



# **Multi-Omics Mining of IncRNAs with Biological and Clinical Relevance in Cancer**

Ivan Salido-Guadarrama <sup>1,\*,†</sup>, Sandra L. Romero-Cordoba <sup>2,3,†</sup> and Bertha Rueda-Zarazua <sup>4</sup>

- <sup>1</sup> Departamento de Bioinformatica y Análisis Estadísticos, Instituto Nacional de Perinatología Isidro Espinosa de los Reyes, Mexico City 11000, Mexico
- <sup>2</sup> Departamento de Medicina Genómica y Toxicología Ambiental, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico; sromero@iibiomedicas.unam.mx
- <sup>3</sup> Biochemistry Department, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City 14080, Mexico
- <sup>4</sup> Posgrado en Ciencias Biológicas, Facultad de Medicina, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico; bgruedazara@gmail.com
- \* Correspondence: silvervann@gmail.com
- <sup>+</sup> These authors contributed equally to this work.

**Abstract:** In this review, we provide a general overview of the current panorama of mining strategies for multi-omics data to investigate lncRNAs with an actual or potential role as biological markers in cancer. Several multi-omics studies focusing on lncRNAs have been performed in the past with varying scopes. Nevertheless, many questions remain regarding the pragmatic application of different molecular technologies and bioinformatics algorithms for mining multi-omics data. Here, we attempt to address some of the less discussed aspects of the practical applications using different study designs for incorporating bioinformatics and statistical analyses of multi-omics data. Finally, we discuss the potential improvements and new paradigms aimed at unraveling the role and utility of lncRNAs in cancer and their potential use as molecular markers for cancer diagnosis and outcome prediction.

Keywords: long non-coding RNAs; cancer; multi-omics



Romero-Cordoba, S.L.;

Citation: Salido-Guadarrama, I.;

Rueda-Zarazua, B. Multi-Omics

Mining of IncRNAs with Biological

and Clinical Relevance in Cancer. Int.

*J. Mol. Sci.* **2023**, *24*, 16600. https:// doi.org/10.3390/ijms242316600

Academic Editor: Alain Couvineau

Received: 28 September 2023

Revised: 14 November 2023

Accepted: 15 November 2023

Published: 22 November 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/

(i)

(cc)

4.0/).

### 1. Introduction

The advent of next-generation sequencing and other revolutionary technologies for the study of omics, such as genomics, transcriptomics, proteomics, and metabolomics, has completely transformed the way basic and clinical cancer research is conducted [1,2]. As a consequence of these fast-moving advances, we have been able to produce large volumes of data from an assortment of multiple omics layers (i.e., genomics, epigenomics, transcriptomics, proteomics, metabolomics, and microbiomics) from different tissue and cell models for the study of basic and translational research in cancer [3,4]. In particular, the exponential advances in sequencing technology have allowed researchers to explore the complexity and diversity of human transcriptomes [5,6]. One population of RNA molecules that has received special attention in recent years are long non-coding RNAs (lncRNAs), which are molecules without protein-coding capacities but with versatile and pleiotropic functions [7,8], with well-defined action mechanisms linked to different hallmarks of cancer [9–11]. lncRNAs are increasingly being recognized as potential pharmacological targets and as diagnostic and prognostic biomarkers [12–14]. Indeed, advances in RNA-seq and other large-scale methodologies represent a valuable resource for deepening our knowledge of the molecular aspects of lncRNAs and their roles in cancer biology [15,16]. So far, most studies have primarily centered on detecting aberrant expression changes of lncRNAs. However, given the complex and pleiotropic functions of lncRNAs, we must recognize that to truly harness the potential of these transcripts to reveal their complex functional aspects and associations with extrinsic and intrinsic factors that influence the susceptibility, incidence, and survival of different types of cancer, we need to incorporate new

frameworks for multidimensional analyses, as have been proposed for the study of other complex diseases [17]. In this context, multi-omics analysis approaches that combine and integrate protein-coding gene expressions with other dimensions of molecular information, such as genomics and epigenomic, have improved our capacity for obtaining significant understanding of the molecular aberrations underpinning the oncologic characteristics in different types and subtypes of tumors (e.g., cancer cell fate and survival) [4,18]. Likewise, a multi-omics framework progressively incorporated into research is helping us to delineate the way lncRNAs perform their functions and how they impact cancer development [19]. As a new paradigm, these approaches might boost research to produce sufficient evidence to ascribe them a causal relationship to the malignancy presence and most importantly, a clinical value as diagnostics, prognostic, and predictive markers in cancer [20,21].

In this review we do not intend to cover an exhaustive compendium of the many IncRNAs and their biological and molecular roles in cancer, nor of the large number of bioinformatics programs that have been developed for the analysis of transcriptomics, genomics, and epigenomics data, both of which have been extensively discussed elsewhere [18,22–24]. Instead, we will try to shed light on some of the less discussed aspects of the practical applications using different integrative, comparative, and multidimensional study designs incorporating bioinformatics and statistical analysis of multi-omics data. Finally, we will discuss the possible improvements and new paradigms aimed at unraveling the activity of lncRNAs in cancer, and their potential use as molecular markers in clinical and biological behaviors. Throughout the following sections, we present and discuss the recent scientific evidence employing different omics strategies for the functional and biological study of lncRNAs in cancer research, with special interest in their informative role in oncology programs. Next, we focus our discussion on machine learning and biostatistics solutions that have been incorporated as part of methodological pipelines for quantifying the performance of lncRNAs signatures in the classification of cancer samples and the prediction of response and survival in cancer patients. Finally, we discuss future scenarios of new technological advances for the discovery and profiling of lncRNAs in cancer, and the challenges of translating these advances into real-world clinical settings.

The complexity of genomic alterations, gene expression, protein–protein interactions, epigenetic mechanisms, cellular secretome, and metabolome is a complex subject, especially when we want to define how their components interact (Figure 1). For this reason, and considering our main interest, in this work, we searched for literature in the PubMed database using the keywords ("lncRNA" or "long non-coding RNA" and "cancer" and "multi-omics") up to October 2023. Subsequently, we manually revised our search to focus mainly on the different approaches for multi-omics integration with the scope of finding and evaluating functional lncRNAs signatures with a clinical prognosis and prediction of patient outcome. The integration of the data can occur from a simple to a more complex view of cellular function. We will start with the simplest view, which focuses on describing the combination of phenotypes observed at different biological levels independently.



**Figure 1.** Representation of the various molecular features with changes and alterations at the DNA, RNA, and protein levels that can occur in cancer cells, representing the spectrum of multi-omics states that can be interrogated by high-throughput methods. DNA is the substrate for somatic aberrations, such as point mutations (arrow pointing at yellow spark) and somatic copy-number alterations (SCNA). DNA is also the molecule where the methylation of 5-methy-cytosine in different patterns has consequences in epigenomic regulation. On the other hand, the transcriptome, which is composed of the many RNA species in a cell, including the coding and non-coding genes (i.e., mRNAs, lncRNAs, and microRNAs), undergoes various dynamic changes that affect the levels of these species in different cells and tissues. Changes in the expression of various RNA molecules, such as lncRNAs, have become a potential repertoire of functional molecular markers for different types of cancer. Similarly, proteins and metabolites can also become informative indicators of the oncogenic processes occurring in a cell, tissue, organ, or physiological system in the body. Although graphically represented as separate compartments, these molecular dimensions are interconnected, and changes in the state in one dimension have consequences in the others and vice versa.

### 2. IncRNAs Associated with Cancer Driver Somatic Mutations Profiles

Cancer is characterized by the somatic acquisition of various cellular alterations that lead to selective advantages, such as unrestricted growth, the suppression of apoptosis,

and enhanced metabolism [25], and genomic instability has been widely recognized as one of the leading hallmarks of cancer [26]. Recently, especially using high-throughput sequencing, research has revealed the complexity of the somatic DNA aberration landscapes of cancer genomes [27]. In fact, defining specific somatic mutational profiles has long been a keystone for characterizing cancer patients and leading to the discovery of specific cancer driver genes [28], which in turn can help estimate the inherited risk, prioritize therapy and prognosis, and understand pharmacologic responses to drug treatments [29]. Not surprisingly, given the protein-centric view that has prevailed in many branches of molecular biology in recent years, cancer studies have focused more on exploring the mutations that occur in protein-coding regions, despite the fact that the exome represents a very small fraction (1-2%) of the human genome (Figure 1). Excitingly, with the expansion of our knowledge about other functional regions within the human genome other than the regions with protein-coding loci, a growing number of discoveries have now emerged demonstrating that recurrent somatic mutations in non-protein-coding regions can also be informative features, paving the way for discovering causal links between lncRNAs functions and oncogenesis [30]. By leveraging whole-genome sequencing data, researchers have been able to better characterize the presence, frequency, and functional impacts of somatic mutations in the non-coding regions of cancer genomes [31]. To our current knowledge, the most robust evidence has shown that cancer driver point mutations and structural variants are less common in non-coding genes and regulatory sequences than in protein-coding genes. The availability of whole-genome sequencing data and capture panels that include non-coding regulatory regions from projects such as TCGA, along with other valuable efforts, has led to the discovery of driving somatic mutations in non-coding regions that affect the function of genes other than the protein-coding genes [32–35]. For example, Rheinbay et al. sequenced 360 primary breast tumors and patient-matched normal samples, and observed cancer driver mutations in the regulatory non-coding regions that could indeed occur at frequencies similar to the coding regions. However, they also found promoter elements with a significant burden or clustering of mutations, including promoters of lncRNA genes, such as RMRP8 and NEAT1. Notably, three of the four mutations present in the NEAT1 promoter element induced a reduction in the NEAT1 expression, thus adding new loss-offunction alterations affecting NEAT1 in breast cancer [36]. This study represents one of the most sophisticated integrative efforts performed to date. The approach used by the authors emphasized the discovery of somatic alterations that did not occur at the loci encoding the transcripts of lncRNAs, but in the regulatory regions that controlled their transcription (Figure 2). A broader view of the state-of-the-art approaches that integrate multiple omics data to characterize the somatic point mutations at loci-encoding lncRNAs that may lead to a gain or a loss of function are discussed in Table 1.



**Figure 2.** Multi-omics approaches used in cancer research and interrogated at a multidimensional level for the characterization of the biological, functional, and pathophysiological properties of lncRNA. In multi-omics, different modalities for the study design have been proposed, mainly combining two or more dimensions of molecular features (i.e., genomics, transcriptomics, epigenomics, proteomics, and metabolomics). For that purpose, bioinformatics and statistical frameworks have been also envisaged by researchers to mine information from the integrative analysis of these multidimensional data structures (e.g., post-analysis and early integration analysis). In combination with bioinformatics pipelines, different statistical and machine learning frameworks have been implemented, which are aimed at constructing and refining models to predict the outcome of cancer patients, such as free survival, time to relapse, and treatment response.

