Advances in Bioinformatics Techniques to Predict Neoantigen: Exploring Tumor Immune Microenvironment and Transforming Data into Therapeutic Insights

Abdulwahed Alrehaily

Biology Department, Faculty of Science, Islamic University of Madinah, Madinah 42351, Saudi Arabia

Email: alrehaily.abdulwahed@gmail.com

ORCID: https://orcid.org/0000-0002-3509-7565

Abstract: The incorporation of bioinformatics into the prediction of neoantigens has greatly enhanced cancer immunotherapy by improving the understanding of tumor-specific antigens that can trigger targeted immune responses. This review emphasizes the vital role of bioinformatics in identifying neoantigens, which are unique antigens arising from somatic mutations, and their significance in customizing cancer treatments like therapeutic vaccines and T-cell therapies. It critically examines advanced sequencing technologies, such as whole-genome (WGS) and whole-exome sequencing (WES), for their role in assessing mutations that lead to neoantigen production. The review also discusses innovative computational methods, including artificial intelligence (AI), machine learning (ML), and deep learning (DL), for their effectiveness in predicting immunogenic neoantigens and tailoring personalized therapies. Case studies illustrate the successes achieved through these bioinformatics advancements, showcasing their potential in developing personalized vaccines that address the specific genetic makeup of tumors. Despite challenges like tumor heterogeneity and the complexities of data analysis, ongoing advancements.

Keywords: Bioinformatics, Machine Learning (ML), Artificial Intelligence (AI), Immuno-oncology, Immunotherapy, Computational Pipelines.

التطورات في تقنيات المعلوماتية الحيوية للتنبؤ بالمستضدات الجديدة: استكشاف البيئة المناعية للورم وتحويل البيانات إلى رؤى علاجية

الملخص: أدى دمج المعلوماتية الحيوية في التنبؤ بالمستضدات الجديدة إلى تعزيز العلاج المناعي للسرطان بشكل كبير من خلال تحسين فهم المستضدات الخاصة بالأورام والتي يمكن أن تؤدي إلى استجابات مناعية مستهدفة. تؤكد هذه المراجعة على الدور الحيوي للمعلوماتية الحيوية في تحديد المستضدات الجديدة، وهي مستضدات فريدة تنشأ عن الطفرات الجسدية، وأهميتها في تخصيص علاجات السرطان مثل اللقاحات العلاجية وعلاجات الخلايا التائية. يتناول هذا البحث بشكل نقدي تقنيات التسلسل المتقدمة، مثل تسلسل الجينوم الكامل (WGS) وتسلسل الإكسوم الكامل (WES)، لدورها في تقييم الطفرات التي تؤدي إلى إنتاج المستضدات الجديدة. كما يناقش البحث الأساليب الحسابية المبتكرة، بما في ذلك الذكاء الاصطناعي (AI)، والتعلم الآلي (ML)، والتعلم العميق (DL)، لفعاليتها في التنبؤ بالمستضدات الجديدة المناعية وتصميم العلاجات الشخصية. توضح دراسات الحالة النجاحات التي تحققت من خلال هذه التطورات في المعلوماتية الحيوية، حيث تعرض إمكاناتها في تطوير لقاحات شخصية تعالج التركيبة الجينية المحددة للأورام. وعلى الرغم من التحديات مثل تباين الورم وتعقيدات تحليل البيانات، فإن التطورات الجارية.

1. Introduction

Cancer is primarily a genetic illness, during its course, it is accompanied by genomic instability leading to point mutations and structural changes [1]. Cancers can be classified into metastatic and nonmetastatic forms, with metastasis arising during the development of tumors. Metastatic dissemination enables cancer cells to evade main tumors and establish colonies in other organs [2]. Tumors are intricate systems consisting of neoplastic cells, extracellular matrix (ECM), and "accessory" nonneoplastic cells, such as resident mesenchymal support cells, endothelial cells, and infiltrated inflammatory immune cells. Tumor growth is influenced by interactions between accessory cells and cancer cells. During tumor growth, the structure of the tissue changes and becomes a specialized microenvironment that is defined by a damaged extracellular matrix (ECM) and long-lasting inflammation [3]. Cancer-related inflammation plays a role in causing genetic instability, modifying epigenetic patterns, promoting the growth of cancer cells, enhancing pathways that prevent cell death, stimulating the formation of new blood vessels, and facilitating the spread of cancer [4]. The role of inflammatory immune cells in cancer-related inflammation is crucial, and several studies have demonstrated how immune cells affect tumor fate at various stages of the disease, such as early neoplastic transformation, clinically detected tumors, metastatic dissemination, and therapeutic intervention [5].

Cells of the innate immune system, such as natural killer (NK) cells, eosinophils, basophils, and phagocytic cells, have a role in suppressing tumors by either directly killing them or by triggering adaptive immunological responses. The adaptive immune system, which comprises lymphocytes, plays a crucial role in both humoral and cell-mediated immune responses. Unfortunately, cancer cells have developed defense mechanisms against immune surveillance, which impairs immune cells' ability to act as effectors thereby rendering immunotherapy less successful [6]. Gaining insight into these intricate interactions inside the tumor microenvironment (TME) is essential for the advancement of more potent cancer treatments. Immunotherapy has made significant improvements in the treatment of a variety of cancer types by utilizing the immune system's capacity to recognize and destroy cancer cells. Recent advancements in single-cell technologies and spatial transcriptomics have yielded valuable information about the diversity and spatial arrangement of immune cells within tumors. This has allowed for a thorough understanding of the tumor microenvironment (TME) and the discovery of new targets for therapeutic intervention [7].

These genetic changes can result in the production of neoantigens, or tumor-specific antigens. The immune system interprets these neoantigens as alien, which sets off cellular immunological responses [8]. One major problem that contributes to treatment failure and disease progression in human primary and metastatic cancers is cancer heterogeneity. The immune system's selective pressure, hierarchical architecture from the start of cancer stem cells, and genetic instability are a few of the factors that can account for this variability [5]. Tumor clonal growth and heterogeneity reduction are facilitated by cancer immune editing, which eradicates immunogenic cancer cells. Nevertheless, the absence of immune selection leads to a greater diversity of neoantigens. Neoantigen heterogeneity in lung and melanoma patients increases their vulnerability to T-cell attacks and responsiveness to tumor checkpoint inhibition [9]. This heterogeneity is heightened due to the inactivation of DNA repair machinery in colorectal, breast, and pancreatic cell lines. As genetic variation within a tumor increases, certain subpopulations of cells may evade the immune system's defenses. Metastatic development and therapeutic resistance often originate from a few clones inside the original tumors. In ovarian cancer, shrinking metastatic tumors were linked to an immune infiltrate characterized by CD4+ and CD8+ T cells, with higher tumor mutation and neoepitope load compared to advancing lesions [9].

Understanding the intricate connections between cancer cells and immune responses is crucial for developing effective cancer treatments.

1.1. Scope and Objectives of the Review

This review specifically examines the critical role of bioinformatics techniques in the prediction of neoantigens, which are essential for advancing personalized cancer immunotherapy. The primary focus is on how computational tools and sequencing technologies enable the identification of tumor-specific neoantigens that can be targeted for tailored cancer treatments, including therapeutic vaccines and T-cell-based therapies. The first domain explores the application of bioinformatics tools in the prediction of neoantigens, which play a vital role in tailored cancer immunotherapies such as vaccinations and T-cell treatments [10]. In addition, the second domain explores, data integration techniques for researching tumor-immune interactions covered in the review, with a focus on the significance of combining various data types from transcriptomics, proteomics, and genomes [11]. The third domain discusses various tools and platforms, such as single-cell RNA sequencing, spatial transcriptomics, CIBERSORT, and TIDE, which are being investigated to gain a thorough comprehension of the interactions between tumors and the immune system [12-14]. These methods aim to uncover possible targets for therapeutic interventions. The review also emphasizes the clinical benefit and ongoing progress in the field of bioinformatics-driven immunotherapy development, including immune checkpoint inhibitors, personalized cancer vaccines, and CAR-T cell therapy.

This review offers an extensive analysis of the transformative impact of bioinformatics on the field of immuno-oncology. The review emphasizes the most recent progress in bioinformatics tools and techniques, providing a clear explanation of cutting-edge methodologies. It also demonstrates the practical significance of bioinformatics in the development of efficient immunotherapies, as evidenced by case studies and success stories. In addition, it addresses potential future research avenues to improve cancer treatment and tackles persistent problems like data integration and complexity. The review seeks to be a beneficial resource for researchers and clinicians by summarizing important discoveries, tools, and methodologies. It aims to assist in the development and implementation of innovative investigations and therapeutic applications.

