.; Gomez, H.L.; Romieu, G.; Manikhas, A.; Kennedy, M.J.; et al. Lapatinib Combined With Letrozole Versus Letrozole and Placebo As First-Line Therapy for Postmenopausal Hormone Receptor-Positive Metastatic Breast Cancer. *JCO* **2009**, *27*, 5538–5546. [[CrossRef](#)]
21. Villanueva, C.; Romieu, G.; Salvat, J.; Chaigneau, L.; Merrouche, Y.; N’Guyen, T.; Vuillemin, A.T.; Demarchi, M.; Dobi, E.; Pivot, X. Phase II Study Assessing Lapatinib Added to Letrozole in Patients with Progressive Disease under Aromatase Inhibitor in Metastatic Breast Cancer—Study BES 06. *Target. Oncol.* **2013**, *8*, 137–143. [[CrossRef](#)] [[PubMed](#)]
22. Basu, A.K. DNA Damage, Mutagenesis and Cancer. *Int. J. Mol. Sci.* **2018**, *19*, 970. [[CrossRef](#)] [[PubMed](#)]
23. Nik-Zainal, S.; Alexandrov, L.B.; Wedge, D.C.; Van Loo, P.; Greenman, C.D.; Raine, K.; Jones, D.; Hinton, J.; Marshall, J.; Stebbings, L.A.; et al. Mutational Processes Molding the Genomes of 21 Breast Cancers. *Cell* **2012**, *149*, 979–993. [[CrossRef](#)] [[PubMed](#)]
24. Alexandrov, L.B.; Nik-Zainal, S.; Wedge, D.C.; Aparicio, S.A.J.R.; Behjati, S.; Biankin, A.V.; Bignell, G.R.; Bolli, N.; Borg, A.; Børresen-Dale, A.-L.; et al. Signatures of Mutational Processes in Human Cancer. *Nature* **2013**, *500*, 415–421. [[CrossRef](#)] [[PubMed](#)]
25. Alexandrov, L.B.; Kim, J.; Haradhvala, N.J.; Huang, M.N.; Tian Ng, A.W.; Wu, Y.; Boot, A.; Covington, K.R.; Gordenin, D.A.; Bergstrom, E.N.; et al. The Repertoire of Mutational Signatures in Human Cancer. *Nature* **2020**, *578*, 94–101. [[CrossRef](#)] [[PubMed](#)]
26. Baez-Ortega, A.; Gori, K. Computational Approaches for Discovery of Mutational Signatures in Cancer. *Brief. Bioinform.* **2019**, *20*, 77–88. [[CrossRef](#)]
27. Omichessan, H.; Severi, G.; Perduca, V. Computational Tools to Detect Signatures of Mutational Processes in DNA from Tumours: A Review and Empirical Comparison of Performance. *PLoS ONE* **2019**, *14*, e0221235. [[CrossRef](#)]
28. Alexandrov, L.B.; Jones, P.H.; Wedge, D.C.; Sale, J.E.; Campbell, P.J.; Nik-Zainal, S.; Stratton, M.R. Clock-like Mutational Processes in Human Somatic Cells. *Nat. Genet.* **2015**, *47*, 1402–1407. [[CrossRef](#)]
29. Lee, D.D.; Seung, H.S. Learning the Parts of Objects by Non-Negative Matrix Factorization. *Nature* **1999**, *401*, 788–791. [[CrossRef](#)]
30. Paatero, P.; Tapper, U. Positive Matrix Factorization: A Non-Negative Factor Model with Optimal Utilization of Error Estimates of Data Values. *Environmetrics* **1994**, *5*, 111–126. [[CrossRef](#)]
31. Ardin, M.; Cahais, V.; Castells, X.; Bouaoun, L.; Byrnes, G.; Herceg, Z.; Zavadil, J.; Olivier, M. MutSpec: A Galaxy Toolbox for Streamlined Analyses of Somatic Mutation Spectra in Human and Mouse Cancer Genomes. *BMC Bioinform.* **2016**, *17*, 170. [[CrossRef](#)] [[PubMed](#)]
32. Fantini, D.; Vidimar, V.; Yu, Y.; Condello, S.; Meeks, J.J. MutSignatures: An R Package for Extraction and Analysis of Cancer Mutational Signatures. *Sci. Rep.* **2020**, *10*, 18217. [[CrossRef](#)] [[PubMed](#)]
33. Kasar, S.; Kim, J.; Improgio, R.; Tiao, G.; Polak, P.; Haradhvala, N.; Lawrence, M.S.; Kiezun, A.; Fernandes, S.M.; Bahl, S.; et al. Whole-Genome Sequencing Reveals Activation-Induced Cytidine Deaminase Signatures during Indolent Chronic Lymphocytic Leukaemia Evolution. *Nat. Commun.* **2015**, *6*, 8866. [[CrossRef](#)] [[PubMed](#)]
34. Kim, J.; Mouw, K.W.; Polak, P.; Braunstein, L.Z.; Kamburov, A.; Kwiatkowski, D.J.; Rosenberg, J.E.; Van Allen, E.M.; D’Andrea, A.; Getz, G. Somatic ERCC2 Mutations Are Associated with a Distinct Genomic Signature in Urothelial Tumors. *Nat. Genet.* **2016**, *48*, 600–606. [[CrossRef](#)]
35. Rosales, R.A.; Drummond, R.D.; Valieris, R.; Dias-Neto, E.; da Silva, I.T. Signer: An Empirical Bayesian Approach to Mutational Signature Discovery. *Bioinformatics* **2017**, *33*, 8–16. [[CrossRef](#)] [[PubMed](#)]
36. Shiraishi, Y.; Tremmel, G.; Miyano, S.; Stephens, M. A Simple Model-Based Approach to Inferring and Visualizing Cancer Mutation Signatures. *PLoS Genet.* **2015**, *11*, e1005657. [[CrossRef](#)] [[PubMed](#)]
37. Fischer, A.; Illingworth, C.J.; Campbell, P.J.; Mustonen, V. EMu: Probabilistic Inference of Mutational Processes and Their Localization in the Cancer Genome. *Genome Biol.* **2013**, *14*, R39. [[CrossRef](#)]

38. Gehring, J.S.; Fischer, B.; Lawrence, M.; Huber, W. SomaticSignatures: Inferring Mutational Signatures from Single-Nucleotide Variants. *Bioinformatics* **2015**, *31*, 3673–3675. [[CrossRef](#)]
39. Carlson, J.; Li, J.Z.; Zöllner, S. Helmsman: Fast and Efficient Mutation Signature Analysis for Massive Sequencing Datasets. *BMC Genom.* **2018**, *19*, 845. [[CrossRef](#)]
40. Rosenthal, R.; McGranahan, N.; Herrero, J.; Taylor, B.S.; Swanton, C. DeconstructSigs: Delineating Mutational Processes in Single Tumors Distinguishes DNA Repair Deficiencies and Patterns of Carcinoma Evolution. *Genome Biol.* **2016**, *17*, 31. [[CrossRef](#)]
41. Huang, X.; Wojtowicz, D.; Przytycka, T.M. Detecting Presence of Mutational Signatures in Cancer with Confidence. *Bioinformatics* **2018**, *34*, 330–337. [[CrossRef](#)]
42. Degasperi, A.; Amarante, T.D.; Czarnecki, J.; Shooter, S.; Zou, X.; Glodzik, D.; Morganello, S.; Nanda, A.S.; Badja, C.; Koh, G.; et al. A Practical Framework and Online Tool for Mutational Signature Analyses Show Intertissue Variation and Driver Dependencies. *Nat. Cancer* **2020**, *1*, 249–263. [[CrossRef](#)]
43. Blokzijl, F.; Janssen, R.; van Boxtel, R.; Cuppen, E. MutationalPatterns: Comprehensive Genome-Wide Analysis of Mutational Processes. *Genome Med.* **2018**, *10*, 33. [[CrossRef](#)] [[PubMed](#)]
44. Di Noia, J.M.; Neuberger, M.S. Molecular Mechanisms of Antibody Somatic Hypermutation. *Annu. Rev. Biochem.* **2007**, *76*, 1–22. [[CrossRef](#)] [[PubMed](#)]
45. Chan, K.; Roberts, S.A.; Klimczak, L.J.; Sterling, J.F.; Saini, N.; Malc, E.P.; Kim, J.; Kwiatkowski, D.J.; Fargo, D.C.; Mieczkowski, P.A.; et al. An APOBEC3A Hypermutation Signature Is Distinguishable from the Signature of Background Mutagenesis by APOBEC3B in Human Cancers. *Nat. Genet.* **2015**, *47*, 1067–1072. [[CrossRef](#)] [[PubMed](#)]
46. Petljak, M.; Alexandrov, L.B.; Brammell, J.S.; Price, S.; Wedge, D.C.; Grossmann, S.; Dawson, K.J.; Ju, Y.S.; Iorio, F.; Tubio, J.M.C.; et al. Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. *Cell* **2019**, *176*, 1282–1294.e20. [[CrossRef](#)] [[PubMed](#)]
47. Chang, L.; Ruiz, P.; Ito, T.; Sellers, W.R. Targeting Pan-Essential Genes in Cancer: Challenges and Opportunities. *Cancer Cell* **2021**, *39*, 466–479. [[CrossRef](#)] [[PubMed](#)]
48. Buisson, R.; Lawrence, M.S.; Benes, C.H.; Zou, L. APOBEC3A and 3B Activities Render Cancer Cells Susceptible to ATR Inhibition. *Cancer Res.* **2017**, *77*, 4567–4578. [[CrossRef](#)] [[PubMed](#)]
49. Moody, S.; Senkin, S.; Islam, S.M.A.; Wang, J.; Nasrollahzadeh, D.; Cortez Cardoso Penha, R.; Fitzgerald, S.; Bergstrom, E.N.; Atkins, J.; He, Y.; et al. Mutational Signatures in Esophageal Squamous Cell Carcinoma from Eight Countries with Varying Incidence. *Nat. Genet.* **2021**, *53*, 1553–1563. [[CrossRef](#)]