		8	I	I		71	
IncRNA of Relevance from the Study	Potential Association Based on Biological or Clinical Data	Goal	Main Insight Regarding the lncRNA	Model Construction	Kaplan–Meier and Log-Rank Test	Receiver Operator Curve AUC	Reference
ESR1, TRPS1, ERG, RUNX1, SNHG16, and HOTAIR	Various cancer types	Molecular and biological characterization of the regulatory mutations on functional lncRNAs in cancer	Somatic mutations on the lncRNA TF binding site are associated with their expression and activity in cancer	NA	NA	NA	[32]
CASC8	Breast cancer	Prioritize and predict lncRNAs with a functional impact through mutation analysis	Somatic mutations on the lncRNA loci site are associated with expression alterations that have a functional impact	NA	NA	NA	[37]
ENSG0000021403, ENSG00000261650, ENSG00000281406, and G001643	Colorectal carcinoma	Molecular and biological characterization of the regulatory mutations on functional lncRNAs in cancer	Somatic mutations on the lncRNA loci site are associated with expression alterations that have a functional impact	NA	NA	NA	[38]
LINC00460, AC156455.1, AC015977.2, 'PRDM16-dt', AL139351.1, AL035661.1, and LINC01606	Renal cell carcinoma	Identify the genome instability-related lncRNAs and their clinical significance	Discovery of 11 lncRNAs related to mutational burden and associated with patient poor overall survival	Univariable and multi-variable COX prognostic regression model	High-risk vs. low-risk groups; p < 0.001	Train AUC = 0.743 Test = 0.770	[39]
LINC00460, LINC01234	Clear cell renal carcinoma	Identify the genome instability-related IncRNAs and their clinical significance	Identification of a IncRNAs signature related to the mutational burden and associated with patient poor overall survival	Univariable and multi-variable COX prognostic regression model	High-risk vs. low-risk groups; p < 0.001	AUC = 0.681	[40]

**Table 1.** Research studies conducted, in which multi-omics approaches were used to identify and characterize the biological and clinico-pathological features of different lncRNAs signatures and evaluate their performance as predictors of patient clinical outcomes for various cancer types. NA: No information available.

	Table 1. Cont.						
IncRNA of Relevance from the Study	Potential Association Based on Biological or Clinical Data	Goal	Main Insight Regarding the IncRNA	Model Construction	Kaplan–Meier and Log-Rank Test	Receiver Operator Curve AUC	Reference
FAM30A, CACNA1C-AS1, LINC02595, LINC00926, AL589863.1, and AP000919.3	AML	Construct a somatic mutation-associated risk index	IncRNAs related to the mutational burden and associated with patient poor overall survival	Selection of candidate lncRNAs using LASSO followed by the univariable and multi-variable COX prognostic regression model	High-risk vs. low-risk groups; p < 0.001	AUC = 0.804	[41]
AC007996.1, AC009237.14, AP0035555.1, and AL590483.1	Colorectal carcinoma	Evaluate the performance of a genome stability-related IncRNA signature as a risk predictor	Identification of a lncRNAs signature related to the mutational burden and associated with patient poor overall survival	Univariable and multi-variable COX prognostic regression model	High-risk vs. low-risk groups at 3 years; <i>p</i> < 0.001	AUC = 0.713	[42]
ZNF503-AS1, AL353747.2, AC129492.1, AP003555.1, and AC009237.14	Colorectal carcinoma	Evaluate the performance of a genome stability-related IncRNA signature as a risk predictor	Identification of a lncRNAs signature related to the mutational burden and immune infiltration that are associated with patient poor overall survival	Univariable and multi-variable COX prognostic regression model	High-risk vs. low-risk groups; p < 0.001; in the validation set	AUC (1 year) = 0.750; AUC (3 years) = 0.757, AUC (5 years) = 0.711; in the validation set	[43]
AC107464.2, MIR100HG, and AP001527.2	Cervical carcinoma	Evaluate the performance of a genome stability-related lncRNA signature as a risk predictor	Identification of a lncRNAs signature related to the mutational burden and associated with patient poor overall survival	Univariable and multi-variable COX prognostic regression model	High-risk (1.7 years) vs. low-risk (1.5 years) groups; p < 0.001; in the validation set	AUC (3 years) = 0.663; in the validation set	[44]

	Table 1. Cont.						
IncRNA of Relevance from the Study	Potential Association Based on Biological or Clinical Data	Goal	Main Insight Regarding the IncRNA	Model Construction	Kaplan–Meier and Log-Rank Test	Receiver Operator Curve AUC	Reference
AC002511.2, LINC00501, LINC02055, LINC02714, LINC01508, LOC105371967, RP11_96A15.1, RP11_305F18.1, RP11_342M1.3, RP11_432J24.3, and U95743.1	Hepatocellular carcinoma	Evaluate the performance of a genome stability-related IncRNA signature as a risk predictor	Identification of a IncRNAs signature related to the mutational burden and associated with patient poor overall survival	Selection of candidate lncRNAs using LASSO followed by the univariable and multi-variable COX prognostic regression model	High-risk vs. low-risk groups; p < 0.001	NA	[45]
C116351.1, ZFPM2-AS1, AC145343.1, and MIR210HG	Hepatocellular carcinoma	Evaluate the performance of a genome stability-related IncRNA signature as a risk predictor	Identification of a IncRNAs signature related to the mutational burden and associated with patient poor overall survival	Univariable and multi-variable COX prognostic regression model	High-risk vs. low-risk groups at 3 and 5 years; <i>p</i> < 0.001	AUC (3 years) = 0.710, AUC (5 years) = 0.707; in the validation set	[46]
BCAL8	Breast cancer	Molecular and biological characterization of the functional regulatory mutations on cancer	Identification of a somatic SCNA-related lncRNAs signature associated with patient prognosis	Selection of candidate lncRNAs using filtering methods	NA	NA	[47]
RUSC1-AS1, LINC01990, LINC01411, LINC02099, H19, LINC00452, ADPGK-AS1, and C1QTNF1-AS1	Cervical carcinoma	Evaluate the performance of a SCNA-related lncRNA signature as a risk predictor	Identification of a somatic SCNA-related lncRNAs signature with a value as independent tumor-free survival	Univariable and multi-variable COX prognostic regression model	High-risk vs. low-risk groups at 1, 3, and 5 years; <i>p</i> < 0.001	AUC (1 year), AUC (3 years), and AUC (5 years) > 0.750	[48]

	Table 1. Cont.						
IncRNA of Relevance from the Study	Potential Association Based on Biological or Clinical Data	Goal	Main Insight Regarding the lncRNA	Model Construction	Kaplan–Meier and Log-Rank Test	Receiver Operator Curve AUC	Reference
PRAL	Tumor suppressor	Molecular and biological characterization of SCNA-related lncRNAs in cancer	Identification of a somatic SCNA-related lncRNA with a value as an independent predictor for reduced tumor-free survival	Univariable and multi-variable COX prognostic regression model	High-risk vs. low-risk groups; p < 0.001	NA	[49]
LOC101927604, LOC105377267, CASC15, LINC-PINT, CLDN10-AS1, C14orf132, LMF1, LINC00675, CCDC144NL-AS1, and LOC284454	Colorectal carcinoma	Evaluate the performance of an SCNA-related lncRNA signature as a risk predictor for CRC	Identification of somatic SCNA-related lncRNA with a value as an independent predictor for reduced tumor-free survival	Clustering based on the gene expression, SCNA, and DNA methylation followed by the univariable and multi-variable COX prognostic regression model	High-risk vs. low-risk groups; p < 0.001	NA	[50]
RP11-571M6.8	Glioblastoma	Molecular and biological characterization of SCNA-related lncRNAs in cancer	Identification of a somatic SCNA-related lncRNA that is significantly predictive of disease-free survival	Selection of candidate lncRNAs using filtering methods followed by the univariable and multi-variable COX prognostic regression model	High-risk vs. low-risk groups; p < 0.001	NA	[51]
RP11-1020A11.1	Bladder carcinoma	Molecular and biological characterization of SCNA-related lncRNAs with risk prediction utility in cancer	Identification of a somatic SCNA-related lncRNA that is significantly predictive of disease-free survival	Selection of candidate lncRNAs using filtering methods followed by the univariable and multi-variable COX prognostic regression model	High-risk vs. low-risk groups; p < 0.001	NA	[51]

9 of 33

	Table 1. Cont.						
IncRNA of Relevance from the Study	Potential Association Based on Biological or Clinical Data	Goal	Main Insight Regarding the IncRNA	Model Construction	Kaplan–Meier and Log-Rank Test	Receiver Operator Curve AUC	Reference
LINC02528, SEMA6A-AS1, EBLN3P, MIR155HG, LYRM4-AS1, and HLA-DQB1-AS1	Skin cutaneous melanoma	Identify a novel prognostic signature using m6A-related IncRNAs and evaluate the prognostic of survival performance	Identification of a IncRNAs signature-related somatic SCNA associated with patient overall survival	Selection of candidate lncRNAs using LASSO followed by the COX prognostic regression model	High-risk vs. low-risk groups at 1, 2, 3, and 5 years; p < 0.001 and construction of a nomogram for the clinical decision risk score	AUC (1, 2, 3, and 5 years) > 0.6	[52]
AL121772.1, BX640514.2, LINC01133, and LYPLAL1-AS1	Pancreatic cancer	Investigate the prognostic performance of a lncRNA signature and its relationship with the tumor immune microenvironment	Identification of a genomic instability-related lncRNAs signature associated with patient overall survival	Selection of candidate lncRNAs using filtering methods followed by the univariable and multi-variable COX prognostic regression model	High-risk vs. low-risk groups at 1 year; <i>p</i> < 0.001	AUC (1 year) = 0.653	[53]
CTD-2256P15.2	Lung adenocarcinoma	Molecular and biological characterization of SCNA-related lncRNAs with risk prediction utility in cancer	Identification of a somatic SCNA-related IncRNAs signature associated with the prognosis of patients after methyl ethyl ketone (MEK) inhibitors treatment	Selection of candidate lncRNAs using filtering methods followed by the univariable and multi-variable COX prognostic regression model	High-risk vs. low-risk groups; p < 0.001	NA	[54]