2. Bioinformatics Tools for Predicting Neoantigens

The study of neoantigens has greatly expedited the progress and regulatory approval of tumor immunotherapies. These include cancer vaccines, adoptive cell therapy, and antibody-based therapies [15]. Neoantigens are newly formed antigens created by tumor cells. They arise due to tumor-specific alterations, such as genomic mutations, dysregulated RNA splicing, and viral open reading frames. These antigens can trigger an immune response that bypasses central and peripheral tolerance mechanisms. Neoantigens are crucial for personalized cancer immunotherapies and have the potential to induce strong immune responses and reduce the likelihood of targeting normal tissues [16]. Their unique characteristics make them potential candidates for immunotherapeutic techniques including customized cancer vaccines and adoptive T-cell treatments. Synthetic peptides that imitate neoantigens are used in personalized cancer vaccines to train the immune system to recognize and destroy cancer cells. Adoptive T-cell therapies design T cells to target neoantigens, thereby improving cancer control [17]. The prediction of neoantigens proves challenging due to tumor heterogeneity, as each tumor has unique mutations and neoantigens can differ greatly between patients. As a result, it is necessary to implement highly personalized strategies to effectively identify and target neoantigens [18].

One of the challenges experienced by researchers in the field of immunology is accurately recognizing immunogenic neoantigens. These are mutations that can trigger an immune response. However, not all mutations result in neoantigens that can effectively elicit this response. To predict which mutations will produce neoantigens that bind to major histocompatibility complex (MHC) molecules and are recognized by T cells, computational tools are used. The stability and affinity with which neoantigen peptides bind to MHC molecules are the main characteristics that computational methods need to consider. Additionally, computational methods must account for how these complexes are recognized by T-cell receptors [19, 20]. The tumor microenvironment plays a crucial role in the presentation of neoantigens and the recognition of these antigens by the immune system.

Tumors can manipulate the immune response by downregulating antigen presentation machinery and creating an immunosuppressive microenvironment. These mechanisms make it challenging to accurately predict neoantigens and develop effective immunotherapies [21, 22]. Additionally, the field of somatic mutation identification is not without its challenges. Significant challenges are posed by technological restrictions, such as the inability to identify low-frequency mutations and differentiate genuine somatic changes from sequencing artifacts. Furthermore, modern bioinformatics tools and skills are required for the technically challenging integration of multi-omics data to provide a thorough understanding of the neoantigen landscape [23]. However, despite these obstacles, significant progress has been made in developing bioinformatics tools for neoantigen prediction. These tools are continuously improving in accuracy and efficiency, and they play a crucial role in advancing personalized cancer immunotherapy. Since somatic mutations lead to neoantigen formation, the first step in neoantigen prediction is the identification of somatic mutation. Some of the sequencing technologies and bioinformatics tools for predicting somatic mutations are discussed below.

To understand the complex mechanism of human diseases, researchers rely on integrating data from multiple omics techniques, including genomics, transcriptomics, epigenomics, and proteomics to utilize next-generation sequencing (NGS) in analyzing DNA. NGS enables the analysis of DNA through various approaches such as whole-genome sequencing (WGS), whole-exome sequencing [24], and targeted sequencing. This powerful tool allows for the sequencing of millions of DNA fragments simultaneously, providing detailed information about the structure of genomes, genetic variations, gene activity, and alterations in gene behavior [25].

2.1 Whole-genome sequencing (WGS)

Whole-genome sequencing (WGS) is demonstrated to be a powerful technique for determining an individual’s DNA sequence. It lists all genes, regulatory areas, and non-coding elements in an individual's genome. It can be applied in plant and animal studies, cancer research, rare genetic diseases, population genetics, and genome assembly. This technique is highly useful in identifying genomic variations from single-nucleotide polymorphisms (SNPs) to structural changes by sequencing the entire genome [26]. There are two methods of WGS: Large and small, which are used to interpret eukaryotic and prokaryotic genomes, respectively. Short-read sequencing is best for mutation calling, and long-read for genome assembly. These two can be applied to accurate genome assembly without a reference sequence [27].

2.2 Whole-Exome Sequencing (WES)

The whole-exome sequencing (WES) technique is centered on sequencing the exome, or the parts of the genome that code for proteins. Though it is a minor portion of the whole genome, the majority of variations that cause disease are found in the exome. WES is applied to identify genetic variations in protein-coding genes, including single-nucleotide variants, insertions, deletions, and copy number changes. Additionally, it can be applied to population and cancer genetics as well as rare clinical illnesses, where it is a more affordable option than WGS. Whole-exome sequencing enriches exonic regions through hybrid capture or target-specific amplification techniques followed by high-throughput sequencing. The Illumina NGS platform can be used with a variety of exome capture assays [28]. Since WES is a component of WGS, the bioinformatic analysis method utilized for WES data is the same as that used for WGS. Therefore, WGS is a valuable technique for identifying genetic variations in protein-coding regions of the genome, making it particularly useful in disease research and clinical applications.

2.3 Targeted sequencing

Targeted sequencing is a successful approach that focuses on particular sections of the gene, enabling the identification of different types of genetic variants linked to disease phenotypes. Although targeted sequencing may have lower exploratory capabilities than WGS or WES, it offers benefits such as cost-effectiveness and manageable data for medical professionals.

This allows for more precise and well-informed clinical decisions based on disease-specific information [29]. In addition, targeted sequencing can offer enhanced coverage for rare alleles in genetic disorders and low-frequency evolving mutant clones in cancer, enabling a more thorough comprehension of tumor heterogeneity and disease progression. In general, targeted sequencing has the potential to advance genomics research significantly, enhance personalized healthcare, and improve our understanding of diseases. The candidate gene approach and commercially available targeted panels come from large-scale WGS/WES projects. These panels can test both inherited (germline) and acquired (somatic) variants. Some examples are listed in Table 1. Targeted panels use region-specific primers to amplify selected DNA regions. The resulting libraries are then sequenced and analyzed with bioinformatics tools. Overall, targeted sequencing is a valuable method for identifying genetic variants linked to diseases, offering cost-effective and focused insights for clinical applications.

Table 1: Some Examples of Targeted Panels in Research and Diagnostics

Disease Condition	Panel Name	Inheritance Type	Sample Type
Cardiovascular defects	Cardiovascular Panel	Germline	Blood
Arrhythmias and cardiomyopathies	Arrhythmia and Cardiomyopathy Panel	Germline	Blood
Drug sensitivity	Pharmacogenomics Panel	Germline	Blood
Antimicrobial treatment efficacy	Antimicrobial Resistance Panel	Microbial Gene	Bacterial Culture
Infertility	Infertility Panel	Germline	Blood
Homologous recombination defects	HRR Gene Panel	Somatic	Tumor Tissue
Myeloid cancers	Myeloid Cancer Panel	Somatic	Blood
HIV drug resistance	HIV-X Gene Panel	Pathogen	Plasma
Antimicrobial resistance in TB	TB Resistance Panel	Pathogen	TB Specimen
Metabolic disorders	Metabolism Error Panel	Germline	DBS/Blood
Hereditary cancers	BRCA and Hereditary Cancer Panel	Germline	Blood

3. Bioinformatics Pipelines

The rapid advancements in next-generation sequencing (NGS) technologies have significantly aided in the identification of mutations within the exomes of individual tumors. These mutations lead to neoantigens, which can be presented by the patient's Human Leukocyte Antigen [30] molecules to the immune system, which triggers an immune response against the tumor. The identification and selection of these neoantigens have become crucial in developing personalized cancer immunotherapies, particularly in the design of cancer vaccines and adoptive cell therapies. The prediction of neoantigens begins with the identification of somatic mutations, which is the critical first step in neoantigen prediction, as these mutations lead to the formation of neoantigens. The advancement in sequencing technologies and bioinformatics tools have greatly added to the identification of genetic differences and differentiate between somatic mutations, which occur in specific cells, and germline variations, which are inherited and present in all cells of an individual [31].

The process of neoantigen identification involves multiple computational steps, each contributing to the accurate prediction of neoantigens. Initially, HLA typing is performed using RNA-seq, WGS, or WES data to determine the specific HLA alleles of the patient, followed by the prediction of mutated peptides resulting from somatic mutations identified in the tumor. The next step involves the identification of neoantigens that are likely to be presented on the surface of tumor cells, which is achieved by predicting the binding affinity of the peptides to the HLA molecules.

Finally, candidate neoantigens are prioritized based on their predicted binding affinities and other factors, such as proteasomal processing and peptide transport [32].

In recent years, there has been significant growth and improvement of bioinformatics pipelines to improve the process of identifying and selecting neoantigens in a more precise and effective way. These pipelines encompass advanced machine learning algorithms and merge various kinds of omics data, like mass spectrometry and RNA-seq, to enhance the predictive power of neoantigen exploration. The existing bioinformatics pipelines are characterized by the incorporation of four principal computational components: HLA typing, mutation-driven peptide deduction, MHC binding and forecast of antigen presentation, and prioritization of neoantigens. The integration of these distinct modules within the bioinformatics pipelines plays a crucial role in advancing the field of neoantigen prediction, thus facilitating the identification of potential targets for immunotherapy [33].

In this section, we discuss bioinformatics tools and pipelines, developed to address neoantigen identification. These tools not only support the identification of potential neoantigens but also facilitate their selection for therapeutic applications, paving the way for personalized cancer treatments. A summary of these pipelines, including their strengths, limitations, and key references, is provided in Table 1.

