

Review

# Advances in Genomic Data and Biomarkers: Revolutionizing NSCLC Diagnosis and Treatment

Juan Carlos Restrepo <sup>1</sup>, Diana Dueñas <sup>1</sup>, Zuray Corredor <sup>2,3</sup>  and Yamil Liscano <sup>1,\*</sup> 

<sup>1</sup> Grupo de Investigación en Salud Integral (GISI), Departamento Facultad de Salud, Universidad Santiago de Cali, Cali 760035, Colombia; juan.restrepo16@usc.edu.co (J.C.R.); diana.duenas00@usc.edu.co (D.D.)

<sup>2</sup> Grupo de Investigaciones en Odontología (GIOD), Facultad de Odontología, Universidad Cooperativa de Colombia, Pasto 520002, Colombia

<sup>3</sup> Facultad de Salud, Departamento de Ciencias Básicas, Universidad Libre, Cali 760026, Colombia; zurayf.corredorm@unilibre.edu.co

\* Correspondence: yamil.liscano00@usc.edu.co

**Simple Summary:** In the field of non-small cell lung cancer (NSCLC), there have been significant advancements in genomic data and bioinformatics tools, which have improved early diagnosis, treatment, and follow-up using biomarkers. Biomarkers provide measurable indicators of disease characteristics and help tailor treatment strategies in precision medicine. The integration of big data and artificial intelligence (AI) further enhances personalized medicine through advanced biomarker analysis. However, challenges exist in terms of limited evidence on the impact of new biomarkers on mortality and treatment efficacy, as well as data analysis and the adoption of precision medicine in clinical practice. Despite these obstacles, the integration of biomarkers into precision medicine has shown promise in improving patient outcomes in NSCLC. Continued research and advancements in biomarker discovery, utilization, and evidence generation are needed to overcome these challenges and further enhance precision medicine's effectiveness in NSCLC.

**Abstract:** Non-small cell lung cancer (NSCLC) is a significant public health concern with high mortality rates. Recent advancements in genomic data, bioinformatics tools, and the utilization of biomarkers have improved the possibilities for early diagnosis, effective treatment, and follow-up in NSCLC. Biomarkers play a crucial role in precision medicine by providing measurable indicators of disease characteristics, enabling tailored treatment strategies. The integration of big data and artificial intelligence (AI) further enhances the potential for personalized medicine through advanced biomarker analysis. However, challenges remain in the impact of new biomarkers on mortality and treatment efficacy due to limited evidence. Data analysis, interpretation, and the adoption of precision medicine approaches in clinical practice pose additional challenges and emphasize the integration of biomarkers with advanced technologies such as genomic data analysis and artificial intelligence (AI), which enhance the potential of precision medicine in NSCLC. Despite these obstacles, the integration of biomarkers into precision medicine has shown promising results in NSCLC, improving patient outcomes and enabling targeted therapies. Continued research and advancements in biomarker discovery, utilization, and evidence generation are necessary to overcome these challenges and further enhance the efficacy of precision medicine. Addressing these obstacles will contribute to the continued improvement of patient outcomes in non-small cell lung cancer.

**Keywords:** biomarkers; miRNA; bioinformatics; non-small cell lung cancer; bioprospecting; machine learning; artificial intelligence



**Citation:** Restrepo, J.C.; Dueñas, D.; Corredor, Z.; Liscano, Y. Advances in Genomic Data and Biomarkers: Revolutionizing NSCLC Diagnosis and Treatment. *Cancers* **2023**, *15*, 3474. <https://doi.org/10.3390/cancers15133474>

Academic Editor: David Wong

Received: 5 June 2023

Revised: 23 June 2023

Accepted: 27 June 2023

Published: 3 July 2023



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

## 1. Introduction

Non-small cell lung cancer (NSCLC) is a serious public health problem, accounting for 84% of diagnosed lung cancers. In 2020, there were 2,206,771 (11.4%) new diagnoses

of lung cancer in both sexes and all ages, second only to breast cancer. However, it had the highest mortality rate with 1.8 million deaths in 2020 [1], mainly due to late diagnosis. With the increase in genomic data, the possibilities for early diagnosis, effective and safe treatment, and follow-up are enhanced by increasing the likelihood of prediction through biomarkers. This is achieved by using bioinformatics tools and automated machine learning algorithms to analyze large amounts of data, including clinical data obtained from harmonized medical records, radiological images, and pathological studies (tissue biopsies), as well as genomic data.

The rapid growth of genomic data in cancer and the development of bioinformatics analysis methods have led to the identification of tumor biomarkers that facilitate early detection, treatment, and prognosis, reducing mortality rates in some types of cancer [2]. Public data sources such as the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) provide opportunities to explore tumor genesis and progression, as well as the identification of new biomarkers for diagnosis, prognosis, and treatment response [3]. Biomarkers, combined with clinical data, have prognostic value and can predict outcomes, guiding specific treatments [4]. To structure and perform classification and prediction tasks with this big data, different machine learning methods have evolved according to the needs and magnitude of the data [5]. In a multi-omic conceptual framework, integrating component analysis, non-parametric combinations, and exploratory analysis contribute to consistent information and enable the application of classification and prediction algorithms such as decision trees, random forests, support vector machines, linear and logistic regression, and deep learning models. Recently, dimensionality reduction techniques such as t-distributed stochastic neighbor embedding (t-SNE), multidimensional scaling (MDS), and uniform manifold approximation and projection (UMAP) have been applied based on data structure and acquisition methods [6].

In the case of non-small cell lung cancer, the impact of a new biomarker on mortality and the efficacy of specific treatment is a challenge due to the limited evidence available. Therefore, the discovery and application of relevant biomarkers, both new and existing, will lead the future of precision medicine in NSCLC [7,8].

## 2. Non-Small Cell Lung Cancer

NSCLC is the most common type of lung cancer, accounting for approximately 85% of all cases [9], and it is the leading cause of cancer-related deaths worldwide, ranking first in men and second in women, with nearly 1.8 million deaths per year [10]. It is a complex and diverse disease that arises from the abnormal growth of cells in the lung tissue, primarily caused by smoking. Early detection and increased utilization of diagnostic tools such as PET scans and the discovery of biomarkers are crucial for improving patient outcomes and reducing lung cancer mortality rates [11,12].

NSCLC encompasses various subtypes, including adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. This type of cancer is often diagnosed at advanced stages, which poses challenges for effective treatment. Adenocarcinoma is the most common type of lung cancer, accounting for approximately 40% of cases. It develops from small airway epithelial cells known as type II alveolar cells, which secrete mucus and other substances. Adenocarcinoma affects both smokers and non-smokers and occurs in men and women of all ages. It tends to occur in the outer regions of the lungs, possibly due to cigarette filters preventing large particles from entering the lungs. Adenocarcinoma generally grows slower and has a higher chance of being detected before it spreads beyond the lungs compared to other types of lung cancer [10,11].

Large cell carcinoma (undifferentiated) represents 5–10% of lung cancers. This type of carcinoma lacks evidence of squamous or glandular maturation and is often diagnosed by excluding other possibilities. Large cell carcinoma typically originates in the central part of the lungs and can spread to nearby lymph nodes, the chest wall, and distant organs. This type of carcinoma is strongly associated with tobacco consumption [9,10].

### 2.1. Risk Factors and Epidemiology

Some factors that indicate a negative outlook include being male, having impaired functional abilities, and being over the age of 70. The use of radiological imaging does not effectively reduce mortality because of the aggressive nature of the disease. The most crucial steps to reduce mortality are quitting smoking and prevention, given that smoking is the primary risk factor for lung cancer. Non-smokers exposed to secondhand tobacco smoke also face an increased risk of developing lung cancer, and living with a smoker can raise the chances by 20–30% [9,10,13].

Tobacco consumption has been identified as the cause of 90% of all lung cancer cases. Current smokers with a history of 40 pack-years have twenty times higher chances of developing lung cancer than a nonsmoker. This risk can increase when additional environmental or lifestyle exposures are combined with tobacco consumption, such as asbestos exposure. It is believed that adenocarcinoma, in particular, originated from the invention of filtered cigarettes in the 1960s, although this has not been proven [14,15].

Another risk factor is radon, a natural carcinogen present in uranium deposits in basements, which is associated with lung cancer and is estimated to have caused approximately 21,000 deaths from this disease in the United States. Occupational exposure to substances such as asbestos, arsenic, beryllium, and other chemicals also increases the risk of lung cancer. Air pollution, especially in areas with heavy traffic and high concentration of pollutants, including polycyclic aromatic hydrocarbons, is identified as a risk factor for lung cancer, with an 8% increase in overall mortality risk from this disease. Personal or family history of lung cancer are also risk factors, and certain genes and chromosomes associated with a higher risk of developing the disease, such as the TP53 gene and a marker on chromosome 15, have been identified [9,10,13].

On a global level, lung cancer is the leading cause of cancer-related death in men and the second most common in women. There is significant variation in the incidence of lung cancer across different populations, primarily driven by the prevalence of tobacco consumption in various countries. The incidence of lung cancer is directly related to the rise or decline in smoking rates among different populations. For instance, it is projected that age-adjusted mortality rates in the United States will decrease by 79% between 2015 and 2065 due to declining tobacco use rates and anti-smoking campaigns. Incidence and mortality rates of lung cancer are higher in developed countries. In contrast, lung cancer rates in underdeveloped geographic areas, such as Central/South America and much of Africa, are lower [14,15].

However, many developing countries lack a centralized reporting system, leading to underreporting of lung cancer cases, making it challenging to determine the true incidence of the disease. The World Health Organization (WHO) estimates that global lung cancer mortality rates will continue to rise, primarily due to the increasing global tobacco consumption, particularly in Asia. In the United States, lung cancer incidence and mortality rates have been declining in men, while initially increasing in women until around the year 2000, and have since stabilized. Due to the shifting incidence of lung cancer in women, mortality rates have decreased in women over a decade after declining in men [14,15].

According to data from the Global Cancer Observatory (GLOBOCAN), lung cancer is the second most common cancer worldwide, following female breast cancer. In terms of mortality, lung cancer is the leading cause of cancer-related death. In South America, the incidence and mortality rates of lung cancer vary among countries. In 2020, the incidence of lung cancer in South America was 17.8 cases per 100,000 population in men and 10.3 in women. In Colombia, lung cancer ranked fifth in incidence in 2020, following breast, prostate, colorectal, and stomach cancer. In terms of mortality, lung cancer is the second leading cause of cancer-related death in Colombia, with approximately 9.2 deaths per 100,000 population each year [16,17]. Adenocarcinoma is the most common histological variant in lung cancer patients in Colombia, followed by squamous cell carcinoma. The majority of patients present at advanced stages of the disease. Mutations in the Epidermal Growth Factor Receptor (EGFR) gene are detected in a significant proportion of patients,

mainly in non-smoking women. ALK rearrangement is also found in a small proportion of cases. It is necessary to implement lung cancer prevention and control strategies to reduce its incidence and mortality in Colombia [16,17].

Lung cancer has a high social and economic impact globally, causing a significant number of disability-adjusted life years lost. In Latin America, lung cancer treatment is costly and limited in terms of the adoption of new technologies and medications. In Colombia, lung cancer mortality rates have shown a decreasing trend since 2005, both in men and women, across all age groups [16,17]. The decline in mortality rates in men aged 35 to 64 was observed in 1985, while in women of the same age group, it was observed in 1991. However, in the group of individuals older than 64, the decline in mortality rates began to be observed in the early 21st century. The study by Giraldo-Osorio et al., from 2022 is the first in over 30 years to determine the trend in lung cancer mortality at the national level in Colombia and uses joinpoint regression analysis. It highlights the increase in lung cancer mortality rates in women over 64 years old until the early 21st century and in women under 65 years old until 1991 [16]. This could be related to exposure to solid fuel smoke, especially in indigenous women over 30 years old living in rural areas. The use of solid fuels for cooking has decreased in Colombia, which may have contributed to the decline in lung cancer mortality rates in women. The decrease in tobacco consumption prevalence in Colombia may explain the reduction in mortality rates, especially in men [16]. Over the decades, there has been a decrease in tobacco consumption prevalence in the general population, although men have higher rates than women. Laws and projects for tobacco control have been implemented, contributing to the reduction in consumption. It is important for Colombia to continue efforts in primary prevention to prevent the initiation of tobacco use and in secondary prevention to assist people in quitting smoking [16].

## 2.2. Pathophysiology, Histology, and Classification

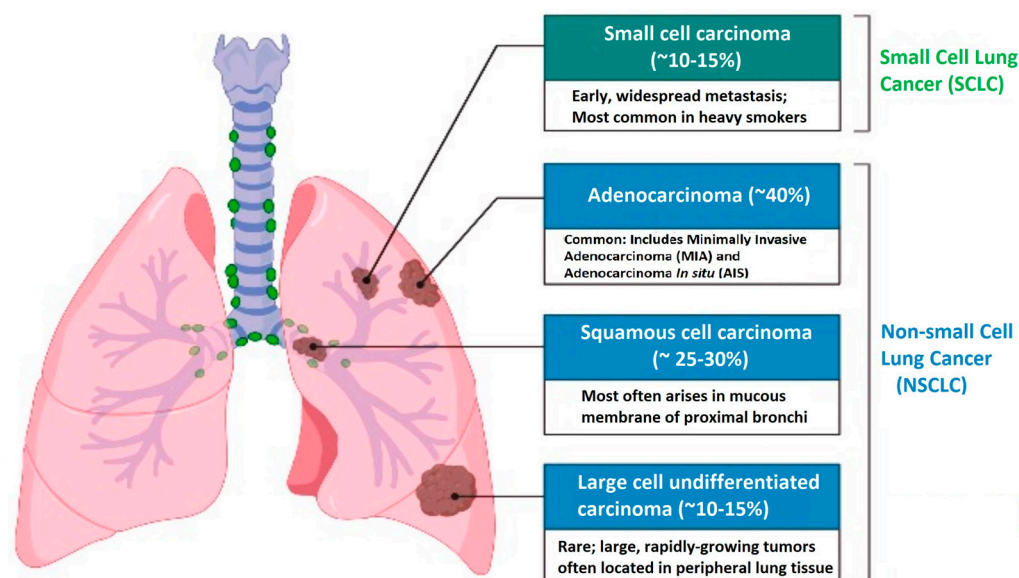
Oncogene mutations play a significant role in NSCLC and can be effectively targeted with specific drugs. Familial clustering of lung cancer has provided evidence for a hereditary component in disease development. Carriers of TP53 germline sequence variations have an increased risk of lung cancer, particularly if they are smokers. A genome-wide linkage study has identified a susceptibility locus at 6q23-25p that influences lung cancer risk [13]. The most common mutation occurs in the EGFR gene, affecting 10–30% of patients. This mutation is associated with downstream signaling pathways, including MAPK/ERK, PI3K/AKT, and Bax/Bcl-2. The presence of the EGFR T790M sequence variation has been identified in families with multiple cases of NSCLC. Additionally, it is linked to acquired resistance to EGFR tyrosine kinase inhibitors (TKIs) [18]. Osteopontin (OPN) and the PI3K/AKT/mTOR pathway also contribute to resistance against EGFR TKIs. Gene rearrangements in the RET gene are detected in approximately 1% to 2% of patients with NSCLC and affect signaling pathways such as PI3K/AKT, JAK-STAT, and RAS/MAPK. These pathways play a role in cell proliferation, invasion, and migration. Mutations in the MET gene lead to abnormal expression of the MET axis, promoting the migration and invasion of tumor cells, as well as resistance to inhibitors targeting EGFR and VEGFR. Rearrangements in the Anaplastic Lymphoma Kinase (ALK) gene occur in 5% to 6% of NSCLC patients, resulting in increased expression of ALK, which triggers epithelial-mesenchymal transition (EMT) and enhances migration and invasion. Resistance to ALK inhibitors is associated with specific mutations such as F1174L and involves the STAT3/ZEB1 signaling pathway [18]. Human epidermal growth factor receptor 2 (HER2) gene mutations, particularly in exon 20, are found in 2–4% of NSCLC patients, especially in women, and are associated with brain metastases. Activation of HER2 triggers downstream signaling through MEK/ERK and PI3K/AKT pathways, promoting lung cancer cell migration and proliferation. BRAF mutations, particularly V600E, confer resistance to BRAF inhibitors but respond better to combined BRAF and MEK inhibitors. ROS1 gene rearrangements are present in 1–2% of NSCLC patients, with CD74-ROS1 being the most common fusion. Activation of the ROS1 gene initiates signaling through the PI3K/AKT/mTOR, JAK/STAT,

and MAPK/ERK pathways, facilitating the proliferation and invasion of tumor cells. In NSCLC patients, KRAS mutations are identified in around 13% of cases, predominantly among smokers, and are linked to drug resistance and unfavorable outcomes. Various KRAS mutations activate specific downstream signaling pathways. Gene fusions involving the Neurotrophic tropomyosin receptor kinase (NTRK) gene occur in less than 1% of NSCLC patients and affect the MEK/ERK and PI3K/AKT pathways, which play a role in cell proliferation, migration, and chemotherapy-induced apoptosis. Early-stage NSCLC patients with NTRK fusions exhibit a significant response rate to TKIs [12,18].

