


Review

Cutting-Edge AI Technologies Meet Precision Medicine to Improve Cancer Care

Peng-Chan Lin ^{1,2} , Yi-Shan Tsai ³, Yu-Min Yeh ¹ and Meng-Ru Shen ^{4,5,6,*}

¹ Department of Oncology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan 704, Taiwan

² Department of Genomic Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan 704, Taiwan

³ Department of Medical Imaging, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan 704, Taiwan

⁴ Institute of Clinical Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan 704, Taiwan

⁵ Department of Obstetrics and Gynecology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan 704, Taiwan

⁶ Department of Pharmacology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan 704, Taiwan

* Correspondence: mrshen@mail.ncku.edu.tw; Tel.: +886-6-235-3535

Abstract: To provide precision medicine for better cancer care, researchers must work on clinical patient data, such as electronic medical records, physiological measurements, biochemistry, computerized tomography scans, digital pathology, and the genetic landscape of cancer tissue. To interpret big biodata in cancer genomics, an operational flow based on artificial intelligence (AI) models and medical management platforms with high-performance computing must be set up for precision cancer genomics in clinical practice. To work in the fast-evolving fields of patient care, clinical diagnostics, and therapeutic services, clinicians must understand the fundamentals of the AI tool approach. Therefore, the present article covers the following four themes: (i) computational prediction of pathogenic variants of cancer susceptibility genes; (ii) AI model for mutational analysis; (iii) single-cell genomics and computational biology; (iv) text mining for identifying gene targets in cancer; and (v) the NVIDIA graphics processing units, DRAGEN field programmable gate arrays systems and AI medical cloud platforms in clinical next-generation sequencing laboratories. Based on AI medical platforms and visualization, large amounts of clinical biodata can be rapidly copied and understood using an AI pipeline. The use of innovative AI technologies can deliver more accurate and rapid cancer therapy targets.

Keywords: artificial intelligence; bioinformatics; next-generation sequencing; high-performance computing; precision medicine; cancer genomics



Citation: Lin, P.-C.; Tsai, Y.-S.; Yeh, Y.-M.; Shen, M.-R. Cutting-Edge AI Technologies Meet Precision Medicine to Improve Cancer Care. *Biomolecules* **2022**, *12*, 1133. <https://doi.org/10.3390/biom12081133>

Academic Editor: Jürg Bähler

Received: 31 May 2022

Accepted: 15 August 2022

Published: 17 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

Artificial intelligence (AI) techniques, platforms, and high-performance computation have been extensively used in cancer genomics precision medicine. Many sources of clinical data of cancer patients exist, such as electronic medical records, physiological measurements, biochemistry, computerized tomography scans, digital pathology, and the genetic landscape of cancer tissue for clinical diagnosis, treatment, and monitoring (Figure 1). Many AI medical platforms, such as NVIDIA, QOCA (Quanta computer incorporated), Advantech, and system analysis program (SAP)/high-performance analytic appliance (HANA), can serve as big biodata sources. Physicians can quickly copy and understand large amounts of clinical biodata by using high-performance computing (HPC) and AI pipelines. Multidisciplinary teams of professionals, such as physicians, biostatisticians, and

information (IT) technicians, are needed to establish operational flows. This article discusses the use of AI techniques in cancer genomics based on high-performance computing.

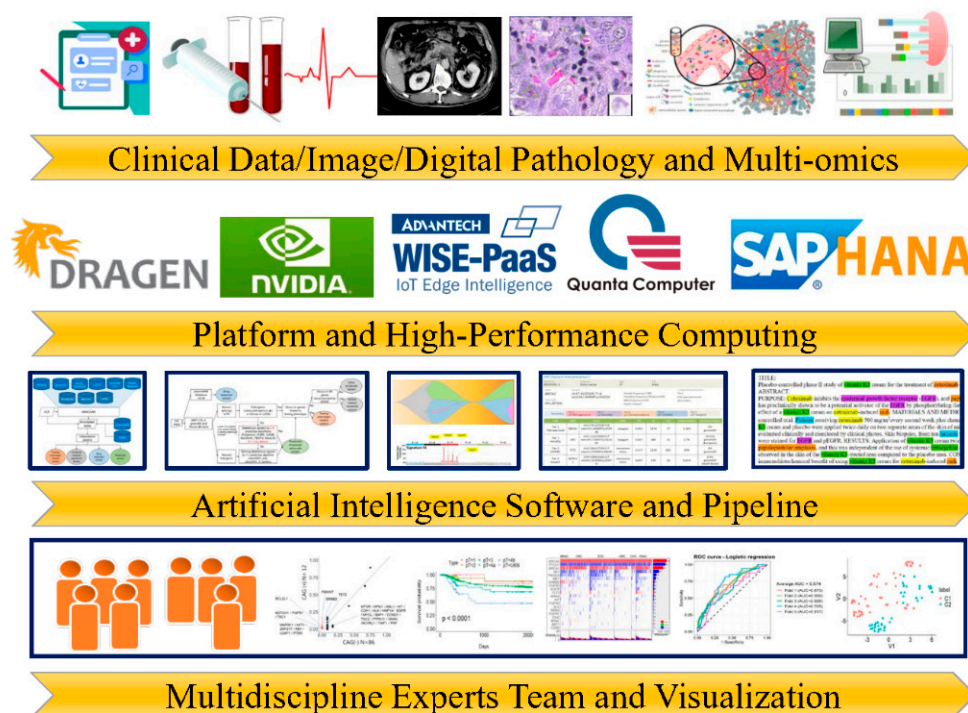


Figure 1. Clinical practice for precision cancer genomics and artificial intelligence-powered bioinformatic technologies. Artificial intelligence techniques, software platforms, and high-performance computation have been used extensively to provide improved cancer care via clinical patient data, such as electronic medical records, physiological measurements, biochemistry, computerized tomography scans, digital pathology, and the genetic landscape of cancer tissue.

In cancer patients, many bioinformatic analysis workflows are used to investigate cancer targets and monitor strategies using DNA whole-genome/whole-exome sequencing (WGS/WES) and circulating tumor DNA, RNA, and tumor deep-targeted sequencing (1). In addition to the use of genetic changes, the clinical use of a tumor mutation burden (TMB), microsatellite instability (MSI), and mutational signature patterns in cancers was reported by Bødker et al. [1]. We previously created an in-house bioinformatic pipeline for data processing and analysis [2]. The Cancer Next-Generation Sequencing (NGS) Laboratory at the National Cheng Kung University Hospital (NCKUH is a hospital in North District, Tainan, Taiwan) provided the genome analysis workflow. DNA was extracted from the blood and tumor tissues of cancer patients and sequenced using next-generation sequencers, such as the Illumina sequencing and Oxford Nanopore systems. We used HPC systems, such as NVIDIA graphics processing units (GPUs) and DRAGEN field programmable gate arrays (FPGAs) systems, to accelerate genomic analysis.

This article reviews new AI technologies for genome analysis, focusing on four primary issues. First, we examine the computational prediction of pathogenic variants and create a pipeline for the interpretation of genetic mutations regarding clinical cancer susceptibility [3]. Second, new cancer therapy options, such as the analysis of somatic genetic mutations, mutational signatures, and cancer evolution, have been developed using AI models. Third, we review the single-cell genomics sequencing technologies and computational techniques in cancer biology. Fourth, we discuss text-mining, which offers a practical approach for identifying cancer gene targets. Finally, to improve performance, we examine how clinical NGS laboratories deploy HPC via NVIDIA, DRAGEN systems, and AI medical cloud platforms. After visualizing and integrating a data-management system, AI has introduced new concepts and clinical practices focusing on genomics precision medicine.

The use of advanced AI technologies can facilitate the more precise and rapid identification of potential cancer-treatment targets.

2. Computational Prediction of Pathogenic Variants of Cancer Susceptibility Genes

Owing to the genome sequencing of germline mutations, the clinical diagnosis of hereditary genetic cancer patients has been translated into treatment options. Immunotherapy and poly (ADP-ribose) polymerase inhibitors are used as first-line treatments in colorectal and ovarian cancers [4,5]. Germline genetic variations may contribute to carcinogenesis, therapeutic efficacy, toxicity, and cancer phenotype. It remains difficult for clinicians to interpret the influence of these mutations on distinct phenotypes.