The Sequence Alignment/Map (SAM)-SAMtools is a suite of utilities for manipulating alignments in the SAM (Sequence Alignment/Map) format, including sorting, merging, indexing, and generating variant calls. It is used for processing next-generation sequencing data. The ‘mpileup’ function in SAMtools is used to call variants, including somatic mutations, by generating a pileup format from BAM files. It is widely adopted, simple to use, and highly efficient for basic operations on sequencing data. However, it has limitations in handling complex variant calling scenarios and lacks advanced algorithms for distinguishing somatic from germline mutations [34]. Another highly effective tool, VarScan2 that detects single nucleotide variants (SNVs) using SAMtools ‘mpileup’ data. It compares tumor and normal sample data to identify germline and somatic mutations. Additionally, it detects copy number analysis and structural variants. One of the main strengths of VarScan2 is its ability to accurately call low-frequency variants. Nevertheless, this tool is constrained by accurate pileup generation and may be less effective for highly heterogeneous samples [35]. Another robust toolkit developed by the Broad Institute for identifying genetic variations in large-scale sequencing data is GATK (Genome Analysis software). It offers a variety of tools for data pre-processing, variant calling, and variant filtering. The HaplotypeCaller tool from GATK is widely recognized for its ability to call germline variants, while MuTect2 accurately recognizes somatic variants. GATK is extensively documented and is supported by the community. Moreover, it is adept at managing intricate genomic regions. Nevertheless, big datasets, need a substantial resource [36]. MuTect is a tool within the GATK suite specifically designed for identifying somatic point mutations in tumor samples. The tool employs a Bayesian classifier to effectively distinguish between somatic mutations and sequencing artifacts or germline variants by using matched normal samples. MuTect possesses high sensitivity and specificity, rendering it highly proficient in the detection of infrequent somatic mutations. However, the tool is focused on point mutations, and minor insertions/deletions, and is less efficient for major structural variations [37]. Strelka is a specialized tool used for identifying single nucleotide variations (SNVs) and small insertions or deletions (indels) in tumor-normal pairs. It utilizes a Bayesian framework to precisely identify genetic variations and is capable of analyzing data from both WGS and WES. Strelka is highly sensitive and accurately identifies low-frequency somatic mutations and small indels. Moreover, it has high computational efficiency.

However, it only emphasizes detecting minor genetic differences and not on larger structural changes [38]. FreeBayes is a variant detector that can identify SNPs, indels, MNVs, and complex events in both diploid and polyploid genomes. It is suitable for both germline and somatic variant calling. One of its strengths is its ability to handle complex variants and mixed ploidy populations. However, it can be computationally intensive and requires high-quality input data [39]. Platypus is a variant caller that detects SNVs and indels from NGS data using local realignment and assembly. It is faster than other tools and generates calls from raw aligned read data without preprocessing. It offers high accuracy and is effective in identifying complex variants. However, it is computationally intensive and requires significant memory resources [40]. Lancet is another pipeline for the somatic variant caller that employs localized micro-assembly to identify SNVs and indels in tumor-normal pairs. It is particularly effective in challenging genomic regions due to its high sensitivity and specificity; however, it is computationally intensive and requires high-quality input data [41].

Another pipeline developed by Google Health is DeepVariant, which is a deep learning-based variant caller. It uses a convolutional neural network (CNN) to call variants from NGS data, treating the variant calling process as an image classification problem. It offers high accuracy and robustness and is capable of handling complex variants and various sequencing technologies. However, requires significant computational resources, particularly GPU power for training and inference [42]. SomaticSniper is a variant caller that detects somatic mutations by comparing tumor and normal samples. It differentiates somatic mutations from germline variants by identifying differences in base calls. It is simple to use and offers effective analyses for paired tumor-normal samples, however, it lacks sensitivity for low-frequency variants [43]. LoFreq is a variant caller that uses a Poisson-based model to detect low-frequency variants in high-throughput sequencing data. It is highly sensitive and suitable for analyzing heterogeneous samples, but may require significant computational resources for large datasets [44]. JBrowse is a genome browser that allows one to visualize and explore genomic data by integrating with multiple variant calling methods. It displays somatic mutations and other genomic variations. Its advantages include a user-friendly interface and compatibility with other bioinformatics tools. As it is essentially a visualization tool, it must be integrated with other pipelines to do variant calling [45]. Germline is a software application that uses a probabilistic model to detect both somatic and germline mutations in whole-genome sequencing data. It is well-suited for in-depth genomic investigations, while conducting whole-genome analyses may necessitate substantial computational resources [46]. HaplotypeCaller is a GATK suite tool for calling variants by building haplotypes in specific genomic regions. It is effective for both germline and somatic variations and provides high precision via local haplotype assembling. However, it is computationally demanding and may necessitate substantial resources for huge datasets [47].

Another robust pipeline is Pisces, which is a precise somatic variant caller optimized for Illumina sequencing data. It offers high specificity and sensitivity for low-frequency mutations. Its strengths include high accuracy and sensitivity; however, it may require significant computational resources [48]. A pipeline, Sentieon TNscope, is a high-performance variant caller that offers accurate and fast identification of somatic mutations. Its performance is better with faster runtimes and supports large-scale genomic studies. However, it requires licensing as commercial software [49]. A micro-assembly-based variant caller for indels, developed in C/C++, is the Scalpel pipeline. It is high in accuracy in calling indels from NGS data. And also offers a pipeline integration, simplifying workflows for researchers working with NGS data. However, it may require more computational resources, potentially slowing down analysis, especially for large datasets [50].

For identifying somatic variants, a deep convolutional neural network-based pipeline, NeuSomatic, was developed to accurately identify somatic mutations from high-throughput sequencing data. It offers high accuracy and robustness; however, it requires significant computational resources, particularly GPU power for training and inference [51].

Another variant caller that uses local haplotype assembly and Bayesian statistical models is Octopus. It is highly accurate in germline and somatic variant calling. However, it is computationally exhaustive and may necessitate substantial resources for large datasets [52]. Strelka2 is an updated version of the Strelka variant caller, offering somatic and germline variant calling in both WGS and WES. It provides high sensitivity and specificity and improves computational efficiency but may not capture larger structural variations [53]. SomaticSignatures is an R package that offers tools for analyzing and visualizing mutational signatures in somatic mutations. It is not a variant caller but can integrate results from other variant callers for comprehensive analysis. The package is user-friendly and designed for data exploration and hypothesis testing within the R environment. However, it requires input from other variant calling tools and is R dependent, which may be a limitation for users not accustomed to the programming language [54]. Maftools is an R package designed for analyzing and visualizing somatic variants in cancer studies. It offers comprehensive visualization, user-friendly interface, and integration capabilities for data from various sources. However, it is not a variant caller and requires variant data from other pipelines. Additionally, it is R dependent, requiring familiarity with R, which may be a limitation for some researchers. Overall, Maftools provides a comprehensive view of cancer genomics [55]. DeconstructSigs is a R package that quantifies the contribution of known mutational processes in cancer genomes. It is user-friendly, offering comprehensive documentation and examples. DeconstructSigs can integrate output from various variant callers, making it versatile for different input data types.

However, it does not perform variant calling and relies on variant data generated by other tools. Additionally, users need to be familiar with R to effectively use DeconstructSigs, which may be a barrier for those not accustomed to working with R [56]. Finally, understanding cancer genomes and personalized treatment relies heavily on identifying somatic mutations. The wide array of bioinformatics pipelines examined, each possessing distinct advantages and drawbacks, underscores the intricate and meticulous nature necessary for the precise identification of mutations. Based on the characteristics of their datasets and their specific requirements, researchers can choose the most suitable tools. The comprehensive explanations and relative advantages and disadvantages of various pipelines are summarized in Table 2, offering a great reference for choosing the most appropriate pipeline for somatic mutation analysis.

Table 2: Bioinformatics Pipelines for Somatic Mutation Identification

Pipeline	Year	Language	Description	Strengths	Limitations	Class	Reference
SAMtools	2009	C	Utilities for manipulating alignments and calling variants.	Simple, widely adopted, efficient for basic operations.	Limited handling of complex variants, no distinction between somatic and germline.	CA	[34]
VarScan2	2009	Java	Detects SNVs and indels, suitable for tumor-normal comparisons.	Effective for low-frequency variants, includes copy number analysis.	Less effective for highly heterogeneous samples.	CA	[35]
GATK	2010	Java	Comprehensive toolkit for variant discovery and genotyping.	High accuracy, robust, extensive documentation.	Computationally intensive.	CA	[36]
MuTect	2013	Java	Identifies somatic point mutations in tumor samples.	High sensitivity and specificity for low-frequency mutations.	Focused on point mutations, less effective for structural variants.	CA	[37]
Strelka	2012	C++	Detects SNVs and indels in tumor-normal pairs.	Highly sensitive to low-frequency variants and small indels.	Limited to small variants.	CA	[38]