	Table 1. Cont.						
lncRNA of Relevance from the Study	Potential Association Based on Biological or Clinical Data	Goal	Main Insight Regarding the IncRNA	Model Construction	Kaplan–Meier and Log-Rank Test	Receiver Operator Curve AUC	Reference
LINC00896, MCM8-AS1, LINC01251, LNX1-AS1, GPRC5D-AS1, CTD-2350J17.1, LINC01133, LINC01121, and AC073130.1	Non-small cell lung cancer	Evaluate the performance of a SCNA-related lncRNA signature as a risk predictor	Identification of a lncRNAs signature related to somatic SCNA and associated with patient poor overall survival	Selection of candidate lncRNAs using LASSO followed by the COX prognostic regression model	High-risk vs. low-risk groups at 1 and 3 years; <i>p</i> < 0.001	AUC (1 and 3 years) = 0.73	[54]
RP11-241F15.10	Osteosarcoma	Evaluate the performance of a SCNA-related lncRNA signature as a risk predictor	Identification of a lncRNAs signature-related somatic SCNA associated with patient disease-free and overall survival	Univariable and multi-variable COX prognostic regression model	High-risk vs. low-risk groups at 1 and 3 years; <i>p</i> < 0.001	NA	[55]
ALAL-1	Non–small cell lung cancer	Molecular and biological characterization of SCNA-related lncRNAs in cancer	Identification of a pro-oncogenic lncRNA that mediates cancer immune evasion, pointing to a new target for immune potentiation	NA	NA	NA	[56]
LOC339803, F11-AS1, and PCAT2 TMEM220-AS1	Hepatocellular carcinoma	Evaluate the performance of a SCNA-related lncRNA signature as a risk predictor	Identification of CNA-related lncRNAs that can better evaluate the prognosis of patients with liver cancer	Selection of candidate lncRNAs using LASSO followed by the COX prognostic regression model	High-risk vs. low-risk groups at 1, 3, and 5 years; <i>p</i> < 0.001	AUC (1, 3, and 5 year) > 0.7	[57]
ENSG00000261582	Lung adenocarcinoma and cervical carcinoma	Evaluate the performance of a SCNA-related lncRNA signature as a risk predictor	Identification of a somatic SCNA-related lncRNAs signature associated with patient overall survival	Univariable and multi-variable COX prognostic regression model	High-risk vs. low-risk groups at 5 years; <i>p</i> < 0.001	NA	[58]

Table 1. Cont.

lncRNA of Relevance from the Study	Potential Association Based on Biological or Clinical Data	Goal	Main Insight Regarding the lncRNA	Model Construction	Kaplan–Meier and Log-Rank Test	Receiver Operator Curve AUC	Reference
PCAN-R1 (Ensembl ID ENSG00000228288) and PCAN-R2 (Ensembl ID ENSG00000231806)	Prostate adenocarcinoma	Evaluate the performance of a SCNA-related lncRNA signature as a risk predictor	Identification of a somatic SCNA-related lncRNAs signature associated with patient overall survival	Univariable and multi-variable COX prognostic regression model	High-risk vs. low-risk groups at 5 years; <i>p</i> < 0.001	NA	[58]
LOC101927151, LINC00861, and LEMD1-AS1	Ovarian cancer	Evaluate the performance of a SCNA-related lncRNA signature as a risk predictor	Identification of a somatic SCNA-related lncRNAs signature associated with patient prognosis	Univariable and multi-variable COX prognostic regression model	High-risk vs. low-risk groups at 5 years; <i>p</i> < 0.001	NA	[59]
TSPOAP1-AS1, CCNT2-AS1, LINC01094, AL033527.2, and LINC00460	Gastric cancer	Evaluate the performance of a mutation-related lncRNA signature as a risk predictor and characterization of functional activity	Identification of an anoikis-related lncRNAs signature associated with patient poor overall survival and immunotherapy response	Selection of candidate lncRNAs using LASSO followed by the COX prognostic regression model	High-risk vs. low-risk groups; p < 0.001	Train AUC = 0.707 Test = 0.646	[53]
NA	Kidney renal clear cell carcinoma	Evaluate the performance of a multi-omics-derived lncRNA signature as a risk predictor	Identification of a lncRNAs signature derived from the analysis of transcriptomics and DNA methylation associated with patient poor overall survival	Selection of candidate lncRNAs using a novel TRS method utilizing multiple omics data and a XGBoost model followed by the COX prognostic regression model	High-risk vs. low-risk groups; p < 0.001	AUC = 0.95; using the best predictor model	[60]

	Table 1. Cont.						
IncRNA of Relevance from the Study	Potential Association Based on Biological or Clinical Data	Goal	Main Insight Regarding the IncRNA	Model Construction	Kaplan–Meier and Log-Rank Test	Receiver Operator Curve AUC	Reference
Twenty-six lncRNAs	Hepatocellular carcinoma	Evaluate the performance of a multi-omics-derived lncRNA signature as a risk predictor and characterization of functional activity	Identification of an exosome-related lncRNAs signature associated with patient poor overall survival and a response to transarterial chemoembolization (TACE) therapy and sorafenib therapy	Selection of candidate lncRNAs using a weighted correlation network analysis followed by the COX prognostic regression model	High-risk vs. low-risk groups; p < 0.001	AUC > 0.7	[61]
Ten lncRNAs, including: LINC00582, MIR205HG and TRG-S1,	Breast cancer	Evaluate the performance of a mutation-related lncRNA signature as a risk predictor and characterization of functional activity	Identification of a TMB-related lncRNAs signature associated with patient poor overall survival and a response to immunotherapy	Clustering based on the gene expression and selection of the gene predictors using LASSO followed by the COX prognostic regression model	High-risk vs. low-risk groups at 1, 3, and 5 years; <i>p</i> < 0.001	AUC (1 year) = 0.722, AUC (3 years) = 0.745, AUC (5 years) = 0.811	[62]
C20orf197, UCA1, MIR17HG, and MIR22HG	Various cancer types	Evaluate the performance of a lncRNA signature as a patient prognosis predictor	Oxiplatin sensitivity-related lncRNAs signature associated with the prognosis of patients given oxaliplatin-based chemotherapy	Selection of candidate lncRNAs using LASSO, decision tree, random forest, and a support vector machine followed by the COX prognostic regression model	High-risk vs. low-risk groups at 1, 3, and 5 years; <i>p</i> < 0.001	AUC (1 year) = 0.76, AUC (3 years) = 0.79, AUC (5 years) = 0.88; using the best predictor model	[63]
18 lncRNAs	Gastric cancer	Evaluate the performance of a lncRNA signature as a patient prognosis predictor	Identification of an immune-related lncRNAs signature that helps predict the prognosis of patients suffering from gastric cancer	Selection of candidate lncRNAs using the integration of multiple machine learning algorithms followed by the COX prognostic regression model	High-risk vs. low-risk groups at 1, 3, and 5 years; <i>p</i> < 0.001	AUC (1 year) = 0.715, AUC (3 years) = 0.80, AUC (5 years) = 0.809; using the best predictor model	[64]

	Table 1. Cont.						
lncRNA of Relevance from the Study	Potential Association Based on Biological or Clinical Data	Goal	Main Insight Regarding the lncRNA	Model Construction	Kaplan–Meier and Log-Rank Test	Receiver Operator Curve AUC	Reference
Various lncRNAs	Breast cancer	Develop an interpretable deep-learning-based network for classifying the recurrence risk and revealing the potential biological mechanisms	Construction of a lncRNa-based model associated with the radiomics of magnetic resonance features for predicting individual recurrence risk after surgery	Construction of a lncRNa-based model using the Cox proportional hazards deep neural network	High-risk vs. low-risk groups at 1, 2, and 3 years; <i>p</i> < 0.001	AUC (1 year) = 0.98, AUC (2 years) = 0.94, AUC (3 years) = 0.92; using the best predictor model	[65]
CRNDE, AC010273.2, MPPED2-AS1, SNHG18, and CYTOR	Lower-grade gliomas	Evaluate the impact of DNA methylation-related lncRNAs with an effect on genome stability and the immune microenvironment on disease progression	Identification of a five DNA methylation-related signature with an independent prognostic value	Selection of lncRNAs using filtering methods followed by the COX prognostic regression model	High-risk vs. low-risk groups at 1, 2, and 3 years; <i>p</i> < 0.001	AUC (1 year) = 0.893, AUC (2 years) = 0.919, AUC (3 years) = 0.866; using the best predictor model	[66]
CRNDE, AC010273.2, MPPED2-AS1, SNHG18, and CYTOR	Various cancer types	Explore the functional effects of lncRNAs related to DNA methylation and evaluate their predictive performance on patient survival	Identification of a DNA methylation and genome stability-related signature with an independent prognostic value	Construction of a lncRNa-based co-expression network of pADM–lncRNA	High-risk vs. low-risk groups; p < 0.001	AUC = 0.6839	[67]
LIFR-AS1	Colorectal carcinoma	Evaluate the impact of DNA methylation-related lncRNAs with an effect on disease progression	Identification of LIFR-AS1 as a tumor suppressor RNA with an independent prognostic value	Selection using filtering methods followed by the COX prognostic regression model	High-risk vs. low-risk groups; p < 0.001	AUC = 0.872	[68]

# 3. Approaches for Prioritizing Cancer Driver IncRNAs Using Somatic Mutations Profiles

From an omics and bioinformatics point of view, we can broadly classify the approaches that have been developed to prioritize lncRNAs with a possible association with cancer into at least two categories: (1) an early integrative analysis of somatic variants with cancer driver potential that have an impact on the functional activity of lncRNAs associated with prognosis in cancer patients [37,38,69–71], and (2) post-analysis integration strategies centered on the relationship between genome instability and lncRNAs with aberrant expressions associated with tumor prognosis [39–45] (Figure 2). A representative study was conducted to evaluate the potential impact of somatic mutations in human lncRNAs (denominated MutLncs) and their functional significance in cancer, interrogate the mutation profiles in genomic regions harboring lncRNA and their vicinity across 17 cancer types, and use an integrative pipeline to describe the significance of the MutLncs contribution to cancer [69]. Interestingly, they discovered different cancer-specific networks of co-occurring lncRNAs in the MutLncs by considering the mixing effects (i.e., the influence of a pairwise combination of multidimensional data, including DNA methylation, TF expression, and miRNA expression). Notably, the study found that a number of MutLncs in co-occurring pairs were correlated with patient survival, particularly in human cutaneous squamous cell carcinoma (SKCM) and glioblastoma (GBM) tumors [69].

The complexity of cancer genomes goes beyond a set of specific targetable driver mutations. Other features, such as the tumor mutational burden (TMB) and mutational signatures, are increasingly being incorporated into bioinformatic analyses aimed at uncovering therapeutic and prognostic insights in cancer [72]. Therefore, it is not unexpected that many studies have integrated somatic mutation profiles and transcriptome expression data to identify lncRNAs [39-42,46] (Table 1). For example, in one study, colon cancer researchers used a post-analysis integration strategy to characterize patients by their somatic mutation profiles and subsequently analyzed the transcriptome data and survival information from the same set of patients. They were able to derive a score calculated on a seven genome instability-associated lncRNA signature that was able to distinguish high-risk patients characterized by high somatic mutation, high microsatellite instability, significantly worse clinical outcomes, and specific tumor immune infiltration [43]. Using a similar approach in cervical endometrial cancer (CESC), Zhang and collaborators created a computational model derived from what they called the mutator hypothesis, in which they first grouped patients based on the cumulative number of somatic mutations, then screened them for differences in the lncRNAs expression, and found a genomic instabilityassociated lncRNAs signature with the potential to determine patient survival and help differentiate high-risk from low-risk CESC tumors [44]. Remarkably, the TMB has become an emerging marker for checkpoint inhibitor-based immunotherapy [44,73]. In a recent study, by integrating RNA sequencing data, mutation profiles, and the corresponding clinical information from a TCGA colorectal cancer dataset (COAD), the authors evaluated the TMB of the COAD samples to classify them into low- and high-TMB groups. They, then, used a machine learning approach to construct a 14-lncRNA-based classifier index related to three immune checkpoints (i.e., PD1, PD-L1, and CTLA-4). Interestingly, the obtained classifier was significantly associated with the TMB levels and could accurately predict the overall survival of the COAD patients [74].