50. Pilati, C.; Shinde, J.; Alexandrov, L.B.; Assié, G.; André, T.; Hélias-Rodzewicz, Z.; Ducoudray, R.; Le Corre, D.; Zucman-Rossi, J.; Emile, J.-F.; et al. Mutational Signature Analysis Identifies MUTYH Deficiency in Colorectal Cancers and Adrenocortical Carcinomas. *J. Pathol.* **2017**, *242*, 10–15. [[CrossRef](#)]
51. Alexandrov, L.B.; Ju, Y.S.; Haase, K.; Van Loo, P.; Martincorena, I.; Nik-Zainal, S.; Totoki, Y.; Fujimoto, A.; Nakagawa, H.; Shibata, T.; et al. Mutational Signatures Associated with Tobacco Smoking in Human Cancer. *Science* **2016**, *354*, 618–622. [[CrossRef](#)] [[PubMed](#)]
52. Haradhvala, N.J.; Kim, J.; Maruvka, Y.E.; Polak, P.; Rosebrock, D.; Livitz, D.; Hess, J.M.; Leshchiner, I.; Kamburov, A.; Mouw, K.W.; et al. Distinct Mutational Signatures Characterize Concurrent Loss of Polymerase Proofreading and Mismatch Repair. *Nat. Commun.* **2018**, *9*, 1746. [[CrossRef](#)] [[PubMed](#)]
53. Behjati, S.; Gundem, G.; Wedge, D.C.; Roberts, N.D.; Tarpey, P.S.; Cooke, S.L.; Van Loo, P.; Alexandrov, L.B.; Ramakrishna, M.; Davies, H.; et al. Mutational Signatures of Ionizing Radiation in Second Malignancies. *Nat. Commun.* **2016**, *7*, 12605. [[CrossRef](#)] [[PubMed](#)]
54. Rose Li, Y.; Halliwill, K.D.; Adams, C.J.; Iyer, V.; Riva, L.; Mamunur, R.; Jen, K.-Y.; del Rosario, R.; Fredlund, E.; Hirst, G.; et al. Mutational Signatures in Tumours Induced by High and Low Energy Radiation in Trp53 Deficient Mice. *Nat. Commun.* **2020**, *11*, 394. [[CrossRef](#)]
55. Riva, L.; Pandiri, A.R.; Li, Y.R.; Droop, A.; Hewinson, J.; Quail, M.A.; Iyer, V.; Shepherd, R.; Herbert, R.A.; Campbell, P.J.; et al. The Mutational Signature Profile of Known and Suspected Human Carcinogens in Mice. *Nat. Genet.* **2020**, *52*, 1189–1197. [[CrossRef](#)] [[PubMed](#)]
56. Huang, R.; Zhou, P.-K. DNA Damage Repair: Historical Perspectives, Mechanistic Pathways and Clinical Translation for Targeted Cancer Therapy. *Signal Transduct. Target.* **2021**, *6*, 254. [[CrossRef](#)] [[PubMed](#)]
57. Alhmod, J.F.; Woolley, J.F.; Al Moustafa, A.-E.; Malki, M.I. DNA Damage/Repair Management in Cancers. *Cancers* **2020**, *12*, 1050. [[CrossRef](#)]
58. Negrini, S.; Gorgoulis, V.G.; Halazonetis, T.D. Genomic Instability—An Evolving Hallmark of Cancer. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 220–228. [[CrossRef](#)]
59. Lengauer, C.; Kinzler, K.W.; Vogelstein, B. Genetic Instabilities in Human Cancers. *Nature* **1998**, *396*, 643–649. [[CrossRef](#)]
60. Turgeon, M.-O.; Perry, N.J.S.; Poulogiannis, G. DNA Damage, Repair, and Cancer Metabolism. *Front. Oncol.* **2018**, *8*. [[CrossRef](#)]
61. Li, L.; Guan, Y.; Chen, X.; Yang, J.; Cheng, Y. DNA Repair Pathways in Cancer Therapy and Resistance. *Front. Pharmacol.* **2021**, *11*. [[CrossRef](#)] [[PubMed](#)]
62. Schreiber, V.; Dantzer, F.; Ame, J.-C.; de Murcia, G. Poly(ADP-Ribose): Novel Functions for an Old Molecule. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 517–528. [[CrossRef](#)] [[PubMed](#)]

63. Murai, J.; Huang, S.N.; Das, B.B.; Renaud, A.; Zhang, Y.; Doroshow, J.H.; Ji, J.; Takeda, S.; Pommier, Y. Trapping of PARP1 and PARP2 by Clinical PARP Inhibitors. *Cancer Res.* **2012**, *72*, 5588–5599. [[CrossRef](#)] [[PubMed](#)]
64. Davies, H.; Glodzik, D.; Morganella, S.; Yates, L.R.; Staaf, J.; Zou, X.; Ramakrishna, M.; Martin, S.; Boyault, S.; Sieuwerts, A.M.; et al. HRDetect Is a Predictor of BRCA1 and BRCA2 Deficiency Based on Mutational Signatures. *Nat. Med.* **2017**, *23*, 517–525. [[CrossRef](#)]
65. Staaf, J.; Glodzik, D.; Bosch, A.; Vallon-Christersson, J.; Reuterswärd, C.; Häkkinen, J.; Degasperri, A.; Amarante, T.D.; Saal, L.H.; Hegardt, C.; et al. Whole-Genome Sequencing of Triple-Negative Breast Cancers in a Population-Based Clinical Study. *Nat. Med.* **2019**, *25*, 1526–1533. [[CrossRef](#)]
66. Nones, K.; Johnson, J.; Newell, F.; Patch, A.M.; Thorne, H.; Kazakoff, S.H.; de Luca, X.M.; Parsons, M.T.; Ferguson, K.; Reid, L.E.; et al. Whole-Genome Sequencing Reveals Clinically Relevant Insights into the Aetiology of Familial Breast Cancers. *Ann. Oncol.* **2019**, *30*, 1071–1079. [[CrossRef](#)]
67. Zhao, E.Y.; Shen, Y.; Pleasance, E.; Kasaian, K.; Leelakumari, S.; Jones, M.; Bose, P.; Ch'ng, C.; Reisle, C.; Eirew, P.; et al. Homologous Recombination Deficiency and Platinum-Based Therapy Outcomes in Advanced Breast Cancer. *Clin. Cancer Res.* **2017**, *23*, 7521–7530. [[CrossRef](#)]
68. Nguyen, L.W.M.; Martens, J.; Van Hoeck, A.; Cuppen, E. Pan-Cancer Landscape of Homologous Recombination Deficiency. *Nat. Commun.* **2020**, *11*, 5584. [[CrossRef](#)]
69. Chopra, N.; Tovey, H.; Pearson, A.; Cutts, R.; Toms, C.; Proszek, P.; Hubank, M.; Dowsett, M.; Dodson, A.; Daley, F.; et al. Homologous Recombination DNA Repair Deficiency and PARP Inhibition Activity in Primary Triple Negative Breast Cancer. *Nat. Commun.* **2020**, *11*, 2662. [[CrossRef](#)]
70. Gulhan, D.C.; Lee, J.J.-K.; Melloni, G.E.M.; Cortés-Ciriano, I.; Park, P.J. Detecting the Mutational Signature of Homologous Recombination Deficiency in Clinical Samples. *Nat. Genet.* **2019**, *51*, 912–919. [[CrossRef](#)]
71. Toh, M.; Ngeow, J. Homologous Recombination Deficiency: Cancer Predispositions and Treatment Implications. *Oncol.* **2021**, *26*, e1526–e1537. [[CrossRef](#)] [[PubMed](#)]
72. Buisson, R.; Joshi, N.; Rodrigue, A.; Ho, C.K.; Kreuzer, J.; Foo, T.K.; Hardy, E.J.-L.; Dellaire, G.; Haas, W.; Xia, B.; et al. Coupling of Homologous Recombination and the Checkpoint by ATR. *Mol. Cell* **2017**, *65*, 336–346. [[CrossRef](#)]
73. Yazinski, S.A.; Comaills, V.; Buisson, R.; Genois, M.-M.; Nguyen, H.D.; Ho, C.K.; Todorova Kwan, T.; Morris, R.; Lauffer, S.; Nussenzweig, A.; et al. ATR Inhibition Disrupts Rewired Homologous Recombination and Fork Protection Pathways in PARP Inhibitor-Resistant BRCA-Deficient Cancer Cells. *Genes Dev.* **2017**, *31*, 318–332. [[CrossRef](#)] [[PubMed](#)]
74. Kim, H.; Xu, H.; George, E.; Hallberg, D.; Kumar, S.; Jagannathan, V.; Medvedev, S.; Kinose, Y.; Devins, K.; Verma, P.; et al. Combining PARP with ATR Inhibition Overcomes PARP Inhibitor and Platinum Resistance in Ovarian Cancer Models. *Nat. Commun.* **2020**, *11*, 3726. [[CrossRef](#)] [[PubMed](#)]
75. Banerjee, S.; Stewart, J.; Porta, N.; Toms, C.; Leary, A.; Lheureux, S.; Khalique, S.; Tai, J.; Attygalle, A.; Vroobel, K.; et al. ATARI Trial: ATR Inhibitor in Combination with Olaparib in Gynecological Cancers with ARID1A Loss or No Loss (ENGOT/GYN1/NCRI). *Int. J. Gynecol. Cancer* **2021**, *31*, 1471–1475. [[CrossRef](#)]
76. FDA. FDA Approves First-Line Immunotherapy for Patients with MSI-H/DMMR Metastatic Colorectal Cancer. Available online: <https://www.fda.gov/news-events/press-announcements/fda-approves-first-line-immunotherapy-patients-msi-hdmmr-metastatic-colorectal-cancer> (accessed on 22 August 2022).