FreeBayes	2012	C++	Haplotype-based variant detection for diploid and polyploid genomes.	Handles complex variants, flexible.	Computationally intensive.	CA	[39]
VarDict	2014	Java, Perl	Detects SNVs, indels, and structural variants in both germline and somatic.	High sensitivity and specificity, effective for complex indels.	Requires parameter tuning.	CA	Lai et al. (2016)
Platypus	2014	Python, C	Detects SNVs and indels using local realignment and assembly.	High accuracy for complex variants.	Computationally intensive.	CA	[40]
Lancet	2017	C++	Somatic variant caller using localized micro-assembly to identify SNVs and indels in tumor-normal pairs.	High sensitivity and specificity, effective in challenging genomic regions.	Computationally intensive, requires high-quality input data.	CA	[41]
DeepVariant	2018	Python, C++	Deep learning-based variant caller using a convolutional neural network to call variants	High accuracy and robustness, effective for complex variants across various sequencing technologies.	Requires significant computational resources, particularly GPU power for training	DL	[42]

			from NGS data.		and inference.		
SomaticSniper	2011	C	Detects somatic mutations by comparing tumor and normal samples.	Effective for paired samples.	Limited to SNV calling.	CA	[43]
LoFreq	2012	C	Uses a Poisson-based model to call low-frequency variants.	High sensitivity for low-frequency variants.	Requires significant computational resources.	SM	[44]
JBrowse	2009	Java Script, Perl	Genome browser for visualizing and exploring genomic data, integrates with various variant calling tools.	User-friendly interface, integrates with other bioinformatics tools for comprehensive visualization.	Primarily a visualization tool, requires integration with variant calling pipelines for analysis.	CA	[45]
Germline	2013	Python	Tool for identifying both somatic and germline mutations in WGS data.	Capable of identifying both somatic and germline mutations, suitable for comprehensive genomic studies.	May require significant computational resources for WGS data.	CA	[46]
HaplotypeCaller	2013	Java	Part of GATK suite, builds haplotypes in regions and calls variants.	High accuracy due to local haplotype assembly, effective for complex	Computationally intensive, requires significant resources	CA	[47]

				genomic regions.	for large datasets.		
Pisces	2015	C#	Detects somatic variants with high specificity and sensitivity.	Optimized for Illumina data.	Limited to Illumina data.	CA	[48]
Sentieon TNscope	2018	C/C++	High-performance variant caller with fast runtimes.	High accuracy, faster than GATK.	Commercial software.	CA	[49]
Scalpel	2014	C/C++, Python	Detects indels using micro-assembly.	High accuracy for indels.	Computationally intensive.	CA	[50]
NeuSomatic	2019	Python	Deep learning-based somatic variant caller.	High accuracy due to deep learning.	Requires significant computational resources.	DL	[51]
Octopus	2019	C++	Uses local haplotype assembly and Bayesian models for variant calling.	High accuracy for both germline and somatic variants.	Computationally intensive.	BM	[52]
Strelka2	2019	C++	Improved version of Strelka for more accurate and efficient variant calling.	High sensitivity and specificity, computationally efficient.	Focused on SNVs and small indels.	CA	[53]
SomaticSignatures	2015	R	Analyzes and	Effective for mutational	Not a variant	SM	[54]

			visualizes mutational signatures in somatic mutations.	signature analysis.	caller, requires input from other tools.		
Maftools	2018	R	Analyzes and visualizes somatic variants in cancer studies.	Comprehensive visualization and analysis tools.	Not a variant caller, requires input from other tools.	SM	[55]
DeconstructSigs	2016	R	Quantifies the contribution of known mutational processes in cancer genomes.	Effective for identifying mutational processes.	Not a variant caller, requires input from other tools.	SM	[56]

Abbreviations: *Classical Algorithms: CA; Deep Learning: DL; Bayesian Methods: BM; Statistical Methods: SM*

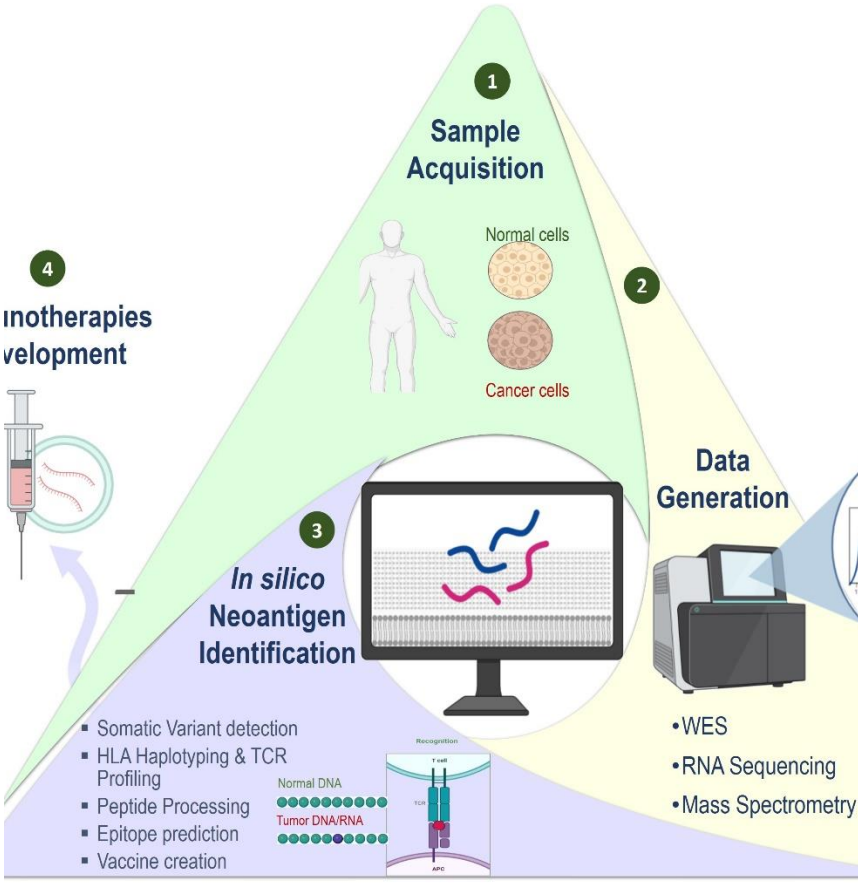
4. Neoantigen Prediction and Bioinformatics Methods and Tools

Neoantigen prediction includes several steps: identifying somatic mutations, predicting peptide sequences, evaluating binding affinities to MHC molecules, and assessing the immunogenicity of the neoantigens. Certain notable case studies of successful neoantigen predictions in cancer treatment are as follows:

A case study by Sahin et al. demonstrated the development of personalized neoantigen vaccines for melanoma patients. The study applied an RNA-based poly-neo-epitope approach that included whole-exome sequencing and RNA sequencing to identify tumor-specific mutations by predicting the binding of the resulting peptides to MHC molecules using the NetMHCpan tool to mobilize immunity against a spectrum of cancer mutations. All patients presented T-cell responses against multiple new epitopes from the vaccine. The personalized vaccines thus developed when administered to the patients led to a significant reduction in the rate of metastatic events, resulting in sustained progression-free survival [57]. This case is a pioneering example of how *in silico* neoantigen prediction can be directly translated into a therapeutic vaccine.

In another study, neoantigens were predicted as personalized immunotherapy for glioblastoma patients. The neoantigens were identified by sequencing the tumor's genome and conducting *in silico* predictions such as Broad Picard Pipeline and NetMHCpan tool [58].

In another study, the team of researchers addressed the shortcomings and constraints associated with non-chemotherapy treatment modalities for non-small cell lung cancer (NSCLC) by introducing a novel approach involving neoantigen vaccines. These vaccines were engineered based on unique and individualized tumor DNA mutations, thereby modulating immune response mechanisms to effectively and precisely target the malignant cells responsible for the disease. Potential neoantigens from lung cancer tumors were predicted and the binding affinity between these mutations and MHC class I molecules (specifically the H-2 Kb allele in LLC cells and C57BL/6 mice) was predicted using NetMHCpan, NetMHC, NetMHCcons, Pick Pocket, MHCflurry, SMM, SMMPBMC and MHCnug getsI [59]. The comprehensive analysis of some of these case studies serves to effectively illustrate the successful implementation of *in silico* techniques to successfully identify neoantigens that can be employed in personalized therapeutic regimens to treat a diverse array of cancer types, Figure 1. Neoantigen prediction and bioinformatics tools have shown promising results in developing personalized cancer therapies through successful case studies.



1: Representation of application of bioinformatics methods in neoantigen predi
› immunotherapies.

5. The Impact of Artificial Intelligence and Machine Learning on Immuno-Oncology

The introduction of innovative technological developments, notably those related to artificial intelligence (AI), machine learning (ML), and quantum computing, creates a powerful opportunity for the progressive refinement of immuno-oncology, a sector dedicated to using the immune system in the fight against cancer. The methodologies driven by AI can analyze extensive datasets with a level of efficiency that far surpasses traditional analytical techniques, thereby unveiling new and transformative insights into the complex biology of tumors and the multifaceted immune responses they elicit. In the following section, the role of different types of technologies in neoantigen prediction and vaccine development is discussed:

5.1 Application of Artificial Intelligence (AI) in Neoantigen Prediction

Recent advancements in technology have significantly improved the prediction of neoantigens by enabling the analysis of large genomic and proteomic datasets. Specifically, deep learning algorithms have shown promise in predicting which mutated peptides are likely to bind to a patient's MHC molecules, which is essential for assessing the immunogenic potential of neoantigens.. Machine learning techniques, as a branch of artificial intelligence, excel at estimating the binding affinities between peptides and Major Histocompatibility Complex molecules, which is essential for determining the potential for an immune response from a neoantigen. Traditional methodologies may lack the feasibility to accurately predict immunogenic peptides as effectively as AI and ML; for instance, AI-enhanced instruments like NetMHCpan leverage deep learning methodologies to refine peptide-MHC binding predictions by integrating data from a broad spectrum of HLA alleles, including those with minimal experimental binding evidence.