The number of studies discussed so far provided a vision of different modalities of integrative multi-omics analyses being employed as powerful strategies for identifying and characterizing lncRNAs as informative features that allow for diagnosis, prediction, and prognosis. Without an integrative genomic approach, it is unlikely that more complex questions can be addressed, for example, the search for the interaction of lncRNAs with other biological and environmental factors, such as sex hormones, smoking, occupational risk, and dietary habits, among others, and their association with the risk of cancer development, patient survival, and response to treatment. For instance, lung adenocarcinoma (LADC), the most common type of lung cancer, is largely caused by chronic tobacco smok-

ing. However, approximately 25% of cases occur in non-smokers. Notably, this proportion of LADC in non-smokers is associated with being female, and the etiology remains an elusive question. In a recent study to uncover genomic evidence of smoking-associated mutagenesis, researchers performed a multilevel integrative analysis using whole-genome sequencing and RNA-seq data to unravel the mutational processes and complex genomic rearrangements that drive the development of LADC. One of the intriguing findings of this multi-omics analysis was the discovery of a recurrent fusion gene, composed of the gene ERBB3 and the lncRNA anti-estrogen-resistance 4 (BCAR4), which was present in two tumors from female cases in which low to no contribution of a mutational signature associated with direct DNA damage caused by tobacco smoke mutagens (COSMIC signature 4) was observed [75] (Figure 3). Recently, other comprehensive integrative genomic analyses have led to the discovery of diverse fusion variants involving BCAR4 in lung adenocarcinoma [76–78]. Interestingly, the CD63–BCAR4 fusion was discovered in a never-smoking female patient by another genome-wide study on non-small cell lung cancer (NSCLC) [79]. Here, the authors found that the expression of both transcripts within this fusion were highly activated. Considering that a higher expression of BCAR4 was also established as an independent predictive factor for tamoxifen resistance and poor progression-free survival in ER+ breast cancer patients [80], the clinical relevance of these complex rearrangements involving BCAR4 in other tumor types, such as lung cancer, may become more apparent and demonstrate the utility of using highly multidimensional molecular analyses to uncover the targetable alterations that would otherwise be overlooked.



**Figure 3.** Summary of an integrative multi-omics analysis in lung adenocarcinoma to address how the complex genomic rearrangements and the resulting gene fusions, which are often copy-number

balanced, appear as a process of early oncogenesis that is often acquired in the first decades of life. We graphically depicted an alteration described in 25% of lung adenocarcinoma cases without a history of smoking, mainly enriched in female patients and whose etiology is unknown (**upper** panel). The integrative multi-omics analysis identified a recurrent oncogenic fusion gene composed of BCAR4:ERBB3 (lncRNA and messenger RNA) that was also overexpressed in a subset of cases (**bottom** panel).

#### 4. Copy-Number Alterations in Genomic Regions Encoding IncRNAs

Indeed, genomics and mathematical analyses of the patterns of somatic alterations have become a powerful strategy for identifying cancer driver genes. In this regard, although somatic point mutations and small insertions and deletions (INDELs) have been the primary focus of cancer genomic studies, copy-number alterations (CNAs) are also important forms of DNA aberrations, also referred to as somatic copy-number alterations (SCNAs), which encompass larger genomic regions and often harbor key genes involved in the development and progression of many cancers [81,82] (Figure 1). Not surprisingly, most large-scale genomic analyses conducted to date have successfully identified almost exclusively protein-coding cancer driver genes located in regions of focal amplification and deletion [83–85], and it is only relatively recently that the first systematic analyses have reported on the identification of lncRNAs that are also contained within focal CNAs in cancer genomes [47,86–89]. Given the evidence that nearly three-quarters of the human genome can be transcribed to RNA [90] and that only 2% of the human transcriptional landscape codes for a protein, the need to understand the functional impact of SCNAs on IncRNAs has become clearer. In one of the first comprehensive characterizations of the impact of cancer driver lncRNAs in regions of SCNA alteration, Hu et al. implemented a bioinformatic integration of SCNAs and an expression analysis across 12 different cancer types. In developing their approach, the researchers scanned large-scale genomic data for focal alterations, which often exhibited high-amplitude variations, and mapped the known lncRNA loci to these regions of a focal gain or loss. Interestingly, approx. 17.8% of the IncRNAs-encoded loci with SCNAs were expressed in 40 cancer cell lines representing five different tumor types, among which they discovered FAL1 as a potentially oncogenic IncRNA associated with clinical outcomes in ovarian cancer patients. In a follow-up study, the authors confirmed the SCNAs in FAL1 in OC and observed that it was also present in the other five tumor types [48-50,82]. The authors also raised a key issue to consider in the characterization of lncRNAs as tumorigenic drivers by distinguishing the driver SCNAs from passenger ones by their levels of presence (i.e., in sufficient quantities to be detected) in the cancer cells of interest. In this regard, the identification of SCNAs from the lncRNA loci that are actively transcribed in a given cell at a given time may help characterize the IncRNA molecules with a real function and role as drivers of tumorigenesis and of potential clinical utility, as further demonstrated by Hu et al. and other studies [47,48,51,54–56,87] (Table 1).

A second consideration that arose in light of the above approaches for understanding the actual contribution of lncRNAs as drivers of cancer and its deregulation was based on estimating the relationship that exists between the copy-number values of the SCNAs at the lncRNA loci and the expression levels of their corresponding transcripts present in a given cancer type. Indeed, this aspect was consistent with our knowledge that the differential gene expressions of the protein-coding cancer driver genes were significantly correlated with the CNAs [82,91]. As a consequence, researchers have begun to delineate the correlation between the SCNAs and expression levels as useful parameters for defining the gain or loss of function of lncRNAs with oncogenic or tumor suppressor functions and for potentially associating them with the clinical outcome for different tumors [47–50,56,57,82].

Cancer is a heterogeneous disease that includes a diversity of tumors from the same or different organs and tissues, displaying differences in cellular compositions and biological and molecular features [92]. Much of our understanding about cancer heterogeneity and

its clinical implications has been improved by new developments in genomics [93]. Some studies have so far addressed the integration of multi-omics data, including the lncRNAs expression and SCNA, with the aim of defining the molecular profiles that capture the diversity of the tumor subtypes with different prognoses and responses to treatment [58,59,94]. As a notable example, in a large-scale genomic analysis of high-grade serous ovarian carcinoma (OV), Akrami et al. investigated the correlation between SCNA and lncRNAs expression profiles and were able to identify the lncRNA OVAL (RP11-522D2.1) as a target of focal DNA amplification. Notably, this alteration was detected specifically in OV and was consistent with the increase in the OVAL transcript expression [95]. It is important to note that this relationship may be more complex, and that the SCNAs present in a tumor may be associated with aberrant expressions of lncRNAs other than those encoded at their loci. For example, other mechanisms may operate to disrupt their activity, such as the occurrence of a breakpoint event within the boundaries of the promoters and regulatory sites of a particular lncRNA, or genomic alterations that directly disrupt the protein-coding gene targets for such lncRNAs. In this regard, further research is still needed to unravel the complicated interactions between lncRNAs, protein-coding genes, miRNAs, and other types of functional elements that are potential drivers of cancer.

### 5. The Revolution of Non-Coding Transcriptome in Cancer Studies

There is increasing evidence that large numbers of lncRNAs are present in the human genome. The latest release of LNCipedia, a public database for lncRNA sequences and annotation, contains 107,039 high-confidence transcripts that show no coding potential belonging to 49,372 lncRNA genes [96]. Undoubtedly, the burgeoning number of lncRNAs over the past few years has aroused the interest of the scientific community in the study of their biological properties and roles [11]. Based on the fact that they play critical roles in various processes that are necessary for body cell functions, many studies have addressed and demonstrated the relationship between lncRNAs and disease [21]. In particular, experimental evidence for the relevance of lncRNAs as cancer markers has been accumulating over the last decades. Among the most prominent examples of lncRNAs whose aberrant expression is associated with cancer development are the descriptions of MALAT1 [97], HOTAIR [98], and PCA3 [99]. Since its discovery, MALAT1 has become the paradigm of functional alterations of lncRNA in cancer and has been proposed as a potential biomarker with a critical role in several tumors, such as non-small cell lung adenocarcinoma (NSCLC) [100], gastric cancer [101], and colorectal cancer [102]. The oncogenic expression of lncRNA HOTAIR also holds a promising value as a biomarker of the response to breast [103] and hepatocellular carcinoma [104]. Meanwhile, the levels of PCA3 in urine has been used as a marker for prostate cancer aggressiveness to guide medical decisions on patient treatment and has been available since its approval by the FDA [105,106]. Aside from PCA3, no other lncRNA has been approved by the FDA for diagnostic, prognostic, or treatment response prediction in any type of cancer. Despite these successful examples discovering more associations between lncRNAs and diseases, characterizing them as biomarkers or targets of therapeutic agents has become an increasingly challenging task. It is at this point that the integrative analysis of large-scale data has become a game changer. In this regard, there are several scientific examples that have used analytical strategies to investigate the relationship between somatic alterations, including somatic mutations, copynumber alteration profiles, and transcriptome alterations, such as the perturbed correlated expression of protein-coding genes, microRNAs, and lncRNAs, to reveal cancer drivers or biomarkers with potential utility in the prognosis and prediction of the treatment response. As shown in Table 1, we enumerated many of the most relevant examples involving multiomics strategies for studying the biology and clinical utility of lncRNAs. In the following sections, we will review some of them in more detail to provide insights into these studies and exemplify the cross-linking, integration, and complementation between the different traits that were analyzed using multiple omics dimensions.

### 19 of 33

## 6. Multi-Omics Network Approaches Reveal IncRNA Biological Relevance on Cancer Biology

As discussed above, the technological and computational advances in recent decades have allowed for the identification of thousands of lncRNAs whose molecular alterations are associated with various types of cancer (Figure 1). A major challenge is that the initial efforts to discover cancer-related lncRNAs took advantage of classical functional genomic approaches, primarily by characterizing the global transcriptomic landscape, evolutionary conservation, or proximity to known cancer genes. While these analytical strategies provided valuable insights, the altered transcriptional profiles alone do not indicate a causal role in cancer programs. Currently, various biological evidence describing known cancer-associated lncRNAs has been presented in the literature, most of which focused on the function of a single lncRNA linked to malignant transformation through its role in gene regulation and its impact on cancer hallmarks.

Despite intense research into the mechanisms underlying altered lncRNAs, our understanding of the global biological impact of lncRNA regulation in tumors remains limited [107]. It is challenging to holistically describe the complex biological processes of many lncRNAs and their cooperative mechanisms using traditional biochemical and molecular approaches. Therefore, the availability of multi-omics data from genomics (mutations and CNV), transcriptomics (mRNA, non-coding RNA), proteomics, epigenetics (chromatin methylation and architecture), and metabolomic studies has opened new strategies for advancing the understanding of the roles of lncRNAs in health and disease by developing tools to integrate these data at the systems level (Figure 1).