A diverse range of AI-based diagnostic tools has been developed using various computational genomic models. The American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) have published guidelines for interpreting sequence variations [6]. Multiple prediction scores can be identified and segregated for missense variants. The ACMG refers to three different prediction algorithms, namely, the Rare Exome Variant Ensemble Learner (REVEL) ensemble method, based on combinations of scores; genetic conservation site predictors [7]; and splicing site predictions. The application of existing silicon models to evaluate pathogenic variants in a particular gene, including alternative splicing with protein disruption, is greatly limited. As a result, we have divided our algorithm into three categories: sequencing-based, amino-acid/protein-based, or ACMG-criteria-provided-based. Table 1 lists AI-based genetic variations for which machine learning methods can be used to estimate pathogenicity. The scoring features and sources of the in silico prediction tools are also categorized.

Table 1. AI-based prediction models for the pathogenicity of genetic variants.

Methods	Categorical Prediction	Algorithms	Author
Sequencing-based prediction			
VEST4	Higher scores are more deleterious	RF	Carter et al., 2013 [8]
MetaSVM	Higher scores are more deleterious	Radial kernel SVM	Dong et al., 2015 [9]
REVEL	Higher scores are more deleterious	Ensemble methods/RF	Ioannidis et al., 2016 [10]
Primate AI	Higher scores are more deleterious	Convolutional neural network	Sundaram et al., 2018 [11]
CADD	Higher scores are more deleterious	Linear kernel SVM	Rentzsch et al., 2019 [12]
Splice AI	Higher scores are more deleterious	Deep neural network	Jaganathan et al., 2019 [13]
3Cnet	Higher scores are more deleterious	Recurrent neural network	Won et al., 2021 [14]
CoLaSp	Higher scores are more deleterious	Latent space matrix factorization	Abdollahi et al., 2021 [15]
MVP	Higher scores are more deleterious	ResNets	Qi et al., 2021 [16]
VARITY	P: Pathogenicity; B: Benign	XGBoost	Wu et al., 2021 [17]
EvoDiagnostics	P: Pathogenicity; B: Benign	RF	Labes et al., 2022 [18]
Amino acid or protein-based prediction			
PROVEAN	D: Deleterious; N: Neutral	Delta alignment score	Choi et al., 2015 [19]
ProtVec	P: Pathogenicity; B: Benign	NLP/SVM	Asgari et al., 2015 [20]
BioSeq-Analysis 2.0	P: Pathogenicity; B: Benign	RF/SVM	Liu et al., 2019 [21]
Rhapsody	Pathogenicity probability	RF	Ponzoni et al., 2020 [22]
LYRUS	P: Pathogenicity; B: Benign	XGBoost	Lai et al., 2021 [23]
LightGBM	P: Pathogenicity; B: Benign	LightGBM	Wu et al., 2021 [24]
ACMG/AMP-based model			
Modelling ACMG	P: Pathogenicity; B: Benign	Bayesian classification framework	Tavtigian et al., 2018 [25]
CharGer	Higher scores are more deleterious	Databases and criteria-based	Scott et al., 2019 [26]
VarSome	P: Pathogenicity; B: Benign	Databases and criteria-based	Kopanos et al., 2019 [27]
Clinvita	P: Pathogenicity; B: Benign	Penalized logistic regression	Nicora et al., 2022 [28]

Notes: RF: random forest; SVM: support vector machine; and NLP: natural language processing.

2.1. Sequencing-Based Prediction

Sequence-based techniques [8–18] are used to construct the most common tools for predicting the pathogenicity of genetic variations. For instance, REVEL [10] employs a random forest method based on ensemble methods with 13 pathogenicity predictors. REVEL scores can identify human pathogenic missense mutations and rare variants. CADD [12] is another ensemble method that integrates several scoring algorithms using a linear kernel support vector machine. VEST4 [8], EvoDiagnostics [18], and MetaSVM [9] are well-known random forest (RF) and support vector machine (SVM) prediction tools. Random forests and SVM may handle linear and non-linear data. The random forests classifier is an ensemble of various distinct decision trees (uncorrelated models). It will outperform each of its individual models. ACMG uses REVEL, random forests, for ensemble computational prediction models. SVM is a supervised learning model for regression analysis and non-probabilistic binary linear classifiers. The algorithms can evaluate genomic data for pathogenic classification. Another ensemble technique is CADD. To increase the forecast accuracy for various numerical pathogenetic scoring tools, CADD uses the ensemble regression technique to aggregate many models.

Artificial neural networks, such as convolutional neural networks (CNN) and ResNets, have been employed by Primate AI [11] and missense variant pathogenicity prediction (MVP) [16] to predict the pathogenicity of missense variations. Splice AI [13] can accurately predict splice junctions from an arbitrary pre-mRNA transcript sequence based on a deep neural network (DNN). Other novel prediction tools based on the utilized recurrent neural network, XGBoost (a variant of the gradient boosted tree), include 3Cnet [14] and VARIETY [17]. The neural network can manage large training datasets and many correlated predictors. The parameters of the aforementioned distribution are modeled by a neural network when trained for a classification or regression task. For example, MVP, based on ResNets, could better prioritize pathogenic missense variants, especially in de novo genetic mutations and tolerance loss of function genetic variants. Splice AI estimates more than 10% of the pathogenic mutations, previously unrecognized mutations, in patients with rare genetic illnesses.

Constructing an AI algorithm to predict phenotype-specific genetic variations remains challenging. We previously developed a matrix factorization-based algorithm to predict the phenotype of chemotherapy-induced neuropathy. The matrix factorization method was evaluated using The Cancer Genome Atlas (TCGA) patient WES data and showed an improved performance (accuracy > 98%) [15].

2.2. Amino-Acid- or Protein-Based Prediction

Amino-acid- or protein-based [19–24] pathogenicity prediction approaches are promising because they consider the context of amino acid sequences and minimize overfitting to prior sequencing-based knowledge. Therefore, it is critical to create single amino-acid- or protein-based prediction models. AI technologies have also been used to predict protein function. For example, PROVEAN [19] predicts pathogenicity using a delta alignment score. Another ensemble method is ProtVec [20], which incorporates natural language processing (NLP) with support vector machines. The authors distribute the representation of biological sequences using the NLP techniques. BioSeq-Analysis 2.0 [21] and Rhapsody [22] are well-known random forest tools. BioSeq-Analysis 2.0 includes a classification algorithm modified from LIBSVM and a sequence-labeling algorithm based on conditional random fields. LYRUS [23] is a new prediction tool, developed by XGBoost (a highly efficient gradient-boosting decision tree, GBDT). We also previously developed a single amino acid variant of the Light Gradient Boosting Machine (LightGBM) [24] based on protein structural energies. Compared with the sequencing-based AI model, amino acid/protein-based techniques use a different database, such as Protein Data Banks (PDBs) or the Database of Protein Disorder (DisProt). Data preprocessing is another critical issue for the prediction model. For example, we use the PDBs and Rosetta energy function for pathogenic prediction [24].

2.3. AI Tools Based on ACMG/AMP and Functional Somatic Mutation

Traditional AI in bioinformatics uses sequence alignment matching, protein–protein interactions, and structure–function analysis to assist in cancer genome research. Such research will aid in the development of molecularly targeted medications. Several cancer drugs that directly target genetic alterations have been evaluated in clinical trials. For example, osimertinib can be used to treat individuals with non-small cell lung cancer (NSCLC) harboring the T790M EGFR mutation [29]. However, cancer is a genetically heterogeneous illness. Therapeutic targets could include somatic genetic mutations, mutational signatures, and cancer evolution. Table 1 shows ACMG-based genetic pathogenic mutation prediction and AI models for mutational signatures and cancer evolution. This section covers AI methods for ACMG prediction tools and the implications of mutational signatures and tumor evolution for drug development, including strategies to lower the likelihood of drug resistance. The classical AI classifier quantitates potentially damaging genes; they did not consider the pathogenicity evaluation in a disease. ACMG guidelines include disease information, such as population allele frequency, functional data of mutation, and segregation analysis in families, to evaluate the pathogenicity of genetic mutations.

For the pathogenicity interpretation of germline variants, the ACMG and the United Kingdom (UK) Association for Clinical Genomic Science (UK-ACGS) both provide updated consensus criteria for the evaluation and classification of pathogenic variations [6,30]. The policy includes 28 attributes with codes addressing different types of evidence to navigate the clinical interpretation of rare diseases. Each variant is assigned a pathogenicity assertion depending on the criteria used. In ClinVar [31], a public dataset, variants are classified as pathogenic, likely pathogenic, having unclear significance, likely benign, or benign, based on review status. The developed machine-learning algorithms were assessed in terms of classification and prioritization. Standard tools for interpreting pathogenicity based on ACMG criteria include modelling ACMG/AMP [25], CharGer [26], VarSome [27], and Clinvitea [28] (Table 1). CharGer [25] and VarSome [27] assign scores to each genetic variant depending on the amount and strength of the evidence provided by the ACMG/AMP criteria. Tavtigian et al. [25] developed a model of ACMG scores using a Bayesian classification system. Clinvitea et al. [28] used penalized logistic regression to prioritize and classify the pathogenicity of genetic variants. We also created an ACMG/AMP-based score computation and designed unique algorithms to assign a score or degree of pathogenicity [14].