77. Zou, X.; Koh, G.C.C.; Nanda, A.S.; Degasperri, A.; Urgo, K.; Roumeliotis, T.I.; Agu, C.A.; Badja, C.; Momen, S.; Young, J.; et al. A Systematic CRISPR Screen Defines Mutational Mechanisms Underpinning Signatures Caused by Replication Errors and Endogenous DNA Damage. *Nat. Cancer* **2021**, *2*, 643–657. [[CrossRef](#)]
78. Brady, S.W.; Gout, A.M.; Zhang, J. Therapeutic and Prognostic Insights from the Analysis of Cancer Mutational Signatures. *Trends Genet.* **2022**, *38*, 194–208. [[CrossRef](#)]
79. Nowak, J.A.; Yurgelun, M.B.; Bruce, J.L.; Rojas-Rudilla, V.; Hall, D.L.; Shivdasani, P.; Garcia, E.P.; Agoston, A.T.; Srivastava, A.; Ogino, S.; et al. Detection of Mismatch Repair Deficiency and Microsatellite Instability in Colorectal Adenocarcinoma by Targeted Next-Generation Sequencing. *J. Mol. Diagn.* **2017**, *19*, 84–91. [[CrossRef](#)]
80. Li, B.; Brady, S.W.; Ma, X.; Shen, S.; Zhang, Y.; Li, Y.; Szlachta, K.; Dong, L.; Liu, Y.; Yang, F.; et al. Therapy-Induced Mutations Drive the Genomic Landscape of Relapsed Acute Lymphoblastic Leukemia. *Blood* **2020**, *135*, 41–55. [[CrossRef](#)]
81. Esteller, M.; Levine, R.; Baylin, S.B.; Ellenson, L.H.; Herman, J.G. MLH1 Promoter Hypermethylation Is Associated with the Microsatellite Instability Phenotype in Sporadic Endometrial Carcinomas. *Oncogene* **1998**, *17*, 2413–2417. [[CrossRef](#)]
82. Picco, G.; Cattaneo, C.M.; van Vliet, E.J.; Crisafulli, G.; Rospo, G.; Consonni, S.; Vieira, S.F.; Rodríguez, I.S.; Cancelliere, C.; Banerjee, R.; et al. Werner Helicase Is a Synthetic-Lethal Vulnerability in Mismatch Repair-Deficient Colorectal Cancer Refractory to Targeted Therapies, Chemotherapy, and Immunotherapy. *Cancer Discov.* **2021**, *11*, 1923–1937. [[CrossRef](#)] [[PubMed](#)]
83. Chan, E.M.; Shibue, T.; McFarland, J.M.; Gaeta, B.; Ghandi, M.; Dumont, N.; Gonzalez, A.; McPartlan, J.S.; Li, T.; Zhang, Y.; et al. WRN Helicase Is a Synthetic Lethal Target in Microsatellite Unstable Cancers. *Nature* **2019**, *568*, 551–556. [[CrossRef](#)] [[PubMed](#)]
84. Connor, A.A.; Denroche, R.E.; Jang, G.H.; Timms, L.; Kalimuthu, S.N.; Selander, I.; McPherson, T.; Wilson, G.W.; Chan-Seng-Yue, M.A.; Borozan, I.; et al. Association of Distinct Mutational Signatures With Correlates of Increased Immune Activity in Pancreatic Ductal Adenocarcinoma. *JAMA Oncol.* **2017**, *3*, 774–783. [[CrossRef](#)] [[PubMed](#)]

85. Jager, M.; Blokzijl, F.; Kuijk, E.; Bertl, J.; Vougioukalaki, M.; Janssen, R.; Besselink, N.; Boymans, S.; de Ligt, J.; Pedersen, J.S.; et al. Deficiency of Nucleotide Excision Repair Is Associated with Mutational Signature Observed in Cancer. *Genome Res.* **2019**, *29*, 1067–1077. [[CrossRef](#)]
86. Mehnert, J.M.; Panda, A.; Zhong, H.; Hirshfield, K.; Damare, S.; Lane, K.; Sokol, L.; Stein, M.N.; Rodriguez-Rodriguez, L.; Kaufman, H.L.; et al. Immune Activation and Response to Pembrolizumab in *POLE*-Mutant Endometrial Cancer. *J. Clin. Investig.* **2016**, *126*, 2334–2340. [[CrossRef](#)]
87. Howitt, B.E.; Shukla, S.A.; Sholl, L.M.; Ritterhouse, L.L.; Watkins, J.C.; Rodig, S.; Stover, E.; Strickland, K.C.; D'Andrea, A.D.; Wu, C.J.; et al. Association of Polymerase ϵ -Mutated and Microsatellite-Unstable Endometrial Cancers With Neoantigen Load, Number of Tumor-Infiltrating Lymphocytes, and Expression of PD-1 and PD-L1. *JAMA Oncol.* **2015**, *1*, 1319–1323. [[CrossRef](#)]
88. Pich, O.; Muiños, F.; Lolkema, M.P.; Steeghs, N.; Gonzalez-Perez, A.; Lopez-Bigas, N. The Mutational Footprints of Cancer Therapies. *Nat. Genet.* **2019**, *51*, 1732–1740. [[CrossRef](#)]
89. Christensen, S.; Van der Roest, B.; Besselink, N.; Janssen, R.; Boymans, S.; Martens, J.W.M.; Yaspo, M.-L.; Priestley, P.; Kuijk, E.; Cuppen, E.; et al. 5-Fluorouracil Treatment Induces Characteristic T>G Mutations in Human Cancer. *Nat. Commun.* **2019**, *10*, 4571. [[CrossRef](#)]
90. Hoang, M.L.; Chen, C.-H.; Sidorenko, V.S.; He, J.; Dickman, K.G.; Yun, B.H.; Moriya, M.; Niknafs, N.; Douville, C.; Karchin, R.; et al. Mutational Signature of Aristolochic Acid Exposure as Revealed by Whole-Exome Sequencing. *Sci. Transl. Med.* **2013**, *5*, 197ra102. [[CrossRef](#)]
91. Poon, S.L.; Huang, M.N.; Choo, Y.; McPherson, J.R.; Yu, W.; Heng, H.L.; Gan, A.; Myint, S.S.; Siew, E.Y.; Ler, L.D.; et al. Mutation Signatures Implicate Aristolochic Acid in Bladder Cancer Development. *Genome Med.* **2015**, *7*, 38. [[CrossRef](#)]
92. Poon, S.L.; Pang, S.-T.; McPherson, J.R.; Yu, W.; Huang, K.K.; Guan, P.; Weng, W.-H.; Siew, E.Y.; Liu, Y.; Heng, H.L.; et al. Genome-Wide Mutational Signatures of Aristolochic Acid and Its Application as a Screening Tool. *Sci. Transl. Med.* **2013**, *5*, 197ra101. [[CrossRef](#)] [[PubMed](#)]
93. Chang, J.; Tan, W.; Ling, Z.; Xi, R.; Shao, M.; Chen, M.; Luo, Y.; Zhao, Y.; Liu, Y.; Huang, X.; et al. Genomic Analysis of Oesophageal Squamous-Cell Carcinoma Identifies Alcohol Drinking-Related Mutation Signature and Genomic Alterations. *Nat. Commun.* **2017**, *8*, 15290. [[CrossRef](#)] [[PubMed](#)]
94. Li, X.C.; Wang, M.Y.; Yang, M.; Dai, H.J.; Zhang, B.F.; Wang, W.; Chu, X.L.; Wang, X.; Zheng, H.; Niu, R.F.; et al. A Mutational Signature Associated with Alcohol Consumption and Prognostically Significantly Mutated Driver Genes in Esophageal Squamous Cell Carcinoma. *Ann. Oncol.* **2018**, *29*, 938–944. [[CrossRef](#)]
95. Letouzé, E.; Shinde, J.; Renault, V.; Couchy, G.; Blanc, J.-F.; Tubacher, E.; Bayard, Q.; Bacq, D.; Meyer, V.; Semhoun, J.; et al. Mutational Signatures Reveal the Dynamic Interplay of Risk Factors and Cellular Processes during Liver Tumorigenesis. *Nat. Commun.* **2017**, *8*, 1315. [[CrossRef](#)] [[PubMed](#)]
96. Wei, R.; Li, P.; He, F.; Wei, G.; Zhou, Z.; Su, Z.; Ni, T. Comprehensive Analysis Reveals Distinct Mutational Signature and Its Mechanistic Insights of Alcohol Consumption in Human Cancers. *Brief. Bioinform.* **2021**, *22*, bbaa066. [[CrossRef](#)] [[PubMed](#)]
97. Secrier, M.; Li, X.; de Silva, N.; Eldridge, M.D.; Contino, G.; Bornschein, J.; MacRae, S.; Grehan, N.; O'Donovan, M.; Miremadi, A.; et al. Mutational Signatures in Esophageal Adenocarcinoma Define Etiologically Distinct Subgroups with Therapeutic Relevance. *Nat. Genet.* **2016**, *48*, 1131–1141. [[CrossRef](#)]
98. Nguyen, L.; Van Hoek, A.; Cuppen, E. Machine Learning-Based Tissue of Origin Classification for Cancer of Unknown Primary Diagnostics Using Genome-Wide Mutation Features. *Nat. Commun.* **2022**, *13*, 4013. [[CrossRef](#)]
99. Wang, K.; Tepper, J.E. Radiation Therapy-Associated Toxicity: Etiology, Management, and Prevention. *CA A Cancer J. Clin.* **2021**, *71*, 437–454. [[CrossRef](#)]
100. Majeed, H.; Gupta, V. Adverse Effects Of Radiation Therapy. In *StatPearls*; StatPearls Publishing: Treasure Island, FL, USA, 2022.