In this section, we will review the progress made in understanding the multi-omics biological features through machine learning and artificial intelligence approaches and provide an overview of the recent advances in uncovering the regulatory basis underlying the functionalities of lncRNAs at various molecular and cellular biological levels.

#### Describing the Novel IncRNAs Drivers in Cancer through Multi-Omics Integration

Apart from a few well-known lncRNAs, the landscape of cancer lncRNAs is far from complete. The diverse functional repertoire of lncRNAs in cancer can be explored through (1) their function as driver genes, resulting from early mutations that are positively selected during tumorigenesis, or (2) as downstream genes, resulting from non-genetic changes in their expression, localization, or molecular interactions [108]. Although both categories contribute to cancer phenotypes, most efforts to discover cancer lncRNAs only take advantage of differential expression approaches. To overcome this, several computational methods have recently been developed to identify lncRNA driver genes by analyzing coordinate omics alterations to detect the signals of positive selection.

One of the first statistical methods for driver gene discovery was OncodriveFML [109], which identifies tumor-associated lncRNAs by interrogating somatic mutations in the coding and non-coding regions and gene expression. Compared to the other methods, OncodriveFML calculates a functional impact score, for which it uses a local mutational background in specific regions to define the positive selection signals in genes across tumor tissues. The method takes into account that the mutational background is influenced by chromatin architecture, replication timing, and transcription factor binding sites. Therefore, by considering this local background as well as the mutational and expression patterns, this computational tool can discover lncRNAs contained in potential genomic driver regions involved in tumorigenesis. Computing this method using whole-genome tumor data sequenced by the TCGA and the Cancer Genome Project, it was possible to identify MALAT and MIAT as two lncRNAs with an excess of high-impact mutations.

Another integrative method for predicting lncRNA drivers that are relevant to tumorigenesis is ExInAtor, which identifies genes with a high somatic single nucleotide variant load across the tumor genomes using the DNA mutation patterns (local trinucleotide background model) and expression data as proxies for their functionality. ExInAtor was modeled on 1112 whole genomes from 23 cancer types deposited in GENCODE, and predicted 15 lncRNA drivers with a high confidence, of which nine were novel lncRNAs and six were known cancer-related transcripts, including PCA3, MALAT1, BCAR4, lncRNA-ATB, and SAMMSON. Most of these lncRNAs were tumor-specific, although NEAT1 and MALAT1 were identified in a pan-cancer context, confirming their role in tumorigenesis. The set of previously unreported driver lncRNAs include MIR100HG, AP000469.2, RP11-308N19.1, RP11-455B3.1, RP11-332J15.1, RP11-707A18.1, RP11-6c14.1, RP11-1101K5.1, RP11-354A14.1, and RP11-189E14.4. These novel candidates are evolutionarily conserved, expressed in normal tissues, and have an increased gene length. They also tend to be proximal to cancer SNPs and are encoded in CNA regions, suggesting a role in tumorigene-sis. The authors highlighted MIR100HG, which was highly conserved and had canonical histone modifications in the promoter region and transcription factor binding sites [110].

More recently, to enhance the discovery of cancer-related lncRNAs and gain insight into their biology, the Cancer lncRNA Census (CLC) was presented as a tool to provide functional or genetic evidence of lncRNAs roles in cancer by integrating genomic and transcriptomic data associated with cancer in different mammalian species [108]. To date, it is not completely clear whether mutated lncRNAs can drive tumorigenesis and whether such altered functions could be conserved during evolution. Therefore, the CLC considers the conserved functions between humans and mice as a relevant feature that could provide strong evidence for the biological role of lncRNAs, both in cancer and under physiological conditions. The application of this computational model revealed the colocalization of cancer lncRNAs with known protein-coding cancer genes. A total of 10 tumor-causing mutations were identified in eight lncRNA orthologs, including DLEU2, GAS5, MONC, NEAT1, PINT, PVT1, SLNCR1, and XIS, some of which have already been reported in cancer.

The integration of DNA, RNA, and protein alterations and the way they cooperatively interact provide new evidence for identifying dysregulated lncRNA in cancer. For example, LongHorn [111], a recently presented computational method, integrates genomic, transcriptomic, and proteomic alterations and predicts dysregulated lncRNA regulatory networks in cancer pathways by modeling their impact on the transcription factors, RNA binding proteins and microRNAs activity, lncRNA promoter binding sites, and post-transcriptional activation/inhibition. The computation of this method for 14 cancer types from the TCGA predicted several lncRNA candidates whose dysregulation affected other known cancer genes and pathways mainly in a tumor-specific context and influenced tumor etiology. OIP5-AS1, TUG1, NEAT1, MALAT1, XIST, and TSIX were predicted to regulate cancer signaling in multiple tumor contexts. In addition, the lncRNA network analyses indicated the enrichment of lncRNA binding sites in the promoter regions of messenger RNAs, which enhanced the transcriptional effects of lncRNAs. The functional experimental analyses confirmed most of the predictions made with LongHorn.

An interesting approach presented by Mitra and collaborators for predicting the biological dependencies of uncharacterized lncRNAs focused on the identification of coessential modules by integrating the copy number, and epigenetic and transcriptomic data of the lncRNA landscape from exogenous knockouts or activation screens that were generated using CRISPR techniques [112]. By applying this model to multi-omics cancer cell line data, the authors estimated 289 lncRNA-gene co-expression networks that recapitulated the known proliferation-regulating lncRNAs and predicted novel lncRNAs associated with proliferative signaling that were poorly characterized, such as PSLR-1/2, which induced G2 arrest through the modulation of the FOXM1 transcriptional network and whose exogenous expression inhibited proliferation and colony formation in cell line models.

Although DNA methylation dysregulation is associated with cancer, the molecular mechanisms of how methylation and transcriptional lncRNA patterns are reciprocally modulated in cancer remain largely unknown. A novel integrative analysis framework, called MeLncTRN (Methylation mediated lncRNA Transcriptional Regulatory Network), integrates transcriptome, DNA methylome, and copy-number variation profiles to identify the regulatory circuits driven by epigenetically driven lncRNAs across 18 cancer types [113]. An analysis of 5970 TCGA tumor samples revealed that the association between lncRNAs

and the DNA methylation mechanisms was common and conserved across multiple cancer types, e.g., a complex interplay between lncRNAs and epigenetic modulators, such as DNA cytosine methyltransferases DNMT1, and histone modification proteins, such as EZH2 (Figure 4). For example, FAM83H-AS1, TUG1, PVT1, and LINC00511 acted as scaffolds to enhance EZH2 or DNMT1 binding and consequently repressed the expression of their mRNA targets. This observation expands the understanding of the role of lncRNAs in transcriptional regulatory circuits in addition to their miRNA sponge activity as competitive endogenous RNA (ceRNA) [114].



**Figure 4.** Summary representation of the multi-dynamic analysis computed by MeLncTRNn, which integrated genomic, transcriptomic, and epigenetic data to identify lncRNA genes and their interaction with epigenetic mechanisms that could serve as markers for classifying cancer patients into subtypes with different prognoses (**top** panel). We described a regulatory network operating in pancreatic cancer defined by MeLncTRNn, integrating lncRNA modulators and target genes related to epigenetic regulation. In particular, the CRNDE module serves as a prognostic information tool for pancreatic cancer patients (**bottom** panel).

Emerging evidence has also revealed the underlying crosstalk between lncRNA and genomic instability, a relevant hallmark of cancer. Novel approaches integrating chip-seq, WGS, and WES data revealed an unexpected relationship between oncogenic lncRNAs and epigenetic alterations that contribute to chromosome fragility in cancer. To characterize the lncRNA-based mechanism by which aberrant epigenetic signatures can be generated, the authors used as a conceptual, the sub-telomeric chromosomal locus 8q24, which contains the cMYC gene and a large histone domain H3 (CENP-A) variant, both of which were altered in cancer cells of various solid tumors. This region also encoded five unique lncRNAs sequences, namely PCAT1, PCAT2, CCAT1, CCAT2, and PVT1, which negatively modulated the occupancy of CENP-A at the chromosomal locus. Their results indicated a competition between the lncRNAs transcription and R-loop occupancy that strongly contributed to the maintenance of CENP-A invasion, which ultimately affected the chromosome stability [115].

An oncology milestone that was already mentioned is the TMB, which is related to the infiltration of various immune cell populations that can enhance or limit cancer programs [116]. Recently, increasing evidence has shown that lncRNAs may play fundamental

roles in the regulation of the immune system, but few immune-related lncRNAs have been described in cancer. Therefore, novel approaches have been developed to shed light on these roles [117,118]. Through an integrative analysis of the lncRNA expression, tumor immune response signatures, and genome-wide DNA methylation data in 9626 tumor samples across 32 cancer types, the lincRNA-based immune response (LIMER) tool revealed 7528 lincRNAs associated with the tumor immune signature [119]. Of interest, EPIC1 was identified as a relevant immune-related lncRNA that was inversely correlated with the MHC expression and CD8+ T activation and infiltration. The in vitro and in vivo models demonstrated that EPIC induced tumor immune evasion and resistance to immunotherapy by epigenetically suppressing the tumor cell antigen presentation through EZH2 interaction. Another interesting tool is ImmLnc, which systematically infers the candidate lncRNA modulators of the immune-related pathways by matching the gene and lncRNA expression profiles. The tool prioritizes cancer-related lncRNAs by comprehensively characterizing the lncRNA landscape and its correlation with the immunome. One of the first findings was that the tumors derived from similar tissues were likely to share lncRNA immune regulators. Furthermore, the novel subtypes identified by ImmLnc showed a distinct mutation burden, immune cell infiltration, expression of immunomodulatory genes, and response to chemotherapy and prognosis [117].

### 7. Challenges and Opportunities of Machine Learning to Deepen the Functional and Biological Roles of IncRNAs in Cancer

In this section, we provide a summary of the approaches we took to address biological data integration and discuss the pitfalls of overcoming and optimizing the use of data from multiple sources and technologies in an analysis pipeline. A first challenge was the difference between the samples, replicates, and technology used, which favored batch effects. Additionally, the different modalities for measuring the cell levels with different measurements posed a significant computational challenge, as there was no common virtual space to integrate the samples.

Machine learning methods have been increasingly used in attempts to better understand the biology of molecules, such as lncRNAs. These methods can be generally classified into supervised and unsupervised learning. The former includes dimension reduction and clustering, while the latter includes regression and classification. In the context of the IncRNA landscape, one of the challenges that has arisen is the accurate identification of true lncRNAs from other types of RNAs, along with the definition of their biological roles. The main approach for the identification of lncRNAs is based on their coding potential and length. Later, other features were integrated, such as the presence of an open reading frame (ORF), nucleotide composition, kmers, secondary structure, codon usage, ribosome release score, conservation scores, and others. Among the most used algorithms for data integration are logistic regression, random forest (RF), support vector machine (SVM), and deep learning (Figure 2). From a mechanistic point of view, to better understand the role of lncRNAs in complex diseases such as cancer, in addition to the coding potential, some studies included computational frameworks based on machine learning methods, such as the Genetic Importance Calculator (GIC), which considers the interaction of lncRNAs with other omics levels. Through these methods, it is possible to accurately identify unknown IncRNA interactions. The number of methods developed for these purposes has increased steadily in recent years and discussing them all is beyond the scope of this review. However, some of the limitations and strengths of these methods have been extensively reviewed elsewhere [84,120–122].