A four-tiered framework has been presented for cancer somatic mutations to characterize somatic sequence variants depending on their clinical importance [32]. In somatic mutations, variants in genes associated with pathogenicity have been characterized. Many cancer-variant databases are available, including the Cancer Genome Interpreter (CGI) [33], Clinical Interpretation of Variants in Cancer (CIViC) [34], the Jackson Laboratory Clinical Knowledgebase (JAX-CKB) [35], OncoKB [36], and the Precision Medicine Knowledgebase (PMKB) [37], which have been used to interpret somatic variants of cancer. Many tools have been developed for modeling somatic mutations, such as MuSE, MuTect, SomaticSniper, Strelka, and VarScan2 [38]. Ng et al. developed a functional genomics platform (FASMIC) [39] to identify the driver mutations for potential clinically actionable genes. FunSeq2 is a tool used for prioritizing and annotating non-coding somatic variants [40]. However, many laboratories have reported functional somatic mutations based on ACMG/AMP criteria.

3. AI model for Mutational Analysis

3.1. Mutational Signatures and AI Tools

To investigate somatic genetic mutagenesis, mutational signatures have revealed the probable mechanisms of cancer etiology and biological processes. The mutational signature landscape can potentially indicate drug-responsive and resistance biomarkers and prognostic factors for cancer. For example, the sensitivity of ovarian cancer to poly (ADP)-ribose polymerase (PARP) inhibitors is linked to mutational signatures related to homologous recombination deficiency. In contrast, APOBEC-related mutational signatures are associ-

ated with responses to ataxia telangiectasia and Rad3-related kinase (ATR) inhibitors. Our previous studies found distinct characteristics of mutational signatures in patients with cancer-associated genetic variations. These mutational signatures provide additional information on the etiology and progression of individual cancers, as well as new biomarkers for cancer treatment [41]. New technologies and machine learning algorithms have increased the feasibility of identifying mutational signatures [42–46] and facilitated the integration of signature analysis into clinical decision-making. The Catalogue of Somatic Mutations in Cancer (COSMIC) Signatures [42], DeaminationSigs [44], and SparseSignatures [45] are standard tools for cancer etiology and quality control based on non-negative matrix factorization (NMF) algorithms (Table 2). For DeconstructSigs, a multiple linear regression model was used [43]. Chevalier et al. established an analysis and visualization tool to characterize and enhance the discovery of mutational signatures [46].

Table 2. AI tools in bioinformatics for mutational analysis.

Methods	DATA	Algorithms	Author
Mutational signatures			
COSMIC Signatures	SNV/indels	Non-negative matrix factorization	Alexandrov et al., 2020 [42]
DeconstructSigs	SNV/indels	Multiple linear regression model	Rosenthal et al., 2016 [43]
DeaminationSigs	SNV/indels	Non-negative matrix factorization	Bhagwate et al., 2019 [44]
SparseSignatures	SNV	Non-negative matrix factorization	Lal et al., 2021 [45]
Musicatk	SNV	Non-negative matrix factorization/LDA	Chevalier et al., 2021 [46]
Tumor evolution model			
LICHeE	SNV/CNV	Directed acyclic graph	Popic et al., 2015 [47]
SCHISM	SNV/CNV	Directed acyclic graph	Niknafs et al., 2015 [48]
Canopy	SNV/CNV	Bayesian mixture models	Jiang et al., 2016 [49]
ClonEvol	SNV/CNV	Bootstrap resampling	Dang et al., 2017 [50]
PACTION	SNV/CNV	Mixed integer linear programming	Sashittal et al., 2022 [51]
DeCiFering	SNV	Descendant cell fraction	Satas et al., 2022 [52]

Notes: ACMG, American College of Medical Genetics and Genomics; SNV, single-nucleotide variant; CNV, copy number variations; and LDA, latent Dirichlet allocation.

3.2. Cancer Evolution and AI Tools

Genetic mutations are characterized by a somatic evolutionary process that contributes to cancer development, progression, and drug resistance. At present, many algorithms are available for the analysis of clonal evolution [47–52]. To solve the problem of intra-tumor heterogeneity (ITH) from bulk DNA sequencing, numerous computational methodologies and instruments have been developed which analyze genome data to reconstruct and describe the clonal evolutionary landscape. There are three prominent roles for the clonal evolutionary development of cancer: clustering genetic variants by cell fraction, reconstructing the cancer clonal evolution tree, and visualizing clonal extension. DeCiFering [52] and ClonEvol [50] were designed for genetic variant clustering based on the descendant cell fraction and bootstrap resampling. LICHeE [47], SCHISM [48], and Canopy [49] used directed acyclic graphs (DAGs) and Bayesian methods to reconstruct phylogeny. PACTION [51] is a straightforward and rapid strategy that reconstructs the clonal architecture of cancer tumors based on mixed-integer linear programming (MILP). Fish plots and timescapes have been used to visualize the evolution of cancer. For the development of cancer therapeutic targets, our study identified genetic subclones, and clusters were identified using SciClone [53] based on a Bayesian clustering method. ClonEvol has been used to visualize the evolution of somatic mutations in cancers. We highlight the significance of cancer evolution models in the development of new methodologies for drug targets.

3.3. Clinical Practice in Mutational Signature and Cancer Evolution

Mutational signatures reveal information about mutagenic processes in cancer patients and the quality of genetic mutation detection in cancer tissues. In addition to homologous

recombination deficiency (mutational signature SBS3) and APOBEC-mutagenesis (mutational signature SBS2), we could detect mutational signatures SBS6, SBS14, SBS15, SBS20, SBS2, SBS26, and SBS44 in cancer patients with mismatch repair deficiencies [41,42]. These could potentially be used as immunotherapy biomarkers. In cancer tissues fixed in formalin, the mutation signature SBS1 was increased (spontaneous deamination of methylated cytosine). To reduce the impact of deamination, a mutational signature analysis should be considered in routine quality reports.

Cancer evolution can reveal the clonal nature of the driver mutations in the evolution process. It can also guide therapy by focusing on clonal and subclonal genetic mutations. Patients with cancer recurrence or drug resistance may benefit from an analysis of the cancer's evolution over time. For example, the *BRAF* clonal mutation remains resistant to *BRAF* inhibitors in some melanomas with co-existing alterations to other clonal genetic mutations. Using the cancer evolution model, we found that concurrent sequential *BRAF* mutations also affected hypermutation status. Compared to *AKT-BRAF* sequential mutations, *PETN-BRAF* sequential mutations were significantly more frequent in hypermutated cancers. The cancer evolution model may guide clinical practice. We also built a cancer evolution model based on the NGS data and applied machine learning analysis to identify potential evaluation therapeutic targets, such as DNA repair, *MYO18A*, and *FBXW7* genetic mutations in CRCs [54]. Using an AI model for mutational analysis could provide us with more detailed clinical cancer information.

4. Single-Cell Genomics and Computational Biology

Among the single-cell genomics technologies, epigenome sequencing, genome sequencing for lineage tracking, spatially resolved transcriptomics, and omics sequencing are the newest developments. Data from single-cell genomics are sparse and high-dimensional, which makes machine-learning analysis challenging. The high-dimensional data were typically reduced using principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), and uniform manifold approximation and projection (UMAP). Table 3 shows the AI methods for computational biology in single-cell genomics, including omics data integration, cell type classification, and trajectory inferences [55–67].

Table 3. Single-cell genomics and computational biology.