101. Cheng, F.; Zhao, J.; Zhao, Z. Advances in Computational Approaches for Prioritizing Driver Mutations and Significantly Mutated Genes in Cancer Genomes. *Brief. Bioinform.* **2016**, *17*, 642–656. [[CrossRef](#)]
102. Tsimberidou, A.M.; Fountzilias, E.; Nikanjam, M.; Kurzrock, R. Review of Precision Cancer Medicine: Evolution of the Treatment Paradigm. *Cancer Treat. Rev.* **2020**, *86*, 102019. [[CrossRef](#)]
103. Moscow, J.A.; Fojo, T.; Schilsky, R.L. The Evidence Framework for Precision Cancer Medicine. *Nat. Rev. Clin. Oncol.* **2018**, *15*, 183–192. [[CrossRef](#)] [[PubMed](#)]
104. Lawrence, M.S.; Stojanov, P.; Polak, P.; Kryukov, G.V.; Cibulskis, K.; Sivachenko, A.; Carter, S.L.; Stewart, C.; Mermel, C.H.; Roberts, S.A.; et al. Mutational Heterogeneity in Cancer and the Search for New Cancer-Associated Genes. *Nature* **2013**, *499*, 214–218. [[CrossRef](#)]
105. Hodis, E.; Watson, I.R.; Kryukov, G.V.; Arold, S.T.; Imielinski, M.; Theurillat, J.-P.; Nickerson, E.; Auclair, D.; Li, L.; Place, C.; et al. A Landscape of Driver Mutations in Melanoma. *Cell* **2012**, *150*, 251–263. [[CrossRef](#)] [[PubMed](#)]
106. Dees, N.D.; Zhang, Q.; Kandoth, C.; Wendl, M.C.; Schierding, W.; Koboldt, D.C.; Mooney, T.B.; Callaway, M.B.; Dooling, D.; Mardis, E.R.; et al. MuSiC: Identifying Mutational Significance in Cancer Genomes. *Genome Res.* **2012**, *22*, 1589–1598. [[CrossRef](#)]
107. Carter, H.; Chen, S.; Isik, L.; Tyekucheva, S.; Velculescu, V.E.; Kinzler, K.W.; Vogelstein, B.; Karchin, R. Cancer-Specific High-Throughput Annotation of Somatic Mutations: Computational Prediction of Driver Missense Mutations. *Cancer Res.* **2009**, *69*, 6660–6667. [[CrossRef](#)] [[PubMed](#)]
108. Wong, W.C.; Kim, D.; Carter, H.; Diekhans, M.; Ryan, M.C.; Karchin, R. CHASM and SNVBox: Toolkit for Detecting Biologically Important Single Nucleotide Mutations in Cancer. *Bioinformatics* **2011**, *27*, 2147–2148. [[CrossRef](#)]

109. Carter, H.; Samayoa, J.; Hruban, R.H.; Karchin, R. Prioritization of Driver Mutations in Pancreatic Cancer Using Cancer-Specific High-Throughput Annotation of Somatic Mutations (CHASM). *Cancer Biol. Ther.* **2010**, *10*, 582–587. [[CrossRef](#)]
110. Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P.; Kondrashov, A.S.; Sunyaev, S.R. A Method and Server for Predicting Damaging Missense Mutations. *Nat. Methods* **2010**, *7*, 248–249. [[CrossRef](#)]
111. Porta-Pardo, E.; Godzik, A. E-Driver: A Novel Method to Identify Protein Regions Driving Cancer. *Bioinformatics* **2014**, *30*, 3109–3114. [[CrossRef](#)]
112. Sim, N.-L.; Kumar, P.; Hu, J.; Henikoff, S.; Schneider, G.; Ng, P.C. SIFT Web Server: Predicting Effects of Amino Acid Substitutions on Proteins. *Nucleic Acids Res.* **2012**, *40*, W452–W457. [[CrossRef](#)]
113. Kumar, P.; Henikoff, S.; Ng, P.C. Predicting the Effects of Coding Non-Synonymous Variants on Protein Function Using the SIFT Algorithm. *Nat. Protoc.* **2009**, *4*, 1073–1081. [[CrossRef](#)] [[PubMed](#)]
114. Jia, P.; Wang, Q.; Chen, Q.; Hutchinson, K.E.; Pao, W.; Zhao, Z. MSEA: Detection and Quantification of Mutation Hotspots through Mutation Set Enrichment Analysis. *Genome Biol.* **2014**, *15*, 489. [[CrossRef](#)] [[PubMed](#)]
115. Ryslik, G.A.; Cheng, Y.; Cheung, K.-H.; Modis, Y.; Zhao, H. Utilizing Protein Structure to Identify Non-Random Somatic Mutations. *BMC Bioinform.* **2013**, *14*, 190. [[CrossRef](#)] [[PubMed](#)]
116. Ryslik, G.A.; Cheng, Y.; Cheung, K.-H.; Modis, Y.; Zhao, H. A Graph Theoretic Approach to Utilizing Protein Structure to Identify Non-Random Somatic Mutations. *BMC Bioinform.* **2014**, *15*, 86. [[CrossRef](#)]
117. Reimand, J.; Bader, G.D. Systematic Analysis of Somatic Mutations in Phosphorylation Signaling Predicts Novel Cancer Drivers. *Mol. Syst. Biol.* **2013**, *9*, 637. [[CrossRef](#)]
118. Leiserson, M.D.M.; Vandin, F.; Wu, H.-T.; Dobson, J.R.; Eldridge, J.V.; Thomas, J.L.; Papoutsaki, A.; Kim, Y.; Niu, B.; McLellan, M.; et al. Pan-Cancer Network Analysis Identifies Combinations of Rare Somatic Mutations across Pathways and Protein Complexes. *Nat. Genet.* **2015**, *47*, 106–114. [[CrossRef](#)]
119. Cho, A.; Shim, J.E.; Kim, E.; Supek, F.; Lehner, B.; Lee, I. MUFFINN: Cancer Gene Discovery via Network Analysis of Somatic Mutation Data. *Genome Biol.* **2016**, *17*, 129. [[CrossRef](#)]
120. Vaske, C.J.; Benz, S.C.; Sanborn, J.Z.; Earl, D.; Szeto, C.; Zhu, J.; Haussler, D.; Stuart, J.M. Inference of Patient-Specific Pathway Activities from Multi-Dimensional Cancer Genomics Data Using PARADIGM. *Bioinformatics* **2010**, *26*, i237–i245. [[CrossRef](#)]
121. Iranzo, J.; Martincorena, I.; Koonin, E.V. Cancer-Mutation Network and the Number and Specificity of Driver Mutations. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E6010–E6019. [[CrossRef](#)]
122. Reyna, M.A.; Haan, D.; Paczkowska, M.; Verbeke, L.P.C.; Vazquez, M.; Kahraman, A.; Pulido-Tamayo, S.; Barenboim, J.; Wadi, L.; Dhingra, P.; et al. Pathway and Network Analysis of More than 2500 Whole Cancer Genomes. *Nat. Commun.* **2020**, *11*, 729. [[CrossRef](#)]
123. Sherman, M.A.; Yaari, A.U.; Priebe, O.; Dietlein, F.; Loh, P.-R.; Berger, B. Genome-Wide Mapping of Somatic Mutation Rates Uncovers Drivers of Cancer. *Nat. Biotechnol.* **2022**, *40*, 1634–1643. [[CrossRef](#)] [[PubMed](#)]
124. Muiños, F.; Martínez-Jiménez, F.; Pich, O.; Gonzalez-Perez, A.; Lopez-Bigas, N. In Silico Saturation Mutagenesis of Cancer Genes. *Nature* **2021**, *596*, 428–432. [[CrossRef](#)] [[PubMed](#)]
125. Chen, H.; Li, J.; Wang, Y.; Ng, P.K.-S.; Tsang, Y.H.; Shaw, K.R.; Mills, G.B.; Liang, H. Comprehensive Assessment of Computational Algorithms in Predicting Cancer Driver Mutations. *Genome Biol.* **2020**, *21*, 43. [[CrossRef](#)] [[PubMed](#)]
126. Raimondi, D.; Tanyalcin, I.; Ferte, J.; Gazzo, A.; Orlando, G.; Lenaerts, T.; Rooman, M.; Vranken, W. DEOGEN2: Prediction and Interactive Visualization of Single Amino Acid Variant Deleteriousness in Human Proteins. *Nucleic Acids Res.* **2017**, *45*, W201–W206. [[CrossRef](#)] [[PubMed](#)]
127. Bailey, M.H.; Tokheim, C.; Porta-Pardo, E.; Sengupta, S.; Bertrand, D.; Weerasinghe, A.; Colaprico, A.; Wendl, M.C.; Kim, J.; Reardon, B.; et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **2018**, *173*, 371–385.e18. [[CrossRef](#)]
128. Sundaram, L.; Gao, H.; Padigepati, S.R.; McRae, J.F.; Li, Y.; Kosmicki, J.A.; Fritzilas, N.; Hakenberg, J.; Dutta, A.; Shon, J.; et al. Predicting the Clinical Impact of Human Mutation with Deep Neural Networks. *Nat. Genet.* **2018**, *50*, 1161–1170. [[CrossRef](#)]
129. Mallik, S.; Zhao, Z. Graph- and Rule-Based Learning Algorithms: A Comprehensive Review of Their Applications for Cancer Type Classification and Prognosis Using Genomic Data. *Brief. Bioinform.* **2020**, *21*, 368–394. [[CrossRef](#)]