This skill is essential for crafting personalized cancer vaccines since it facilitates the choice of the most viable neoantigens to hone in on the individual tumor makeup of the patient. Additionally, AI frameworks are capable of modeling and optimizing the immune reaction, potentially discovering the ideal set of neoantigens to be included in a vaccine, thereby improving its overall efficacy [60, 61]. AI significantly enhances the development of personalized cancer vaccines by improving neoantigen prediction and optimizing immune responses.

5.2 Application of Quantum Computing in Neoantigen Prediction

Quantum computing adds to the advancement in computational capabilities, by potential to transform neoantigen prediction and vaccine formulation. In contrast to classical computing systems, which utilize bits for information processing, quantum computing uses quantum bits, or qubits, enabling multi-fold execution of intricate calculations [62]. This capability is especially valuable in analyzing the vast multi-dimensional datasets required for accurate neoantigen prediction.. Quantum algorithms improve the efficacy of bioinformatics workflows by accelerating the detection of candidate neoantigens and enhancing the precision of predicting MHC-binding. This advancement not only expedites the formulation of personalized cancer vaccines but also permits more advanced modeling of the immune response. Additionally, combining quantum computing with artificial intelligence enables the creation of superior predictive models, which can provide detailed dynamics between neoplastic cells and the immune system. This integration ultimately forges a path for more potent immunotherapies [63]. Quantum computing significantly enhances neoantigen prediction and vaccine formulation in immunotherapy, paving the way for more effective treatments.

5.3 Application of Deep Learning in Neoantigen Prediction

Deep learning (DL), a subset of AI, has become a transformative tool in neoantigen prediction and cancer vaccine development. DL models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), are designed to automatically learn patterns from large datasets. These models can also learn representations from data. These models are implemented in image recognition, natural language processing, and bioinformatics [64]. Since deep learning models have the capability to analyze complex biological patterns, therefore, in healthcare, it is applied to make algorithms for disease diagnosis and personalized treatment [65]. Furthermore, deep learning is revolutionizing immunotherapy by enhancing neoantigen prediction and facilitating the development of cancer vaccines through advanced pattern recognition in biological data.

5.4 Application of Natural Language Processing (NLP) in Neoantigen Prediction

Natural Language Processing (NLP) plays a crucial role in managing and analyzing the extensive unstructured biomedical text data present in immuno-oncology research. It is particularly effective in extracting and processing information pertinent to neoantigen prediction from various sources, including scientific literature, clinical trial reports, and genomic databases.These core processes include: Information Retrieval (IR), which identify and retrieve relevant documents from large datasets or databases in response to specific queries; Semantics and Information Extraction, which accurately interprets the text. NLP systems are designed to perform semantic analysis, which involves recognizing the relationships between words, their definitions, and their syntactic roles within a sentence. This semantic understanding is crucial for tasks such as information extraction, where the goal is to identify specific entities (e.g., genes, proteins, mutations) and their interactions within a text; Information Extraction, which involves identifying and categorizing specific pieces of information from unstructured text. In the context of immuno-oncology, IE is used to automatically extract data about potential neoantigens, patient-specific mutations, and immune response markers from clinical reports and research articles. This process is essential for building comprehensive databases of neoantigens, which can be integrated into bioinformatics pipelines for personalized vaccine development [66].

NLP techniques are used to extract relevant information from scientific literature, clinical trial reports, and patient records, which can then be integrated into bioinformatics pipelines for neoantigen prediction [67]. For instance, NLP can be used to mine databases of scientific publications to search studies that report on neoantigen discovery and validation, thus accelerating the research process by quickly bringing relevant findings to the researchers [68]. Additionally, NLP algorithms can assist in the annotation of genetic sequences by identifying and categorizing mutations that may produce neoantigens. This automated processing of text data not only speeds up the research process but also ensures that no critical information is overlooked, thereby enhancing the accuracy of neoantigen predictions and the subsequent development of personalized cancer vaccines. Also, NLP significantly enhances immunotherapy research by efficiently processing unstructured text data to extract vital information for neoantigen development and personalized vaccine creation.

6. Applications of Neoantigen Prediction in Personalized Cancer Immunotherapy

In silico neoantigen prediction plays a crucial role in the development of personalized cancer vaccines, such as NeoVax for melanoma. By employing advanced bioinformatics tools, researchers can analyze whole-exome sequencing data to identify somatic mutations unique to an individual's tumor. These mutations generate neoantigens—tumor-specific peptides that the immune system recognizes as foreign. Utilizing algorithms like NetMHCpan, scientists predict which neoantigens are most likely to bind to the patient's MHC molecules, thereby facilitating a targeted immune response. The synthesized neoantigens are then formulated into a personalized peptide-based vaccine, which has shown promising results in clinical trials, demonstrating safety and the ability to elicit strong T-cell responses. The success of NeoVax underscores the transformative potential of *in silico* neoantigen prediction in tailoring vaccines that effectively target the distinct mutational profiles of individual tumors, particularly in cases with high mutational burdens [69, 70]. Another personalized vaccine based on in silico neoantigen is exemplified by platforms like iNeST (Individualized Neoantigen-Specific Immunotherapy) developed by BioNTech and Genentech [71]. This innovative approach utilizes whole-exome sequencing and RNA sequencing to identify somatic mutations in a patient's tumor, followed by bioinformatics tools that predict immunogenic neoantigens based on their binding potential to the patient's MHC molecules. Unlike traditional peptide-based vaccines, iNeST employs an mRNA delivery system that encodes these predicted neoantigens, allowing for the direct translation into neoantigenic proteins that elicit a robust immune response. Clinical trials, such as the Phase I study on advanced melanoma patients, have demonstrated the efficacy of this method, showing significant T-cell responses and tumor shrinkage. The advantages of mRNA technology, including the ability to encode multiple neoantigens and rapid production, underscore the transformative potential of in silico neoantigen prediction in advancing personalized cancer immunotherapy [72]. Another example is, TG4050, created by Transgene. This innovative vaccine leverages whole-exome sequencing and RNA sequencing to identify somatic mutations in tumors, followed by a bioinformatics pipeline that predicts the most immunogenic neoantigens. By utilizing these predicted neoantigens, TG4050 employs a viral vector-based approach to enhance the immune response, effectively presenting these neoantigens to the immune system. Early clinical trials have demonstrated TG4050's ability to elicit strong T-cell responses, indicating its potential effectiveness in treating various solid tumors, particularly those with lower mutational burdens. This highlights the significance of in silico prediction in tailoring vaccines to individual patients, ultimately advancing personalized cancer immunotherapy [73]. Additionally, In silico neoantigen prediction plays a crucial role in the development of neoantigen-based therapies, particularly in the engineering of chimeric antigen receptor (CAR)-T cells. By utilizing WES and advanced computational algorithms, researchers can identify patient-specific neoantigens that are unique to individual tumors. This personalized approach allows for the precise engineering of CARs that specifically target these neoantigens, leading to the expansion of T cells *ex vivo* before re-infusion into the patient. The effectiveness of this strategy has been demonstrated in clinical studies, such as the work by Tran et al. (2016), which showcased significant tumor reduction in patients with epithelial cancer.

Ultimately, in silico neoantigen prediction enhances the specificity and efficacy of CAR-T therapies, minimizing off-target effects and providing a promising avenue for treating solid tumors with tailored immunotherapeutic options [74]. Therefore, neoantigen predictions have transformed immuno-oncology with personalized cancer vaccines and therapies like NeoVax and iNeST.

These bioinformatics-driven therapies have enabled precision medicine, where tumors are treated according to their genetics. As bioinformatics tools improve and new technologies like AI and quantum computing are integrated into research, neoantigen prediction will improve, leading to more effective and personalized cancer treatments.