Methods	Goal	Algorithms	Author
Single-Cell Omics Data Integration			
MOFA+	Sparse data	Stochastic version of the algorithm	Argelaguet et al., 2020 [55]
sCCA	Sparse data	Sparse canonical correlation analysis (CCA)	Rodosthenous et al., 2020 [56]
Unicom	Distance matrix	Unsupervised topological alignment	Cao et al., 2020 [57]
sGCN	High-dimensional data	Graph convolutional networks	Song et al., 2020 [58]
PIntMF	Sparse data	Penalized integrative matrix factorization	Pierre-Jean et al., 2021 [59]
Cell type classification			
ACTINN	Immune cell	Neural Networks	Ma et al., 2020 [60]
Ikarus	Tumor cell	Logistic regression/network propagation	Dohmen et al., 2022 [61]
Trajectory inference			
CellRouter	Tree methods	Context likelihood of relatedness	Lummertz et al., 2018 [62]
STREAM	Graph methods	Gaussian process latent variable model	Chen et al., 2019 [63]
TinGA	Graph methods	Growing neural graph algorithm	Todorov et al., 2020 [64]
ELPIgraphy	Cyclic methods	Elastic energy functional and topological graph	Albergante et al., 2020 [65]
CStreet	Graph methods	k-nearest neighbors graph	Zhao et al., 2021 [66]
	Tree methods	Euclidean minimum spanning tree	Tenha et al., 2022 [67]

MOFA+ is a statistical framework for data integration. It reconstructs a low-dimensional data representation using stochastic variational inference that is amenable to GPU computations [55]. Sparse canonical correlation analysis (sCCA) also computed sparse latent

variables to predict complex traits [56]. These AI models support flexible sparsity data management in the same way as Penalized Integrating Matrix Factorization (PIntMF) [59]. Cao et al. use unsupervised topological alignment for single-cell multi-omics integration [57]. A graph convolutional networks algorithm was used to integrate disparate and interaction datasets [58]. Two examples were shown for cell type classification. The ACTINN [60] and Ikarus [61] used a neural network or logistic regression model to distinguish the immune cell or tumor cell.

The developmental trajectories could be computationally inferred using trajectory inference algorithms in single-cell genomics. CellRouter [62] uses tree methods to model the trajectory based on the context likelihood of relatedness. The STREAM [63] and TinGA [64] used graph methods for trajectories based on the gaussian process latent variable model or growing neural graph algorithm, respectively. The ELPIgraphy [65] used cyclic methods for trajectories based on elastic energy functional and topological graphs. CStreet also used the k-nearest neighbor graph for trajectories [66]. Tenha et al. [67] use Euclidean minimum spanning tree methods to model the trajectory based on single-cell biology. Based on the computational biology of signal cell techniques, we accelerated the identification of new cancer cell types and understood the disease trajectories. This may help us to define new cancer subtypes and monitor therapeutic responses.

5. Text Mining for Identifying Genes Targets in Cancers

The biomedical literature has presented far-reaching findings for drug-target identification and cancer treatment, including their biological significance (molecular and cell activities and signal pathways). Data mining is a machine-learning technique that works with AI technologies or statistical methods to identify optimal cancer targets in biomedical science. The link between disease and genetic alterations is critical to obtaining a better understanding of cancer biological mechanisms. Gene2Vec, a study that explored the idea of gene embedding in the spirit of word embedding, is one of the new forms of text-mining [68]. However, we could not explain the biological significance of the vector in the neural-embedding model. Table 4 [69–74] shows that various data-mining techniques have been used to classify gene-mutation diseases.

Table 4. Text-mining model for cancer-associated genes.

Relationships	Name	Algorithms	Author
Mutation–Gene	MuGeX	Naïve Bayes/Rocchio algorithm-TF-IDF	Erdogmus et al., 2007 [69]
Disease–Mutation		C4.5 decision tree	Singhal et al., 2016 [70]
Protein–Mutation	EnzyMiner	Probabilistic indexing	Yeniterzi et al., 2009 [71]
Variants–Literature	tmVar 2.0	Conditional random fields	Wei et al., 2018 [72]
Variant–Disease–Gene	MAGPEL	Sentence co-occurrence scoring	Saberian et al., 2020 [73]
Cancer–Genes		Hypergeometric test	Chen et al., 2021 [74]

Notes: TF: term frequency; and IDF: inverse document frequency.

Several text-mining methods have been developed for mutation–disease relationships. For example, MuGeX [69] employs the Nave Bayes/Rocchio algorithm-IDF to retrieve mutation–gene combinations from Medline abstracts in response to a disease query. tmVar 2.0 [72] is an approach that integrates genetic variant information from the literature with the Single Nucleotide Polymorphism Database (dbSNP) and ClinVar using conditional random fields (CRF). Several ML classifiers were tested, such as the C4.5, decision tree, multilayer perceptron, and Bayesian logistic regression. Singhal et al. [70] employed the C4.5 decision tree to create an automated pipeline that uses the full-text biomedical literature and is validated using evidence-based gene panels. This approach is focused on disease–mutation relationships. To infer variant-driven gene panels, Saberian et al. [73] integrated GNormPlus [75], tmVar 2.0 [72], and DNorm [76] into MAGPEL for variant–genotype–phenotype prediction. EnzyMiner [71] used probabilistic indexing for protein

mutation prediction, automatically identifying information on protein stability or enzyme activity from PubMed abstracts.

Our own work contextualized the genes for clinical precision medicine, presenting druggable targets, hereditary cancer syndrome mutations, and illness subgroups [74]. The hypergeometric test was used to construct the mutational landscape of the actionable cancer genome from the biomedical literature, which was then confirmed using the NGS database. Our platform may enable the development of a cancer gene panel recommendation system for precise cancer therapy.

6. The NVIDIA GPUs, DRAGEN FPGAs Systems, and AI Medical Cloud Platforms in the Clinical NGS Lab

6.1. Using NVIDIA GPUs, DRAGEN FPGAs Systems in Bioinformatic Analysis

Recent developments in HPC and biological data analysis technologies have resulted in the rapid growth in biological analyses. The use of HPC in bioinformatic analysis enables the efficient processing of large amounts of data in everyday clinical practice. We previously used hardware accelerators, such as GPUs and FPGAs, to speed up and maximize throughput of cancer genomics. In clinical services, we utilized the NVIDIA Parabricks and Illumina dynamic read analysis for genomics (DRAGEN) platforms (Table 5) [77–80].

Table 5. High-performance computing systems for cancer genome research.

Name	Computing System	Clinical Practice	Author
NVIDIA	GPUs	Mutational signature Critical care	Haradhvala et al., 2018 [77] Gorzynski et al., 2022 [78]
DRAGEN	FPGAs	TSO500 FFPE pipeline TSO500 ctDNA pipeline	Wei et al., 2022 [79] Pommergaard et al., 2022 [80]

Notes: GPUs: graphics processing units; FPGAs: field-programmable gate arrays; and TSO 500: TruSight Oncology 500 assay.

The NVIDIA Parabricks (GPUs) software suite analyzes whole-genome and exome sequencing data. It significantly improves throughput times for common genomic investigations, such as germline and somatic research. The NVIDIA Clara Parabricks toolkit includes germline Deepvariants, somatic, RNA, and human population pipelines. Precise and clear results were obtained for RNA analysis. A Signature Analyzer-GPUs has been used for mutational signature analysis [77]. Haradhvala et al. discovered mutational signatures linked to loss of POLE proofreading and mismatch repair [77]. Such information may help to inform clinical decisions concerning immunotherapy targets for cancer treatment. Gorzynski et al. [78] connected long-read sequencing (nanopore technology) and GPUs in an acute scenario to enable the real-time analysis of ultrarapids. GPU-accelerated tools on NVIDIA Clara Parabricks pipelines for cancer and germline analyses are helpful in clinical situations.

The Illumina DRAGEN Bio-IT Platform (FPGAs) enables precise, comprehensive, and rapid analyzes of NGS data. Updated algorithms for genetic data analysis can be provided using FPGAs-based bioinformatic acceleration devices. TruSight Oncology 500 Assay tumor profiling [79] and liquid biopsy NGS [80], analyzed using FPGAs, were recently developed for cancer treatment-monitoring techniques. These utilize tumor mutation burden, microsatellite instability, and genetic alteration information for cancer diagnosis, prognosis, and treatment. They will also provide for practical use of the homologous recombination deficiency (HRD) score in future treatment of ovarian cancer. We built a workflow in NCKUH to speed up the study of the complete exosome genome and tumor deep-target sequencing for clinical cancer management. These tools can be used for reliable and timely genetic diagnostics (Figure 1).

O’Connell et al. benchmarked two germline variant callers and four somatic variant callers. They compared traditional x86 CPU algorithms with GPU-accelerated algorithms implemented with NVIDIA Parabricks on Amazon Web Services (AWS) and

Google Cloud Platform (GCP). For germline callers, the author observed speedups of up to $65\times$ (GATK haplotype caller). Alternatively, somatic variant callers achieved speedups of up to $56.8\times$ (Mutect2 algorithm) [81]. For emergency use for hospitalized patients, Clark et al. built a pipeline based on the DRAGEN platform to analyze genome sequencing data. A median delivery time of less than 24 h was observed from blood samples to provisional findings. High accuracy and sensitivity were also observed [82].