130. Zhang, W.; Chien, J.; Yong, J.; Kuang, R. Network-Based Machine Learning and Graph Theory Algorithms for Precision Oncology. *npj Precis. Oncol.* **2017**, *1*, 25. [[CrossRef](#)]
131. Hofree, M.; Shen, J.P.; Carter, H.; Gross, A.; Ideker, T. Network-Based Stratification of Tumor Mutations. *Nat. Methods* **2013**, *10*, 1108–1115. [[CrossRef](#)]
132. Patterson, A.; Auslander, N. Mutated Processes Predict Immune Checkpoint Inhibitor Therapy Benefit in Metastatic Melanoma. *Nat. Commun.* **2022**, *13*, 5151. [[CrossRef](#)]
133. Zolotovskaia, M.A.; Sorokin, M.L.; Emelianova, A.A.; Borisov, N.M.; Kuzmin, D.V.; Borger, P.; Garazha, A.V.; Buzdin, A.A. Pathway Based Analysis of Mutation Data Is Efficient for Scoring Target Cancer Drugs. *Front. Pharmacol.* **2019**, *10*. [[CrossRef](#)] [[PubMed](#)]
134. Kuijjer, M.L.; Paulson, J.N.; Salzman, P.; Ding, W.; Quackenbush, J. Cancer Subtype Identification Using Somatic Mutation Data. *Br. J. Cancer* **2018**, *118*, 1492–1501. [[CrossRef](#)] [[PubMed](#)]
135. Auslander, N.; Wolf, Y.I.; Koonin, E.V. Interplay between DNA Damage Repair and Apoptosis Shapes Cancer Evolution through Aneuploidy and Microsatellite Instability. *Nat. Commun.* **2020**, *11*, 1234. [[CrossRef](#)] [[PubMed](#)]

136. Zhang, L.; Cao, L.; Li, S.; Wang, L.; Song, Y.; Huang, Y.; Xu, Z.; He, J.; Wang, M.; Li, K. Biologically Interpretable Deep Learning to Predict Response to Immunotherapy in Advanced Melanoma Using Mutations and Copy Number Variations. *Res. Sq.* **2022**, *preprint*.
137. Zou, J.; Huss, M.; Abid, A.; Mohammadi, P.; Torkamaani, A.; Telenti, A. A Primer on Deep Learning in Genomics. *Nat. Genet.* **2019**, *51*, 12–18. [[CrossRef](#)] [[PubMed](#)]
138. Dash, S.; Kinney, N.A.; Varghese, R.T.; Garner, H.R.; Feng, W.; Anandakrishnan, R. Differentiating between Cancer and Normal Tissue Samples Using Multi-Hit Combinations of Genetic Mutations. *Sci. Rep.* **2019**, *9*, 1005. [[CrossRef](#)] [[PubMed](#)]
139. Leiserson, M.D.; Wu, H.-T.; Vandin, F.; Raphael, B.J. CoMEt: A Statistical Approach to Identify Combinations of Mutually Exclusive Alterations in Cancer. *Genome Biol.* **2015**, *16*, 160. [[CrossRef](#)]
140. Ciriello, G.; Cerami, E.; Sander, C.; Schultz, N. Mutual Exclusivity Analysis Identifies Oncogenic Network Modules. *Genome Res.* **2012**, *22*, 398–406. [[CrossRef](#)]
141. van de Haar, J.; Canisius, S.; Yu, M.K.; Voest, E.E.; Wessels, L.F.A.; Ideker, T. Identifying Epistasis in Cancer Genomes: A Delicate Affair. *Cell* **2019**, *177*, 1375–1383. [[CrossRef](#)]
142. Gussow, A.B.; Koonin, E.V.; Auslander, N. Identification of Combinations of Somatic Mutations That Predict Cancer Survival and Immunotherapy Benefit. *NAR Cancer* **2021**, *3*, zcab017. [[CrossRef](#)]
143. Vural, S.; Wang, X.; Guda, C. Classification of Breast Cancer Patients Using Somatic Mutation Profiles and Machine Learning Approaches. *BMC Syst. Biol.* **2016**, *10*, 62. [[CrossRef](#)] [[PubMed](#)]
144. Jiao, W.; Atwal, G.; Polak, P.; Karlic, R.; Cuppen, E.; Danyi, A.; de Ridder, J.; van Herpen, C.; Lolkema, M.P.; Steeghs, N.; et al. A Deep Learning System Accurately Classifies Primary and Metastatic Cancers Using Passenger Mutation Patterns. *Nat. Commun.* **2020**, *11*, 728. [[CrossRef](#)] [[PubMed](#)]
145. Gerstung, M.; Jolly, C.; Leshchiner, I.; Dentre, S.C.; Gonzalez, S.; Rosebrock, D.; Mitchell, T.J.; Rubanova, Y.; Anur, P.; Yu, K.; et al. The Evolutionary History of 2,658 Cancers. *Nature* **2020**, *578*, 122–128. [[CrossRef](#)] [[PubMed](#)]
146. Jolly, C.; Van Loo, P. Timing Somatic Events in the Evolution of Cancer. *Genome Biol.* **2018**, *19*, 95. [[CrossRef](#)] [[PubMed](#)]
147. Attolini, C.S.-O.; Cheng, Y.-K.; Beroukhim, R.; Getz, G.; Abdel-Wahab, O.; Levine, R.L.; Mellinghoff, I.K.; Michor, F. A Mathematical Framework to Determine the Temporal Sequence of Somatic Genetic Events in Cancer. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 17604–17609. [[CrossRef](#)]
148. Cheng, Y.-K.; Beroukhim, R.; Levine, R.L.; Mellinghoff, I.K.; Holland, E.C.; Michor, F. A Mathematical Methodology for Determining the Temporal Order of Pathway Alterations Arising during Gliomagenesis. *PLoS Comput. Biol.* **2012**, *8*, e1002337. [[CrossRef](#)]
149. Desper, R.; Jiang, F.; Kallioniemi, O.P.; Moch, H.; Papadimitriou, C.H.; Schäffer, A.A. Distance-Based Reconstruction of Tree Models for Oncogenesis. *J. Comput. Biol.* **2000**, *7*, 789–803. [[CrossRef](#)]
150. Bozic, I.; Nowak, M.A. Timing and Heterogeneity of Mutations Associated with Drug Resistance in Metastatic Cancers. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 15964–15968. [[CrossRef](#)]
151. Huang, Y.; Wang, J.; Jia, P.; Li, X.; Pei, G.; Wang, C.; Fang, X.; Zhao, Z.; Cai, Z.; Yi, X.; et al. Clonal Architectures Predict Clinical Outcome in Clear Cell Renal Cell Carcinoma. *Nat. Commun.* **2019**, *10*, 1245. [[CrossRef](#)]
152. Little, P.; Lin, D.-Y.; Sun, W. Associating Somatic Mutations to Clinical Outcomes: A Pan-Cancer Study of Survival Time. *Genome Med.* **2019**, *11*, 37. [[CrossRef](#)]
153. Auslander, N.; Wolf, Y.I.; Koonin, E.V. In Silico Learning of Tumor Evolution through Mutational Time Series. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 9501–9510. [[CrossRef](#)] [[PubMed](#)]
154. Yoo, B.C.; Kim, K.-H.; Woo, S.M.; Myung, J.K. Clinical Multi-Omics Strategies for the Effective Cancer Management. *J. Proteom.* **2018**, *188*, 97–106. [[CrossRef](#)] [[PubMed](#)]
155. Dorman, S.N.; Baranova, K.; Knoll, J.H.M.; Urquhart, B.L.; Mariani, G.; Carcangiu, M.L.; Rogan, P.K. Genomic Signatures for Paclitaxel and Gemcitabine Resistance in Breast Cancer Derived by Machine Learning. *Mol. Oncol.* **2016**, *10*, 85–100. [[CrossRef](#)] [[PubMed](#)]
156. Freeman, S.S.; Sade-Feldman, M.; Kim, J.; Stewart, C.; Gonye, A.L.K.; Ravi, A.; Arniella, M.B.; Gushterova, I.; LaSalle, T.J.; Blaum, E.M.; et al. Combined Tumor and Immune Signals from Genomes or Transcriptomes Predict Outcomes of Checkpoint Inhibition in Melanoma. *Cell Rep. Med.* **2022**, *3*, 100500. [[CrossRef](#)]
157. Cheng, B.; Zhou, P.; Chen, Y. Machine-Learning Algorithms Based on Personalized Pathways for a Novel Predictive Model for the Diagnosis of Hepatocellular Carcinoma. *BMC Bioinform.* **2022**, *23*, 248. [[CrossRef](#)]
158. Kim, D.; Li, R.; Lucas, A.; Verma, S.S.; Dudek, S.M.; Ritchie, M.D. Using Knowledge-Driven Genomic Interactions for Multi-Omics Data Analysis: Metadimensional Models for Predicting Clinical Outcomes in Ovarian Carcinoma. *J. Am. Med. Inform. Assoc.* **2017**, *24*, 577–587. [[CrossRef](#)]
159. Wang, Y.; Yang, Y.; Chen, S.; Wang, J. DeepDRK: A Deep Learning Framework for Drug Repurposing through Kernel-Based Multi-Omics Integration. *Brief. Bioinform.* **2021**, *22*, bbab048. [[CrossRef](#)]