The genomic characteristics of lung cancer differ significantly between smokers and never-smokers. Smokers display a higher frequency of mutations, including non-actionable mutations in genes such as KRAS and TP53, characterized by specific nucleotide changes. On the other hand, never-smokers more commonly exhibit specific actionable gene alterations such as activating EGFR mutations and ROS1 and ALK translocations. These genetic events that drive tumor development also impact the composition of the tumor microenvironment (TME). NSCLC, in particular, demonstrates a high number of somatic tumor mutations, especially in smokers, and metastatic tumors tend to have more mutations than primary lung lesions. Certain mutations result in the formation of neoantigens, which can be recognized by immune cells infiltrating the tumor. An increased burden of these neoantigens is associated with an inflamed TME, enriched with activated immune cells and the expression of immune-related proteins. Additionally, defects in DNA mismatch repair and microsatellite instability contribute to a high tumor mutation burden and favorable responses to immune checkpoint inhibitors (ICBs). Furthermore, genetic alterations can influence the cellular composition and functions of the TME. For instance, the inactivation of the STK11 tumor suppressor gene in KRAS-mutated lung adenocarcinoma shifts the TME towards immunosuppressive neutrophils and reduced expression of PD-L1. Further investigations are required to fully understand the intricate relationship between genotypes of NSCLC and the cellular makeup of the TME [12].

Histology plays a crucial role in cancer diagnosis, allowing for the distinction of different subtypes of lung tumors. In adenocarcinoma, evidence of neoplastic gland formation, expression of markers such as TTF-1 and napsin, or the presence of intracytoplasmic mucin is required. Squamous cell carcinoma is diagnosed by detecting keratin production and intercellular desmosomes. Immunohistochemistry (IHC) enables the identification of specific markers such as p40, p63, CK5, or desmoglein in squamous cell carcinoma. On the other hand, large cell carcinoma is a diagnosis of exclusion as it can exhibit features of squamous, glandular, or neuroendocrine differentiation. In cases of poorly differentiated carcinoma, it will be classified as large cell carcinoma only if specific markers indicating another subtype of lung cancer are not found [15].

There are two main types of lung cancer: NSCLC, which accounts for 85% of cases, and small cell lung cancer (SCLC), which makes up 15% of cases. NSCLC is further classified into three main types: adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. Adenocarcinoma is the most common subtype and arises from alveolar cells in the smaller airway epithelium. Squamous cell carcinomas arise from cells in the airway epithelium, and large cell cancers are typically poorly differentiated and composed of large cells. Immunohistochemical markers are used to identify these subtypes (Figure 1) [14,15].



**Figure 1.** Histological classification of lung cancer.

### 2.3. Resistance to Medications and Immunotherapy in NSCLC: Mechanisms and Therapeutic Strategies

Drug resistance is one of the main causes of therapeutic failure in NSCLC, leading to tumor recurrence and disease progression. Resistance in NSCLC is a significant clinical problem as it diminishes treatment efficacy and may contribute to disease progression and the development of metastasis. The clinical relevance of resistance in NSCLC lies in the need to develop more effective therapeutic approaches to overcome it. Research is focused on identifying the mechanisms responsible for resistance and finding strategies to counteract them. This includes the development of therapies specifically targeting genetic mutations present in tumor cells, as well as the use of combinations of different treatments to address resistance more effectively [19,20].

NSCLC is a heterogeneous disease that exhibits various genetic alterations. Some of the most relevant therapeutic targets in NSCLC include EGFR, ALK, ROS1, BRAF, MET, RET, and KRAS. The intrinsic mechanisms of cellular resistance in this disease encompass changes in drug transporter expression and activation of pro-survival and anti-apoptotic pathways, along with non-intrinsic influences from the tumor microenvironment [19]. TKIs have proven to be effective in the treatment of NSCLC in patients with activating mutations in these genes. For instance, EGFR TKIs such as erlotinib and gefitinib have shown promising results in patients with EGFR mutations. Similarly, ALK TKIs such as crizotinib and alectinib have demonstrated activity in patients with ALK fusions. Additionally, targeted drugs have been developed for ROS1, BRAF, MET, RET, and KRAS, thereby expanding treatment options for NSCLC patients harboring these genetic alterations [20].

The genetic alterations and tumor heterogeneity in NSCLC are driven by subpopulations of tumor cells called cancer stem cells (CSCs), which possess tumor-initiating capabilities, high self-renewal capacity, and the ability to differentiate into multiple lineages. CSCs have been identified in NSCLC and have been associated with resistance to chemotherapy, radiation therapy, and immunotherapy [19,21].

Resistance to immunotherapy can manifest at different levels of dysfunction and be related to specific tumor phenotypes. One mechanism of resistance involves the absence of specific T lymphocytes in the tumor microenvironment or their inability to express the necessary T cell receptor (TCR) to recognize the tumor antigen. Another mechanism is related to the lack of infiltration of activated T lymphocytes into the tumor tissue due to the presence of substances such as VEGF that promote the formation of barriers for immune cells [21].

In CPCNP, resistance to immunotherapy can also arise, such as in the case of immune checkpoint inhibitors. The tumor microenvironment and interactions between cancer cells and immune cells play a crucial role in determining the response to immunotherapy. Tumors with a high population of CSCs have been found to exhibit immunosuppressive characteristics and evade immune surveillance, leading to resistance to immunotherapy [19]. CSCs can modulate immune responses through various mechanisms, including the expression of immune inhibitory molecules, secretion of immunosuppressive factors, and induction of dysfunction in immune cells [19]. On the other hand, resistance to EGFR inhibitors is common in patients with lung adenocarcinoma. Various mechanisms of resistance have been identified, including EGFR mutations, oncogenic changes, alterations in apoptosis, and epithelial-mesenchymal transition. EGFR inhibition can lead to the enrichment of the cancer stem cell subpopulation in non-small cell lung cancer, whether mutated or non-mutated in EGFR, through a NOTCH3-dependent process. Additionally, this inhibition can increase cell death in non-stem cells and enhance the formation of pulmonary spheres [19]. Inhibiting NOTCH1 and HES1 can reverse resistance to gefitinib by increasing apoptosis. Dual inhibition of EGFR and NOTCH2/3 reduces EGFR and radiation-induced cancer stem cell subpopulations, as well as the expression of DNA repair genes. Similarly, resistance to ALK inhibitors in NSCLC may be mediated by mechanisms associated with NOTCH signaling. Erlotinib faces resistance due to mutations such as T790M, D761Y, L747S, and T854A in the EGFR gene, as well as EGFR amplification, PIK3CA gene mutations, and MET gene amplification. The MET gene encodes a receptor tyrosine kinase (RTK) protein that is activated by binding with the HGF ligand, triggering a signaling cascade that promotes cellular progression and proliferation in various malignant tumors. This amplification can occur early in tumorigenesis or as a mechanism of acquired resistance to tyrosine kinase inhibitors. However, specific MET inhibitors and multi-kinase inhibitors have been developed, showing promising results in treating tumors with MET gene amplification, thereby improving clinical outcomes [22]. Studies have shown that both overexpression and amplification of MET are associated with poor prognosis in patients with NSCLC, and MET amplification appears to be an independent marker of poor outcome after surgical resection of NSCLC. In a series of 687 Asian patients with resected NSCLC, MET alterations were unfavorable prognostic factors for overall survival [23].

Regarding ALK inhibitors, mutations such as G1269A, C1156Y, I1171T/N/S, S1206C, E1210K, L1152P/R, V1180L, G1128A, F1174V, and L1196M in the ALK kinase domain, along with bypass pathway activation, KIT amplification, and MAPK signaling, contribute to resistance. ROS1 inhibitors face resistance due to point mutations in the ROS1 kinase domain, such as D2033N, G2032, L2026M, L2155S, and S1986F/Y, as well as the involvement of other receptors and effectors in the MAPK pathway. Lastly, BRAF inhibitors are affected by mutations such as BRAF(V600E), BRAF(D594G), and BRAF(G469A/V), as well as bypass pathway activation and restoration of MAPK signaling [20].

Regarding anti-angiogenic therapy targeting VEGF and DLL4-NOTCH, resistance can arise due to the activation of alternative pro-angiogenic signals. In KRAS-driven tumors, NOTCH inhibition can improve treatment sensitivity and suppress RAS and PI3K signaling pathways [19]. The KRAS gene is frequently mutated in NSCLC, and although it has long been considered elusive, specific inhibitors such as KRASG12C inhibitors have been developed. However, resistance mechanisms have also been identified, both KRAS-dependent and KRAS-independent, involving acquired alterations in KRAS, activation of alternative signaling pathways, and histological transformation [24].

In the field of therapy based on immune checkpoint inhibitors (ICIs), which target PD-1 and PD-L1, a revolution has been observed in the treatment of NSCLC. However, not all patients respond to this therapy, and different resistance mechanisms have been identified. Primary resistance is related to the inability of the immune system to activate an adequate response against cancer cells, influenced by intrinsic and extrinsic factors to the tumor. Acquired resistance occurs when tumor cells adapt to the immune system and evade its response. Furthermore, cancer progression can also be observed after discontinuation of

treatment, which may involve elements of both primary and acquired resistance [20,24]. Intrinsic resistance to immunotherapy can be attributed to various genetic aberrations in oncogenes and tumor suppressor genes, which affect the immune response and alter cytokine profiles and immune cell composition. For example, certain genetic alterations, such as RET gene rearrangements and HER2 gene mutations, are associated with low PD-L1 expression, while activating mutations in the EGFR gene and ALK gene rearrangements are associated with high PD-L1 expression but a limited response to ICIs. Other genetic factors, such as KRAS and STK11/LKB1 gene mutations, have been linked to both sensitivity and resistance to ICIs. The STK11/AMPK axis is an intracellular signaling pathway that plays a crucial role in the regulation of metabolism and energy homeostasis in cells. STK11 (also known as LKB1) is a protein kinase that acts as a key regulator in this pathway, while AMPK (AMP-activated protein kinase) is its main substrate. When cellular AMP levels increase, as occurs during situations of energy stress or intense physical exercise, AMPK is activated through phosphorylation by STK11. Once activated, AMPK triggers a series of adaptive responses to restore energy balance. This includes the inhibition of protein and lipid synthesis, as well as the activation of catabolic pathways that generate ATP, such as glycolysis and fatty acid oxidation [25]. LKB1 deficiency, observed in tumors such as lung adenocarcinomas, is associated with lower overall survival and the formation of aggressive tumors with metastatic characteristics. Additionally, the absence of LKB1 promotes EMT, affects cell polarity, and creates an immunosuppressive microenvironment around tumor cells. In the treatment of advanced NSCLC, STK11 mutations are associated with increased resistance to radiotherapy and anti-PD-L1 therapy. Loss of LKB1 due to mutations in STK11 contributes to cancer progression and resistance to certain therapies, making it a potential target for the development of more effective lung cancer treatments [25].

Although ICI has significantly improved overall survival in patients with certain types of cancer, it still has limitations and challenges. One of the main obstacles is acquired resistance to therapy, which has been associated with decreased MHC-I in the presentation of tumor antigens to T cells. However, several therapeutic strategies are being explored to increase MHC-I expression and improve the effectiveness of immunotherapy. These approaches include activation of the interferon-gamma signaling pathway, stimulation of the NF- $\kappa$ B signaling pathway, use of chemotherapeutic agents, and inhibition of autophagy. Although further research is still needed, these strategies could provide new opportunities to enhance the immunotherapeutic response in cancer patients [26].

On the other hand, mutations in the BRAF gene have been reported in approximately 4% of NSCLC cases, being more common in non-small cell lung adenocarcinoma. The BRAF gene encodes a kinase protein that plays a crucial role in the regulation of cell growth and proliferation. Mutations in the BRAF gene, especially the V600E mutation, have been identified in various types of cancer, including melanoma and NSCLC. Activation of the BRAF gene in NSCLC has been associated with uncontrolled cell proliferation and resistance to apoptosis, contributing to tumor growth. Although BRAF mutant inhibition has proven effective in treating melanoma, its efficacy in NSCLC is limited due to the development of resistance. Resistance to BRAF inhibition in NSCLC can be primary or acquired. Primary resistance is due to intrinsic mechanisms that limit the effectiveness of BRAF inhibition. On the other hand, acquired resistance develops during treatment and can be caused by the activation of alternative signaling pathways, such as the EGFR pathway or the PI3K/AKT/mTOR signaling pathway. Several strategies have been proposed to overcome resistance to BRAF inhibition in NSCLC. These include the combination of BRAF inhibitors with MEK inhibitors to block multiple signaling pathways, the use of EGFR inhibitors in combination with BRAF inhibitors, and the combination of targeted therapies with immunotherapy to enhance the immune response against tumor cells. Although significant progress has been made in the treatment of resistance to BRAF inhibition in NSCLC, further research is still needed to fully understand the resistance mechanisms and develop more effective therapeutic strategies [27,28].



Another resistance found in NSCLC is EMT. EMT is a biological process that can drive metastasis in various types of cancer, including NSCLC. EMT has been associated with immune exclusion and resistance to immunotherapy in melanoma and other types of cancer. EMT has been found to be associated with the exclusion of critical immune cells in the immune response against cancer, such as CD4 and CD8 T cells. Additionally, certain immunosuppressive cytokines and immune checkpoint molecules have been identified to be associated with EMT in NSCLC, suggesting potential mechanisms of resistance to immunotherapy. It is also important to note that EMT is considered a key factor in the metastasis of lung cancer. EMT involves the transformation of epithelial cells into mesenchymal cells, promoting cancer progression and transforming cells into cancer stem cells (CSCs). Various signaling pathways and transcription factors, such as Snail, TWIST, ZEB, and FOXC2, play a significant role in regulating EMT in lung cancer. These factors contribute to the loss of cell adhesion, activation of proteolysis, and increased cell motility, leading to tumor metastasis. EMT is induced by signaling pathways such as TGF- $\beta$ , which activates the SMAD, PI3K-AKT, and MAPK pathways. Overall, these findings highlight the importance of understanding how EMT affects the tumor microenvironment and immune response in NSCLC to develop more effective therapeutic strategies [29,30].

In conclusion, it is crucial to continue research in NSCLC to identify new markers of resistance and develop effective therapeutic strategies. The identification of these markers and the development of appropriate treatment approaches are essential for improving patient outcomes and overcoming the challenges posed by this disease.

#### 2.4. Diagnosis and Treatment

Early detection of lung cancer through screening is of utmost importance as it can significantly improve clinical outcomes. During the initial stages, most patients with lung cancer do not exhibit noticeable symptoms, apart from coughing. This lack of symptoms can lead to missed opportunities for early diagnosis and treatment. Diagnostic approaches encompass imaging tests, biopsies, and biomarker tests. Imaging tests, including computed tomography (CT), positron emission tomography (PET), and magnetic resonance imaging (MRI), are vital in diagnosing lung cancer. CT screening, in particular, has shown a 20% reduction in lung cancer mortality and a decrease in overall mortality rates [14,18]. Advancements in diagnostic technologies play a critical role in saving lives and extending the lifespan of patients. Among these technologies, computed tomography (CT) is widely used as a diagnostic method, enabling the determination of tumor size and the identification of nodules in individuals with lung cancer. In order to detect lung cancer at an early stage, the US Preventive Services Task Force recommends annual low-dose CT screening for individuals at high risk. Furthermore, the NELSON trial demonstrated a noteworthy 26% decrease in lung cancer mortality among high-risk individuals who underwent low-dose CT screening [11,13,14]. It can also detect metastases, especially in mediastinal lymph nodes. However, biopsy is still necessary to determine if the nodules are benign or malignant, which can be tissue- or liquid-based. Tissue biopsy is invasive and considered the gold standard in clinical practice. It can determine the different histological types of lung cancer and detect mutations. However, it can have complications. Liquid biopsy is less invasive and utilizes peripheral blood samples. It can detect circulating tumor DNA, circulating tumor cells, exosomes, and other molecules such as miRNA and circRNA. Although it is more sensitive and practical than tissue biopsy, its application is limited [14,18].