7. Challenges and Future Perspectives

Bioinformatics in immuno-oncology encounters various technical obstacles due to the intricate nature and volume of the data involved. A primary challenge is the precise identification and prediction of neoantigens. Regardless of the advancements in sequencing technology and computational techniques, the precision of predictions remains limited, particularly regarding MHC class II-restricted epitopes, which demonstrate increased variability and longer peptide lengths. Additionally, there are technical constraints concerning the sensitivity and specificity of algorithms utilized for HLA typing, mutation detection, and neoantigen prediction. These tools frequently yield inconsistent results depending on the quality of input data and the specific algorithms applied, resulting in discrepancies across different studies. Another hurdle faced is the tremendous data output resulting from next-generation sequencing (NGS) techniques. The processing and examination of these extensive datasets necessitate considerable computational power, including high-performance computing (HPC) systems. Therefore, there is a pressing need for more accessible tools that researchers can utilize without requiring specialized bioinformatics expertise, as current tools often demand substantial knowledge in computational biology. The future of bioinformatics in immuno-oncology is dependent on the ongoing improvement of computational approaches. ML and DL algorithms can be implemented with further advancement to boost the precision of neoantigen prediction and HLA typing, since these algorithms are capable of learning from extensive datasets, allowing them to analyze intricate patterns that cannot be identified by the conventional methods. When spatial transcriptomics is combined with single-cell RNA sequencing (scRNA-seq), it enhances the ability to point to the location of neoantigens in the tumor microenvironment. Such an analysis offers valuable insights into the spatial dynamics governing immune responses. Further, the construction of hybrid models can involve the synthesis of data-informed techniques alongside mechanistic models related to immune responses. These hybrid frameworks will synthesize multi-omics data to simulate the interactions occurring between tumors and the immune system. This methodology will enable researchers to predict the outcomes associated with various immunotherapeutic strategies. Personalized immunotherapy denotes the significant advancement in cancer treatment, wherein therapies are customized to the unique genetic and immunological characteristics of each individual. The comprehension of tumor immunology and growth and advancement in bioinformatics tools for neoantigen prediction is aligned with the design of personalized vaccines and adoptive cell therapies. Moreover, the amalgamation of multi-omics data and real-time monitoring technologies will significantly improve the identification of biomarkers and therapeutic adjustments, thereby optimizing treatment efficacy.

8. Conclusion

In conclusion, bioinformatics approaches have gained paramount importance in the field of immuno-oncology, driving significant advancements in personalized cancer treatment. AI, ML, and quantum computing have enabled the development of more effective immunotherapies, such as personalized vaccines and adoptive T-cell therapies.

These technologies have not only improved the development of bioinformatics pipelines to accurately predict the neoantigen prediction but also facilitated the analysis of complex biological data, leading to a deeper understanding of the tumor microenvironment. Neoantigen prediction plays a crucial role in personalized immunotherapy, as demonstrated by therapies like NeoVax and neoantigen-targeted CAR-T cells. These advancements showcase the effectiveness of tailored treatments that target individual tumor mutations, leading to better clinical outcomes and extended survival for patients. However, the field still faces challenges, including the need for more sensitive and specific algorithms and the computational power required to process plethora of heterogenous datasets. Future research will likely focus on overcoming these challenges by refining computational approaches, integrating multi-omics data, and developing more accessible tools for researchers. The ongoing evolution of bioinformatics in immuno-oncology holds great potential for enhancing the efficacy of cancer treatments, ultimately leading to better patient outcomes.

Statements and Declarations

Conflict of Interests

The author shares no conflict of interests

Funding sources

Not available

Acknowledgment

The author acknowledges the support of the Department of Biology, Islamic University of Madinah, KSA.

REFERENCES

[1] Podlaha, O.; M. Riester; S. De, and F. Michor, Evolution of the cancer genome. Trends Genet, 2012. 28(4): p. 155-63.

[2] Hosseini, H.; M.M.S. Obradovic; M. Hoffmann; K.L. Harper; M.S. Sosa; M. Werner-Klein; L.K. Nanduri; C. Werno; C. Ehrl; M. Maneck; N. Patwary; G. Haunschild; M. Guzvic; C. Reimelt; M. Grauvogl; N. Eichner; F. Weber; A.D. Hartkopf; F.A. Taran; S.Y. Brucker; T. Fehm; B. Rack; S. Buchholz; R. Spang; G. Meister; J.A. Aguirre-Ghiso, and C.A. Klein, Early dissemination seeds metastasis in breast cancer. Nature, 2016. 540(7634): p. 552-558.

[3] Coussens, L.M. and Z. Werb, Inflammation and cancer. Nature, 2002. 420(6917): p. 860-7.

[4] Hanahan, D. and R.A. Weinberg, Hallmarks of cancer: the next generation. Cell, 2011. 144(5): p. 646-74.

[5] Gonzalez, H.; C. Hagerling, and Z. Werb, Roles of the immune system in cancer: from tumor initiation to metastatic progression. Genes Dev, 2018. 32(19-20): p. 1267-1284.

[6] Wei, S.C.; C.R. Duffy, and J.P. Allison, Fundamental Mechanisms of Immune Checkpoint Blockade Therapy. Cancer Discov, 2018. 8(9): p. 1069-1086.

[7] Zhang, Y. and Z. Zhang, The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. Cellular & Molecular Immunology, 2020. 17(8): p. 807-821.

[8] Matsushita, H.; M.D. Vesely; D.C. Koboldt; C.G. Rickert; R. Uppaluri; V.J. Magrini; C.D. Arthur; J.M. White; Y.S. Chen; L.K. Shea; J. Hundal; M.C. Wendl; R. Demeter; T. Wylie; J.P. Allison; M.J. Smyth; L.J. Old; E.R. Mardis, and R.D. Schreiber, Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. Nature, 2012. 482(7385): p. 400-4.

[9] McGranahan, N.; A.J. Furness; R. Rosenthal; S. Ramskov; R. Lyngaa; S.K. Saini; M. Jamal-Hanjani; G.A. Wilson; N.J. Birkbak; C.T. Hiley; T.B. Watkins; S. Shafi; N. Murugaesu; R. Mitter; A.U. Akarca; J. Linares; T. Marafioti; J.Y. Henry; E.M. Van Allen; D. Miao; B. Schilling; D. Schadendorf; L.A. Garraway; V. Makarov; N.A. Rizvi; A. Snyder; M.D. Hellmann; T. Merghoub; J.D. Wolchok; S.A. Shukla; C.J. Wu; K.S. Peggs; T.A. Chan; S.R. Hadrup; S.A. Quezada, and C. Swanton, Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science, 2016. 351(6280): p. 1463-9.

[10] Richters, M.M.; H. Xia; K.M. Campbell; W.E. Gillanders; O.L. Griffith, and M. Griffith, Best practices for bioinformatic characterization of neoantigens for clinical utility. Genome Med, 2019. 11(1): p. 56.

[11] Athieniti, E. and G.M. Spyrou, A guide to multi-omics data collection and integration for translational medicine. Computational and Structural Biotechnology Journal, 2023. 21: p. 134-149.

[12] Jovic, D.; X. Liang; H. Zeng; L. Lin; F. Xu, and Y. Luo, Single-cell RNA sequencing technologies and applications: A brief overview. Clin Transl Med, 2022. 12(3): p. e694.

[13] Wang, Y.; B. Liu; G. Zhao; Y. Lee; A. Buzdin; X. Mu; J. Zhao; H. Chen, and X. Li, Spatial transcriptomics: Technologies, applications and experimental considerations. Genomics, 2023. 115(5): p. 110671.

[14] Chen, B.; M.S. Khodadoust; C.L. Liu; A.M. Newman, and A.A. Alizadeh, Profiling Tumor Infiltrating Immune Cells with CIBERSORT. Methods Mol Biol, 2018. 1711: p. 243-259.

[15] Xie, N.; G. Shen; W. Gao; Z. Huang; C. Huang, and L. Fu, Neoantigens: promising targets for cancer therapy. Signal Transduction and Targeted Therapy, 2023. 8(1): p. 9.

[16] Zhang, T.; E. Kurban, and Z. Wang, Neoantigens: The Novel Precision Cancer Immunotherapy. Biologics, 2023. 3(4): p. 321-334.

[17] Pearlman, A.H.; M.S. Hwang; M.F. Konig; E.H.-C. Hsiue; J. Douglass; S.R. DiNapoli; B.J. Mog; C. Bettegowda; D.M. Pardoll; S.B. Gabelli; N. Papadopoulos; K.W. Kinzler; B. Vogelstein, and S. Zhou, Targeting public neoantigens for cancer immunotherapy. Nature Cancer, 2021. 2(5): p. 487-497.

[18] Alexandrov, L.B.; S. Nik-Zainal; D.C. Wedge; S.A. Aparicio; S. Behjati; A.V. Biankin; G.R. Bignell; N. Bolli; A. Borg, and A.-L. Børresen-Dale, Signatures of mutational processes in human cancer. nature, 2013. 500(7463): p. 415-421.

[19] Yadav, M.; S. Jhunjhunwala; Q.T. Phung; P. Lupardus; J. Tanguay; S. Bumbaca; C. Franci; T.K. Cheung; J. Fritsche; T. Weinschenk; Z. Modrusan; I. Mellman; J.R. Lill, and L. Delamarre, Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. Nature, 2014. 515(7528): p. 572-576.

[20] Hundal, J.; S. Kiwala; J. McMichael; C.A. Miller; H. Xia; A.T. Wollam; C.J. Liu; S. Zhao; Y.Y. Feng; A.P. Graubert; A.Z. Wollam; J. Neichin; M. Neveau; J. Walker; W.E. Gillanders; E.R. Mardis; O.L. Griffith, and M. Griffith, pVACtools: A Computational Toolkit to Identify and Visualize Cancer Neoantigens. Cancer Immunol Res, 2020. 8(3): p. 409-420.