6.2. AI Medical Cloud Platforms for Cancer Care

Many AI medical cloud platforms have been developed, including QOCA, Advantech, and SAP/HANA, that could integrate AI technologies for genome visualization and precision medicine for cancer care. In cancer clinics, medical imaging data must be integrated to visualize genomic data, select cancer-target drugs, and predict cancer survival. Integration of the system analysis program (SAP) cloud platform for genomic and clinical data enables practitioners to quickly evaluate and make sense of the data. The Variant Browser visualizes genetic variations and integrates patient clinical data stored in a clinical data warehouse as variant information, as well as genomic interactions relating to specific patients.

The Advantech edge visualization solution efficiently iterates using data-intensive visualizations. We previously collaborated with Advantech to create a cancer clinic dashboard that visualizes multiomics data and clinical information. Real-time recommendations could be provided to patients using advanced platform and visualization technology. QOCA is an AI-assisted platform for medical imaging and autonomous inference; it now plays an essential role in intelligent medical solutions. For example, by using QOCA electrocardiography (ECG) monitoring devices, clinicians might effectively monitor cancer patients receiving portable cancer treatment at home.

For the integration of multi-omics and medical images, we recently demonstrated that utilization of the covariate-adjusted tensor classification in the high-dimensional (CATCH) model could accurately classify recurrent colorectal cancer by combining adjusted radiomics-based CT images with RNA immune genome expression data. We integrated medical images and genome data into an operational flow for recurrent stage III colorectal cancer and provided individualized treatment strategies [83].

7. Conclusions

Five main viewpoints should be considered when providing precision medicine and cancer care. Before treatment, it is essential to identify inherited cancers with (i) cancer susceptibility genes and (ii) AI models for mutation analysis. For example, we identified MLH1 germline genetic mutations based on whole-genome sequencing in colorectal cancer patients [3]. For therapeutic strategies, immunotherapy, instead of traditional chemotherapy, may be the first choice for first-line treatment in metastatic colorectal cancer patients [4]. Another example is a cancer patient with cardiovascular *KCNH2* genetic variants; EKG showed the QTc prolonged during the chemotherapy. For better cancer care, we should carefully monitor the EKG during chemotherapy if the patient carries *KCNH2* genetic variants. (iii) Single-cell genomics may provide data for disease surveillance. We can suggest cancer gene panels to patients based on (iv) text-mining findings for cancer patients with refractory treatment. Mutation analysis, such as somatic mutation, mutational signature, and cancer evolution, could provide therapeutic strategies or targets for cancer patients. The intelligent hospital must set up (iv) telemedicine devices and high-performance computing for real-time, in-person patient care.

For high-risk stage II and stage III colorectal cancer patients, NCKUH developed an AI precision medicine platform to manage Big Biodata. We developed an AI tracking and alarm system for the electronic medical record, biochemistry data, genetics, and CT scan image analysis to improve survival and quality of life. We demonstrated that germline susceptibility and deletion structural variants can have an impact on the survival and therapeutic strategies for stage III colorectal cancer. For example, patients with germline DNA repair genetic variants and *CEP72* deletion structural variants have better survival in

CRCs [84]. Using AI model analysis, we could stratify the risk of cancer recurrence. For the application of the cancer evolutionary model, we identify potential evaluation therapeutic targets, such as *MYO18A*, and *FBXW7* genetic mutations in CRCs [54]. To determine the oncology image biomarker, we integrated adjusted CT images into genome data to accurately classify recurrent CRC. We could use this AI model to provide individualized cancer therapeutic strategies based on adjusted radiomic features in recurrent stage III CRC [83]. To improve the long-term quality of life, we will establish the AI model to predict chemotherapy side-effects, such as neuropathy and sarcopenia, for CRCs in the future.

Cancer care can be provided via telemedicine using AI-based technology. For example, the QOCA provides the AI medical cloud platform (QOCA aim), AI health care platform (QOCA apc), and AI telemedicine platform (QOCA atm). AI medical cloud platform (QOCA aim) can provide the best clinical decision assistance and accurate AI prediction with medical images and structured data analysis. For example, we established the AI model to predict cancer recurrence in stage II and III CRCs via standardized pathology reports. This can be a powerful tool for sharing the decision-making between physicians and patients. The QOCA apc, a case manager, provides a platform including the daily activities and physiological monitors, such as the heart rate, O₂ saturation, body temperature, and glucose levels of cancer patients. Cancer patients who received home-based chemotherapy can be closely monitored via QOCA apc. QOCA atm, a hospital-to-home platform, could help us manage patient needs, such as nutrition supplements and the adverse effects of chemotherapy, via real-time face-to-face interaction.

Choosing the proper service model at the end of life (EOL) for physicians and patients, such as hospice share care (HSC), hospice inpatient care (HIC), and hospice home care (HHC), is a typical challenge. For hospice home care, we set up an artificial intelligence services platform and built an AI model to suggest a services model based on the patient's characteristics, symptoms, and hospice care needs [85]. Our artificial intelligence hospice services deliver goal-directed care to alleviate symptoms and provide holistic care to terminal patients. We installed a computerized detection system for hospice home care with the QOCA apc. We identify the health conditions of patients in the end-of-life period using the EKG and O₂ sensing system. We applied QOCA atm to home-based hospice and palliative care. We could watch the patients and easily inform their family about their terminal status early in the process. We could deliver personalized end-of-life care and reduce the burden of care for family members through telemedicine and AI technology. With intelligent remote medicine technology, we can improve quality of life for terminal patients and respond to the core value of the need for dignity while dying at home.

The extensive use of WGS/WES has completely changed the diagnostic procedures in medical genetics, particularly for cancer, non-invasive prenatal screening, childhood development, and rare disorders. The incidence of cancer has dramatically increased over the last decade. WGS/WES in germline and somatic mutations can provide cancer diagnosis and the etiology of cancer. For example, the mutational signature of urothelial carcinoma showed that aristolochic acid exposure plays a vital role in Taiwan [86]. This could be a screening tool for cancer etiology to determine the public health policy. Concerning public health issues, we can screen for cancer, create a preventive procedure for cancer, and promote lifestyle changes in inherited or high-risk cancer populations [87]. Contrary to the single or multiple gene panels, we performed WGS/WES-based genomic analyses using AI-based high-performance computing methods. More quick and accurate diagnoses could be reached. Pharmacological side-effects (pharmacogenomics) and the ACMG non-oncogenic phenotype are other significant public health issues, particularly for cancer patients. Implementing AI-based algorithms in high-performance computing is most urgent due to the public health concerns.

AI-powered analytical techniques with HPC platforms are widely used in clinical practice for precision cancer genomics. Clinicians must grasp the concept of big data within the healthcare field and translate it into usable knowledge for real-time patient care. Future research should, therefore, focus on the fast-evolving fields of bioinformatics, AI medical

clouds, and visualization platforms. AI-powered bioinformatics technologies are expected to routinely provide clinical diagnostic and treatment services in the future.