PET, unlike CT, exhibits greater sensitivity and specificity due to its utilization of abnormal glucose markers. This imaging technique is capable of identifying malignant lesions characterized by abnormal glucose metabolism, determining their benign or malignant nature, and distinguishing between various types and stages of lung cancer. On the other hand, MRI is employed to detect metastases in the brain and bones of individuals with NSCLC. Through advancements in high-performance gradient systems, phased-array receiver coils, and optimized imaging sequences, MRI can now detect even small nodules in lung tissues, reaching a size as minute as 3 mm [31].

When it comes to biomarker tests, it is advisable to perform molecular tests on patients diagnosed with NSCLC to assess potential mutations. Several techniques are employed for this purpose, including polymerase chain reaction (PCR), fluorescence in situ hybridization (FISH), immunohistochemistry (IHC), and next-generation sequencing (NGS). These tests play a significant role in identifying various mutations associated with lung cancer and are instrumental in the development of targeted therapies. Precise determination of the TNM stage of lung cancer is essential for selecting the most appropriate treatment approach [14,18,32].

In the age of precision medicine, immunohistochemistry (IHC) plays a vital role in categorizing tumors into distinct subtypes and evaluating biomarkers to guide effective therapeutic decision-making. IHC offers several advantages over other techniques, such as its widespread availability, technical simplicity, cost-effectiveness, and rapid turnaround time. It has the potential to serve as a practical screening tool for identifying targetable genetic alterations and assessing biomarkers for molecular-targeted therapies. Furthermore, IHC provides valuable information about the cellular distribution of proteins within tumor tissues, enabling a more comprehensive understanding of the disease (see Table 1) [32]. Despite the benefits, challenges to lung cancer screening adoption include insurance coverage, false-positive results, radiation exposure concerns, patient distress, and the risk of overdiagnosis [14,18].

**Table 1.** Advantages and disadvantages of different diagnostic methods for NSCLC (MRI—Magnetic Resonance Imaging, PET—Positron Emission Tomography, CT—Computed Tomography, FISH—Fluorescence In Situ Hybridization, PCR—Polymerase Chain Reaction, IHC—Immunohistochemistry, FDG—Fludeoxyglucose F18).

Diagnostic Method	Advantages	Disadvantages
MRI	No ionizing radiation exposure	Limited availability and restricted access
	Detailed imaging of soft tissues	Lower sensitivity for detecting small lesions
PET	Detects metabolic and molecular changes	Higher cost and limited availability
	High sensitivity for detecting metastasis	Potential for false positives due to FDG accumulation
CT	Widely available and rapid access	Exposes the patient to ionizing radiation
	High spatial resolution and early tumor detection	Potential for false positives due to benign lesions
FISH Biomarkers	Provides genetic information about specific cancer subtypes	Requires specialized laboratory analysis
PCR Biomarkers	Provides genetic information about specific cancer subtypes,	Requires specialized laboratory analysis
IHC Biomarkers	Provides protein expression information, Helps differentiate cancer subtypes	Requires specialized personnel and equipment, results may vary depending on the method used
Next Generation Sequencing Biomarkers	Provides comprehensive genetic information	Requires specialized laboratory analysis
Liquid Biopsy	Non-invasive and lower risk for the patient	Lower sensitivity compared to tissue biopsy
	Enables monitoring of genetic changes over time	Potential for false negatives due to low concentration
Tissue Biopsy	Provides tissue samples for histopathological analysis	Invasive procedure with associated risks
	High precision and detection of genetic mutations	Potential complications such as bleeding or infection

### 3. Precision Medicine and Biomarkers

#### 3.1. The Impact of Precision Medicine

Precision medicine is a groundbreaking discipline that is transforming the way medicine is practiced. It emerges as a response to the growing understanding that each patient is unique, with genetic and molecular characteristics that can influence their susceptibility to diseases, response to treatments, and prognosis. Precision medicine is important today because it allows for a more individualized and personalized approach to the diagnosis, treatment, and prevention of diseases. Unlike traditional approaches that rely on

general protocols, precision medicine takes into account the unique characteristics of each individual, such as their genetics, lifestyle, environment, and other factors, to make more informed medical decisions. Precision medicine relies on the gathering and examination of diverse data sets, such as genetic, molecular, clinical, and lifestyle information, to gain a deeper understanding of the fundamental biology of a disease and to make well-informed choices regarding diagnosis and treatment options [33,34].

Precision medicine has also driven significant advances in the field of oncology. Through genomic and molecular analysis of tumors, specific alterations in a patient's DNA can be identified, allowing for the selection of therapies targeted at these alterations. This has revolutionized cancer treatment as we can now use specific drugs that act on the altered molecular signaling pathways in cancer cells, improving treatment efficacy and reducing side effects. Additionally, precision medicine also plays a crucial role in disease prevention. By better understanding the genetic and environmental risk factors of each individual, proactive measures can be taken to prevent the onset of diseases or detect them at early stages when they are more treatable. This enables personalized preventive interventions such as lifestyle changes, more frequent medical screenings, or even the use of preventive medications in selected cases [35,36].

A fundamental component of precision medicine is the utilization of biomarkers, which play a crucial role in identifying and understanding specific characteristics in a patient. Biomarkers are measurable and objective indicators that can encompass a wide range of aspects, from genetic mutations and gene expression profiles to protein levels and distinctive clinical features. These biomarkers offer valuable insights into the underlying biological aspects of a disease and enable healthcare professionals to obtain a more comprehensive and detailed understanding of a patient's medical condition. For example, specific genetic mutations may indicate the presence of an inherited disease or susceptibility to certain types of cancer. Gene expression profiles can reveal molecular patterns associated with a positive or negative response to a particular treatment. Protein levels in the body can indicate the presence or progression of a disease. By analyzing these biomarkers, physicians can tailor treatment more precisely to each patient. This means that therapeutic approaches can be customized based on the specific molecular and genetic characteristics of each individual, increasing the likelihood of a successful outcome [37–39].

Big data and artificial intelligence (AI) are becoming increasingly crucial in the analysis and interpretation of biomarkers within the realm of precision medicine. With the advancement of technology, there has been an explosion in the generation and collection of large-scale health data, leading to the formation of large datasets, known as big data. These datasets include genetic information, clinical records, medical images, laboratory test results, and much more. AI, in combination with big data analytics, has the potential to extract valuable information and hidden patterns from these large datasets [34]. Machine learning algorithms and artificial intelligence (AI) have the capability to process and analyze vast amounts of data in an efficient and accurate manner. They can uncover hidden correlations and patterns that may not be readily apparent to humans. This allows for the discovery of new relationships between biomarkers and diseases, as well as the identification of patterns in response to specific treatments. In the context of biomarkers, the application of big data and AI can significantly enhance predictive and diagnostic capabilities. Machine learning models can analyze multiple biomarkers simultaneously and generate predictive profiles that help identify diseases at early stages or predict the likelihood of a positive or negative response to a specific treatment. For example, in the field of cancer, AI algorithms can analyze genomic and molecular data from patients to identify patterns that indicate tumor aggressiveness or the probability of recurrence [40].

Furthermore, AI can also help improve the accuracy and efficiency in the interpretation of biomarkers. AI algorithms have the ability to analyze medical images, including CT scans and MRI, to identify distinctive characteristics that could be linked to particular diseases or prognoses. This can assist doctors in making more informed and precise decisions regarding diagnosis and treatment [34,35,41].

### 3.2. Types of Biomarkers

Biomarkers are any measurable characteristic that indicates the presence or absence of a disease or the biological response to a stimulus, and they are used to assess health status, diagnose diseases, track disease progression, and evaluate treatment response and other aspects related to health [42,43]. The FDA-NIH Biomarker Working Group has established seven categories to classify biomarkers according to their clinical application: susceptibility and risk, diagnosis, monitoring, prognosis, predictability, pharmacodynamics, and treatment response and safety. In addition to this clinical classification, biomarkers can also be classified based on their bodily source and type of measurement [44,45]. Here are some common types of classification for biomarkers [44–47]:

According to their biochemical nature:

- Genetic biomarkers: include variations in DNA, such as genetic mutations or single nucleotide polymorphisms (SNPs).
- Protein biomarkers: refer to specific proteins or protein profiles that can be measured in biological samples, such as blood or urine.
- Lipid biomarkers: are related to the lipids or fats present in the body and may be associated with cardiovascular or metabolic diseases.
- Metabolic biomarkers: refer to products or intermediates of metabolism that can be measured in biological samples.

According to their clinical application:

- Diagnostic biomarkers: used to confirm or rule out the presence of a specific disease or medical condition.
- Prognostic biomarkers: used to predict the progression of a disease or response to treatment.
- Treatment response biomarkers: used to evaluate the effectiveness of a specific treatment and adjust its dosage or duration.
- Disease progression biomarkers: used to monitor the advancement of a disease and evaluate its severity.
- Predictive biomarkers: used to predict the likelihood of a patient's response to a specific treatment before it is administered. These biomarkers are based on biological characteristics or signals that indicate the probability of a patient responding favorably to a particular treatment.

According to their location:

- Blood biomarkers: found in the blood and easily accessible through blood analysis.
- Urinary biomarkers: found in urine and used to assess renal function and detect urinary tract diseases.
- Tissue biomarkers: found in specific tissues and may require biopsies or imaging studies for their retrieval.

### 3.3. NSCLC Cancer Biomarkers

Biomarkers are of utmost importance in lung cancer, serving critical roles in the diagnosis, prognosis stratification, and selection of treatment options. Given the varied nature and complexity of this disease, biomarkers are instrumental in identifying distinct subtypes of lung cancer, guiding the selection of targeted therapies, and assessing the effectiveness of treatments [48,49].

One of the extensively studied and utilized biomarkers in lung cancer is EGFR. Mutations in the EGFR gene are prevalent in specific subtypes of NSCLC and can serve as predictors for the response to EGFR tyrosine kinase inhibitors such as erlotinib or gefitinib. Molecular tests, such as polymerase chain reaction (PCR) or genetic sequencing, are commonly employed to detect these mutations. Another significant biomarker is the ALK gene (anaplastic lymphoma kinase). ALK gene fusions are observed in a subset of non-small cell lung cancers and are associated with a heightened response to ALK inhibitors such as crizotinib, ceritinib, and alectinib. Molecular testing is also used to

detect ALK gene fusions, similar to EGFR [9,14]. Moreover, to these biomarkers, other biomarkers are being investigated in NSCLC, such as mutations in the KRAS gene, fusion of the RET gene, and expression of ROS1 and BRAF, among others. These biomarkers can provide additional information about the molecular characteristics and treatment response in NSCLC [27,50,51].

Besides genetic biomarkers, protein biomarkers also hold substantial importance in lung cancer. One such example is the assessment of programmed death-ligand 1 (PD-L1) expression in tumor cells, which serves as a biomarker to identify patients who might respond favorably to PD-1/PD-L1 inhibitor immunotherapy. In addition to biomarkers employed for treatment guidance, ongoing research focuses on identifying biomarkers for early diagnosis and detecting cancer recurrence in lung cancer patients. For instance, analysis of microRNAs (small non-coding ribonucleic acids) in tissue or bodily fluid samples can provide information about the presence and stage of lung cancer [32,52].

**Diagnostic biomarkers by immunohistochemistry.** These protein biomarkers, such as TTF-1 (Thyroid Transcription Factor-1) and p40, are directly analyzed in tissue biopsies and exhibit high specificity. TTF-1, encoded by the NKX2-1 gene, is involved in regulating the transcription of specific genes in the thyroid, lung, and diencephalon. Its presence provides valuable information about the origin of the tumor. On the other hand, p40 is a protein predominantly expressed in squamous cell carcinoma of the lung. When co-expressed, it allows for the differentiation between squamous cell carcinoma and adenocarcinoma, particularly in cases where there is poor differentiation in the initial pathological diagnosis [53]. A slightly less specific marker is Napsin A, which is a protein encoded by the NAPS1 gene and is useful for distinguishing adenocarcinoma from other lung cancers. For example, a tumor that is negative for TTF-1/p40 is unlikely to be a NSCLC [42].

**Circulating Biomarkers.** They are easy to measure, but studies have not shown high specificity, but rather sensitivity, and this has increased with the discovery of new protein and miRNA-type biomarkers (non-coding RNA, single-stranded molecules of 20 to 25 nucleotides in length, which can alter gene expression after transcription and play an important role in various biological functions and are highly stable). The latter have high potential as new specific and easily measurable biomarkers in liquid biopsy [54]. Among these, miR-205 and miR-375 have been identified as specific for squamous cell carcinoma and adenocarcinoma, respectively, and miR-93, miR-221, and miR-100 have been identified as potential biomarkers in adenocarcinoma [55]. A review has shown that miR-106a, miR-125a-5p, miR-129-3p, miR-205, miR-21, miR-29b, miR-375, and miR-7 may contribute to the differentiation between small cell carcinoma and non-small cell carcinoma. There is also clear and specific evidence for miRNAs in NSCLC, such as miR-19b-3p, which is related to the regulation of the HOXA9 gene (down-regulation), implicated in a negative prognosis [55], and miR-199a-5p, which is related to the regulation of the SLC2A1 gene (down-regulation) [56].

**Predictive Biomarkers.** These biomarkers are primarily associated with treatment considerations. Prominent organizations such as the College of American Pathologists (CAP), the International Association for the Study of Lung Cancer (IASLC), the Association for Molecular Pathology (AMP), the American Society of Clinical Oncology (ASCO), the European Society for Medical Oncology (ESMO), and the National Comprehensive Cancer Network (NCCN) have issued recommendations regarding the predictive value of biomarkers for patient monitoring. According to their guidelines, a minimum panel of biomarkers should be included, comprising EGFR, ROS1, ALK, BRAF, NTRK, MET, and RET, to assess the suitability of targeted therapies [55,57,58].

**Predictive Biomarkers for Immunotherapy.** Due to the evasion of the immune system by cancer cells, one of the primary mechanisms involves immune inhibitory pathways, particularly immune checkpoint proteins present on immune cells' surface. An example of such a protein is PD-L1 (Programmed Death Ligand 1) [52], which acts as a receptor on macrophages, antigen-presenting cells, B cells, T lymphocytes, and tumor cells, particularly in lung cancer. The upregulation of PD-L1 receptor expression deceives the immune system,

leading to the inhibition of the inflammatory response. Consequently, PD-L1 serves as a preferred biomarker for immunotherapy treatment [50].

**Biomarkers for Targeted Therapy.** Mutations affecting the KRAS gene, particularly G12C, are implicated in the proliferation of cells and the inhibition of apoptosis in NSCLC, primarily in adenocarcinoma [59]. Another biomarker of interest is Fibroblast Growth Factor Receptor 1 (FGFR1), a tyrosine kinase that regulates various cellular processes such as proliferation, differentiation, migration, and survival. Mutation in FGFR1 contributes to the promotion of mesenchymal-predominant NSCLC [60]. Additionally, Discoidin domain Receptor 2 (DDR2), a tyrosine kinase, is considered as a biomarker, as its mutation promotes tumor growth, particularly in 30% of squamous NSCLC [61]. Furthermore, the human epidermal growth factor receptor 2 (HER2), involved in activating cell proliferation pathways such as PI3K-AKT and MEK-ERK, is found in 10 to 30% of NSCLC cases, especially in adenocarcinoma [62].