[21] Spranger, S.; D. Dai; B. Horton, and T.F. Gajewski, Tumor-Residing Batf3 Dendritic Cells Are Required for Effector T Cell Trafficking and Adoptive T Cell Therapy. Cancer Cell, 2017. 31(5): p. 711-723 e4.

[22] Jardim, D.L.; A. Goodman; D. de Melo Gagliato, and R. Kurzrock, The Challenges of Tumor Mutational Burden as an Immunotherapy Biomarker. Cancer Cell, 2021. 39(2): p. 154-173.

[23] Pereira, R.; J. Oliveira, and M. Sousa, Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. J Clin Med, 2020. 9(1).
[24] Burley, S.K.; C. Bhikadiya; C. Bi; S. Bittrich; L. Chen; G.V. Crichlow; C.H. Christie; K. Dalenberg; L. Di Costanzo; J.M. Duarte; S. Dutta; Z. Feng; S. Ganesan; D.S. Goodsell; S. Ghosh; R.K. Green; V. Guranović; D. Guzenko; B.P. Hudson; Catherine L. Lawson; Y. Liang; R. Lowe; H. Namkoong; E. Peisach; I. Persikova; C. Randle; A. Rose; Y. Rose; A. Sali; J. Segura; M. Sekharan; C. Shao; Y.-P. Tao; M. Voigt; John D. Westbrook; J.Y. Young; C. Zardecki, and M. Zhuravleva, RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. Nucleic Acids Research, 2020. 49(D1): p. D437-D451.
[25] Satam, H.; K. Joshi; U. Mangrolia; S. Waghoo; G. Zaidi; S. Rawool; R.P. Thakare; S. Banday; A.K. Mishra; G. Das, and S.K. Malonia, Next-Generation Sequencing Technology: Current Trends and Advancements. Biology (Basel), 2023. 12(7).
[26] Costain, G.; R.D. Cohn; S.W. Scherer, and C.R. Marshall, Genome sequencing as a diagnostic test. CMAJ, 2021. 193(42): p. E1626-E1629.
[27] Logsdon, G.A.; M.R. Vollger, and E.E. Eichler, Long-read human genome sequencing and its applications. Nat Rev Genet, 2020. 21(10): p. 597-614.
[28] Warr, A.; C. Robert; D. Hume; A. Archibald; N. Deeb, and M. Watson, Exome Sequencing: Current and Future Perspectives. G3 (Bethesda), 2015. 5(8): p. 1543-50.
[29] Williams, M.J.; A. Sottoriva, and T.A. Graham, Measuring Clonal Evolution in Cancer with Genomics. Annu Rev Genomics Hum Genet, 2019. 20: p. 309-329.
[30] Li, H.; A. Coghlan; J. Ruan; L.J. Coin; J.-K. Heriche; L. Osmotherly; R. Li; T. Liu; Z. Zhang, and L. Bolund, TreeFam: a curated database of phylogenetic trees of animal gene families. Nucleic acids research, 2006. 34(suppl_1): p. D572-D580.
[31] Robbins, P.F.; Y.-C. Lu; M. El-Gamil; Y.F. Li; C. Gross; J. Gartner; J.C. Lin; J.K. Teer; P. Cliften; E. Tycksen; Y. Samuels, and S.A. Rosenberg, Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. Nature Medicine, 2013. 19(6): p. 747-752.
[32] De Mattos-Arruda, L.; M. Vazquez; F. Finotello; R. Lepore; E. Porta; J. Hundal; P. Amengual-Rigo; C.K.Y. Ng; A. Valencia; J. Carrillo; T.A. Chan; V. Guallar; N. McGranahan; J. Blanco, and M. Griffith, Neoantigen prediction and computational perspectives towards clinical benefit: recommendations from the ESMO Precision Medicine Working Group. Annals of Oncology, 2020. 31(8): p. 978-990.
[33] Hackl, H.; P. Charoentong; F. Finotello, and Z. Trajanoski, Computational genomics tools for dissecting tumour–immune cell interactions. Nature Reviews Genetics, 2016. 17(8): p. 441-458.
[34] Li, H.; B. Handsaker; A. Wysoker; T. Fennell; J. Ruan; N. Homer; G. Marth; G. Abecasis; R. Durbin, and S. Genome Project Data Processing, The Sequence Alignment/Map format and SAMtools. Bioinformatics, 2009. 25(16): p. 2078-9.
[35] Koboldt, D.C.; Q. Zhang; D.E. Larson; D. Shen; M.D. McLellan; L. Lin; C.A. Miller; E.R. Mardis; L. Ding, and R.K. Wilson, VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res, 2012. 22(3): p. 568-76.
[36] McKenna, A.; M. Hanna; E. Banks; A. Sivachenko; K. Cibulskis; A. Kernytsky; K. Garimella; D. Altshuler; S. Gabriel; M. Daly, and M.A. DePristo, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res, 2010. 20(9): p. 1297-303.
[37] Cibulskis, K.; M.S. Lawrence; S.L. Carter; A. Sivachenko; D. Jaffe; C. Sougnez; S. Gabriel; M. Meyerson; E.S. Lander, and G. Getz, Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol, 2013. 31(3): p. 213-9.

[38] Saunders, C.T.; W.S. Wong; S. Swamy; J. Becq; L.J. Murray, and R.K. Cheetham, Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics, 2012. 28(14): p. 1811-7.
[39] Garrison, E. and G. Marth, Haplotype-based variant detection from short-read sequencing. arXiv, 2016. 1207.
[40] Rimmer, A.; H. Phan; I. Mathieson; Z. Iqbal; S.R.F. Twigg; W.G.S. Consortium; A.O.M. Wilkie; G. McVean, and G. Lunter, Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet, 2014. 46(8): p. 912-918.
[41] Narzisi, G.; A. Corvelo; K. Arora; E.A. Bergmann; M. Shah; R. Musunuri; A.-K. Emde; N. Robine; V. Vacic, and M.C. Zody, Genome-wide somatic variant calling using localized colored de Bruijn graphs. Communications Biology, 2018. 1(1): p. 20.
[42] Poplin, R.; P.-C. Chang; D. Alexander; S. Schwartz; T. Colthurst; A. Ku; D. Newburger; J. Dijamco; N. Nguyen; P.T. Afshar; S.S. Gross; L. Dorfman; C.Y. McLean, and M.A. DePristo, A universal SNP and small-indel variant caller using deep neural networks. Nature Biotechnology, 2018. 36(10): p. 983-987.
[43] Larson, D.E.; C.C. Harris; K. Chen; D.C. Koboldt; T.E. Abbott; D.J. Dooling; T.J. Ley; E.R. Mardis; R.K. Wilson, and L. Ding, SomaticSniper: identification of somatic point mutations in whole genome sequencing data. Bioinformatics, 2012. 28(3): p. 311-7.
[44] Wilm, A.; P.P. Aw; D. Bertrand; G.H. Yeo; S.H. Ong; C.H. Wong; C.C. Khor; R. Petric; M.L. Hibberd, and N. Nagarajan, LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res, 2012. 40(22): p. 11189-201.
[45] Skinner, M.E.; A.V. Uzilov; L.D. Stein; C.J. Mungall, and I.H. Holmes, JBrowse: a next-generation genome browser. Genome Res, 2009. 19(9): p. 1630-8.
[46] Gusev, A.; J.K. Lowe; M. Stoffel; M.J. Daly; D. Altshuler; J.L. Breslow; J.M. Friedman, and I. Pe'er, Whole population, genome-wide mapping of hidden relatedness. Genome research, 2009. 19(2): p. 318-326.
[47] Lin, Y.-L.; P.-C. Chang; C. Hsu; M.-Z. Hung; Y.-H. Chien; W.-L. Hwu; F. Lai, and N.-C. Lee, Comparison of GATK and DeepVariant by trio sequencing. Scientific Reports, 2022. 12(1): p. 1809.
[48] Obradovic, A.; L. Vlahos; P. Laise; J. Worley; X. Tan; A. Wang, and A. Califano, PISCES: A pipeline for the systematic, protein activity-based analysis of single cell RNA sequencing data. Biorxiv, 2021. 6: p. 22.
[49] Freed, D.; R. Pan, and R. Aldana, TNscope: accurate detection of somatic mutations with haplotype-based variant candidate detection and machine learning filtering. biorxiv, 2018: p. 250647.
[50] Fang, H.; E.A. Bergmann; K. Arora; V. Vacic; M.C. Zody; I. Iossifov; J.A. O'Rawe; Y. Wu; L.T. Jimenez Barron; J. Rosenbaum; M. Ronemus; Y.-h. Lee; Z. Wang; E. Dikoglu; V. Jobanputra; G.J. Lyon; M. Wigler; M.C. Schatz, and G. Narzisi, Indel variant analysis of short-read sequencing data with Scalpel. Nature Protocols, 2016. 11(12): p. 2529-2548.
[51] Sahraeian, S.M.E.; R. Liu; B. Lau; K. Podesta; M. Mohiyuddin, and H.Y. Lam, Deep convolutional neural networks for accurate somatic mutation detection. Nature communications, 2019. 10(1): p. 1041.
[52] Cooke, D.; G. Lunter, and D. Wedge, Accurate genotyping of single cells with Octopus. 2021.
[53] Kim, S.; K. Scheffler; A.L. Halpern; M.A. Bekritsky; E. Noh; M. Källberg; X. Chen; Y. Kim; D. Beyter, and P. Krusche, Strelka2: fast and accurate calling of germline and somatic variants. Nature methods, 2018. 15(8): p. 591-594.
[54] Gehring, J.S.; B. Fischer; M. Lawrence, and W. Huber, SomaticSignatures: inferring mutational signatures from single-nucleotide variants. Bioinformatics, 2015. 31(22): p. 3673-3675.