In recent years, a number of strategies based on the use of deep learning and complex network analyses have emerged to uncover the relevant dysregulations of lncRNAs and their potential value in identifying their relationships with cancer diagnosis, prognosis, and treatment [123–127]. One of the limitations of such algorithms is the lack of reliable and verified negative samples. In addition, they use known biomolecule–disease associations that cannot be applied to new diseases. Finally, these data integration tools, while useful

for efficiently learning the structure of high-dimensional omics data, generally decrease the interpretability of the model, which is a limitation in a biomedical application where we wish to understand the relationship between the learned latent variables and the observed variables, such as disease. In recent years, formal work has begun to overcome this obstacle [128,129].

In addition, all the data have different calculation rules, which can lead to inaccuracies. Moreover, the link between lncRNAs and cancer goes beyond identifying the molecules that may be actively involved in the disease. Other aspects have been explored, such as the identification of biomarkers that are capable of predicting prognosis or even a response to drugs [130,131]. For this area, the combination of different approaches has been proposed, such as the least absolute shrinkage and selector operation (LASSO) and the support vector machine recursive feature elimination (SVM-RFE) [132,133].

In general, the methods for lncRNA identification, the prediction of their functions, and their relationship to disease, even as part of precision medicine tools for prognosis and drug response prediction, continue to improve by using higher level features and generating more and new information to feed the algorithms for better classifications. Despite these efforts, the integration of all the available information in a biologically meaningful way remains a challenge. During biological analysis, and more specifically in omics data analysis, subsequent functional validation steps are essential to provide new, more complete, and useful insights for future applications that can be transferred to the clinic. Thus, designs should include a holistic view integrating computational and experimental characterization.

### 8. Perspective on the New Directions in lncRNA Research and Its Implication in Cancer through Multi-Omics Analysis

What is the best way to answer new questions and discover the novel and controversial mechanisms of lncRNAs in a disease state through multidimensional analysis? A good example is the milestone discovery that lncRNAs are completely unable to produce peptides, even though there is evidence that some lncRNAs are found in the cytoplasm and can associate with ribosomes, suggesting the possibility of translation and the production of microproteins [134].

Unlike other protein by-products, such as insulin, which are derived from larger proteins, microproteins are small and translated directly from short open reading frames (sORFs) of less than 300 nucleotides that produce a protein of up to 100 amino acids. These sORFs are often found within the lncRNAs sequence. Despite the difficulties, researchers are making progress in characterizing the functions of the microproteins generated by lncRNAs that are altered in cancer and are related to the clinical behavior of tumors [135–137].

Although the complete repertoire of functional sORFs is unknown, new techniques and computational approaches are being developed based on the search for conserved DNA sequences, natural selection, ancient phylogenetic origins, and ribosomal profiling—a tool based on deep sequencing that allows for detailed measurements of translation [138–140]. Combining these methods, the lncRNA sORFs capable of encoding a microprotein have been identified. The data have shown that these microproteins are signaling molecules that regulate enzymes and act as receptor ligands and critical transmembrane components. More microproteins are being described which have implications in cancer. Here, we list some of the microproteins produced by lncRNAs linked to cancer (Table 2).

lncRNA	Micropeptide	Biological Effect	Method	Citation
lncRNA AC025154.2	MIAC	Inhibits the actin cytoskeleton by interacting with Aquaporin 2 to suppress tumor growth and metastasis	RNA-seq data from kidney renal clear cell carcinoma patients	[141]
LINC-PINT	PINT87aa	Suppresses glioblastoma cell proliferation by interacting with PAF1c and preventing the transcriptional elongation of cancer-related genes	RNA-seq of circRNAs (transcriptome sequencing) and RNC-RNAs (translatome sequencing)	[142]
LINC02381	LINC02381-aa	LINC02381-aa contained in exosomes enhances ferroptosis by promoting the glucose transporter SLC2A10 in glioblastoma	Machine learning to integrate the multi-omics data to identify microproteins	[143]
NR_029453	CASIMO1	Interacts with the enzyme squalene epoxidase and increases the phosphorylation of ERK, a relevant actor of the MAPK pathway affecting cell proliferation	Transcriptome analysis and DNA conservation evaluation	[144]
lncRNA CTD-2256P15.2	РАСМР	PACMP acts as an activator of the PARP1-dependent pathways (DNA repair process), enhancing tumor growth and limiting the cell response to PARP inhibitors	RNA-seq	[145]
HOXB-AS3	HOXB-AS3 peptide	Suppresses cancer growth by limiting PKM splicing and, subsequently, the metabolic reprogramming	Transcriptome analysis and RNA affinity purification analysis	[146]
LINC00665	CIP2A-BP	Decreases cell invasion in vivo and correlates with better patient survival. Competes with PP2A for CIP2A binding, suppressing the oncogenic P13K/AKT/NFkB pathway	Ribosome profiling and RNA sequencing data analysis	[147]

**Table 2.** Studies focused on describing the microproteins translated from the lncRNA sequences with relevant implications in cancer.

### 9. Perspective on the Future Use of lncRNAs for Therapeutic Purposes through Multi-Omics Oncology

The field of lncRNA research has developed rapidly over the past decade, moving from understanding their basic biological properties to exploiting their clinical relevance. As a result, lncRNAs have been proposed as biomarkers and therapeutic targets that are being actively explored. In this section, we review the emerging strategies for the therapeutic exploitation of lncRNA.

It would be appropriate to begin by discussing the value of circulating lncRNAs as biomedical tools for the detection and monitoring of various diseases. Despite the important progress in translating the clinical utility of circulating molecules as biomarkers, such as proteins, metabolites, or free mRNA, there are still relevant limitations. This is why the study of lncRNA has been postulated as a new source of biological information that presents relevant advantages over other circulating analytes, such as a greater resistance to degradation, stability in biofluids due to their secondary structures, and their transport in extracellular vesicles. These characteristics present them as reliable cancer biomarkers [148,149].

The case of PCA3 is a good example to illustrate the potential use of lncRNAs as informative biomarkers in cancer clinics. In 1999, Bussemakers and colleagues discovered that the ncRNA prostate cancer antigen 3 (PCA3), which they initially named differential display code 3 (DD3), was overexpressed in prostate cancer tissues compared to non-neoplastic prostate tissues [106]. Interestingly, PCA3 non-coding RNA is involved in the

control of prostate cancer cell survival and modulates androgen receptor signaling [99]. Although PCA3 is the only lncRNA that has received FDA approval for the diagnosis, prognosis, or prediction of treatment response in any type of cancer, there are other reports describing new potential lncRNA biomarkers with high sensitivity and specificity for detecting specific neoplasms [150]. Thus, the diagnostic and monitoring utility of circulating lncRNAs biomarkers has not yet reached its full potential. For example, MIR205HG, the host gene of miR-205, is involved in prostate cancer cell differentiation and has been described as a potential marker that is capable of differentiating prostate cancer samples from normal ones [151].

A rapidly growing area of interest in the field of oncology is the phenomenon of drug resistance, which is the major limiting factor in achieving a cure for cancer patients. Emerging preclinical evidence supports that the lncRNA expression patterns predict a response to anticancer drugs [130,152]. By comparing the publicly available transcriptional profiles of different RNA species at baseline and after drug treatment with hundreds of compounds in the cancer cell lines, many lncRNAs such as GAS5 and ZEB2AS1 were shown to be highly predictive of the sensitivity to various anticancer drugs. Therefore, lncRNAs could explain new signals of how cancer cells become resistant to anticancer therapies and represent a new source of biomarkers. In addition, recent evidence showed that the lncRNAs EGFR-AS1 and MIR205HG could significantly improve the response prediction to erlotinib and gefinitib, better than EGFR somatic mutations and amplifications, suggesting a critical role for these lncRNAs in cancer precision medicine [153].

Given that lncRNAs can act as competitive endogenous RNAs (e.g., siRNAs or miRNA sponges), they could potentially control resistance-related biological processes. By constructing a ceRNA network, including lncRNA and mRNA, several drug resistance-related modules were identified as novel drug resistance markers. For example, the GAS5-RPL8 ceRNA pair regulated the drug resistance, as the GAS5 down-expression enhanced the RPL8 miRNA inhibition, which was reported to be associated with chemotherapy resistance [154]. The integration of multi-omics data revealed another relevant ceRNA module, consisting of HOXA-AS2, which was regulated by EIF4A3, FMR1, and HNRNPA2B1 and was down-expressed mainly in breast cancer patients, leading to adriamycin resistance. These alterations also down-expressed the expression of miR-107 [155]. Taken together, these studies demonstrated the potential viability of lncRNAs as complementary biomarkers and drug targets.

The therapeutic targeting of non-coding RNAs is an attractive approach for the treatment of cancer. Although no lncRNA-based therapy has been introduced into clinics, the functional diversity of lncRNAs provides an opportunity for their therapeutic modulation through transcriptional and post-transcriptional inhibition, the steric hindrance of promoters or secondary structures, exogenous synthetic lncRNAs, and editing tools such as CRISPR-Cas systems [122,123,156,157]. Each of these approaches has its own challenges, and future studies are needed to demonstrate their therapeutic efficacies.

### **10. Conclusions and Future Directions**

With the advent of new technological resources, such as single-cell and spatial genomics and transcriptomics and the popularization of and ease of access to machine learning frameworks, multi-omics data integration approaches will further establish themselves as a powerful paradigm in the study of lncRNAs in cancer. The current state of knowledge includes a variety of studies and reviews that offer different perspectives on the topic. These range from studies that focus on elucidating the biological role of lncR-NAs in cancer, to their proposed utility as molecular markers in clinical cancer research and the revision of computational methods for predicting lncRNA interactions with other molecules. However, many questions remain regarding the pragmatic application of the different molecular technologies and bioinformatics algorithms for mining multi-omics data. Furthermore, there is no clear consensus on how to integrate the different omics platforms and what information can be obtained from such approaches that can be used for the biological and pathological characterization of lncRNAs. Although many papers in the literature claim to have discovered the biological role or association between different lncRNAs and cancer using multi-omics strategies, only a fraction of them integrated information from different molecular platforms. Others explored different aspects of the same molecular features interrogated by each platform (e.g., RNA expression (gene expression, microRNAs expression), DNA somatic mutations, copy-number variations (CNV), or DNA methylation). In this review, we referred to multi-omics as the paradigm of using data obtained from and including two or more of the aforementioned molecular features. There is little doubt that these methods will complement our set of resources to unravel the relationships between lncRNAs and other biological factors that critically modulate cancer cell fitness and demonstrate new resources for studying tumor-relevant lncRNAs using dedicated multi-omics computational tools. It is highly desirable that, in the foreseeable future, researchers make efforts to become native users of these technologies to keep up with the current and future challenges in the field of cancer lncRNAs.