Exosomal PD-L1 has emerged as a significant biomarker in the field of NSCLC immunotherapy monitoring. It has proven to be instrumental in advancing the effectiveness of immunotherapy by serving as an indicator of the tumor cells' adaptive response to T cells. By analyzing exosomal PD-L1, clinicians can stratify patients into responders and non-responders, aiding in treatment decision-making and prognosis assessment [63]. N6-methyladenine (m6A) is the main epigenetic RNA modification in eukaryotes, and there is significant evidence that m6A-associated proteins act as oncogenes and tumor suppressors, being abnormally expressed in NSCLC cell lines and tissues [64]. SFTA2, whose down-regulation indicates a poor prognosis in NSCLC patients, is related to increased inflammatory cells in some parts of the EMT, tumor heterogeneity, treatment efficiency, and the immune microenvironment of NSCLC [65]. Table 2 provides information on various biomarkers analyzed in NSCLC.

**Table 2.** Biomarkers in NSCLC.

Biomarkers in Nsclc					
Diagnostic Biomarkers in NSCLC					
Immunohistochemistry TTF-1 (Thyroid Transcription Factor-1) p40 Napsin A			Circulating tumor proteins Cytokeratin 19 fragment (CYFRA 21-1) Carcinoembryonic Antigen (CEA) Squamous cell carcinoma antigen (SCCA) Carbohydrate antigen 125 (CA125)		
microRNA (miRNA)					
miR-205	miR-106a	miR-29b			
miR-375	miR-125a-5p	miR-375			
miR-93	miR-129-3p	miR-7			
miR-221	miR-205	miR-19b-3p			
miR-100	miR-21	miR-199a-5p			
Predictive Biomarkers in NSCLC					
Targeted Therapy		For Immunotherapy		Novel predictive biomarkers	
Biomarker	Therapy	Biomarker	Antibody	Targeted therapy	Immunotherapy
EGFR	Afatinib	PD-1	Atezolizumab	KRAS	Exosome PD-L1 m6A methylation SFTA2 TIL's TIM-3 TMB
	Erlotinib		Pembrolizumab	FGFR1	
	Gefitinib		Durvalumab	DDR2	
	Osimertinib		Nivolumab	HER2	
ROS1	Entrectinib				
	Ceritinib				
	Crizotinib				
ALK	Alectinib				
	Crizotinib				
	Lorlatinib				
MET	Tepotinib				
	Capmatinib				
RET	Selpercatinib				
	Praseltinib				
BRAF	Trametinib				
	Dabrafenib+				
NTRK (1,2,3)	Larotrectinib				

### 3.4. Novel NSCLC Cancer Biomarkers

The identification of new biomarkers in the field of NSCLC has revolutionized the classification of the disease, leading to more accurate and personalized treatment strategies. In the past, NSCLC classification relied primarily on histological subtypes such as adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. However, this approach failed to capture the extensive heterogeneity and molecular diversity observed in NSCLC [66,67].

The advent of genomic technologies and high-throughput screening techniques has enabled the identification of specific genetic mutations, gene expression patterns, and protein markers that are associated with NSCLC. These biomarkers provide insights into the underlying molecular mechanisms driving tumor growth, metastasis, and treatment resistance. They also hold promise for predicting patient outcomes, guiding treatment decisions, and developing novel therapies. Furthermore, biomarkers associated with drug resistance, tumor recurrence, and metastasis are being actively investigated. Understanding the mechanisms underlying treatment resistance and disease progression is crucial for developing effective therapies and improving patient outcomes. Biomarkers such as TP53 mutations, VEGF expression, and TUBB3 levels, as mentioned earlier, have been linked to increased resistance to therapy, poor prognosis, and aggressive tumor behavior in NSCLC [66–70].

The incorporation of these biomarkers into clinical practice has the potential to revolutionize the management of NSCLC. They offer valuable insights into early detection, risk assessment, treatment selection, and treatment monitoring. By identifying patients who are likely to benefit from specific therapies or require alternative approaches, biomarkers enable a personalized and targeted approach to NSCLC management. It is important to acknowledge that the discovery and validation of biomarkers is an ongoing process. Rigorous research studies and clinical trials are necessary to establish their clinical usefulness, reliability, and reproducibility. Furthermore, the development of standardized testing methods and guidelines for biomarker evaluation is vital to ensure accurate and consistent outcomes across different healthcare settings and laboratories [67,71].

According to the analysis of Šutić et al. (2021), a group of predictive biomarkers is being studied, regarding the alteration of the KRAS gene, specifically the G12C mutation. Different drugs such as AMG510 (Sotorasib) and MRTX849 (Adagrasib) are being evaluated in patients with previously treated or metastatic NSCLC. These clinical trials, known as CodeBreak200 and KRYSTAL-12, aim to compare the effectiveness of these drugs with standard chemotherapy, such as Docetaxel, in patients with NSCLC. Another promising approach is the combination of selumetinib with docetaxel for patients with specific KRAS mutations in codons 12 or 13. The SELECT-1 trial is investigating the efficacy of this combination compared to placebo plus docetaxel in patients with locally advanced or metastatic NSCLC. Abemaciclib (LY2835219) is another drug currently being evaluated in the JUNIPER clinical trial. This study focuses on stage IV NSCLC patients who have experienced disease progression after receiving platinum-based chemotherapy and aims to compare the effectiveness of abemaciclib with erlotinib. Moreover, combinations of drugs are being explored in patients with specific genetic alterations. For instance, the NCT02743923 trial is comparing the efficacy of carboplatin, paclitaxel, and bevacizumab with cisplatin and pemetrexed in stage IIIB or IV NSCLC patients who are eligible for platinum-based chemotherapy and have not undergone prior chemotherapy. The amplification of the FGFR1 gene is also being investigated, and the use of dovitinib is being studied in previously treated advanced squamous cell NSCLC patients in the NCT01861197 trial, specifically targeting this subgroup of patients. Additionally, several ongoing clinical trials are evaluating the use of different drugs in patients with mutations or overexpression of the HER2 gene. For example, the NICHE trial is investigating the effectiveness of afatinib in previously treated advanced NSCLC patients. Another trial, DESTINYLung01, is assessing the use of trastuzumab deruxtecan (DS-8201a) in patients with unresectable and/or metastatic NSCLC with HER2 mutations or overexpression [71].

Prognostic biomarkers play a crucial role in predicting the outcome of patients with various diseases, including NSCLC. Table 3 provides a summary of some novel prognostic biomarkers and their associated outcomes [71].

**Table 3.** Novel prognostic biomarkers for NSCLC.

Biomarker	Outcome
TP53	Resistance to Therapy increased.
VEGF	Poor prognosis, metastasis, tumor recurrence.
TUBB3	Poor prognosis
KIAA1522	Poor prognosis and lower response rate
nm23-H1	Low levels poor prognosis
TGF- $\beta$	Poor prognosis
LAG-3	Better survival
NLR&PLR	Worse overall survival
Ki-67	Poor prognosis

**TP53:** The TP53 gene, which acts as a critical tumor suppressor gene, has been the subject of extensive research in different types of cancers, including NSCLC. Mutations in the TP53 gene have been found to be associated with heightened resistance to therapy in patients with NSCLC. Additionally, this gene plays a crucial role in regulating cellular growth, DNA repair processes, and programmed cell death (apoptosis). Mutations in this gene can disrupt these functions, leading to uncontrolled cell growth and impaired response to standard therapies. NSCLC patients with TP53 mutations may have a higher likelihood of treatment resistance and poorer treatment outcomes [68].

**VEGF** is a protein that plays a role in the process of angiogenesis, which is the formation of new blood vessels. In the context of NSCLC, elevated levels of VEGF have been linked to a less favorable prognosis. Elevated VEGF expression promotes the growth of blood vessels that supply tumors with essential nutrients and oxygen. This process facilitates tumor growth, metastasis, and recurrence. Measurement of this protein levels can provide insights into the aggressiveness of the disease and help guide treatment decisions [72].

**TUBB3 (Class III Beta-Tubulin):** TUBB3 is a protein involved in the organization and function of microtubules, which are essential components of the cell's structural framework and which are crucial for cell division and intracellular transport. Increased expression of this protein in NSCLC tumors has been linked to a poor prognosis. Elevated levels of TUBB3 have been correlated with a more aggressive tumor phenotype and resistance to specific chemotherapy drugs. As a result, TUBB3 holds promise as a potential biomarker for identifying patients who may require alternative treatment approaches or targeted therapies [69].

**KIAA1522:** Elevated expression of the KIAA1522 gene has been correlated with a poorer prognosis and lower response rates in NSCLC. KIAA1522, a gene involved in multiple cellular processes such as cell growth and proliferation, has been associated with a more aggressive tumor behavior and diminished responsiveness to therapy. Higher expression levels of KIAA1522 may serve as an indicator of a tumor's aggressive nature and decreased sensitivity to treatment options. Monitoring this gene levels can aid in predicting treatment response and guiding personalized treatment approaches [73].

**nm23-H1:** The presence of nm23-H1 has prognostic significance in NSCLC. The expression of this gene is responsible for regulating cell motility and the ability of cancer cells to spread and form metastases. When this biomarker is expressed at low levels in NSCLC patients, it has been associated with a less favorable prognosis. This suggests a higher risk of disease progression and the potential for metastasis to occur. Assessing nm23-H1 expression can help in determining the aggressiveness of the disease and informing treatment decisions [74].



## 4. Methodology for Discovering Novel Biomarkers

### 4.1. Artificial Intelligence Algorithms

Artificial intelligence algorithms have emerged as promising tools in the prediction of biomarkers in NSCLC. These algorithms rely on the ability to process large amounts of clinical, genetic, and molecular data to identify hidden patterns and relationships that may be relevant for biomarker prediction [75–77]. For example, the description of disease subgroups based on genomic and clinical data allows for accurate stratification through the application of machine learning tools and algorithms to the datasets, representing a revolution in disease stratification [78]. The identification of overall survival or relapse will require a robust and accurate predictive model, which in turn requires a large volume of complex data that need to be analyzed through machine learning algorithms to detect patterns and relationships among variables and generate predictive models that contribute to the outcome, such as susceptibility to a disease such NSCLC, its recurrence, and/or morbidity-mortality [76,79,80].

Machine learning models have been used for just over 30 years in cancer detection and diagnosis, primarily artificial neural networks and decision trees. In the past decade, they have also been employed in cancer prediction and prognosis (cancer susceptibility, recurrence, and survival). Some of the frequently employed supervised algorithms include logistic regression, decision trees, support vector machines, random forest, naive Bayes, k-nearest neighbors, and neural networks. On the other hand, unsupervised algorithms such as K-means, Principal Component Analysis (PCA), and nonlinear dimensionality reduction algorithms such as t-SNE (t-distributed stochastic neighbor embedding) and UMAP (Uniform Manifold approximation and projection) have been widely utilized [78,80–82] (see Table 4).

The application of artificial intelligence algorithms in NSCLC involves several stages. Firstly, clinical and molecular data from patients are collected, such as demographic information, medical histories, laboratory test results, and genetic profiles. These data are used to construct datasets that contain relevant features for biomarker prediction. Next, a suitable artificial intelligence algorithm is selected for the data analysis. This may include supervised machine learning algorithms such as logistic regression, support vector machines (SVM), or neural networks, which can learn from labeled data and make predictions based on those labels. Unsupervised learning algorithms such as clustering or dimensionality reduction can also be employed to discover patterns and clusters in the data without the need for prior labels. Once the algorithm is selected, it is trained using the collected dataset. During training, the algorithm analyzes the data and adjusts its parameters to find patterns and relationships that enable the prediction of the target biomarkers. This process may involve optimizing the algorithm’s hyperparameters and cross-validation to evaluate its performance on independent datasets. Once the algorithm is trained, it is used to make predictions on new data. This may involve evaluating new patients’ genetic profiles to predict the presence or absence of specific biomarkers, estimating the risk of disease progression or treatment response, or classifying patients into subgroups based on gene expression [76,79,80].

**Table 4.** Supervised and unsupervised machine learning algorithms used in biological and life sciences.

Algorithm	Advantages	Limitations	References
Supervised Algorithms			
Logistic Regression	High power for supervised classification with a dichotomous variable	Not useful for continuous variables	Yang, 2022 [83]
Support Vector Machine	Applied in non-linear models and survival prediction in cancer and demographic studies, among others. Good control of overfitting and good classifier	Complex algorithm structure. Training is slower.	Huang, 2022 [84]

Table 4. Cont.

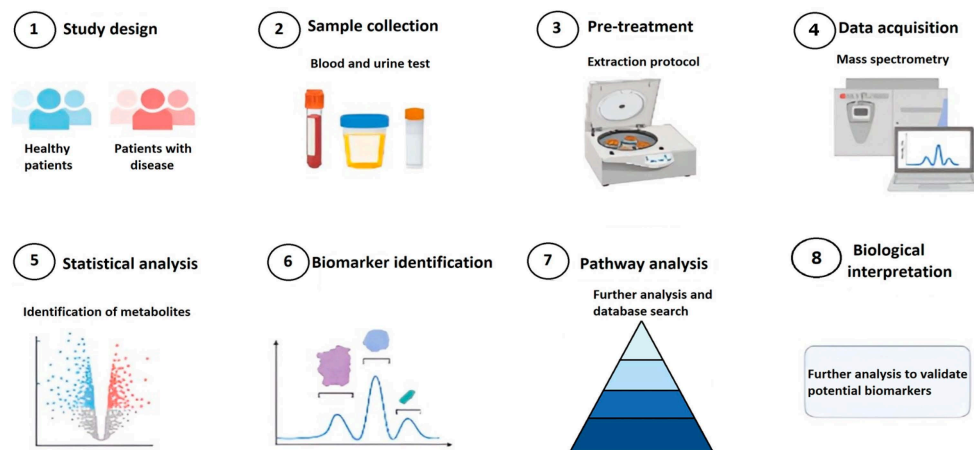
Algorithm	Advantages	Limitations	References
Decision Trees	Easy algorithm for data training. Used in diagnostic protocols	Can have overfitting problems, especially when there is a significant increase in branching in internal nodes	Lai, 2020 [75]; Batra, 2022 [85], 2022 [7]
Random Forest	Good predictive algorithm used in medicine in different imaging studies and recently in biomarker studies	May have overfitting problems	Batra, 2022 [85]; Handelman, 2018 [80]
Naïve Bayes	Still used in symptom characterization, complication prediction, imaging data, and demographic data.	As it is based on probabilistic statistical models, it can assume that attributes are independent. Redundant attributes can induce classification errors	Yang, 2023 [86]
K-Nearest Neighbor	Used as a classification and prediction algorithm in demographic models and genomic data, among others. Tolerant to noisy and missing data	Can assume that data attributes are equally important and may have similar classifications. Computationally complex with increasing data and attributes	Podolsky, 2016 [82]
Artificial Neural Networks	Algorithmic model capable of classifying and predicting based on a combination of parameters and applying it at the same time.	May have overfitting with too many attributes, and the optimal network structure is determined for experimentation	Lian, 2022 [87]; Civit-Masot, 2022 [88]
Unsupervised Algorithms			
K-Means	Widely used algorithm in biological and medical research and is easy to adapt and understand. Performs well on large datasets	The number of K needs to be manually assigned. Outliers can generate incorrect clusters. Scaling issues with the number of dimensions	Huang, 2021 [89]
Principal Component Analysis (PCA)	Linear dimensionality reduction algorithm that allows pattern observation and generates independent variables called principal components. Widely used in biological and genomic data observation	Does not allow non-linear dimensionality reduction. Lack of data standardization can be detrimental to results and information loss	Shin, 2018 [90]
t-SNE	Algorithm that enables visualization of high-dimensional datasets. Frequently used with PCA in biological and life sciences, primarily in omics analysis	Some issues when applied to non-linear parameter dimensionality reduction	Islam, 2021 [91]; Wang, 2021 [92]
UMAP	Next-generation algorithm that, similar to t-SNE, enables visualization of high-dimensional datasets. Offers higher accuracy when working with non-linear structures. Widely used in omics analysis	Currently limited to dimensional reduction due to its relative lack of familiarity	Islam, 2021 [91]; Nascimben, 2022 [93]

Machine learning algorithms have evolved according to the quantity and characteristics of the data and the objective being pursued, aiming for improved classification, prediction, and visualization. There are numerous publications in biological and life sciences that include one or multiple algorithms in their methodology and results, indicating an increasing possibility of obtaining easily interpretable predictive outcomes for individual and population health [90].