160. Chaudhary, K.; Poirion, O.B.; Lu, L.; Garmire, L.X. Deep Learning-Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer. *Clin. Cancer Res.* **2018**, *24*, 1248–1259. [[CrossRef](#)]
161. Wang, J.; Chen, P.; Su, M.; Zhong, G.; Zhang, S.; Gou, D. Integrative Modeling of Multiomics Data for Predicting Tumor Mutation Burden in Patients with Lung Cancer. *BioMed Res. Int.* **2022**, *2022*, e2698190. [[CrossRef](#)]

162. Olivier, M.; Asmis, R.; Hawkins, G.A.; Howard, T.D.; Cox, L.A. The Need for Multi-Omics Biomarker Signatures in Precision Medicine. *Int. J. Mol. Sci.* **2019**, *20*, 4781. [[CrossRef](#)]
163. Lewis, J.; Breeze, C.E.; Charlesworth, J.; Maclaren, O.J.; Cooper, J. Where next for the Reproducibility Agenda in Computational Biology? *BMC Syst. Biol.* **2016**, *10*, 52. [[CrossRef](#)] [[PubMed](#)]
164. Garijo, D.; Kinnings, S.; Xie, L.; Xie, L.; Zhang, Y.; Bourne, P.E.; Gil, Y. Quantifying Reproducibility in Computational Biology: The Case of the Tuberculosis Drugome. *PLoS ONE* **2013**, *8*, e80278. [[CrossRef](#)] [[PubMed](#)]
165. Niven, D.J.; McCormick, T.J.; Straus, S.E.; Hemmelgarn, B.R.; Jeffs, L.; Barnes, T.R.M.; Stelfox, H.T. Reproducibility of Clinical Research in Critical Care: A Scoping Review. *BMC Med.* **2018**, *16*, 26. [[CrossRef](#)]
166. Greener, J.G.; Kandathil, S.M.; Moffat, L.; Jones, D.T. A Guide to Machine Learning for Biologists. *Nat. Rev. Mol. Cell Biol.* **2022**, *23*, 40–55. [[CrossRef](#)]
167. Cook, J.A.; Ranstam, J. Overfitting. *Br. J. Surg.* **2016**, *103*, 1814. [[CrossRef](#)]
168. Chicco, D. Ten Quick Tips for Machine Learning in Computational Biology. *BioData Min.* **2017**, *10*, 35. [[CrossRef](#)] [[PubMed](#)]
169. Papin, J.A.; Gabhann, F.M.; Sauro, H.M.; Nickerson, D.; Rampadarath, A. Improving Reproducibility in Computational Biology Research. *PLoS Comput. Biol.* **2020**, *16*, e1007881. [[CrossRef](#)]
170. Sandve, G.K.; Nekrutenko, A.; Taylor, J.; Hovig, E. Ten Simple Rules for Reproducible Computational Research. *PLoS Comput. Biol.* **2013**, *9*, e1003285. [[CrossRef](#)]
171. Piccolo, S.R.; Frampton, M.B. Tools and Techniques for Computational Reproducibility. *GigaScience* **2016**, *5*, 30. [[CrossRef](#)]
172. Heil, B.J.; Hoffman, M.M.; Markowetz, F.; Lee, S.-I.; Greene, C.S.; Hicks, S.C. Reproducibility Standards for Machine Learning in the Life Sciences. *Nat. Methods* **2021**, *18*, 1132–1135. [[CrossRef](#)]
173. Beam, A.L.; Manrai, A.K.; Ghassemi, M. Challenges to the Reproducibility of Machine Learning Models in Health Care. *JAMA* **2020**, *323*, 305–306. [[CrossRef](#)] [[PubMed](#)]
174. McDermott, M.B.A.; Wang, S.; Marinsek, N.; Ranganath, R.; Foschini, L.; Ghassemi, M. Reproducibility in Machine Learning for Health Research: Still a Ways to Go. *Sci. Transl. Med.* **2021**, *13*, eabb1655. [[CrossRef](#)] [[PubMed](#)]
175. Doshi-Velez, F.; Kim, B. Considerations for Evaluation and Generalization in Interpretable Machine Learning. In *Explainable and Interpretable Models in Computer Vision and Machine Learning*; Escalante, H.J., Escalera, S., Guyon, I., Baró, X., Güçlütürk, Y., Güçlü, U., van Gerven, M., Eds.; The Springer Series on Challenges in Machine Learning; Springer International Publishing: Cham, Switzerland, 2018; pp. 3–17. ISBN 978-3-319-98131-4.
176. Pasolli, E.; Truong, D.T.; Malik, F.; Waldron, L.; Segata, N. Machine Learning Meta-Analysis of Large Metagenomic Datasets: Tools and Biological Insights. *PLoS Comput. Biol.* **2016**, *12*, e1004977. [[CrossRef](#)] [[PubMed](#)]
177. Barbiero, P.; Squillero, G.; Tonda, A. Modeling Generalization in Machine Learning: A Methodological and Computational Study. *arXiv* **2020**, arXiv:2006.15680.
178. Liu, J.; Tripathi, S.; Kurup, U.; Shah, M. Auptimizer—An Extensible, Open-Source Framework for Hyperparameter Tuning. In Proceedings of the IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019.
179. Liang, J.; Meyerson, E.; Hodjat, B.; Fink, D.; Mutch, K.; Miikkulainen, R. Evolutionary Neural AutoML for Deep Learning. In Proceedings of the Genetic and Evolutionary Computation Conference, Prague, Czech Republic, 13–17 July 2019.
180. Chen, B.; Wu, H.; Mo, W.; Chattopadhyay, I.; Lipson, H. Autostacker: A Compositional Evolutionary Learning System. In Proceedings of the Genetic and Evolutionary Computation Conference, Kyoto, Japan, 15–19 July 2018.
181. Olson, R.S.; Urbanowicz, R.J.; Andrews, P.C.; Lavender, N.A.; Kidd, L.C.; Moore, J.H. Automating Biomedical Data Science through Tree-Based Pipeline Optimization. In *Applications of Evolutionary Computation: 19th European Conference, EvoApplications 2016, Porto, Portugal, 30 March–1 April 2016*; Springer International Publishing: Berlin/Heidelberg, Germany, 2016.
182. Olson, R.S.; Bartley, N.; Urbanowicz, R.J.; Moore, J.H. Evaluation of a Tree-Based Pipeline Optimization Tool for Automating Data Science. In Proceedings of the Genetic and Evolutionary Computation Conference, Denver, CO, USA, 20–24 July 2016.
183. Li, L.; Jamieson, K.; DeSalvo, G.; Rostamizadeh, A.; Talwalkar, A. Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. *J. Mach. Learn. Res.* **2018**, *18*, 6765–6816.
184. Xanthopoulos, I.; Tsamardinos, I.; Christophides, V.; Simon, E.; Salinger, A. Putting the Human Back in the AutoML Loop. In Proceedings of the Workshops of the EDBT/ICDT 2020 Joint Conference, Copenhagen, Denmark, 30 March 2020.
185. Baker, B.; Gupta, O.; Raskar, R.; Naik, N. Accelerating Neural Architecture Search Using Performance Prediction. *arXiv* **2017**, arXiv:1705.10823.
186. Errington, T.M.; Iorns, E.; Gunn, W.; Tan, F.E.; Lomax, J.; Nosek, B.A. An Open Investigation of the Reproducibility of Cancer Biology Research. *eLife* **2014**, *3*, e04333. [[CrossRef](#)]
187. Nosek, B.A.; Errington, T.M. Making Sense of Replications. *eLife* **2017**, *6*, e23383. [[CrossRef](#)]
188. Quang, D.; Xie, X. DanQ: A Hybrid Convolutional and Recurrent Deep Neural Network for Quantifying the Function of DNA Sequences. *Nucleic Acids Res.* **2016**, *44*, e107. [[CrossRef](#)]
189. Azodi, C.B.; Tang, J.; Shiu, S.-H. Opening the Black Box: Interpretable Machine Learning for Geneticists. *Trends Genet.* **2020**, *36*, 442–455. [[CrossRef](#)]
190. Molnar, C.; Casalicchio, G.; Bischl, B. Interpretable Machine Learning—A Brief History, State-of-the-Art and Challenges. In Proceedings of the ECML PKDD 2020 Workshops, Ghent, Belgium, 14–18 September 2020; Koprinska, I., Kamp, M., Appice, A., Loglisci, C., Antonie, L., Zimmermann, A., Guidotti, R., Özgöbek, Ö., Ribeiro, R.P., Gavaldà, R., et al., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 417–431.

191. Ahmad, M.A.; Eckert, C.; Teredesai, A. Interpretable Machine Learning in Healthcare. In Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, Washington, DC, USA, 29 August–1 September 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 559–560.