Author Contributions: Conceptualization, P.-C.L., Y.-S.T., Y.-M.Y. and M.-R.S.; funding acquisition, M.-R.S.; supervision, P.-C.L. and M.-R.S.; and writing—original draft and review/editing, P.-C.L., Y.-S.T., Y.-M.Y. and M.-R.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Ministry of Science and Technology (MOST), Taiwan under Research Grant of MOST 111-2634-F-006-002 and MOST 111-2634-F-006-007, the Ministry of Health and Welfare (MOHW111-TDU-B-221-114005), and the National Cheng Kung University Hospital (NCKUH-11102061).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Bødker, J.S.; Sønderkær, M.; Vesteghem, C.; Schmitz, A.; Brøndum, R.F.; Sommer, M.; Rytter, A.S.; Nielsen, M.M.; Madsen, J.; Jensen, P.; et al. Development of a Precision Medicine Workflow in Hematological Cancers, Aalborg University Hospital, Denmark. *Cancers* **2020**, *12*, 312. [[CrossRef](#)] [[PubMed](#)]
2. Conceição, S.I.R.; Couto, F.M. Text Mining for Building Biomedical Networks Using Cancer as a Case Study. *Biomolecules* **2021**, *11*, 1430. [[CrossRef](#)] [[PubMed](#)]
3. Lin, P.C.; Yeh, Y.M.; Wu, P.Y.; Hsu, K.F.; Chang, J.Y.; Shen, M.R. Germline susceptibility variants impact clinical outcome and therapeutic strategies for stage III colorectal cancer. *Sci. Rep.* **2019**, *9*, 3931. [[CrossRef](#)] [[PubMed](#)]
4. André, T.; Shiu, K.K.; Kim, T.W.; Jensen, B.V.; Jensen, L.H.; Punt, C.; Smith, D.; Garcia-Carbonero, R.; Benavides, M.; Gibbs, P.; et al. Pembrolizumab in Microsatellite-Instability-High Advanced Colorectal Cancer. *N. Engl. J. Med.* **2020**, *383*, 2207–2218. [[CrossRef](#)] [[PubMed](#)]
5. Moore, K.; Colombo, N.; Scambia, G.; Kim, B.G.; Oaknin, A.; Friedlander, M.; Lisianskaya, A.; Floquet, A.; Leary, A.; Sonke, G.S.; et al. Maintenance Olaparib in Patients with Newly Diagnosed Advanced Ovarian Cancer. *N. Engl. J. Med.* **2018**, *379*, 2495–2505. [[CrossRef](#)]
6. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W.W.; Hegde, M.; Lyon, E.; Spector, E.; et al. Standards and guidelines for interpreting sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **2015**, *17*, 405–424. [[CrossRef](#)]
7. Valenzuela-Palomo, A.; Bueno-Martínez, E.; Sanoguera-Miralles, L.; Lorca, V.; Fraile-Bethencourt, E.; Esteban-Sánchez, A.; Gómez-Barrero, S.; Carvalho, S.; Allen, J.; García-Álvarez, A.; et al. Splicing predictions, minigene analyses, and ACMG-AMP clinical classification of 42 germline PALB2 splice-site variants. *J. Pathol.* **2022**, *256*, 321–334. [[CrossRef](#)]
8. Carter, H.; Douville, C.; Stenson, P.D.; Cooper, D.N.; Karchin, R. Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genom.* **2013**, *14*, S3. [[CrossRef](#)]
9. Dong, C.; Wei, P.; Jian, X.; Gibbs, R.; Boerwinkle, E.; Wang, K.; Liu, X. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* **2015**, *24*, 2125–2137. [[CrossRef](#)]
10. Ioannidis, N.M.; Rothstein, J.H.; Pejaver, V.; Middha, S.; McDonnell, S.K.; Baheti, S.; Musolf, A.; Li, Q.; Holzinger, E.; Karyadi, D.; et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* **2016**, *99*, 877–885. [[CrossRef](#)]
11. 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](#)] [[PubMed](#)]
12. Rentzsch, P.; Witten, D.; Cooper, G.M.; Shendure, J.; Kircher, M. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **2019**, *47*, D886–D894. [[CrossRef](#)] [[PubMed](#)]
13. Jaganathan, K.; Kyriazopoulou Panagiotopoulou, S.; McRae, J.F.; Darbandi, S.F.; Knowles, D.; Li, Y.I.; Kosmicki, J.A.; Arbelaez, J.; Cui, W.; Schwartz, G.B.; et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell.* **2019**, *176*, 535–548. [[CrossRef](#)] [[PubMed](#)]
14. Won, D.G.; Kim, D.W.; Woo, J.; Lee, K. 3Cnet: Pathogenicity prediction of human variants using multitask learning with evolutionary constraints. *Bioinformatics* **2021**, *37*, 4626–4634. [[CrossRef](#)] [[PubMed](#)]
15. Abdollahi, S.; Lin, P.C.; Shen, M.R.; Chiang, J.H. Precise uncertain significance prediction using latent space matrix factorization models: Genomics variant and heterogeneous clinical data-driven approaches. *Brief. Bioinform.* **2021**, *22*, bbaa281. [[CrossRef](#)] [[PubMed](#)]

16. Qi, H.; Zhang, H.; Zhao, Y.; Chen, C.; Long, J.J.; Chung, W.K.; Guan, Y.; Shen, Y. MVP predicts the pathogenicity of missense variants by deep learning. *Nat. Commun.* **2021**, *12*, 510. [[CrossRef](#)]
17. Wu, Y.; Liu, H.; Li, R.; Sun, S.; Weile, J.; Roth, F.P. Improved pathogenicity prediction for rare human missense variants. *Am. J. Hum. Genet.* **2021**, *108*, 1891–1906. [[CrossRef](#)]
18. Labes, S.; Stupp, D.; Wagner, N.; Bloch, I.; Lotem, M.; Lahad, E.L.; Polak, P.; Pupko, T.; Tabach, Y. Machine-learning of complex evolutionary signals improves classification of SNVs. *NAR Genom. Bioinform.* **2022**, *4*, lqac025. [[CrossRef](#)]
19. Choi, Y.; Chan, A.P. PROVEAN /btv195.web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **2015**, *31*, 2745–2747. [[CrossRef](#)]
20. Asgari, E.; Mofrad, M.R. Continuous Distributed Representation of Biological Sequences for Deep Proteomics and Genomics. *PLoS ONE* **2015**, *10*, e0141287. [[CrossRef](#)]
21. Liu, B.; Gao, X.; Zhang, H. BioSeq-Analysis2.0: An updated platform for analyzing DNA, RNA and protein sequences at sequence level and residue level based on machine learning approaches. *Nucleic Acids Res.* **2019**, *47*, e127. [[CrossRef](#)] [[PubMed](#)]
22. Ponzoni, L.; Peñaherrera, D.A.; Oltvai, Z.N.; Bahar, I. Rhapsody: Predicting the pathogenicity of human missense variants. *Bioinformatics* **2020**, *36*, 3084–3092. [[CrossRef](#)] [[PubMed](#)]
23. Lai, J.; Yang, J.; Gamsiz Uzun, E.D.; Rubenstein, B.M.; Sarkar, I.N. LYRUS: A machine learning model for predicting the pathogenicity of missense variants. *Bioinform. Adv.* **2021**, *2*, vbab045. [[CrossRef](#)] [[PubMed](#)]
24. Wu, T.H.; Lin, P.C.; Chou, H.H.; Shen, M.R.; Hsieh, S.Y. Pathogenicity Prediction of Single Amino Acid Variants with Machine Learning Model Based on Protein Structural Energies. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**. [[CrossRef](#)] [[PubMed](#)]
25. Tavtigian, S.V.; Greenblatt, M.S.; Harrison, S.M.; Nussbaum, R.L.; Prabhu, S.A.; Boucher, K.M.; Biesecker, L.G. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet. Med.* **2018**, *20*, 1054–1060. [[CrossRef](#)]
26. Scott, A.D.; Huang, K.L.; Weerasinghe, A.; Mashl, R.J.; Gao, Q.; Martins Rodrigues, F.; Wyczalkowski, M.A.; Ding, L. CharGer: Clinical Characterization of Germline variants. *Bioinformatics* **2019**, *35*, 865–867. [[CrossRef](#)]
27. Kopanos, C.; Tsiolkas, V.; Kouris, A.; Chapple, C.E.; Aguilera, M.A.; Meyer, R.; Massouras, A. VarSome: The human genomic variant search engine. *Bioinformatics* **2019**, *35*, 1978–1980. [[CrossRef](#)]
28. Nicora, G.; Zucca, S.; Limongelli, I.; Bellazzi, R.; Magni, P. A machine learning approach based on ACMG/AMP guidelines for genomic variant classification and prioritization. *Sci. Rep.* **2022**, *12*, 2517. [[CrossRef](#)]
29. Mok, T.S.; Wu, Y.; Ahn, M.; Garassino, M.C.; Kim, H.R.; Ramalingam, S.S.; Shepherd, F.A.; He, Y.; Akamatsu, H.; Theelen, W.S.; et al. Osimertinib or platinum–pemetrexed in EGFR T790M–positive lung cancer. *N. Engl. J. Med.* **2017**, *376*, 629–640. [[CrossRef](#)]
30. Garrett, A.; Durkie, M.; Callaway, A.; Burghel, G.J.; Robinson, R.; Drummond, J.; Torr, B.; Cubuk, C.; Berry, I.R.; Wallace, A.J.; et al. Combining evidence for and against pathogenicity for variants in cancer susceptibility genes: CanVIG-UK consensus recommendations. *J. Med. Genet.* **2021**, *58*, 297–304. [[CrossRef](#)]
31. Landrum, M.J.; Lee, J.M.; Benson, M.; Brown, G.; Chao, C.; Chitipiralla, S.; Gu, B.; Hart, J.; Hoffman, D.; Hoover, J.; et al. ClinVar: Public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **2016**, *44*, D862–D868. [[CrossRef](#)] [[PubMed](#)]
32. Li, M.M.; Datto, M.; Duncavage, E.J.; Kulkarni, S.; Lindeman, N.I.; Roy, S.; Tsimberidou, A.M.; Vnencak-Jones, C.L.; Wolff, D.J.; Younes, A.; et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J. Mol. Diagn.* **2017**, *19*, 4–23. [[CrossRef](#)] [[PubMed](#)]
33. Tamborero, D.; Rubio-Perez, C.; Deu-Pons, J.; Schroeder, M.P.; Vivancos, A.; Rovira, A.; Tusquets, I.; Albanell, J.; Rodon, J.; Taberero, J.; et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med.* **2018**, *10*, 25. [[CrossRef](#)] [[PubMed](#)]
34. Griffith, M.; Spies, N.C.; Krysiak, K.; McMichael, J.F.; Coffman, A.C.; Danos, A.M.; Ainscough, B.J.; Ramirez, C.A.; Rieke, D.T.; Kujan, L.; et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat. Genet.* **2018**, *49*, 170–174. [[CrossRef](#)] [[PubMed](#)]
35. Patterson, S.E.; Liu, R.; Statz, C.M.; Durkin, D.; Lakshminarayana, A.; Mockus, S.M. The clinical trial landscape in oncology and connectivity of somatic mutational profiles to targeted therapies. *Hum. Genom.* **2016**, *10*, 4. [[CrossRef](#)] [[PubMed](#)]
36. Chakravarty, D.; Gao, J.; Phillips, S.M.; Kundra, R.; Zhang, H.; Wang, J.; Rudolph, J.E.; Yaeger, R.; Soumerai, T.; Nissan, M.H.; et al. OncoKB: A Precision Oncology Knowledge Base. *JCO Precis. Oncol.* **2017**, *2017*, PO.17.00011. [[CrossRef](#)]
37. Huang, L.; Fernandes, H.; Zia, H.; Tavassoli, P.; Rennert, H.; Pisapia, D.; Imielinski, M.; Sboner, A.; Rubin, M.A.; Kluk, M.; et al. The cancer precision medicine knowledge base for structured clinical-grade mutations and interpretations. *J. Am. Med. Assoc.* **2017**, *24*, 513–519. [[CrossRef](#)]
38. Fan, Y.; Xi, L.; Hughes, D.S.; Zhang, J.; Zhang, J.; Futreal, P.A.; Wheeler, D.A.; Wang, W. MuSE: Accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. *Genome Biol.* **2016**, *17*, 178. [[CrossRef](#)]
39. Ng, P.K.; Li, J.; Jeong, K.J.; Shao, S.; Chen, H.; Tsang, Y.H.; Sengupta, S.; Wang, Z.; Bhavana, V.H.; Tran, R.; et al. Systematic Functional Annotation of Somatic Mutations in Cancer. *Cancer Cell* **2018**, *33*, 450–462.e10. [[CrossRef](#)]
40. Dhingra, P.; Fu, Y.; Gerstein, M.; Khurana, E. Using FunSeq2 for Coding and Non-Coding Variant Annotation and Prioritization. *Curr. Protoc. Bioinform.* **2017**, *57*, 15.11.1–15.11.17. [[CrossRef](#)]