#### 4.2. Bioinformatics for Biomarkers Prediction

Bioinformatics plays a critical role in the identification and discovery of novel biomarkers by facilitating the efficient analysis and interpretation of vast amounts of biological data. It enables researchers to navigate through various phases of biomarker development, starting from the initial discovery phase, followed by analytical validation, evaluation of clinical utility, and ultimately leading to the clinical application of biomarkers (as shown in Figure 2). In the biomarker discovery phase, biomaterials obtained from cell assays, animal models, and clinical trials are useful, and their information (genome, transcriptome, epigenome, proteome, metabolome, and microbiome) is stored in databases such as the Gene Expression Omnibus (GEO). Statistical and bioinformatic methods are developed using this information to identify candidate biomarkers for further analytical validation using samples from cancer patients. In the validation phase, *in vitro* or *in vivo* molecular

tests such as qPCR or RT-PCR and immunoassays are performed to detect the biomarker. Additionally, statistical significance tests are conducted to validate and ensure reproducibility in its detection. In the clinical evaluation phase, randomized clinical trials are conducted to test whether these biomarkers have diagnostic, prognostic, or predictive utility. Once the biomarker's utility is demonstrated, it proceeds to the clinical use phase where it is commercialized with its respective protocol and complies with the corresponding regulatory processes [49,94–96].



**Figure 2.** Methodology for biomarker identification.

In the biomarker discovery phase, bioinformatics methods have gained relevance due to their ability to identify differentially expressed genes, proteins, microRNAs, and peptides in healthy and affected tissues, with the aim of characterizing which of these expressed genes can potentially be a cancer biomarker. The storage of information in databases such as GEO has facilitated computational bioprospecting for the discovery of new biomarkers. Within the information provided by the GEO database, it includes the type of data, number of samples, biological type, cell line type, and whether or not they received treatment. Prior to data analysis, the data must be preprocessed by removing duplicate data and normalizing using packages such as LIMMA in the R software [66,67,95,97]. In the discovery of biomarkers, not only the analysis of differential gene expression has been important but also the use of clustering algorithms such as K-means, co-expression analysis, and investigation of central genes using tools such as Cytoscape (<https://cytoscape.org/> accessed on 20 March 2023) or STRING (<https://string-db.org/> accessed on 20 March 2023), gene ontology analysis (DAVID: <https://david.ncifcrf.gov> accessed on 20 March 2023), KEGG (<https://www.genome.jp/kegg/pathway.html> accessed on 20 March 2023), Gene Ontology (GO: <http://geneontology.org/> accessed on 20 March 2023), and detection of associated metabolic pathways and co-expression/abundance of genes. The validation of candidate biomarker genes is performed using databases such as Gene Expression Profiling Interactive Analysis or GEPIA (<http://gepia.cancer-pku.cn/> accessed on 20 March 2023), The Cancer Genome Atlas or TCGA (<https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga> accessed on 20 March 2023), and The Human Protein Atlas or HPA (<https://www.proteinatlas.org/> accessed on 20 March 2023) [94,98,99]. The evaluation of biomarkers for cancer diagnosis can also be achieved by creating a prediction model based on the Support Vector Machine (SVM) classification algorithm, which creates a hyperplane that differentiates between two classes (presence and absence of the tumor) using two datasets, the training dataset and the test dataset. Survival analysis using Kaplan–Meier is often performed to assess the prognostic capacity of each biomarker gene and distinguish between low and high expression groups [99,100].

Below is a general methodology for discovering new biomarkers using bioinformatics tools [79,101,102]:

1. Genomic data identification: The first step is to gather relevant genomic data for the study. This may include publicly available datasets from genomic databases, such as GenBank or the Human Genome Project Repository, or internally generated data from experiments.
2. Data preprocessing: Once the genomic data are obtained, they need to undergo preprocessing to ensure data quality and prepare the data for analysis. This may involve data cleaning, normalization, and the removal of irrelevant or noisy data.
3. Gene expression analysis: One of the most common ways to discover biomarkers is through gene expression analysis. This involves comparing gene expression levels between different sample groups, such as samples from patients with and without a particular disease. Techniques such as microarrays or RNA sequencing can be used to measure gene expression.
4. Genetic variant analysis: Another strategy is to analyze genetic variants, such as mutations or polymorphisms, and their association with a specific condition. This can be achieved through the analysis of DNA sequencing data, where differences in genetic sequences among different sample groups are sought.
5. Data mining and statistical analysis: Once the preprocessed data are available, data mining and statistical analysis techniques can be applied to identify significant patterns and associations. This may include correlation analysis, enrichment analysis of biological pathways, gene interaction network analysis, and classification or clustering analysis.
6. Experimental validation: Biomarkers identified through bioinformatic analysis need to be experimentally validated to confirm their clinical relevance. This may involve the use of molecular biology techniques such as real-time PCR, Western blotting, or immunohistochemistry to verify the expression of the biomarkers in additional samples.
7. Clinical application: Once validated, the new biomarkers can be used in clinical studies to assess their utility in the diagnosis, prognosis, or treatment response of a specific disease. They can also be valuable for developing personalized therapies based on the presence or absence of certain biomarkers.

The following are the bioinformatics methodologies addressed in each research on biomarker discovery for various types of cancer (Table 5).

**Table 5.** Methodologies using bioinformatics for the prediction of biomarkers for different types of cancer. TCGA: The Cancer Genome Atlas, GEO: Gene Expression Omnibus, LIMMA: Linear Models for Microarray Data, STRING: Search Tool for the Retrieval of Interacting Genes/Proteins, GO: Gene Ontology, KEGG: Kyoto Encyclopedia of Genes and Genomes, FUNRICH: Functional Enrichment Analysis Tool, DAVID: Database for Annotation, Visualization, and Integrated Discovery, GEPIA: Gene Expression Profiling Interactive Analysis, Affy: Affymetrix, HMDD: Human MicroRNA Disease Database, BLAST: Basic Local Alignment Search Tool, miRDB: MicroRNA Target Prediction Database.

Author	Type of Cancer	Type of Data	Database	Data Preprocessing and Differentially Expressed Genes (DEGs)	MicroRNA Target Prediction	Protein–Protein Interaction (PPI)	Functional Enrichment Analysis	Validation
Zhang et al., 2020 [103]	Bladder	Gene	TCGA-BLCA, GEO	R software	Not Realized	CytoHubba, STRING	GO, KEGG, FUNRICH, DAVID	Oncomine database, GEPIA
Sarafidis et al., 2022 [95]	Bladder	Gene	GEO (metanalysis)	Affy, LIMMA R packages, outlier removal quality control	Not Realized	STRING, Cytoscape	GO, KEGG, Disease Ontology (DO), Reactome	GEPIA2, TCGA, Human Protein Atlas (proteomics, RNA-Seq), survival analysis
Pandi et al., 2022 [96]	Breast	Gene	GEO	R LIMMA package, GEO2R	Not Realized	STRING, Cytoscape	Enrichr, KEGG	TCGA-BRCA, GEPIA, survival analysis
Xu et al., 2022 [104]	Breast	Gene	GEO	R LIMMA package	Not Realized	STRING	GO, KEGG	GEPIA, survival analysis
Wu et al., 2022 [105]	Breast	Gene	TCGA	R LIMMA package	Not Realized	STRING	GO, clusterProfiler R tpackage	Breast Cancer Gene-Expression Miner v4.8 (bc-GenExMiner v4.8), survival analysis

Table 5. Cont.

Author	Type of Cancer	Type of Data	Database	Data Preprocessing and Differentially Expressed Genes (DEGs)	MicroRNA Target Prediction	Protein–Protein Interaction (PPI)	Functional Enrichment Analysis	Validation
Fadaka et al., 2019 [106]	Colon	microRNA	miRBase ( <a href="https://www.mirbase.org/">https://www.mirbase.org/</a> accessed on 20 March 2023), miR2Disease ( <a href="http://www.mir2disease.org/">http://www.mir2disease.org/</a> accessed on 20 March 2023), HMDD ( <a href="http://www.cuilab.cn/hmdd">http://www.cuilab.cn/hmdd</a> accessed on 20 March 2023), miRCancer ( <a href="http://mircancer.ecu.edu/">http://mircancer.ecu.edu/</a> accessed on 20 March 2023), BLAST	R software	miRDB ( <a href="http://www.mirdb.org/index.html">http://www.mirdb.org/index.html</a> accessed on 20 March 2023), TargetScan ( <a href="https://www.targetscan.org/vert_72/">https://www.targetscan.org/vert_72/</a> accessed on 20 March 2023) y mirDIP ( <a href="http://ophid.utoronto.ca/mirDIP/index.jsp">http://ophid.utoronto.ca/mirDIP/index.jsp</a> accessed on 20 March 2023)	STRING, Cytoscape	DAVID, GO, KEGG	Gene correlation (gbCRC) at <a href="http://gbcr.bioinformaticsminzhao.org/">http://gbcr.bioinformaticsminzhao.org/</a> accessed on 20 March 2023
Dai et al., 2019 [107]	Colon	Gene	GEO (systematic review)	R software, BRB array tools	Not Realized	STRING, Cytoscape	FunRich ( <a href="http://www.funrich.org/">http://www.funrich.org/</a> accessed on 20 March 2023), KEGG, DAVID	The Human Protein Atlas (HPA), The Cancer Genome Atlas (TCGA), survival analysis
Li et al., 2021 [108]	Colon	Gene	GEO	R limma package	Not Realized	STRING, Cytoscape	KEGG, DAVID	TCGA, GEPIA, survival analysis
Hammad et al., 2021 [77]	Colon	Gene	GEO	R software	Not Realized	STRING, Cytoscape	KEGG, DAVID	GEPIA, survival analysis (PROGgene)
Paksoy and Hilal, 2022 [109]	Colon	Gene	<a href="https://figshare.com/articles/dataset/The_microarray_dataset_of_colon_cancer_in_csv_format_/13658790/1">https://figshare.com/articles/dataset/The_microarray_dataset_of_colon_cancer_in_csv_format_/13658790/1</a> accessed on 20 March 2023	Synthetic Minority Oversampling Technique, or SMOTE method	Not Realized	Not Realized	Not Realized	Not Realized
Wang et al., 2021 [110]	Gastric	Gene	TCGA	R limma package	Not Realized	Cytoscape	GO, KEGG	GEPIA, survival analysis, Jiangsu Province Yixing People’s Hospital
Liu et al., 2022 [111]	Gastric	Gene	GEO, TCGA	R limma package, clustering analysis (Bioconductor)	Not Realized	STRING, Cytoscape	KEGG, DAVID	TCGA, survival analysis
Lvu et al., 2020 [97]	Glioma	mRNAsi	TCGA	EdgeR method	Not Realized	Not Realized	GO, KEGG	Survival analysis, Chinese Glioma Genome Atlas (CGGA) ( <a href="http://www.cgga.org.cn/">http://www.cgga.org.cn/</a> accessed on 20 March 2023)
Liao et al., 2020 [112]	Lung	mRNAsi	TCGA	R LIMMA package	Not Realized	STRING, Cytoscape	GO, KEGG, DAVID	GEPIA, survival analysis
Gong et al., 2021 [49]	NSCLC	Gene	GEO	GEO2R	Not Realized	STRING, Cytoscape	KEGG, DAVID	GEPIA, survival analysis, Oncomine database
Wu et al., 2019 [100]	Ovarian	microRNA	GEO (Systematic review and Metanalysis)	edgeR package of R	Not Realized		GO, KEGG, DAVID	Survival analysis
Chen et al., 2020 [99]	Ovarian	Gene	GEO	R software	Not Realized	GeneMANIA ( <a href="https://genemania.org/">https://genemania.org/</a> accessed on 20 March 2023)	KEGG, DAVID	Survival analysis, Dataset GSE9891
Zahra et al., 2022 [113]	Ovarian	Gene	TCGA, UK BioBank, cBioPortal	R software	Not Realized	Not Realized	Not Realized	Not Realized
Shi et al., 2022 [114]	Pancreas	Gene	GEO	GEO2R		STRING, Cytoscape	GO, KEGG, DAVID	GEPIA, Survival analysis
Yuan et al., 2017 [98]	Prostate	Gene	GEO	Affy, LIMMA R packages		STRING, Cytoscape	GO, DAVID	Protein Atlas Database, Oncomine database
Lombe et al., 2022 [94]	Prostate	microRNA	GEO	GENT2 ( <a href="http://gent2.appex.kr/gent2/">http://gent2.appex.kr/gent2/</a> accessed on 20 March 2023)	TargetScan Human, miRDB, DIANA microT	STRING, Cytoscape	GO, KEGG, DAVID	GEPIA, survival analysis
Author	Prediction of drug-gene interaction		Evaluation of pronostic biomarkers		Protein acquisition, 3D modeling and protein visualizer		In vitro Validation	
Zhang et al., 2020 [103]	Not Realized		Not Realized		Not Realized		Not Realized	
Sarafidis et al., 2022 [95]	Not Realized		Least Absolute Shrinkage and Selection Operator (LASSO) regression		Not Realized		Not Realized	
Pandi et al., 2022 [96]	Not Realized		Not Realized		Not Realized		Not Realized	
Xu et al., 2022 [104]	Not Realized		Not Realized		Not Realized		Not Realized	
Wu et al., 2022 [105]	DrugBank, Cytoscape		Not Realized		Not Realized		Not Realized	
Fadaka et al., 2019 [106]	Not Realized		PrognoScan ( <a href="http://dna00.bio.kyutech.ac.jp/PrognoScan/">http://dna00.bio.kyutech.ac.jp/PrognoScan/</a> accessed on 20 March 2023)		Not Realized		Not Realized	
Dai et al., 2020 [107]	Not Realized		Not Realized		Not Realized		Not Realized	
Li et al., 2021 [108]	Not Realized		Not Realized		Not Realized		Not Realized	
Hammad et al., 2021 [77]	Not Realized		Prediction model with Support Vector Machine (SVM classifier)		Not Realized		Not Realized	
Paksoy and Hilal, 2022 [109]	Not Realized		Random Forest, Desicion Trees, Gaussian Bayes		Not Realized		Not Realized	

Table 5. Cont.

Author	Type of Cancer	Type of Data	Database	Data Preprocessing and Differentially Expressed Genes (DEGs)	MicroRNA Target Prediction	Protein-Protein Interaction (PPI)	Functional Enrichment Analysis	Validation
Wang et al., 2021 [110]	Not Realized			Not Realized		Not Realized		Gastric cell lines (AGS, HGC27 and MKN45) and normal gastric mucosa cells, FISH, RNA extraction, qRT-PCR
Liu et al., 2022 [111]	Drug-Gene Interaction database (DGIdb), Cytoscape			Not Realized		Not Realized		Not Realized
Lvu et al., 2020 [97]	Not Realized			Estimation of mRNA <sub>si</sub> using one-class logistic regressionmachine learning (OCLR), Least Absolute Shrinkage and Selection Operator (LASSO) regression		Not Realized		Not Realized
Liao et al., 2020 [112]	Not Realized			Estimation of mRNA <sub>si</sub> using one-class logistic regressionmachine learning (OCLR)		Not Realized		Not Realized
Gong et al., 2021 [49]	Not Realized			Not Realized		Not Realized		A549 and HBE normal cells, via qPCR
Wu et al., 2019 [100]	Not Realized			Not Realized		Not Realized		Not Realized
Chen et al., 2020 [99]	Not Realized			Not Realized		Not Realized		Not Realized
Zahra et al., 2022 [113]	Not Realized			Not Realized		Uniprot, RSCB PDB (Protein Data Bank),Phyre2, Swissmodel, Alpha fold, Missense3D tool, YASARA, PYMOL, PROVEAN		Not Realized
Shi et al., 2022 [114]	Not Realized			Not Realized		Not Realized		Four PDA cell lines (AsPC-1, SW1990, PANC-1, and BxPC-3) and a normal human pancreatic ductal epithelial cell line (HPDE), qRT-CPR
Yuan et al., 2017 [98]	Not Realized			Not Realized		Not Realized		Not Realized
Lombe et al., 2022 [94]	Not Realized			Not Realized		Not Realized		MicroRNAs via qPCR

The study by Wang et al. (2022), using the methodology presented above, discovered the genes ABCA8, ADAMTS8, ASPA, CEP55, FHL1, PYCR1, RAMP3, and TPX2 as potential diagnostic biomarkers for NSCLC. Therefore, these genes can be considered as predictive biomarkers that may help predict the presence or risk of NSCLC. However, further studies are needed to determine if these genes can also be useful as prognostic biomarkers to predict the progression or treatment outcome of NSCLC [79].