A crucial step in the study of lncRNAs as disease markers from a statistical point of view is the validation of their performance as predictors. However, there is still vague information and discrepancies about the biostatistical and machine learning approaches that have been used to discover and evaluate the performance of lncRNAs signatures, the power of these approaches, and the recall and sensitivity they report. In this regard, it is in the best interest of the entire research community to reach a consensus on the best practices for producing a more robust statistical model and description of the many steps in the selection of these models and validation methods. For example, the studies that used feature selection strategies to select lncRNAs as the variables of interest followed by the use of univariable or multi-variable regression models (e.g., Cox proportional hazards regression analysis) should provide additional rationales for the chosen method. In order to increase the scientific validity of statistical or machine learning approaches based on multi-omics data, it is important that researchers ensure the reproducibility of their findings by incorporating cross-validation information, which would give researchers a clearer perspective on the bias and general application of such models.

As the pace of technological advancement accelerates, an even greater amount of data will be generated from which researchers can extract novel features to describe the biological, physiological, and pharmacological properties of lncRNAs. Perhaps the two most notorious state-of-the-art paradigms in transcriptomic analysis are single-cell and spatial RNA-seq technologies, which seek to explore the repertoire of RNA molecules and their differential expressions at a single cell level and histological coordinates. Beyond the obvious cost of computational resources, this scenario comes at the expense of simple modeling, as incorporating these data into our research will add new layers of complexity. In this context, the lack of parsimony could become a more latent risk when generating new models explaining the associations between lncRNAs and different cancer hallmarks and their observable pathophysiological consequences (i.e., tumor invasion, metastasis, immune evasion, immune infiltration, etc.). By integrating single-cell and spatial transcriptomics data into multi-omics frameworks, researchers would be able to obtain new features such as the proportion and identity of cells within a given tumor that express a particular lncRNA. Even more, these data could be used to refine the association of lncRNA molecules with features such as the TMB, which is currently of value in cancer prognosis, the proportion of specialized cells harboring a somatic aberration in a locus containing functional lncRNAs, or the specific wiring of epigenomics interactions and signals in which lncRNAs are involved within such a cell. From a clinical point of view, having these features as the predictor variables associated with a particular outcome could indeed make the modeling process more demanding and complex, with the risk of overfitting. However, this risk can be mitigated if the precautions mentioned in the previous paragraphs are taken into account. Ultimately, the addition of these new dimensions could enhance the potential use of IncRNAs as diagnostic, prognostic, and predictive factors with more objective real-world applications in cancer clinics.

Author Contributions: Conceptualization, I.S-G. and S.L.R.-C.; investigation, I.S-G., S.L.R.-C. and B.R.-Z.; writing—original draft preparation, I.S-G., S.L.R.-C. and B.R.-Z.; writing—review and editing, I.S.-G., S.L.R.-C. and B.R.-Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

**Informed Consent Statement:** Not applicable.

Data Availability Statement: Not applicable.

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

#### References

- Francies, H.E.; McDermott, U.; Garnett, M.J. Genomics-Guided Pre-Clinical Development of Cancer Therapies. *Nat. Cancer* 2020, 1, 482–492. [CrossRef] [PubMed]
- Malone, E.R.; Oliva, M.; Sabatini, P.J.B.; Stockley, T.L.; Siu, L.L. Molecular Profiling for Precision Cancer Therapies. *Genome Med.* 2020, 12, 8. [CrossRef]
- Perakakis, N.; Yazdani, A.; Karniadakis, G.E.; Mantzoros, C. Omics, Big Data and Machine Learning as Tools to Propel Understanding of Biological Mechanisms and to Discover Novel Diagnostics and Therapeutics. *Metabolism* 2018, 87, A1–A9. [CrossRef] [PubMed]
- Vlachavas, E.I.; Bohn, J.; Ückert, F.; Nürnberg, S. A Detailed Catalogue of Multi-Omics Methodologies for Identification of Putative Biomarkers and Causal Molecular Networks in Translational Cancer Research. *Int. J. Mol. Sci.* 2021, 22, 2822. [CrossRef] [PubMed]
- 5. Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A Revolutionary Tool for Transcriptomics. *Nat. Rev. Genet.* 2009, 10, 57–63. [CrossRef]
- 6. Hong, M.; Tao, S.; Zhang, L.; Diao, L.-T.; Huang, X.; Huang, S.; Xie, S.-J.; Xiao, Z.-D.; Zhang, H. RNA Sequencing: New Technologies and Applications in Cancer Research. *J. Hematol. Oncol.* **2020**, *13*, 166. [CrossRef]
- 7. Adelman, K.; Egan, E. Non-Coding RNA: More Uses for Genomic Junk. Nature 2017, 543, 183–185. [CrossRef]
- 8. Moore, J.E.; Purcaro, M.J.; Pratt, H.E.; Epstein, C.B.; Shoresh, N.; Adrian, J.; Kawli, T.; Davis, C.A.; Dobin, A.; ENCODE Project Consortium; et al. Expanded Encyclopaedias of DNA Elements in the Human and Mouse Genomes. *Nature* **2020**, *583*, 699–710.
- 9. Iyer, M.K.; Niknafs, Y.S.; Malik, R.; Singhal, U.; Sahu, A.; Hosono, Y.; Barrette, T.R.; Prensner, J.R.; Evans, J.R.; Zhao, S.; et al. The Landscape of Long Noncoding RNAs in the Human Transcriptome. *Nat. Genet.* **2015**, *47*, 199–208. [CrossRef]
- 10. Zhang, X.; Meyerson, M. Illuminating the Noncoding Genome in Cancer. Nat. Cancer 2020, 1, 864–872. [CrossRef]
- 11. Kopp, F.; Mendell, J.T. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* **2018**, 172, 393–407. [CrossRef] [PubMed]
- 12. Matsui, M.; Corey, D.R. Non-Coding RNAs as Drug Targets. Nat. Rev. Drug Discov. 2017, 16, 167–179. [CrossRef]
- Amodio, N.; Stamato, M.A.; Juli, G.; Morelli, E.; Fulciniti, M.; Manzoni, M.; Taiana, E.; Agnelli, L.; Cantafio, M.E.G.; Romeo, E.; et al. Drugging the lncRNA MALAT1 via LNA gapmeR ASO Inhibits Gene Expression of Proteasome Subunits and Triggers Anti-Multiple Myeloma Activity. *Leukemia* 2018, *32*, 1948–1957. [CrossRef] [PubMed]
- 14. Tang, Q.; Zheng, F.; Liu, Z.; Wu, J.; Chai, X.; He, C.; Li, L.; Hann, S.S. Novel Reciprocal Interaction of lncRNA HOTAIR and miR-214-3p Contribute to the Solamargine-Inhibited PDPK1 Gene Expression in Human Lung Cancer. *J. Cell. Mol. Med.* **2019**, *23*, 7749–7761. [CrossRef]
- 15. Cieślik, M.; Chinnaiyan, A.M. Cancer Transcriptome Profiling at the Juncture of Clinical Translation. *Nat. Rev. Genet.* **2018**, *19*, 93–109. [CrossRef] [PubMed]
- 16. Zhou, M.; Hu, L.; Zhang, Z.; Wu, N.; Sun, J.; Su, J. Recurrence-Associated Long Non-Coding RNA Signature for Determining the Risk of Recurrence in Patients with Colon Cancer. *Mol. Ther. Nucleic Acids* **2018**, *12*, 518–529. [CrossRef] [PubMed]
- 17. Yuan, L.; Zhao, J.; Sun, T.; Shen, Z. A Machine Learning Framework That Integrates Multi-Omics Data Predicts Cancer-Related lncRNAs. *BMC Bioinform.* 2021, 22, 332. [CrossRef] [PubMed]
- Heo, Y.J.; Hwa, C.; Lee, G.-H.; Park, J.-M.; An, J.-Y. Integrative Multi-Omics Approaches in Cancer Research: From Biological Networks to Clinical Subtypes. *Mol. Cells* 2021, 44, 433–443. [CrossRef]
- Qian, Y.; Shi, L.; Luo, Z. Long Non-Coding RNAs in Cancer: Implications for Diagnosis, Prognosis, and Therapy. *Front. Med.* 2020, 7, 612393. [CrossRef] [PubMed]
- 20. Yan, C.; Zhang, Z.; Bao, S.; Hou, P.; Zhou, M.; Xu, C.; Sun, J. Computational Methods and Applications for Identifying Disease-Associated lncRNAs as Potential Biomarkers and Therapeutic Targets. *Mol. Ther. Nucleic Acids* **2020**, *21*, 156–171. [CrossRef]
- Chen, G.; Wang, Z.; Wang, D.; Qiu, C.; Liu, M.; Chen, X.; Zhang, Q.; Yan, G.; Cui, Q. IncRNADisease: A Database for Long-Non-Coding RNA-Associated Diseases. *Nucleic Acids Res.* 2013, 41, D983–D986. [CrossRef]
- 22. Aprile, M.; Katopodi, V.; Leucci, E.; Costa, V. IncRNAs in Cancer: From Garbage to Junk. Cancers 2020, 12, 3220. [CrossRef]