192. Wang, F.; Kaushal, R.; Khullar, D. Should Health Care Demand Interpretable Artificial Intelligence or Accept “Black Box” Medicine? *Ann. Intern. Med.* **2020**, *172*, 59–60. [[CrossRef](#)] [[PubMed](#)]
193. Watson, D.S.; Krutzinna, J.; Bruce, I.N.; Griffiths, C.E.; McInnes, I.B.; Barnes, M.R.; Floridi, L. Clinical Applications of Machine Learning Algorithms: Beyond the Black Box. *BMJ* **2019**, *364*, 1886. [[CrossRef](#)] [[PubMed](#)]
194. Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy* **2021**, *23*, 18. [[CrossRef](#)] [[PubMed](#)]
195. Palatnik de Sousa, I.; Maria Bernardes Rebutti Vellasco, M.; Costa da Silva, E. Local Interpretable Model-Agnostic Explanations for Classification of Lymph Node Metastases. *Sensors* **2019**, *19*, 2969. [[CrossRef](#)] [[PubMed](#)]
196. Gabbay, F.; Bar-Lev, S.; Montano, O.; Hadad, N. A LIME-Based Explainable Machine Learning Model for Predicting the Severity Level of COVID-19 Diagnosed Patients. *Appl. Sci.* **2021**, *11*, 10417. [[CrossRef](#)]
197. Shrikumar, A.; Greenside, P.; Kundaje, A. Learning Important Features Through Propagating Activation Differences. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017.
198. Washburn, J.D.; Mejia-Guerra, M.K.; Ramstein, G.; Kremling, K.A.; Valluru, R.; Buckler, E.S.; Wang, H. Evolutionarily Informed Deep Learning Methods for Predicting Relative Transcript Abundance from DNA Sequence. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 5542–5549. [[CrossRef](#)]
199. Zuallaert, J.; Godin, F.; Kim, M.; Soete, A.; Saeys, Y.; De Neve, W. SpliceRover: Interpretable Convolutional Neural Networks for Improved Splice Site Prediction. *Bioinformatics* **2018**, *34*, 4180–4188. [[CrossRef](#)]
200. Kim, J.-S.; Gao, X.; Rzhetsky, A. RIDDLE: Race and Ethnicity Imputation from Disease History with Deep Learning. *PLoS Comput. Biol.* **2018**, *14*, e1006106. [[CrossRef](#)]
201. Kong, L.; Chen, Y.; Xu, F.; Xu, M.; Li, Z.; Fang, J.; Zhang, L.; Pian, C. Mining Influential Genes Based on Deep Learning. *BMC Bioinform.* **2021**, *22*, 27. [[CrossRef](#)]
202. Chen, L.; Capra, J.A. Learning and Interpreting the Gene Regulatory Grammar in a Deep Learning Framework. *PLoS Comput. Biol.* **2020**, *16*, e1008334. [[CrossRef](#)]
203. Elmarakeby, H.A.; Hwang, J.; Arafeh, R.; Crowdis, J.; Gang, S.; Liu, D.; AlDubayan, S.H.; Salari, K.; Kregel, S.; Richter, C.; et al. Biologically Informed Deep Neural Network for Prostate Cancer Discovery. *Nature* **2021**, *598*, 348–352. [[CrossRef](#)] [[PubMed](#)]
204. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
205. Galdos, F.X.; Xu, S.; Goodyer, W.R.; Duan, L.; Huang, Y.V.; Lee, S.; Zhu, H.; Lee, C.; Wei, N.; Lee, D.; et al. DevCellPy Is a Machine Learning-Enabled Pipeline for Automated Annotation of Complex Multilayered Single-Cell Transcriptomic Data. *Nat. Commun.* **2022**, *13*, 5271. [[CrossRef](#)] [[PubMed](#)]
206. Elbasir, A.; Mall, R.; Kunji, K.; Rawi, R.; Islam, Z.; Chuang, G.-Y.; Kolatkar, P.R.; Bensmail, H. BCrystal: An Interpretable Sequence-Based Protein Crystallization Predictor. *Bioinformatics* **2020**, *36*, 1429–1438. [[CrossRef](#)] [[PubMed](#)]
207. Jiang, B.; Mu, Q.; Qiu, F.; Li, X.; Xu, W.; Yu, J.; Fu, W.; Cao, Y.; Wang, J. Machine Learning of Genomic Features in Organotropic Metastases Stratifies Progression Risk of Primary Tumors. *Nat. Commun.* **2021**, *12*, 6692. [[CrossRef](#)] [[PubMed](#)]
208. Li, H.; Guan, Y. Asymmetric Predictive Relationships across Histone Modifications. *Nat. Mach. Intell.* **2022**, *4*, 288–299. [[CrossRef](#)]
209. Wang, D.; Zhang, C.; Wang, B.; Li, B.; Wang, Q.; Liu, D.; Wang, H.; Zhou, Y.; Shi, L.; Lan, F.; et al. Optimized CRISPR Guide RNA Design for Two High-Fidelity Cas9 Variants by Deep Learning. *Nat. Commun.* **2019**, *10*, 4284. [[CrossRef](#)]
210. Camacho, D.M.; Collins, K.M.; Powers, R.K.; Costello, J.C.; Collins, J.J. Next-Generation Machine Learning for Biological Networks. *Cell* **2018**, *173*, 1581–1592. [[CrossRef](#)]
211. Auslander, N.; Gussow, A.B.; Koonin, E.V. Incorporating Machine Learning into Established Bioinformatics Frameworks. *Int. J. Mol. Sci.* **2021**, *22*, 2903. [[CrossRef](#)]
212. Yang, X.; Wang, W.; Ma, J.-L.; Qiu, Y.-L.; Lu, K.; Cao, D.-S.; Wu, C.-K. BioNet: A Large-Scale and Heterogeneous Biological Network Model for Interaction Prediction with Graph Convolution. *Brief. Bioinform.* **2022**, *23*, bbab491. [[CrossRef](#)]
213. Peng, W.; Tang, Q.; Dai, W.; Chen, T. Improving Cancer Driver Gene Identification Using Multi-Task Learning on Graph Convolutional Network. *Brief. Bioinform.* **2022**, *23*, bbab432. [[CrossRef](#)]
214. Chu, Y.; Wang, X.; Dai, Q.; Wang, Y.; Wang, Q.; Peng, S.; Wei, X.; Qiu, J.; Salahub, D.R.; Xiong, Y.; et al. MDA-GCNFTG: Identifying miRNA-Disease Associations Based on Graph Convolutional Networks via Graph Sampling through the Feature and Topology Graph. *Brief. Bioinform.* **2021**, *22*, bbab165. [[CrossRef](#)] [[PubMed](#)]
215. Ying, R.; Bourgeois, D.; You, J.; Zitnik, M.; Leskovec, J. GNNExplainer: Generating Explanations for Graph Neural Networks. *arXiv* **2019**, arXiv:1903.03894.
216. Carvalho, D.V.; Pereira, E.M.; Cardoso, J.S. Machine Learning Interpretability: A Survey on Methods and Metrics. *Electronics* **2019**, *8*, 832. [[CrossRef](#)]
217. Wyatt, K.D.; Branda, M.E.; Anderson, R.T.; Pencille, L.J.; Montori, V.M.; Hess, E.P.; Ting, H.H.; LeBlanc, A. Peering into the Black Box: A Meta-Analysis of How Clinicians Use Decision Aids during Clinical Encounters. *Implement. Sci.* **2014**, *9*, 26. [[CrossRef](#)]
218. Doshi-Velez, F.; Kim, B. Towards A Rigorous Science of Interpretable Machine Learning. *arXiv* **2017**, arXiv:1702.08608.