On the other hand, the study of Lai et al. (2023) to identify key genes and potential biomarkers for NSCLC used integrated microarray datasets and various bioinformatics analyses. Four microarray datasets were combined, and differentially expressed genes (DEGs) were identified. Functional clustering and pathway enrichment analyses were performed, revealing enrichment in processes such as mitotic nuclear division and cell cycle regulation. Protein-protein interaction (PPI) network analysis identified central node genes, and their potential prognostic value was confirmed through survival analysis. The markers predicted in this study are listed below [101]:

- ANLN: ANLN is observed to be overexpressed in multiple tumor types, including pancreatic, brain, breast, and lung cancers. It is involved in cell proliferation, and its inhibition can impede cancer cell division, migration, and invasion. Overexpression of ANLN has been associated with lung adenocarcinoma metastasis, making it a potential target for cancer therapy.
- CDKN3: CDKN3 exhibits overexpression in glioma and cervical cancer and is linked to poorer survival outcomes. Its expression levels fluctuate during the cell cycle, peaking during mitosis. High levels of mitotic CDKN3 expression are often observed in various human cancers.
- CCNB1 and CCNB2: These genes play essential roles in meiotic resumption and have been implicated in tumor cell division, proliferation, and tumor growth in several cancer types, including colorectal, pancreatic, breast, hepatocellular carcinoma, and NSCLC.
- KIF4A: KIF4A is involved in DNA replication and repair processes and promotes cell proliferation. It is associated with tumor size in oral carcinoma and has potential prognostic value in various solid tumors.

- KIF11 and MELK: Both KIF11 and MELK have been identified as oncogenes in multiple cancers and are being investigated as potential targets for cancer treatment in ongoing phase I/II clinical trials.
- CEP55: CEP55 is considered a promising cancer vaccine candidate and serves as a marker for predicting cancer invasion risk, metastasis, and therapeutic response.
- HMMR: HMMR is a microtubule-associated protein that regulates mitosis and meiosis. Aberrant expression of HMMR disrupts the cell division process and is associated with cancer risk and progression across various tumor types.
- ASPM: ASPM has emerged as a predictor of tumor aggressiveness and prognosis in bladder, prostate, and endometrial cancers.
- CENPF: CENPF serves as a proliferative marker for malignant tumor cell growth.
- BUB1: BUB1 is a serine/threonine-protein kinase that plays a critical role in oncogenesis, chromosome arrangement, and spindle assembly.

## 5. Challenges and Future Perspectives

In the field of cancer biomarkers, a key challenge is the need for a predefined panel of diagnostic markers that can be efficiently tested on small tumor biopsies with a quick turnaround time. Additionally, the emergence of new biomarkers and limited awareness of available testing options contribute to gaps in testing. Expanding the testing to include a broader range of biomarkers is crucial for personalized treatment options and access to clinical trials. Another challenge lies in the standardization of ctDNA analysis, a minimally invasive approach for tumor profiling. While ctDNA analysis offers advantages such as easier blood sampling and faster results, there are issues regarding standardization, sensitivity, and technical aspects that need to be addressed, including the detection of gene fusions in plasma and cost-effectiveness of the testing [115].

The current trend in lung cancer biomarker detection aims to achieve more accurate results through less invasive procedures. However, this approach presents new challenges for pathologists. In practical scenarios, it can be difficult to combine all the necessary techniques when working with the biological samples received in the laboratory. The analysis of phenotypic and genotypic characteristics in small tissue specimens becomes particularly complex due to the tumor's heterogeneity, especially when using multiplex immunohistochemistry (IHC) or next-generation sequencing (NGS) approaches that involve comprehensive panels. Tumor heterogeneity is a dynamic phenomenon that can complicate the assessment of certain biomarkers during the follow-up of lung cancer patients. While automated tests offer the advantage of delivering rapid results, the cost associated with performing certain analyses systematically, such as NGS, is significant compared to the level of care provided based on the obtained results [116].

Ensuring access to novel diagnostic tools for all patients is another significant challenge. Barriers such as limited testing capabilities, geographic limitations, and guidelines hinder the availability of accurate and timely molecular diagnoses. Overcoming these barriers requires regional coordination, collaboration among accredited laboratories, and consideration of factors like turnaround time and cost-effectiveness. Reimbursement and financial issues must also be resolved to support the adoption of innovative diagnostic techniques. Lastly, the harmonization of molecular diagnostics in NSCLC involves collaboration and consensus among various stakeholders, including physicians, pathologists, molecular laboratories, and healthcare institutions. Achieving consensus on standard diagnostic tools and approaches, addressing performance variations, and involving the pharmaceutical and diagnostics industry through education and pilot projects are crucial for harmonization. This long-term process requires evidence-based innovation and shared responsibility to improve the implementation of biomarker testing and enhance patient outcomes in cancer management [115].

The adoption of AI in healthcare faces several challenges. The lack of transparency in how deep learning systems generate predictions, especially weakly supervised models, is a major concern. Techniques such as attention weights and knowledge distillation of-

fer potential solutions by providing insights into the network's decision-making process. Ensuring the generalizability of AI systems to specific data types is another challenge, as models may exhibit inflated performance for low-frequency events and unintended biases can arise from data generation. Addressing these issues requires careful assessment of predictions and consideration of potential sources of noise. The requirement for labeled data is also a hurdle, but crowdsourcing annotations and utilizing self-supervised or unsupervised learning techniques can alleviate the labeling burden. The specificity of domain knowledge in healthcare further complicates AI adoption, necessitating the development of networks that mimic medical practitioners' expertise and incorporating multi-modal data for more accurate predictions. Overcoming these challenges will contribute to the successful integration of AI in healthcare, leading to improved clinical outcomes [117].

## 6. Conclusions

NSCLC is a significant public health concern with high mortality rates. Advances in genomic data and bioinformatics tools have improved the possibilities for early diagnosis, effective treatment, and follow-up through the use of biomarkers. Biomarkers combined with clinical data can predict outcomes and guide specific treatments. However, the impact of new biomarkers on mortality and treatment efficacy in NSCLC is challenging due to limited evidence. The identification and application of relevant biomarkers, both new and existing, will shape the future of precision medicine in NSCLC.

Precision medicine, which focuses on personalized and targeted treatments based on individual genetic and molecular characteristics, has transformed the field of healthcare. It has led to the development of targeted therapies and improved treatment outcomes in oncology. Biomarkers play a crucial role in precision medicine by providing measurable indicators of disease characteristics and helping tailor treatment approaches. The integration of big data and AI in biomarker analysis and interpretation further enhances the potential for personalized medicine. However, challenges remain in terms of data analysis, interpretation, and the adoption of precision medicine approaches in clinical practice. Continued research and advancements in biomarker discovery and utilization are essential for further improving patient outcomes in lung cancer and other diseases.

**Author Contributions:** Conceptualization, Y.L. and J.C.R.; methodology, Y.L., J.C.R. and Z.C.; validation, Y.L. and Z.C.; formal analysis, Y.L., J.C.R., D.D. and Z.C.; investigation, Y.L., J.C.R. and Z.C.; resources, Y.L.; data curation, Y.L. and J.C.R.; writing—original draft preparation, Y.L. and J.C.R.; writing—review and editing, Y.L., D.D. and Z.C.; visualization J.C.R.; supervision Y.L.; project administration, Y.L.; funding acquisition Y.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research has been funded by Dirección General de Investigaciones de la Universidad Santiago de Cali, Convocatoria Interna No. 02–2023.

**Acknowledgments:** We thank Universidad Santiago de Cali for funding this manuscript.

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

## Abbreviations

NSCLC	Non-Small Cell Lung Cancer
PET	Positron Emission Tomography
GEO	Gene Expression Omnibus
TCGA	The Cancer Genome Atlas
MDS	Multidimensional Scaling
UMAP	Uniform Manifold Approximation and Projection
TP53	Tumor Protein 53
EGFR	Epidermal Growth Factor Receptor
ALK	Anaplastic Lymphoma Kinase
MRI	Magnetic Resonance Imaging
CT	Computed Tomography



FISH	Fluorescence In Situ Hybridization
PCR	Polymerase Chain Reaction
IHC	Immunohistochemistry
TKIs	Tyrosine Kinase Inhibitors
OPN	Osteopontin
PI3K/AKT/mTOR	Phosphoinositide 3-kinase/Protein Kinase B/Mammalian Target of Rapamycin
RET	Rearranged during Transfection
JAK-STAT	Janus Kinase-Signal Transducer and Activator of Transcription
RAS/MAPK	Rat Sarcoma/Mitogen-Activated Protein Kinase
MET	Mesenchymal Epithelial Transition
EMT	Epithelial-Mesenchymal Transition
HER2	Human Epidermal Growth Factor Receptor 2
MEK/ERK	Mitogen-Activated Protein Kinase Kinase/Extracellular Signal-Regulated Kinase
BRAF	B-Raf Proto-Oncogene, Serine/Threonine Kinase
ROS1	ROS Proto-Oncogene 1, Receptor Tyrosine Kinase
KRAS	Kirsten Rat Sarcoma Viral Oncogene Homolog
NTRK	Neurotrophic Tropomyosin Receptor Kinase
TME	Tumor Microenvironment
ICBs	Immune Checkpoint Inhibitors
STK11	Serine/Threonine Kinase 11
NGS	Next-Generation Sequencing
TNM	Tumor, Node, Metastasis
FDA	U.S. Food and Drug Administration
NIH	National Institutes of Health
DNA	Deoxyribonucleic Acid
SNPs	Single Nucleotide Polymorphisms
FDA-NIH	FDA-NIH Biomarker Working Group
PD-L1	Programmed Death-Ligand 1
PD-1	Programmed Death-1
NKX2-1	Thyroid Transcription Factor-1
TTF-1	Thyroid Transcription Factor-1
NAPSA	Napsin A
CYFRA 21-1	Cytokeratin 19 fragment
CEA	Carcinoembryonic Antigen
SCCA	Squamous cell carcinoma antigen
CA125	Carbohydrate antigen 125
miRNA	microRNA
FGFR1	Fibroblast Growth Factor Receptor 1
DDR2	Discoidin domain Receptor 2
m6A	N6-methyladenine
SFTA2	Surfactant Protein A2
ROS1	ROS Proto-Oncogene 1
ALK	Anaplastic Lymphoma Kinase
BRAF	B-Raf Proto-Oncogene
NTRK	Neurotrophic Tyrosine Kinase Receptor
TIM-3	T-cell Immunoglobulin and Mucin Domain Containing-3
TMB	Tumor Mutational Burden
AMP	Association for Molecular Pathology
CAP	College of American Pathologists
IASLC	International Association for the Study of Lung Cancer
ASCO	American Society of Clinical Oncology
ESMO	European Society for Medical Oncology
NCCN	National Comprehensive Cancer Network
VEGF	Vascular Endothelial Growth Factor
TUBB3	Tubulin Beta-3
KIAA1522	KIAA1522 protein
TGF- $\beta$	Transforming Growth Factor Beta
LAG-3	Lymphocyte Activation Gene-3
NLR	Neutrophil-to-Lymphocyte Ratio

PLR	Platelet-to-Lymphocyte Ratio
PCA	Principal Component Analysis
t-SNE	t-Distributed Stochastic Neighbor Embedding
SVM	Support Vector Machine
qPCR	Quantitative Polymerase Chain Reaction
RT-PCR	Reverse Transcription Polymerase Chain Reaction
R	Programming language/software for statistical computing and graphics
LIMMA	Linear Models for Microarray Data Analysis
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
Cytoscape	Software platform for visualizing molecular interaction networks
DAVID	Database for Annotation, Visualization, and Integrated Discovery
KEGG	Kyoto Encyclopedia of Genes and Genomes
GO	Gene Ontology
GEPIA	Gene Expression Profiling Interactive Analysis
HPA	The Human Protein Atlas

## References

1. Cancer Today. Available online: <http://gco.iarc.fr/today/home> (accessed on 31 May 2023).
2. Chen, L.; Li, H.; Xie, L.; Zuo, Z.; Tian, L.; Liu, C.; Guo, X. Editorial: Big Data and Machine Learning in Cancer Genomics. *Front. Genet.* **2021**, *12*, 749584. [[CrossRef](#)]
3. Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Holko, M.; et al. NCBI GEO: Archive for Functional Genomics Data Sets—Update. *Nucleic Acids Res.* **2012**, *41*, D991–D995. [[CrossRef](#)]
4. Meng, L.; Tian, Z.; Long, X.; Diao, T.; Hu, M.; Wang, M.; Zhang, W.; Zhang, Y.; Wang, J.; He, Y. Caspase 4 Overexpression as a Prognostic Marker in Clear Cell Renal Cell Carcinoma: A Study Based on the Cancer Genome Atlas Data Mining. *Front. Genet.* **2021**, *11*, 600248. [[CrossRef](#)] [[PubMed](#)]
5. Zhang, X.; Wang, J.; Zhuang, J.; Liu, C.; Gao, C.; Li, H.; Ma, X.; Li, J.; Sun, C. A Novel Glycolysis-Related Four-mRNA Signature for Predicting the Survival of Patients with Breast Cancer. *Front. Genet.* **2021**, *12*, 606937. [[CrossRef](#)] [[PubMed](#)]
6. Becht, E.; McInnes, L.; Healy, J.; Dutertre, C.-A.; Kwok, I.W.H.; Ng, L.G.; Ginhoux, F.; Newell, E.W. Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP. *Nat. Biotechnol.* **2019**, *37*, 38–44. [[CrossRef](#)]
7. Lai, G.G.Y.; Guo, R.; Drilon, A.; Shao Weng Tan, D. Refining Patient Selection of MET-Activated Non-Small Cell Lung Cancer through Biomarker Precision. *Cancer Treat. Rev.* **2022**, *110*, 102444. [[CrossRef](#)]
8. Yang, S.-R.; Schultheis, A.M.; Yu, H.; Mandelker, D.; Ladanyi, M.; Büttner, R. Precision Medicine in Non-Small Cell Lung Cancer: Current Applications and Future Directions. *Semin. Cancer Biol.* **2022**, *84*, 184–198. [[CrossRef](#)] [[PubMed](#)]
9. Zappa, C.; Mousa, S.A. Non-Small Cell Lung Cancer: Current Treatment and Future Advances. *Transl. Lung Cancer Res.* **2016**, *5*, 288–300. [[CrossRef](#)]
10. Wang, S.; Zimmermann, S.; Parikh, K.; Mansfield, A.S.; Adjei, A.A. Current Diagnosis and Management of Small-Cell Lung Cancer. *Mayo Clin. Proc.* **2019**, *94*, 1599–1622. [[CrossRef](#)]
11. Mithoowani, H.; Febbraro, M. Non-Small-Cell Lung Cancer in 2022: A Review for General Practitioners in Oncology. *Curr. Oncol.* **2022**, *29*, 1828–1839. [[CrossRef](#)]
12. Herbst, R.S.; Morgensztern, D.; Boshoff, C. The Biology and Management of Non-Small Cell Lung Cancer. *Nature* **2018**, *553*, 446–454. [[CrossRef](#)] [[PubMed](#)]
13. Molina, J.R.; Yang, P.; Cassivi, S.D.; Schild, S.E.; Adjei, A.A. Non-Small Cell Lung Cancer: Epidemiology, Risk Factors, Treatment, and Survivorship. *Mayo Clin. Proc.* **2009**, *83*, 584–594. [[CrossRef](#)]
14. Duma, N.; Santana-Davila, R.; Molina, J.R. Non-Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. *Mayo Clin. Proc.* **2019**, *94*, 1623–1640. [[CrossRef](#)]
15. Clark, S.B.; Alsubait, S. Non Small Cell Lung Cancer. In *StatPearls*; StatPearls Publishing: Tampa, FL, USA, 2022.
16. Giraldo-Osorio, A.; Ruano-Ravina, A.; Rey-Brandariz, J.; Arias-Ortiz, N.; Candal-Pedreira, C.; Pérez-Ríos, M. Tendencias en la mortalidad por cáncer de pulmón en Colombia, 1985–2018. *Rev. Panam. Salud Pública* **2022**, *46*, 1. [[CrossRef](#)]
17. Alarcón, M.-L.; Brugés, R.; Carvajal, C.; Vallejo, C.; Beltrán, R. Características de los pacientes con cáncer de pulmón de célula no pequeña en el Instituto Nacional de Cancerología de Bogotá. *Rev. Colomb. Cancerol.* **2021**, *25*, 103–109. [[CrossRef](#)]
18. Guo, Q.; Liu, L.; Chen, Z.; Fan, Y.; Zhou, Y.; Yuan, Z.; Zhang, W. Current Treatments for Non-Small Cell Lung Cancer. *Front. Oncol.* **2022**, *12*, 945102. [[CrossRef](#)]
19. Sosa Iglesias, V.; Giuranno, L.; Dubois, L.J.; Theys, J.; Vooijs, M. Drug Resistance in Non-Small Cell Lung Cancer: A Potential for NOTCH Targeting? *Front. Oncol.* **2018**, *8*, 267. [[CrossRef](#)] [[PubMed](#)]
20. Wu, J.; Lin, Z. Non-Small Cell Lung Cancer Targeted Therapy: Drugs and Mechanisms of Drug Resistance. *IJMS* **2022**, *23*, 15056. [[CrossRef](#)]
21. Błach, J.; Wojas-Krawczyk, K.; Nicoś, M.; Krawczyk, P. Failure of Immunotherapy-The Molecular and Immunological Origin of Immunotherapy Resistance in Lung Cancer. *Int. J. Mol. Sci.* **2021**, *22*, 9030. [[CrossRef](#)]