41. Koh, G.; Degasperi, A.; Zou, X.; Momen, S.; Nik-Zainal, S. Mutational signatures: Emerging concepts, caveats and clinical applications. *Nat. Rev. Cancer* **2021**, *21*, 619–637. [[CrossRef](#)] [[PubMed](#)]
42. 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](#)]
43. 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](#)] [[PubMed](#)]
44. Bhagwate, A.V.; Liu, Y.; Winham, S.J.; McDonough, S.J.; Stallings-Mann, M.L.; Heinzen, E.P.; Davila, J.I.; Vierkant, R.A.; Hoskin, T.L.; Frost, M.; et al. Bioinformatics and DNA-extraction strategies to reliably detect genetic variants from FFPE breast tissue samples. *BMC Genom.* **2019**, *20*, 689. [[CrossRef](#)] [[PubMed](#)]
45. Lal, A.; Liu, K.; Tibshirani, R.; Sidow, A.; Ramazzotti, D. De novo mutational signature discovery in tumor genomes using SparseSignatures. *PLoS Comput. Biol.* **2021**, *17*, e1009119. [[CrossRef](#)]
46. Chevalier, A.; Yang, S.; Khurshid, Z.; Sahelijo, N.; Tong, T.; Huggins, J.H.; Yajima, M.; Campbell, J.D. The Mutational Signature Comprehensive Analysis Toolkit (musicatk) for the Discovery, Prediction, and Exploration of Mutational Signatures. *Cancer Res.* **2021**, *81*, 5813–5817. [[CrossRef](#)]
47. Popic, V.; Salari, R.; Hajirasouliha, I.; Kashef-Haghighi, D.; West, R.B.; Batzoglou, S. Fast and scalable inference of multi-sample cancer lineages. *Genome Biol.* **2015**, *16*, 91. [[CrossRef](#)]
48. Niknafs, N.; Beleva-Guthrie, V.; Naiman, D.Q.; Karchin, R. SubClonal Hierarchy Inference from Somatic Mutations: Automatic Reconstruction of Cancer Evolutionary Trees from Multi-region Next Generation Sequencing. *PLoS Comput. Biol.* **2015**, *11*, e1004416. [[CrossRef](#)]
49. Jiang, Y.; Qiu, Y.; Minn, A.J.; Zhang, N.R. Assessing intratumor heterogeneity and tracking longitudinal and spatial clonal evolutionary history by next-generation sequencing. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E5528–E5537. [[CrossRef](#)]
50. Dang, H.X.; White, B.S.; Foltz, S.M.; Miller, C.A.; Luo, J.; Fields, R.C.; Maher, C.A. ClonEvol: Clonal ordering and visualization in cancer sequencing. *Ann. Oncol.* **2017**, *28*, 3076–3082. [[CrossRef](#)]
51. Sashittal, P.; Zaccaria, S.; El-Kebir, M. Parsimonious Clone Tree Integration in cancer. *Algorithms Mol. Biol.* **2022**, *17*, 3. [[CrossRef](#)] [[PubMed](#)]
52. Satas, G.; Zaccaria, S.; El-Kebir, M.; Raphael, B.J. DeCiFering the elusive cancer cell fraction in tumor heterogeneity and evolution. *Cell Syst.* **2021**, *12*, 1004–1018.e10. [[CrossRef](#)] [[PubMed](#)]
53. Miller, C.A.; White, B.S.; Dees, N.D.; Griffith, M.; Welch, J.S.; Griffith, O.L.; Vij, R.; Tomasson, M.H.; Graubert, T.A.; Walter, M.J.; et al. SciClone: Inferring Clonal Architecture and Tracking the Spatial and Temporal Patterns of Tumor Evolution. *PLoS Comput. Biol.* **2014**, *10*, e1003665. [[CrossRef](#)] [[PubMed](#)]
54. Lin, P.C.; Yeh, Y.M.; Lin, B.W.; Lin, S.C.; Chan, R.H.; Chen, P.C.; Shen, M.R. Intratumor Heterogeneity of MYO18A and FBXW7 Variants Impact the Clinical Outcome of Stage III Colorectal Cancer. *Front. Oncol.* **2020**, *10*, 588557. [[CrossRef](#)] [[PubMed](#)]
55. Argelaguet, R.; Arnol, D.; Bredikhin, D.; Deloro, Y.; Velten, B.; Marioni, J.C.; Stegle, O. MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* **2020**, *21*, 111. [[CrossRef](#)] [[PubMed](#)]
56. Rodosthenous, T.; Shahrezaei, V.; Evangelou, M. Integrating multi-OMICS data through sparse Canonical Correlation Analysis for the prediction of complex traits: A comparison study. *Bioinformatics* **2020**, *36*, 4616–4625. [[CrossRef](#)] [[PubMed](#)]
57. Cao, K.; Bai, X.; Hong, Y.; Wan, L. Unsupervised topological alignment for single-cell multi-omics integration. *Bioinformatics* **2020**, *36*, i48–i56. [[CrossRef](#)]
58. Song, Q.; Su, J.; Zhang, W. scGCN: A graph convolutional networks algorithm for knowledge transfer in single cell Omics. *bioRxiv.* **2020**. [[CrossRef](#)]
59. Pierre-Jean, M.; Mauger, F.; Deleuze, J.F.; Le Floch, E. PIntMF: Penalized Integrative Matrix Factorization method for Multi-omics data. *Bioinformatics* **2021**, *38*, 900–907. [[CrossRef](#)]
60. Ma, F.; Pellegrini, M. ACTINN: Automated identification of cell types in single cell RNA sequencing. *Bioinformatics* **2020**, *36*, 533–538. [[CrossRef](#)]
61. Dohmen, J.; Baranovskii, A.; Ronen, J.; Uyar, B.; Franke, V.; Akalin, A. Identifying tumor cells at the single-cell level using machine learning. *Genome Biol.* **2022**, *23*, 123. [[CrossRef](#)] [[PubMed](#)]
62. Lummertz da Rocha, E.; Rowe, R.G.; Lundin, V.; Malleshaiah, M.; Jha, D.K.; Rambo, C.R.; Li, H.; North, T.E.; Collins, J.J.; Daley, G.Q. Reconstruction of complex single-cell trajectories using CellRouter. *Nat. Commun.* **2018**, *9*, 892. [[CrossRef](#)] [[PubMed](#)]
63. Chen, H.; Albergante, L.; Hsu, J.Y.; Lareau, C.A.; Lo Bosco, G.; Guan, J.; Zhou, S.; Gorban, A.N.; Bauer, D.E.; Aryee, M.J.; et al. Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. *Nat. Commun.* **2019**, *10*, 1903. [[CrossRef](#)] [[PubMed](#)]
64. Todorov, H.; Cannoodt, R.; Saelens, W.; Saeys, Y. TinGa: Fast and flexible trajectory inference with Growing Neural Gas. *Bioinformatics* **2020**, *36*, i66–i74. [[CrossRef](#)] [[PubMed](#)]
65. Albergante, L.; Mirkes, E.; Bac, J.; Chen, H.; Martin, A.; Faure, L.; Barillot, E.; Pinello, L.; Gorban, A.; Zinovyev, A. Robust and Scalable Learning of Complex Intrinsic Dataset Geometry via EIPiGraph. *Entropy* **2020**, *22*, 296. [[CrossRef](#)]
66. Zhao, C.; Xiu, W.; Hua, Y.; Zhang, N.; Zhang, Y. CStreet: A computed Cell State trajectory inference method for time-series single-cell RNA sequencing data. *Bioinformatics* **2021**, *37*, 3774–3780. [[CrossRef](#)]