[55] Mayakonda, A.; D.-C. Lin; Y. Assenov; C. Plass, and H.P. Koeffler, Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome research, 2018. 28(11): p. 1747-1756.
[56] Rosenthal, R.; N. McGranahan; J. Herrero; B.S. Taylor, and C. Swanton, DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome biology, 2016. 17: p. 1-11.
[57] Sahin, U.; E. Derhovanessian; M. Miller; B.-P. Kloke; P. Simon; M. Löwer; V. Bukur; A.D. Tadmor; U. Luxemburger; B. Schrörs; T. Omokoko; M. Vormehr; C. Albrecht; A. Paruzynski; A.N. Kuhn; J. Buck; S. Heesch; K.H. Schreeb; F. Müller; I. Ortseifer; I. Vogler; E. Godehardt; S. Attig; R. Rae; A. Breitkreuz; C. Tolliver; M. Suchan; G. Martic; A. Hohberger; P. Sorn; J. Diekmann; J. Ciesla; O. Waksmann; A.-K. Brück; M. Witt; M. Zillgen; A. Rothermel; B. Kasemann; D. Langer; S. Bolte; M. Diken; S. Kreiter; R. Nemecek; C. Gebhardt; S. Grabbe; C. Höller; J. Utikal; C. Huber; C. Loquai, and Ö. Türeci, Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. Nature, 2017. 547(7662): p. 222-226.
[58] Keskin, D.B.; A.J. Anandappa; J. Sun; I. Tirosh; N.D. Mathewson; S. Li; G. Oliveira; A. Giobbie-Hurder; K. Felt; E. Gjini; S.A. Shukla; Z. Hu; L. Li; P.M. Le; R.L. Allesoe; A.R. Richman; M.S. Kowalczyk; S. Abdelrahman; J.E. Geduldig; S. Charbonneau; K. Pelton; J.B. Iorgulescu; L. Elagina; W. Zhang; O. Olive; C. McCluskey; L.R. Olsen; J. Stevens; W.J. Lane; A.M. Salazar; H. Daley; P.Y. Wen; E.A. Chiocca; M. Harden; N.J. Lennon; S. Gabriel; G. Getz; E.S. Lander; A. Regev; J. Ritz; D. Neuberg; S.J. Rodig; K.L. Ligon; M.L. Suva; K.W. Wucherpfennig; N. Hacohen; E.F. Fritsch; K.J. Livak; P.A. Ott; C.J. Wu, and D.A. Reardon, Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. Nature, 2019. 565(7738): p. 234-239.
[59] Lin, X.; S. Tang; Y. Guo; R. Tang; Z. Li; X. Pan; G. Chen; L. Qiu; X. Dong; L. Zhang; X. Liu; Z. Cai, and B. Xie, Personalized neoantigen vaccine enhances the therapeutic efficacy of bevacizumab and anti-PD-1 antibody in advanced non-small cell lung cancer. Cancer Immunol Immunother, 2024. 73(2): p. 26.
[60] Gevaert, C.M.; M. Carman; B. Rosman; Y. Georgiadou, and R. Soden, Fairness and accountability of AI in disaster risk management: Opportunities and challenges. Patterns, 2021. 2(11).
[61] Ahmadi, A., Quantum Computing and Artificial Intelligence: The Synergy of Two Revolutionary Technologies. Asian Journal of Electrical Sciences, 2023. 12(2): p. 15-27.
[62] Martonosi, M. and M. Roetteler, Next steps in quantum computing: Computer science's role. arXiv preprint arXiv:1903.10541, 2019.
[63] Niraula, D.; J. Jamaluddin; M.M. Matuszak; R.K.T. Haken, and I.E. Naqa, Quantum deep reinforcement learning for clinical decision support in oncology: application to adaptive radiotherapy. Scientific reports, 2021. 11(1): p. 23545.
[64] Mustapha, M.T.; I. Ozsahin, and D.U. Ozsahin, Chapter 2 - Convolution neural network and deep learning, in Artificial Intelligence and Image Processing in Medical Imaging, W.A. Zgallai and D.U. Ozsahin, Editors. 2024, Academic Press. p. 21-50.
[65] Johnson, K.B.; W.Q. Wei; D. Weeraratne; M.E. Frisse; K. Misulis; K. Rhee; J. Zhao, and J.L. Snowdon, Precision Medicine, AI, and the Future of Personalized Health Care. Clin Transl Sci, 2021. 14(1): p. 86-93.
[66] Yandell, M.D. and W.H. Majoros, Genomics and natural language processing. Nature Reviews Genetics, 2002. 3(8): p. 601-610.
[67] Rayhan, A.; R. Kinzler, and R. Rayhan, NATURAL LANGUAGE PROCESSING: TRANSFORMING HOW MACHINES UNDERSTAND HUMAN LANGUAGE. 2023.
[68] Velupillai, S.; H. Suominen; M. Liakata; A. Roberts; A.D. Shah; K. Morley; D. Osborn; J. Hayes; R. Stewart; J. Downs; W. Chapman, and R. Dutta, Using clinical Natural Language Processing for health

outcomes research: Overview and actionable suggestions for future advances. Journal of Biomedical Informatics, 2018. 88: p. 11-19.

[69] Linette, G.P. and B.M. Carreno, Neoantigen Vaccines Pass the Immunogenicity Test. Trends Mol Med, 2017. 23(10): p. 869-871.

[70] Ott, P.A.; Z. Hu; D.B. Keskin; S.A. Shukla; J. Sun; D.J. Bozym; W. Zhang; A. Luoma; A. Giobbie-Hurder; L. Peter; C. Chen; O. Olive; T.A. Carter; S. Li; D.J. Lieb; T. Eisenhaure; E. Gjini; J. Stevens; W.J. Lane; I. Javeri; K. Nellaiappan; A.M. Salazar; H. Daley; M. Seaman; E.I. Buchbinder; C.H. Yoon; M. Harden; N. Lennon; S. Gabriel; S.J. Rodig; D.H. Barouch; J.C. Aster; G. Getz; K. Wucherpfennig; D. Neuberg; J. Ritz; E.S. Lander; E.F. Fritsch; N. Hacohen, and C.J. Wu, An immunogenic personal neoantigen vaccine for patients with melanoma. Nature, 2017. 547(7662): p. 217-221.

[71] Braiteh, F.; P. LoRusso; A. Balmanoukian; S. Klempner; D.R. Camidge; M. Hellmann; M. Gordon; J. Bendell; L. Mueller, and R. Sabado, Abstract CT169: A phase Ia study to evaluate RO7198457, an individualized Neoantigen Specific immunoTherapy (iNeST), in patients with locally advanced or metastatic solid tumors. Cancer Research, 2020. 80(16_Supplement): p. CT169-CT169.

[72] Sahin, U.; P. Oehm; E. Derhovanessian; R.A. Jabulowsky; M. Vormehr; M. Gold; D. Maurus; D. Schwarck-Kokarakis; A.N. Kuhn; T. Omokoko; L.M. Kranz; M. Diken; S. Kreiter; H. Haas; S. Attig; R. Rae; K. Cuk; A. Kemmer-Brück; A. Breitkreuz; C. Tolliver; J. Caspar; J. Quinkhardt; L. Hebich; M. Stein; A. Hohberger; I. Vogler; I. Liebig; S. Renken; J. Sikorski; M. Leierer; V. Müller; H. Mitzel-Rink; M. Miederer; C. Huber; S. Grabbe; J. Utikal; A. Pinter; R. Kaufmann; J.C. Hassel; C. Loquai, and Ö. Türeci, An RNA vaccine drives immunity in checkpoint-inhibitor-treated melanoma. Nature, 2020. 585(7823): p. 107-112.

[73] Delord, J.-P.; M.S. Block; C. Ottensmeier; G. Colon-Otero; C. Le Tourneau; A. Lalanne; C. Jamet; O. Lantz; K.L. Knutson, and G. Lacoste, Phase 1 studies of personalized neoantigen vaccine TG4050 in ovarian carcinoma (OC) and head and neck squamous cell carcinoma (HNSCC). 2022, American Society of Clinical Oncology.

[74] Tran, E.; P.F. Robbins, and S.A. Rosenberg, 'Final common pathway' of human cancer immunotherapy: targeting random somatic mutations. Nature Immunology, 2017. 18(3): p. 255-262.