219. Vidyasagar, M. Machine Learning Methods in the Computational Biology of Cancer. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2014**, *470*, 20140081. [[CrossRef](#)] [[PubMed](#)]
220. Danyi, A.; Jager, M.; de Ridder, J. Cancer Type Classification in Liquid Biopsies Based on Sparse Mutational Profiles Enabled through Data Augmentation and Integration. *Life* **2022**, *12*, 1. [[CrossRef](#)]
221. Myers, M.A.; Zaccaria, S.; Raphael, B.J. Identifying Tumor Clones in Sparse Single-Cell Mutation Data. *Bioinformatics* **2020**, *36*, i186–i193. [[CrossRef](#)]
222. Sason, I.; Chen, Y.; Leiserson, M.D.M.; Sharan, R. A Mixture Model for Signature Discovery from Sparse Mutation Data. *Genome Med.* **2021**, *13*, 173. [[CrossRef](#)]
223. Ji, J.; He, D.; Feng, Y.; He, Y.; Xue, F.; Xie, L. JDINAC: Joint Density-Based Non-Parametric Differential Interaction Network Analysis and Classification Using High-Dimensional Sparse Omics Data. *Bioinformatics* **2017**, *33*, 3080–3087. [[CrossRef](#)]
224. Xu, B.; Li, X.; Gao, X.; Jia, Y.; Liu, J.; Li, F.; Zhang, Z. DeNOPA: Decoding Nucleosome Positions Sensitively with Sparse ATAC-Seq Data. *Brief. Bioinform.* **2022**, *23*, bbab469. [[CrossRef](#)]
225. Ramamoorthy, D.; Severson, K.; Ghosh, S.; Sachs, K.; Glass, J.D.; Fournier, C.N.; Herrington, T.M.; Berry, J.D.; Ng, K.; Fraenkel, E. Identifying Patterns in Amyotrophic Lateral Sclerosis Progression from Sparse Longitudinal Data. *Nat. Comput. Sci.* **2022**, *2*, 605–616. [[CrossRef](#)]
226. Suresh, S.; Saraswathi, S.; Sundararajan, N. Performance Enhancement of Extreme Learning Machine for Multi-Category Sparse Data Classification Problems. *Eng. Appl. Artif. Intell.* **2010**, *23*, 1149–1157. [[CrossRef](#)]
227. Ransohoff, D.F. Rules of Evidence for Cancer Molecular-Marker Discovery and Validation. *Nat. Rev. Cancer* **2004**, *4*, 309–314. [[CrossRef](#)]
228. Fang, J. A Critical Review of Five Machine Learning-Based Algorithms for Predicting Protein Stability Changes upon Mutation. *Brief. Bioinform.* **2020**, *21*, 1285–1292. [[CrossRef](#)] [[PubMed](#)]
229. Giudice, G.; Petsalaki, E. Proteomics and Phosphoproteomics in Precision Medicine: Applications and Challenges. *Brief. Bioinform.* **2019**, *20*, 767–777. [[CrossRef](#)]
230. Li, J.; Liu, L.; Le, T.D.; Liu, J. Accurate Data-Driven Prediction Does Not Mean High Reproducibility. *Nat. Mach. Intell.* **2020**, *2*, 13–15. [[CrossRef](#)]
231. Kim, E.; Ilic, N.; Shrestha, Y.; Zou, L.; Kamburov, A.; Zhu, C.; Yang, X.; Lubonja, R.; Tran, N.; Nguyen, C.; et al. Systematic Functional Interrogation of Rare Cancer Variants Identifies Oncogenic Alleles. *Cancer Discov.* **2016**, *6*, 714–726. [[CrossRef](#)]
232. Dogruluk, T.; Tsang, Y.H.; Espitia, M.; Chen, F.; Chen, T.; Chong, Z.; Appadurai, V.; Dogruluk, A.; Eterovic, A.K.; Bonnen, P.E.; et al. Identification of Variant-Specific Functions of PIK3CA by Rapid Phenotyping of Rare Mutations. *Cancer Res.* **2015**, *75*, 5341–5354. [[CrossRef](#)]
233. Kumar, P.; Gangal, A.; Kumari, S. Prognosis of Breast Cancer by Implementing Machine Learning Algorithms Using Modified Bootstrap Aggregating. In *Innovations in Computational Intelligence and Computer Vision*; Sharma, M.K., Dhaka, V.S., Perumal, T., Dey, N., Tavares, J.M.R.S., Eds.; Springer: Singapore, 2021; pp. 561–569.
234. Roth, H.R.; Lu, L.; Liu, J.; Yao, J.; Seff, A.; Cherry, K.; Kim, L.; Summers, R.M. Improving Computer-Aided Detection Using Convolutional Neural Networks and Random View Aggregation. *IEEE Trans. Med. Imaging* **2016**, *35*, 1170–1181. [[CrossRef](#)]
235. Suthar, B.; Patel, H.; Goswami, A. A Survey: Classification of Imputation Methods in Data Mining. *Int. J. Emerg. Technol. Adv. Eng.* **2012**, *2*, 309–312.
236. Houari, R.; Bounceur, A.; Tari, A.K.; Kecha, M.T. Handling Missing Data Problems with Sampling Methods. In Proceedings of the 2014 International Conference on Advanced Networking Distributed Systems and Applications, Bejaia, Algeria, 17–19 June 2014; pp. 99–104.
237. Ayilara, O.F.; Zhang, L.; Sajobi, T.T.; Sawatzky, R.; Bohm, E.; Lix, L.M. Impact of Missing Data on Bias and Precision When Estimating Change in Patient-Reported Outcomes from a Clinical Registry. *Health Qual. Life Outcomes* **2019**, *17*, 106. [[CrossRef](#)] [[PubMed](#)]
238. Ludbrook, J. Outlying Observations and Missing Values: How Should They Be Handled? *Clin. Exp. Pharmacol. Physiol.* **2008**, *35*, 670–678. [[CrossRef](#)]
239. Langkamp, D.L.; Lehman, A.; Lemeshow, S. Techniques for Handling Missing Data in Secondary Analyses of Large Surveys. *Acad. Pediatr.* **2010**, *10*, 205–210. [[CrossRef](#)] [[PubMed](#)]
240. Donders, A.R.T.; van der Heijden, G.J.M.G.; Stijnen, T.; Moons, K.G.M. Review: A Gentle Introduction to Imputation of Missing Values. *J. Clin. Epidemiol.* **2006**, *59*, 1087–1091. [[CrossRef](#)] [[PubMed](#)]
241. Baraldi, A.N.; Enders, C.K. An Introduction to Modern Missing Data Analyses. *J. Sch. Psychol.* **2010**, *48*, 5–37. [[CrossRef](#)]
242. Graham, J.W. Missing Data Analysis: Making It Work in the Real World. *Annu. Rev. Psychol.* **2009**, *60*, 549–576. [[CrossRef](#)]
243. Lin, J.; Li, N.; Alam, M.A.; Ma, Y. Data-Driven Missing Data Imputation in Cluster Monitoring System Based on Deep Neural Network. *Appl. Intell.* **2020**, *50*, 860–877. [[CrossRef](#)]
244. Choudhury, A.; Kosorok, M.R. Missing Data Imputation for Classification Problems. *arXiv* **2020**, arXiv:2002.10709.
245. Khan, S.I.; Hoque, A.S.M.L. SICE: An Improved Missing Data Imputation Technique. *J. Big Data* **2020**, *7*, 37. [[CrossRef](#)]
246. Al-Helali, B.; Chen, Q.; Xue, B.; Zhang, M. A New Imputation Method Based on Genetic Programming and Weighted KNN for Symbolic Regression with Incomplete Data. *Soft. Comput.* **2021**, *25*, 5993–6012. [[CrossRef](#)]
247. Peng, D.; Zou, M.; Liu, C.; Lu, J. RESI: A Region-Splitting Imputation Method for Different Types of Missing Data. *Expert Syst. Appl.* **2021**, *168*, 114425. [[CrossRef](#)]

248. Greaves, M.; Maley, C.C. Clonal evolution in cancer. *Nature* **2012**, *481*, 306–313. [[CrossRef](#)] [[PubMed](#)]
249. Su, X.; Zhao, L.; Shi, Y.; Zhang, R.; Long, Q.; Bai, S.; Luo, Q.; Lin, Y.; Zou, X.; Ghazanfar, S.; et al. Clonal Evolution in Liver Cancer at Single-Cell and Single-Variant Resolution. *J. Hematol. Oncol.* **2021**, *14*, 22. [[CrossRef](#)] [[PubMed](#)]
250. Biermann, J.; Parris, T.Z.; Nemes, S.; Danielsson, A.; Engqvist, H.; Werner Rönnerman, E.; Forssell-Aronsson, E.; Kovács, A.; Karlsson, P.; Helou, K. Clonal Relatedness in Tumour Pairs of Breast Cancer Patients. *Breast Cancer Res.* **2018**, *20*, 96. [[CrossRef](#)]
251. Hu, Z.; Li, Z.; Ma, Z.; Curtis, C. Multi-Cancer Analysis of Clonality and the Timing of Systemic Spread in Paired Primary Tumors and Metastases. *Nat. Genet.* **2020**, *52*, 701–708. [[CrossRef](#)]
252. Wang, E.; Zou, J.; Zaman, N.; Beitel, L.K.; Trifiro, M.; Paliouras, M. Cancer Systems Biology in the Genome Sequencing Era: Part 2, Evolutionary Dynamics of Tumor Clonal Networks and Drug Resistance. *Semin. Cancer Biol.* **2013**, *23*, 286–292. [[CrossRef](#)]
253. Zare, H.; Wang, J.; Hu, A.; Weber, K.; Smith, J.; Nickerson, D.; Song, C.; Witten, D.; Blau, C.A.; Noble, W.S. Inferring Clonal Composition from Multiple Sections of a Breast Cancer. *PLoS Comput. Biol.* **2014**, *10*, e1003703. [[CrossRef](#)]
254. Ha, G.; Roth, A.; Khattra, J.; Ho, J.; Yap, D.; Prentice, L.M.; Melnyk, N.; McPherson, A.; Bashashati, A.; Laks, E.; et al. TITAN: Inference of Copy Number Architectures in Clonal Cell Populations from Tumor Whole-Genome Sequence Data. *Genome Res.* **2014**, *24*, 1881–1893. [[CrossRef](#)]
255. Roth, A.; Khattra, J.; Yap, D.; Wan, A.; Laks, E.; Biele, J.; Ha, G.; Aparicio, S.; Bouchard-Côté, A.; Shah, S.P. PyClone: Statistical Inference of Clonal Population Structure in Cancer. *Nat. Methods* **2014**, *11*, 396–398. [[CrossRef](#)]
256. Chkhaidze, K.; Heide, T.; Werner, B.; Williams, M.J.; Huang, W.; Caravagna, G.; Graham, T.A.; Sottoriva, A. Spatially Constrained Tumour Growth Affects the Patterns of Clonal Selection and Neutral Drift in Cancer Genomic Data. *PLoS Comput. Biol.* **2019**, *15*, e1007243. [[CrossRef](#)] [[PubMed](#)]
257. Yadav, V.K.; De, S. An Assessment of Computational Methods for Estimating Purity and Clonality Using Genomic Data Derived from Heterogeneous Tumor Tissue Samples. *Brief. Bioinform.* **2015**, *16*, 232–241. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.