67. Tenha, L.; Song, M. Inference of trajectory presence by tree dimension and subset specificity by subtree cover. *PLoS Comput. Biol.* **2022**, *18*, e1009829. [[CrossRef](#)]
68. Du, J.; Jia, P.; Dai, Y.; Tao, C.; Zhao, Z.; Zhi, D. Gene2vec: Distributed Representation of Genes Based on Co-expression. *BMC Genom.* **2019**, *20*, 82. [[CrossRef](#)]
69. Erdogmus, M.; Sezerman, O.U. Application of Automatic Mutation- Gene Pair Extraction to Diseases. *J. Bioinform. Comput. Biol.* **2007**, *5*, 1261–1275. [[CrossRef](#)]
70. Singhal, A.; Simmons, M.; Lu, Z. Text Mining for Precision Medicine: Automating Disease-Mutation Relationship Extraction from Biomedical Literature. *J. Am. Med. Inform. Assoc.* **2016**, *23*, 766–772. [[CrossRef](#)]
71. Yeniterzi, S.; Sezerman, U. EnzyMiner: Automatic Identification of Protein Level Mutations and Their Impact on Target Enzymes from PubMed Abstracts. *BMC Bioinform.* **2009**, *10*, S2. [[CrossRef](#)] [[PubMed](#)]
72. Wei, C.H.; Phan, L.; Feltz, J.; Maiti, R.; Hefferon, T.; Lu, Z. tmVar 2.0: Integrating genomic variant information from literature with dbSNP and ClinVar for precision medicine. *Bioinformatics* **2018**, *34*, 80–87. [[CrossRef](#)] [[PubMed](#)]
73. Saberian, N.; Shafi, A.; Peyvandipour, A.; Draghici, S. MAGPEL: An autoMated Pipeline for Inferring vAriant-Driven Gene PanEls from the Full-Length Biomedical Literature. *Sci. Rep.* **2020**, *10*, 12365. [[CrossRef](#)] [[PubMed](#)]
74. Chen, H.O.; Lin, P.C.; Liu, C.R.; Wang, C.S.; Chiang, J.H. Contextualizing Genes by Using Text-Mined Co-Occurrence Features for Cancer Gene Panel Discovery. *Front. Genet.* **2021**, *12*, 771435. [[CrossRef](#)] [[PubMed](#)]
75. Wei, C.H.; Kao, H.Y.; Lu, Z. GNormPlus: An Integrative Approach for Tagging Genes, Gene Families, and Protein Domains. *BioMed Res. Int.* **2015**, *2015*, 918710. [[CrossRef](#)]
76. Leaman, R.; Islamaj Dogan, R.; Lu, Z. DNorm: Disease name normalization with pairwise learning to rank. *Bioinformatics* **2013**, *29*, 2909–2917. [[CrossRef](#)] [[PubMed](#)]
77. 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](#)]
78. Gorzynski, J.E.; Goenka, S.D.; Shafin, K.; Jensen, T.D.; Fisk, D.G.; Grove, M.E.; Spiteri, E.; Pesout, T.; Monlong, J.; Baid, G.; et al. Ultrarapid Nanopore Genome Sequencing in a Critical Care Setting. *N. Engl. J. Med.* **2022**, *386*, 700–702. [[CrossRef](#)]
79. Wei, B.; Kang, J.; Kibukawa, M.; Arreaza, G.; Maguire, M.; Chen, L.; Qiu, P.; Lang, L.; Aurora-Garg, D.; Cristescu, R.; et al. Evaluation of the TruSight Oncology 500 Assay for Routine Clinical Testing of Tumor Mutational Burden and Clinical Utility for Predicting Response to Pembrolizumab. *J. Mol. Diagn.* **2022**, *24*, 600–608. [[CrossRef](#)]
80. Pommergaard, H.C.; Yde, C.W.; Ahlborn, L.B.; Andersen, C.L.; Henriksen, T.V.; Hasselby, J.P.; Rostved, A.A.; Sørensen, C.L.; Rohrberg, K.S.; Nielsen, F.C.; et al. Personalized circulating tumor DNA in patients with hepatocellular carcinoma: A pilot study. *Mol. Biol. Rep.* **2022**, *49*, 1609–1616. [[CrossRef](#)]
81. O’Connell, K.A.; Yosufzai, Z.B.; Campbell, R.A.; Lobb, C.J.; Engelken, H.T.; Gorrell, L.M.; Carlson, T.B.; Catana, J.J.; Mikdadi, D.; Bonazzi, V.R.; et al. Accelerating genomic workflows using NVIDIA Parabricks. *bioRxiv* **2022**, *7*, 498972. [[CrossRef](#)]
82. Clark, M.M.; Hildreth, A.; Batalov, S.; Ding, Y.; Chowdhury, S.; Watkins, K.; Ellsworth, K.; Camp, B.; Kint, C.I.; Yacoubian, C.; et al. Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation. *Sci. Transl. Med.* **2019**, *11*, eaat6177. [[CrossRef](#)] [[PubMed](#)]
83. Huang, Y.-C.; Tsai, Y.-S.; Li, C.-I.; Chan, R.-H.; Yeh, Y.-M.; Chen, P.-C.; Shen, M.-R.; Lin, P.-C. Adjusted CT Image-Based Radiomic Features Combined with Immune Genomic Expression Achieve Accurate Prognostic Classification and Identification of Therapeutic Targets in Stage III Colorectal Cancer. *Cancers* **2022**, *14*, 1895. [[CrossRef](#)] [[PubMed](#)]
84. Lin, P.C.; Chen, H.O.; Lee, C.J.; Yeh, Y.M.; Shen, M.R.; Chiang, J.H. Comprehensive assessments of germline deletion structural variants reveal the association between prognostic MUC4 and CEP72 deletions and immune response gene expression in colorectal cancer patients. *Hum. Genom.* **2021**, *15*, 3. [[CrossRef](#)] [[PubMed](#)]
85. Lai, W.S.; Liu, I.T.; Tsai, J.H.; Su, P.F.; Chiu, P.H.; Huang, Y.T.; Chiu, G.L.; Chen, Y.Y.; Lin, P.C. Hospice delivery models and survival differences in the terminally ill: A large cohort study. *BMJ Support. Palliat. Care* **2021**, *11*. [[CrossRef](#)]
86. 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](#)]
87. Knerr, S.; Guo, B.; Mittendorf, K.F.; Feigelson, H.S.; Gilmore, M.J.; Jarvik, G.P.; Kauffman, T.L.; Keast, E.; Lynch, F.L.; Muessig, K.R.; et al. Risk-reducing surgery in unaffected individuals receiving cancer genetic testing in an integrated health care system. *Cancer* **2022**, *128*, 3090–3098. [[CrossRef](#)]