22. Schubart, C.; Stöhr, R.; Tögel, L.; Fuchs, F.; Sirbu, H.; Seitz, G.; Seggewiss-Bernhardt, R.; Leistner, R.; Sterlacci, W.; Vieth, M.; et al. MET Amplification in Non-Small Cell Lung Cancer (NSCLC)—A Consecutive Evaluation Using Next-Generation Sequencing (NGS) in a Real-World Setting. *Cancers* **2021**, *13*, 5023. [[CrossRef](#)]
23. Rosellini, P.; Amintas, S.; Caumont, C.; Veillon, R.; Galland-Girodet, S.; Cuguillière, A.; Nguyen, L.; Domblides, C.; Gouverneur, A.; Merlio, J.-P.; et al. Clinical Impact of STK11 Mutation in Advanced-Stage Non-Small Cell Lung Cancer. *Eur. J. Cancer* **2022**, *172*, 85–95. [[CrossRef](#)] [[PubMed](#)]
24. Su, C. Emerging Insights to Lung Cancer Drug Resistance. *Cancer Drug Resist.* **2022**, *5*, 534–540. [[CrossRef](#)] [[PubMed](#)]
25. Sumbly, V.; Landry, I. Unraveling the Role of STK11/LKB1 in Non-Small Cell Lung Cancer. *Cureus* **2022**, *14*, e21078. [[CrossRef](#)] [[PubMed](#)]
26. Taylor, B.C.; Balko, J.M. Mechanisms of MHC-I Downregulation and Role in Immunotherapy Response. *Front. Immunol.* **2022**, *13*, 844866. [[CrossRef](#)] [[PubMed](#)]
27. Yan, N.; Guo, S.; Zhang, H.; Zhang, Z.; Shen, S.; Li, X. BRAF-Mutated Non-Small Cell Lung Cancer: Current Treatment Status and Future Perspective. *Front. Oncol.* **2022**, *12*, 863043. [[CrossRef](#)] [[PubMed](#)]
28. Zhang, L.; Zheng, L.; Yang, Q.; Sun, J. The Evolution of BRAF Activation in Non-Small-Cell Lung Cancer. *Front. Oncol.* **2022**, *12*, 882940. [[CrossRef](#)]
29. Nam, M.-W.; Kim, C.-W.; Choi, K.-C. Epithelial-Mesenchymal Transition-Inducing Factors Involved in the Progression of Lung Cancers. *Biomol. Ther.* **2022**, *30*, 213–220. [[CrossRef](#)]
30. Chae, Y.K.; Chang, S.; Ko, T.; Anker, J.; Agte, S.; Iams, W.; Choi, W.M.; Lee, K.; Cruz, M. Epithelial-Mesenchymal Transition (EMT) Signature Is Inversely Associated with T-Cell Infiltration in Non-Small Cell Lung Cancer (NSCLC). *Sci. Rep.* **2018**, *8*, 2918. [[CrossRef](#)]
31. Greenspan, B.S. Role of PET/CT for Precision Medicine in Lung Cancer: Perspective of the Society of Nuclear Medicine and Molecular Imaging. *Transl. Lung Cancer Res.* **2017**, *6*, 617–620. [[CrossRef](#)]
32. Inamura, K. Update on Immunohistochemistry for the Diagnosis of Lung Cancer. *Cancers* **2018**, *10*, 72. [[CrossRef](#)]
33. Ginsburg, G.S.; Phillips, K.A. Precision Medicine: From Science To Value. *Health Aff.* **2018**, *37*, 694–701. [[CrossRef](#)] [[PubMed](#)]
34. Denny, J.C.; Collins, F.S. Precision Medicine in 2030—Seven Ways to Transform Healthcare. *Cell* **2021**, *184*, 1415–1419. [[CrossRef](#)] [[PubMed](#)]
35. Naithani, N.; Sinha, S.; Misra, P.; Vasudevan, B.; Sahu, R. Precision Medicine: Concept and Tools. *Med. J. Armed Forces India* **2021**, *77*, 249–257. [[CrossRef](#)] [[PubMed](#)]
36. König, I.R.; Fuchs, O.; Hansen, G.; Von Mutius, E.; Kopp, M.V. What Is Precision Medicine? *Eur. Respir. J.* **2017**, *50*, 1700391. [[CrossRef](#)]
37. George, A.P.; Kuzel, T.M.; Zhang, Y.; Zhang, B. The Discovery of Biomarkers in Cancer Immunotherapy. *Comput. Struct. Biotechnol. J.* **2019**, *17*, 484–497. [[CrossRef](#)]
38. Ielapi, N.; Andreucci, M.; Licastro, N.; Faga, T.; Grande, R.; Buffone, G.; Mellace, S.; Sapienza, P.; Serra, R. Precision Medicine and Precision Nursing: The Era of Biomarkers and Precision Health. *IJGM* **2020**, *13*, 1705–1711. [[CrossRef](#)]
39. Slikker, W. Biomarkers and Their Impact on Precision Medicine. *Exp. Biol. Med.* **2018**, *243*, 211–212. [[CrossRef](#)]
40. Hulsen, T.; Jamuar, S.S.; Moody, A.R.; Karnes, J.H.; Varga, O.; Hedensted, S.; Spreafico, R.; Hafler, D.A.; McKinney, E.F. From Big Data to Precision Medicine. *Front. Med.* **2019**, *6*, 34. [[CrossRef](#)]
41. Johnson, K.B.; Wei, W.; Weeraratne, D.; Frisse, M.E.; Misulis, K.; Rhee, K.; Zhao, J.; Snowdon, J.L. Precision Medicine, AI, and the Future of Personalized Health Care. *Clin. Transl. Sci.* **2021**, *14*, 86–93. [[CrossRef](#)]
42. Califf, R.M. Biomarker Definitions and Their Applications. *Exp. Biol. Med.* **2018**, *243*, 213–221. [[CrossRef](#)]
43. Henry, N.L.; Hayes, D.F. Cancer Biomarkers. *Mol. Oncol.* **2012**, *6*, 140–146. [[CrossRef](#)] [[PubMed](#)]
44. Shah, A.; Grimberg, D.C.; Inman, B.A. Classification of Molecular Biomarkers. *Société Int. Urol. J.* **2020**, *1*, 8–15. [[CrossRef](#)]
45. Berk, M. The Classification of Biomarkers. *JAMA Psychiatry* **2015**, *72*, 1056–1057. [[CrossRef](#)] [[PubMed](#)]
46. Rifai, N.; Gillette, M.A.; Carr, S.A. Protein Biomarker Discovery and Validation: The Long and Uncertain Path to Clinical Utility. *Nat. Biotechnol.* **2006**, *24*, 971–983. [[CrossRef](#)]
47. Qiu, S.; Cai, Y.; Yao, H.; Lin, C.; Xie, Y.; Tang, S.; Zhang, A. Small Molecule Metabolites: Discovery of Biomarkers and Therapeutic Targets. *Signal Transduct. Target. Ther.* **2023**, *8*, 132. [[CrossRef](#)]
48. Hoseok, I.; Cho, J.-Y. Lung Cancer Biomarkers. In *Advances in Clinical Chemistry*; Elsevier: Amsterdam, The Netherlands, 2015; Volume 72, ISBN 978-0-12-803314-2.
49. Gong, K.; Zhou, H.; Liu, H.; Xie, T.; Luo, Y.; Guo, H.; Chen, J.; Tan, Z.; Yang, Y.; Xie, L. Identification and Integrate Analysis of Key Biomarkers for Diagnosis and Prognosis of Non-Small Cell Lung Cancer Based on Bioinformatics Analysis. *Technol. Cancer Res. Treat.* **2021**, *20*, 107–170. [[CrossRef](#)]
50. Román, M.; Baraibar, I.; López, I.; Nadal, E.; Rolfo, C.; Vicent, S.; Gil-Bazo, I. KRAS Oncogene in Non-Small Cell Lung Cancer: Clinical Perspectives on the Treatment of an Old Target. *Mol. Cancer* **2018**, *17*, 33. [[CrossRef](#)]
51. Gendarme, S.; Bylicki, O.; Chouaid, C.; Guisier, F. ROS-1 Fusions in Non-Small-Cell Lung Cancer: Evidence to Date. *Curr. Oncol.* **2022**, *29*, 641–658. [[CrossRef](#)]
52. Grigg, C.; Rizvi, N.A. PD-L1 Biomarker Testing for Non-Small Cell Lung Cancer: Truth or Fiction? *J. Immunother. Cancer* **2016**, *4*, 48. [[CrossRef](#)]

53. Affandi, K.A.; Tizen, N.M.S.; Mustangin, M.; Zin, R.R.M.R.M. P40 Immunohistochemistry Is an Excellent Marker in Primary Lung Squamous Cell Carcinoma. *J. Pathol. Transl. Med.* **2018**, *52*, 283–289. [[CrossRef](#)]
54. Pan, J.; Zhou, C.; Zhao, X.; He, J.; Tian, H.; Shen, W.; Han, Y.; Chen, J.; Fang, S.; Meng, X.; et al. A Two-MiRNA Signature (MiR-33a-5p and MiR-128-3p) in Whole Blood as Potential Biomarker for Early Diagnosis of Lung Cancer. *Sci. Rep.* **2018**, *8*, 16699. [[CrossRef](#)]
55. Xu, Y.; Chai, B.; Wang, X.; Wu, Z.; Gu, Z.; Liu, X.; Zhao, Y.; Chen, T.; Ma, Z.; Sun, Q. MiRNA-199a-5p/SLC2A1 Axis Regulates Glucose Metabolism in Non-Small Cell Lung Cancer. *J. Cancer* **2022**, *13*, 2352–2361. [[CrossRef](#)] [[PubMed](#)]
56. Pennell, N.A.; Arcila, M.E.; Gandara, D.R.; West, H. Biomarker Testing for Patients with Advanced Non-Small Cell Lung Cancer: Real-World Issues and Tough Choices. *Am. Soc. Clin. Oncol. Educ. Book* **2019**, 531–542. [[CrossRef](#)]
57. Ettinger, D.S.; Wood, D.E.; Aisner, D.L.; Akerley, W.; Bauman, J.R.; Bharat, A.; Bruno, D.S.; Chang, J.Y.; Chirieac, L.R.; D’Amico, T.A.; et al. NCCN Guidelines Insights: Non-Small Cell Lung Cancer, Version 2.2021: Featured Updates to the NCCN Guidelines. *J. Natl. Compr. Cancer Netw.* **2021**, *19*, 254–266. [[CrossRef](#)] [[PubMed](#)]
58. Wu, Y.-L.; Planchard, D.; Lu, S.; Sun, H.; Yamamoto, N.; Kim, D.-W.; Tan, D.S.W.; Yang, J.C.-H.; Azrif, M.; Mitsudomi, T.; et al. Pan-Asian Adapted Clinical Practice Guidelines for the Management of Patients with Metastatic Non-Small-Cell Lung Cancer: A CSCO-ESMO Initiative Endorsed by JSMO, KSMO, MOS, SSO and TOS. *Ann. Oncol.* **2019**, *30*, 171–210. [[CrossRef](#)]
59. Moes-Sosnowska, J.; Skupinska, M.; Lechowicz, U.; Szczepulska-Wojcik, E.; Skronska, P.; Rozy, A.; Stepniewska, A.; Langfort, R.; Rudzinski, P.; Orłowski, T.; et al. FGFR1–4 RNA-Based Gene Alteration and Expression Analysis in Squamous Non-Small Cell Lung Cancer. *IJMS* **2022**, *23*, 10506. [[CrossRef](#)] [[PubMed](#)]
60. Dhakar, R.; Dakal, T.C.; Sharma, A. Genetic Determinants of Lung Cancer: Understanding the Oncogenic Potential of Somatic Missense Mutations. *Genomics* **2022**, *114*, 110401. [[CrossRef](#)]
61. Uy, N.F.; Merkhofer, C.M.; Baik, C.S. HER2 in Non-Small Cell Lung Cancer: A Review of Emerging Therapies. *Cancers* **2022**, *14*, 4155. [[CrossRef](#)]
62. Wang, N.; Song, X. Research Advances of Immunotherapy of Exosome PD-L1 in Non-Small Cell Lung Cancer. *Chin. J. Lung Cancer* **2022**, *25*, 689–695. [[CrossRef](#)]
63. Qiu, F.-S.; He, J.-Q.; Zhong, Y.-S.; Guo, M.-Y.; Yu, C.-H. Implications of M6A Methylation and Microbiota Interaction in Non-Small Cell Lung Cancer: From Basics to Therapeutics. *Front. Cell. Infect. Microbiol.* **2022**, *12*, 972655. [[CrossRef](#)]
64. Li, N.; Zhai, Z.; Chen, Y.; Li, X. Transcriptomic and Immunologic Implications of the Epithelial–Mesenchymal Transition Model Reveal a Novel Role of SFTA2 in Prognosis of Non-Small-Cell Lung Carcinoma. *Front. Genet.* **2022**, *13*, 911801. [[CrossRef](#)]
65. Tutar, Y. MiRNA and Cancer; Computational and Experimental Approaches. *Curr. Pharm. Biotechnol.* **2014**, *15*, 429. [[CrossRef](#)]
66. Wang, N.; Zhang, H.; Li, D.; Jiang, C.; Zhao, H.; Teng, Y. Identification of Novel Biomarkers in Breast Cancer via Integrated Bioinformatics Analysis of Potential Biomarkers for Pancreatic Cancer and Experimental Validation. *Bioengineered* **2021**, *12*, 12431–12446. [[CrossRef](#)]
67. Wu, Q.; Zhang, B.; Sun, Y.; Xu, R.; Hu, X.; Ren, S.; Ma, Q.; Chen, C.; Shu, J.; Qi, F.; et al. Identification of Novel Biomarkers and Candidate Small Molecule Drugs in Non-Small-Cell Lung Cancer by Integrated Microarray Analysis. *OncoTargets Ther.* **2019**, *12*, 3545–3563. [[CrossRef](#)]
68. Canale, M.; Andrikou, K.; Priano, I.; Cravero, P.; Pasini, L.; Urbini, M.; Delmonte, A.; Crinò, L.; Bronte, G.; Ulivi, P. The Role of TP53 Mutations in EGFR-Mutated Non-Small-Cell Lung Cancer: Clinical Significance and Implications for Therapy. *Cancers* **2022**, *14*, 1143. [[CrossRef](#)]
69. McCarroll, J.A.; Gan, P.P.; Erlich, R.B.; Liu, M.; Dwarto, T.; Sagnella, S.S.; Akerfeldt, M.C.; Yang, L.; Parker, A.L.; Chang, M.H.; et al. TUBB3/BIII-Tubulin Acts through the PTEN/AKT Signaling Axis to Promote Tumorigenesis and Anoikis Resistance in Non-Small Cell Lung Cancer. *Cancer Res.* **2015**, *75*, 415–425. [[CrossRef](#)] [[PubMed](#)]
70. Kong, X.-Q.; Huang, Y.-X.; Li, J.-L.; Zhang, X.-Q.; Peng, Q.-Q.; Tang, L.-R.; Wu, J.-X. Prognostic Value of Vascular Endothelial Growth Factor Receptor 1 and Class III  $\beta$ -Tubulin in Survival for Non-Metastatic Rectal Cancer. *WJGO* **2018**, *10*, 351–359. [[CrossRef](#)] [[PubMed](#)]
71. Šutić, M.; Vukić, A.; Baranašić, J.; Försti, A.; Džubur, F.; Samaržija, M.; Jakopović, M.; Brčić, L.; Knežević, J. Diagnostic, Predictive, and Prognostic Biomarkers in Non-Small Cell Lung Cancer (NSCLC) Management. *JPM* **2021**, *11*, 1102. [[CrossRef](#)] [[PubMed](#)]
72. Villaruz, L.C.; Socinski, M.A. The Role of Anti-Angiogenesis in Non-Small-Cell Lung Cancer: An Update. *Curr. Oncol. Rep.* **2015**, *17*, 26. [[CrossRef](#)]
73. Liu, Y.-Z.; Yang, H.; Cao, J.; Jiang, Y.-Y.; Hao, J.-J.; Xu, X.; Cai, Y.; Wang, M.-R. KIAA1522 Is a Novel Prognostic Biomarker in Patients with Non-Small Cell Lung Cancer. *Sci. Rep.* **2016**, *6*, 24786. [[CrossRef](#)]
74. Hung, C.-Y.; Wang, Y.-C.; Chuang, J.-Y.; Young, M.-J.; Liaw, H.; Chang, W.-C.; Hung, J.-J. Nm23-H1-Stabilized HnRNPA2/B1 Promotes Internal Ribosomal Entry Site (IRES)-Mediated Translation of Sp1 in the Lung Cancer Progression. *Sci. Rep.* **2017**, *7*, 9166. [[CrossRef](#)]
75. Lai, Y.-H.; Chen, W.-N.; Hsu, T.-C.; Lin, C.; Tsao, Y.; Wu, S. Overall Survival Prediction of Non-Small Cell Lung Cancer by Integrating Microarray and Clinical Data with Deep Learning. *Sci. Rep.* **2020**, *10*, 4679. [[CrossRef](#)]
76. Chen, J.H.; Asch, S.M. Machine Learning and Prediction in Medicine—Beyond the Peak of Inflated Expectations. *N. Engl. J. Med.* **2017**, *376*, 2507–2509. [[CrossRef](#)] [[PubMed](#)]
77. Hammad, A.; Elshaer, M.; Tang, X. Identification of Potential Biomarkers with Colorectal Cancer Based on Bioinformatics Analysis and Machine Learning. *Math. Biosci. Eng.* **2021**, *18*, 8997–9015. [[CrossRef](#)] [[PubMed](#)]

78. Arslan, S.; Mehrotra, D.; Schmidt, J.; Geraldles, A.; Singhal, S.; Hense, J.; Li, X.; Bass, C.; Kather, J.N.; Pandya, P.; et al. Deep Learning Can Predict Multi-Omic Biomarkers from Routine Pathology Images: A Systematic Large-Scale Study. *bioRxiv* **2022**. *preprint*.
79. Wang, F.; Su, Q.; Li, C. Identification of Novel Biomarkers in Non-Small Cell Lung Cancer Using Machine Learning. *Sci. Rep.* **2022**, *12*, 16693. [[CrossRef](#)]
80. Handelman, G.S.; Kok, H.K.; Chandra, R.V.; Razavi, A.H.; Lee, M.J.; Asadi, H. EDoctor: Machine Learning and the Future of Medicine. *J. Intern. Med.* **2018**, *284*, 603–619. [[CrossRef](#)]
81. Margolis, R.; Derr, L.; Dunn, M.; Huerta, M.; Larkin, J.; Sheehan, J.; Guyer, M.; Green, E.D. The National Institutes of Health's Big Data to Knowledge (BD2K) Initiative: Capitalizing on Biomedical Big Data. *J. Am. Med. Inform. Assoc.* **2014**, *21*, 957–958. [[CrossRef](#)]
82. Podolsky, M.D.; Barchuk, A.A.; Kuznetsov, V.I.; Gusarova, N.F.; Gaidukov, V.S.; Tarakanov, S.A. Evaluation of Machine Learning Algorithm Utilization for Lung Cancer Classification Based on Gene Expression Levels. *Asian Pac. J. Cancer Prev.* **2016**, *17*, 835–838. [[CrossRef](#)]
83. Yang, Y.; Xu, L.; Sun, L.; Zhang, P.; Farid, S.S. Machine Learning Application in Personalised Lung Cancer Recurrence and Survivability Prediction. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 1811–1820. [[CrossRef](#)]
84. Huang, Y.; Huang, S.; Liu, Z. Multi-Task Learning-Based Feature Selection and Classification Models for Glioblastoma and Solitary Brain Metastases. *Front. Oncol.* **2022**, *12*, 1000471. [[CrossRef](#)] [[PubMed](#)]
85. Batra, U.; Nathany, S.; Sharma, M.; Mehta, A.; Dhanda, S.; Jose, J.T. Machine Learning-Based Algorithm Demonstrates Differences in Del19 and L858R EGFR Subgroups in Non-Small Cell Lung Cancer: A Single Center Experience. *Am. J. Transl. Res.* **2022**, *14*, 2677.
86. Yang, J.; Yee, P.L.; Khan, A.A.; Khan, M.S.; Karamti, H.; Aldweesh, A.; Hussain, L.; Omar, A. Improving Lung Cancer Detection Based on Hybrid Features and Employing Machine Learning Techniques. *Res. Sq.* **2023**. *preprint*. [[CrossRef](#)]
87. Lian, J.; Deng, J.; Hui, E.S.; Koohi-Moghadam, M.; She, Y.; Chen, C.; Vardhanabhuti, V. Early Stage NSCLC Patients' Prognostic Prediction with Multi-Information Using Transformer and Graph Neural Network Model. *eLife* **2022**, *11*, e80547. [[CrossRef](#)] [[PubMed](#)]
88. Civit-Masot, J.; Bañuls-Beaterio, A.; Domínguez-Morales, M.; Rivas-Pérez, M.; Muñoz-Saavedra, L.; Rodríguez Corral, J.M. Non-Small Cell Lung Cancer Diagnosis Aid with Histopathological Images Using Explainable Deep Learning Techniques. *Comput. Methods Programs Biomed.* **2022**, *226*, 107108. [[CrossRef](#)]
89. Huang, Z.; Wang, S.; Zhang, H.-J.; Zhou, Y.L.; Tang, X.; Shi, J.-H. Characteristics of Hypoxic Tumor Microenvironment in Non-Small Cell Lung Cancer, Involving Molecular Patterns and Prognostic Signature. *Transl. Lung Cancer Res.* **2021**, *10*, 2132–2147. [[CrossRef](#)]
90. Shin, H.; Jeong, H.; Park, J.; Hong, S.; Choi, Y. Correlation between Cancerous Exosomes and Protein Markers Based on Surface-Enhanced Raman Spectroscopy (SERS) and Principal Component Analysis (PCA). *ACS Sens.* **2018**, *3*, 2637–2643. [[CrossRef](#)]
91. Islam, M.T.; Xing, L. Geometry and Statistics-Preserving Manifold Embedding for Nonlinear Dimensionality Reduction. *Pattern Recognit. Lett.* **2021**, *151*, 155–162. [[CrossRef](#)]
92. Wang, F.; Zhang, Y.; Hao, Y.; Li, X.; Qi, Y.; Xin, M.; Xiao, Q.; Wang, P. Characterizing the Metabolic and Immune Landscape of Non-Small Cell Lung Cancer Reveals Prognostic Biomarkers Through Omics Data Integration. *Front. Cell Dev. Biol.* **2021**, *9*, 702112. [[CrossRef](#)]
93. Nascimben, M.; Rimondini, L.; Corà, D.; Venturin, M. Polygenic Risk Modeling of Tumor Stage and Survival in Bladder Cancer. *BioData Min.* **2022**, *15*, 23. [[CrossRef](#)]
94. Lombe, C.P.; Meyer, M.; Pretorius, A. Bioinformatics Prediction and Analysis of MicroRNAs and Their Targets as Biomarkers for Prostate Cancer: A Preliminary Study. *Mol. Biotechnol.* **2022**, *64*, 401–412. [[CrossRef](#)]
95. Sarafidis, M.; Lambrou, G.I.; Zoumpourlis, V.; Koutsouris, D. An Integrated Bioinformatics Analysis towards the Identification of Diagnostic, Prognostic, and Predictive Key Biomarkers for Urinary Bladder Cancer. *Cancers* **2022**, *14*, 3358. [[CrossRef](#)]
96. Pandi, J.; Arulprakasam, A.; Dhandapani, R.; Ramanathan, S.; Thangavelu, S.; Chinnappan, J.; Vidhya Rajalakshmi, V.; Alghamdi, S.; Shesha, N.T.; Prasath, S. Biomarkers for Breast Adenocarcinoma Using In Silico Approaches. *Evid.-Based Complement. Altern. Med.* **2022**, *2022*, 825272. [[CrossRef](#)]
97. Lv, W.; Fei, X.; Chen, C.; Zhang, B. In Silico Identification of the Prognostic Biomarkers and Therapeutic Targets Associated with Cancer Stem Cell Characteristics of Glioma. *Biosci. Rep.* **2020**, *40*, BSR20201037. [[CrossRef](#)]
98. Yuan, L.; Cao, X.; Chen, L.; Zhu, Y.; Qian, G.; Xiao, Y. Identification of Potential Biomarkers in Human Prostate Cancer Using *Bioinform. Anal.* **2017**, *1*, 6.
99. Chen, J.; Cai, Y.; Xu, R.; Pan, J.; Zhou, J.; Mei, J. Identification of Four Hub Genes as Promising Biomarkers to Evaluate the Prognosis of Ovarian Cancer in Silico. *Cancer Cell. Int.* **2020**, *20*, 270. [[CrossRef](#)]
100. Wu, L.; Shang, W.; Zhao, H.; Rong, G.; Zhang, Y.; Xu, T.; Zhang, J.; Huang, P.; Wang, F. In Silico Screening of Circulating MicroRNAs as Potential Biomarkers for the Diagnosis of Ovarian Cancer. *Dis. Markers* **2019**, *2019*, 1–12. [[CrossRef](#)]
101. Lai, L.-T.; Ren, Y.-H.; Huai, Y.-J.; Liu, Y.; Liu, Y.; Wang, S.-S.; Mei, J.-H. Identification and Validation of Novel Prognostic Biomarkers and Therapeutic Targets for Non-Small Cell Lung Cancer. *Front. Genet.* **2023**, *14*, 1139994. [[CrossRef](#)]
102. Lavanya, C.; Pooja, S.; Kashyap, A.H.; Rahaman, A.; Niranjana, S.; Niranjana, V. Novel Biomarker Prediction for Lung Cancer Using Random Forest Classifiers. *Cancer Inform.* **2023**, *22*, 117693512311679. [[CrossRef](#)]

103. Zhang, C.; Berndt-Paetz, M.; Neuhaus, J. Bioinformatics Analysis Identifying Key Biomarkers in Bladder Cancer. *Data* **2020**, *5*, 38. [[CrossRef](#)]
104. Xu, Y.; Niu, C. Screening of Breast Cancer Biomarkers Based on Bioinformatics Analysis. *Res. Sq.* **2022**. preprint. [[CrossRef](#)]
105. Wu, Z.; Wan, J.; Wang, J.; Meng, X.; Qian, H. Identification of Prognostic Biomarkers for Breast Cancer Brain Metastases Based on the Bioinformatics Analysis. *Biochem. Biophys. Rep.* **2022**, *29*, 101203. [[CrossRef](#)]
106. Fadaka, A.O.; Klein, A.; Pretorius, A. In Silico Identification of MicroRNAs as Candidate Colorectal Cancer Biomarkers. *Tumour Biol.* **2019**, *41*, 101042831988372. [[CrossRef](#)]
107. Dai, G.; Wang, L.; Wen, Y.; Ren, X.; Zuo, S. Identification of Key Genes for Predicting Colorectal Cancer Prognosis by Integrated Bioinformatics Analysis. *Oncol. Lett.* **2019**, *19*, 388–398. [[CrossRef](#)]
108. Li, H.; Liu, L.; Jiang, Y.; Jiang, H.; Niu, C.; Xia, J.; Liu, Z.; Guo, Y. Integrated Bioinformatics Revealed 2 Novel Prognostic Biomarkers in Colorectal Cancer. *Int. J. Clin. Exp. Med.* **2021**, *14*, 2250–2258.
109. Paksoy, N.; Yağın, F.H. Artificial Intelligence-Based Colon Cancer Prediction by Identifying Genomic Biomarkers. *Med. Rec.* **2022**, *4*, 196–202. [[CrossRef](#)]
110. Wang, W.; Min, K.; Chen, G.; Zhang, H.; Deng, J.; Lv, M.; Cao, Z.; Zhou, Y. Use of Bioinformatic Database Analysis and Specimen Verification to Identify Novel Biomarkers Predicting Gastric Cancer Metastasis. *J. Cancer* **2021**, *12*, 5967–5976. [[CrossRef](#)]
111. Liu, Y.; Wang, D.-X.; Wan, X.-J.; Meng, X.-H. Identification of Candidate Biomarkers Associated with Gastric Cancer Prognosis Based on an Integrated Bioinformatics Analysis. *J. Gastrointest. Oncol.* **2022**, *13*, 1690–1700. [[CrossRef](#)]
112. Liao, Y.; Xiao, H.; Cheng, M.; Fan, X. Bioinformatics Analysis Reveals Biomarkers with Cancer Stem Cell Characteristics in Lung Squamous Cell Carcinoma. *Front. Genet.* **2020**, *11*, 427. [[CrossRef](#)]
113. Zahra, A.; Hall, M.; Chatterjee, J.; Sisu, C.; Karteris, E. In Silico Study to Predict the Structural and Functional Consequences of SNPs on Biomarkers of Ovarian Cancer (OC) and BPA Exposure-Associated OC. *IJMS* **2022**, *23*, 1725. [[CrossRef](#)]
114. Shi, H.; Xu, H.; Chai, C.; Qin, Z.; Zhou, W. Integrated Bioinformatics Analysis of Potential Biomarkers for Pancreatic Cancer. *J. Clin. Lab. Anal.* **2022**, *36*, e24381. [[CrossRef](#)]
115. Van Den Broek, D.; Hiltermann, T.J.N.; Biesma, B.; Dinjens, W.N.M.; 'T Hart, N.A.; Hinrichs, J.W.J.; Leers, M.P.G.; Monkhorst, K.; Van Oosterhout, M.; Scharnhorst, V.; et al. Implementation of Novel Molecular Biomarkers for Non-Small Cell Lung Cancer in the Netherlands: How to Deal with Increasing Complexity. *Front. Oncol.* **2020**, *9*, 1521. [[CrossRef](#)]
116. Hofman, P. The Challenges of Evaluating Predictive Biomarkers Using Small Biopsy Tissue Samples and Liquid Biopsies from Non-Small Cell Lung Cancer Patients. *J. Thorac. Dis.* **2019**, *11*, S57–S64. [[CrossRef](#)]
117. Mandair, D.; Reis-Filho, J.S.; Ashworth, A. Biological Insights and Novel Biomarker Discovery through Deep Learning Approaches in Breast Cancer Histopathology. *NPJ Breast Cancer* **2023**, *9*, 21. [[CrossRef](#)]

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