- 23. Sheng, N.; Huang, L.; Gao, L.; Cao, Y.; Xie, X.; Wang, Y. A Survey of Computational Methods and Databases for lncRNA-MiRNA Interaction Prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2023**, *20*, 2810–2826. [CrossRef]
- 24. Raufaste-Cazavieille, V.; Santiago, R.; Droit, A. Multi-Omics Analysis: Paving the Path toward Achieving Precision Medicine in Cancer Treatment and Immuno-Oncology. *Front. Mol. Biosci.* **2022**, *9*, 962743. [CrossRef]
- 25. Hanahan, D.; Weinberg, R.A. Hallmarks of Cancer: The next Generation. Cell 2011, 144, 646–674. [CrossRef]
- Negrini, S.; Gorgoulis, V.G.; Halazonetis, T.D. Genomic Instability—An Evolving Hallmark of Cancer. *Nat. Rev. Mol. Cell Biol.* 2010, 11, 220–228. [CrossRef]
- Vogelstein, B.; Papadopoulos, N.; Velculescu, V.E.; Zhou, S.; Diaz, L.A., Jr.; Kinzler, K.W. Cancer Genome Landscapes. *Science* 2013, 339, 1546–1558. [CrossRef]
- Lawrence, M.S.; Stojanov, P.; Polak, P.; Kryukov, G.V.; Cibulskis, K.; Sivachenko, A.; Carter, S.L.; Stewart, C.; Mermel, C.H.; Roberts, S.A.; et al. Mutational Heterogeneity in Cancer and the Search for New Cancer-Associated Genes. *Nature* 2013, 499, 214–218. [CrossRef]
- 29. Korf, B.R.; Rehm, H.L. New Approaches to Molecular Diagnosis. JAMA 2013, 309, 1511–1521. [CrossRef]
- 30. Piraino, S.W.; Furney, S.J. Beyond the Exome: The Role of Non-Coding Somatic Mutations in Cancer. *Ann. Oncol.* **2016**, 27, 240–248. [CrossRef]
- Rheinbay, E.; Nielsen, M.M.; Abascal, F.; Wala, J.A.; Shapira, O.; Tiao, G.; Hornshøj, H.; Hess, J.M.; Juul, R.I.; Lin, Z.; et al. Analyses of Non-Coding Somatic Drivers in 2658 Cancer Whole Genomes. *Nature* 2020, 578, 102–111. [CrossRef]
- Rheinbay, E.; Parasuraman, P.; Grimsby, J.; Tiao, G.; Engreitz, J.M.; Kim, J.; Lawrence, M.S.; Taylor-Weiner, A.; Rodriguez-Cuevas, S.; Rosenberg, M.; et al. Recurrent and Functional Regulatory Mutations in Breast Cancer. *Nature* 2017, 547, 55–60. [CrossRef] [PubMed]
- Fujimoto, A.; Furuta, M.; Totoki, Y.; Tsunoda, T.; Kato, M.; Shiraishi, Y.; Tanaka, H.; Taniguchi, H.; Kawakami, Y.; Ueno, M.; et al. Whole-Genome Mutational Landscape and Characterization of Noncoding and Structural Mutations in Liver Cancer. *Nat. Genet.* 2016, 48, 500–509. [CrossRef] [PubMed]
- Gasic, V.; Karan-Djurasevic, T.; Pavlovic, D.; Zukic, B.; Pavlovic, S.; Tosic, N. Diagnostic and Therapeutic Implications of Long Non-Coding RNAs in Leukemia. *Life* 2022, 12, 1770. [CrossRef]
- 35. Puente, X.S.; Beà, S.; Valdés-Mas, R.; Villamor, N.; Gutiérrez-Abril, J.; Martín-Subero, J.I.; Munar, M.; Rubio-Pérez, C.; Jares, P.; Aymerich, M.; et al. Non-Coding Recurrent Mutations in Chronic Lymphocytic Leukaemia. *Nature* **2015**, *526*, 519–524. [CrossRef]
- Nik-Zainal, S.; Davies, H.; Staaf, J.; Ramakrishna, M.; Glodzik, D.; Zou, X.; Martincorena, I.; Alexandrov, L.B.; Martin, S.; Wedge, D.C.; et al. Author Correction: Landscape of Somatic Mutations in 560 Breast Cancer Whole-Genome Sequences. *Nature* 2019, 566, E1. [CrossRef]
- Rezaie, N.; Bayati, M.; Hamidi, M.; Tahaei, M.S.; Khorasani, S.; Lovell, N.H.; Breen, J.; Rabiee, H.R.; Alinejad-Rokny, H. Somatic Point Mutations Are Enriched in Non-Coding RNAs with Possible Regulatory Function in Breast Cancer. *Commun. Biol.* 2022, 5, 556. [CrossRef]
- Iraola-Guzmán, S.; Brunet-Vega, A.; Pegueroles, C.; Saus, E.; Hovhannisyan, H.; Casalots, A.; Pericay, C.; Gabaldón, T. Target Enrichment Enables the Discovery of lncRNAs with Somatic Mutations or Altered Expression in Paraffin-Embedded Colorectal Cancer Samples. *Cancers* 2020, *12*, 2844. [CrossRef] [PubMed]
- Fang, X.; Liu, X.; Lu, L.; Liu, G. Identification of a Somatic Mutation-Derived Long Non-Coding RNA Signatures of Genomic Instability in Renal Cell Carcinoma. Front. Oncol. 2021, 11, 728181. [CrossRef]
- 40. Wang, Y.; Yan, K.; Wang, L.; Bi, J. Genome Instability-Related Long Non-Coding RNA in Clear Renal Cell Carcinoma Determined Using Computational Biology. *BMC Cancer* **2021**, *21*, 727. [CrossRef]
- Xu, Q.; Guo, T. Somatic Mutation-Associated Risk Index Based on IncRNA Expression for Predicting Prognosis in Acute Myeloid Leukemia. *Hematology* 2022, 27, 659–671. [CrossRef]
- 42. Yun, D.; Yang, Z. Identification of a Four-IncRNA Prognostic Signature for Colon Cancer Based on Genome Instability. *J. Oncol.* **2021**, 2021, 7408893. [CrossRef]
- Yin, T.; Zhao, D.; Yao, S. Identification of a Genome Instability-Associated lncRNA Signature for Prognosis Prediction in Colon Cancer. Front. Genet. 2021, 12, 679150. [CrossRef] [PubMed]
- Zhang, J.; Ding, N.; He, Y.; Tao, C.; Liang, Z.; Xin, W.; Zhang, Q.; Wang, F. Bioinformatic Identification of Genomic Instability-Associated lncRNAs Signatures for Improving the Clinical Outcome of Cervical Cancer by a Prognostic Model. *Sci. Rep.* 2021, 11, 20929. [CrossRef] [PubMed]
- Jin, C.; Zhao, J.-S.; Huang, X.-Q.; Yang, X.-Z.; Niu, F.-Y.; Lin, J.-R.; Ma, L.; Shi, Y.-X.; Li, X.-S.; Jiang, P.; et al. A Somatic Mutation-Derived lncRNA Signatures of Genomic Instability Predicts the Prognosis and Tumor Microenvironment Immune Characters in Hepatocellular Carcinoma. *Hepatol. Int.* 2022, *16*, 1220–1233. [CrossRef]
- Wu, J.; Ren, X.; Wang, N.; Zhou, R.; Chen, M.; Cai, Y.; Lin, S.; Zhang, H.; Xie, X.; Dang, C.; et al. A Mutation-Related Long Noncoding RNA Signature of Genome Instability Predicts Immune Infiltration and Hepatocellular Carcinoma Prognosis. *Front. Genet.* 2021, 12, 779554. [CrossRef] [PubMed]
- 47. Yan, X.; Hu, Z.; Feng, Y.; Hu, X.; Yuan, J.; Zhao, S.D.; Zhang, Y.; Yang, L.; Shan, W.; He, Q.; et al. Comprehensive Genomic Characterization of Long Non-Coding RNAs across Human Cancers. *Cancer Cell* **2015**, *28*, 529–540. [CrossRef]

- Zhong, Q.; Lu, M.; Yuan, W.; Cui, Y.; Ouyang, H.; Fan, Y.; Wang, Z.; Wu, C.; Qiao, J.; Hang, J. Eight-IncRNA Signature of Cervical Cancer Were Identified by Integrating DNA Methylation, Copy Number Variation and Transcriptome Data. *J. Transl. Med.* 2021, 19, 58. [CrossRef]
- Zhou, C.-C.; Yang, F.; Yuan, S.-X.; Ma, J.-Z.; Liu, F.; Yuan, J.-H.; Bi, F.-R.; Lin, K.-Y.; Yin, J.-H.; Cao, G.-W.; et al. Systemic Genome Screening Identifies the Outcome Associated Focal Loss of Long Noncoding RNA PRAL in Hepatocellular Carcinoma. *Hepatology* 2016, 63, 850–863. [CrossRef]
- 50. Liu, H.; Gu, X.; Wang, G.; Huang, Y.; Ju, S.; Huang, J.; Wang, X. Copy Number Variations Primed lncRNAs Deregulation Contribute to Poor Prognosis in Colorectal Cancer. *Aging* **2019**, *11*, 6089–6108. [CrossRef]
- 51. Zhang, Y.; Liao, G.; Bai, J.; Zhang, X.; Xu, L.; Deng, C.; Yan, M.; Xie, A.; Luo, T.; Long, Z.; et al. Identifying Cancer Driver lncRNAs Bridged by Functional Effectors through Integrating Multi-Omics Data in Human Cancers. *Mol. Ther. Nucleic Acids* **2019**, *17*, 362–373. [CrossRef] [PubMed]
- 52. Lin, W.; Tan, Z.-Y.; Fang, X.-C. Identification of m6A-Related lncRNAs-Based Signature for Predicting the Prognosis of Patients with Skin Cutaneous Melanoma. *SLAS Technol.* **2023**, *in press.* [CrossRef] [PubMed]
- Zhu, X.; Yu, R.; Peng, Y.; Miao, Y.; Jiang, K.; Li, Q. Identification of Genomic Instability Related lncRNA Signature with Prognostic Value and Its Role in Cancer Immunotherapy in Pancreatic Cancer. *Front. Genet.* 2022, 13, 990661. [CrossRef]
- 54. Ning, J.; Wang, F.; Zhu, K.; Li, B.; Shu, Q.; Liu, W. Characterizing the Copy Number Variation of Non-Coding RNAs Reveals Potential Therapeutic Targets and Prognostic Markers of LUSC. *Front. Genet.* **2021**, *12*, 779155. [CrossRef] [PubMed]
- 55. Luo, Z.; Xiao, L.; Li, J.; Dong, B.; Wang, C. Integrative Analysis Reveals Driver Long Non-Coding RNAs in Osteosarcoma. *Medicine* **2019**, *98*, e14302. [CrossRef]
- Athie, A.; Marchese, F.P.; González, J.; Lozano, T.; Raimondi, I.; Juvvuna, P.K.; Abad, A.; Marin-Bejar, O.; Serizay, J.; Martínez, D.; et al. Analysis of Copy Number Alterations Reveals the lncRNA ALAL-1 as a Regulator of Lung Cancer Immune Evasion. *J. Cell Biol.* 2020, 219, e201908078. [CrossRef]
- 57. Cheng, Z.; Guo, Y.; Sun, J.; Zheng, L. Four-Copy Number Alteration (CNA)-Related lncRNA Prognostic Signature for Liver Cancer. *Sci. Rep.* **2022**, *12*, 14261. [CrossRef]
- Du, Z.; Fei, T.; Verhaak, R.G.W.; Su, Z.; Zhang, Y.; Brown, M.; Chen, Y.; Liu, X.S. Integrative Genomic Analyses Reveal Clinically Relevant Long Noncoding RNAs in Human Cancer. *Nat. Struct. Mol. Biol.* 2013, 20, 908–913. [CrossRef]
- Zheng, M.; Hu, Y.; Gou, R.; Nie, X.; Li, X.; Liu, J.; Lin, B. Identification Three lncRNA Prognostic Signature of Ovarian Cancer Based on Genome-Wide Copy Number Variation. *Biomed. Pharmacother.* 2020, 124, 109810. [CrossRef]
- 60. Hou, X.; Ma, B.; Liu, M.; Zhao, Y.; Chai, B.; Pan, J.; Wang, P.; Li, D.; Liu, S.; Song, F. The Transcriptional Risk Scores for Kidney Renal Clear Cell Carcinoma Using XGBoost and Multiple Omics Data. *Math. Biosci. Eng.* **2023**, *20*, 11676–11687. [CrossRef]
- Peng, W.; Bai, S.; Zheng, M.; Chen, W.; Li, Y.; Yang, Y.; Zhao, Y.; Xiong, S.; Wang, R.; Cheng, B. An Exosome-Related lncRNA Signature Correlates with Prognosis, Immune Microenvironment, and Therapeutic Responses in Hepatocellular Carcinoma. *Transl. Oncol.* 2023, *31*, 101651. [CrossRef]
- 62. Yu, Z.; Zhu, Y.; Ji, J. Multi-Omics Analysis Revealing a Senescence-Relevant lncRNAs Signature for the Assessment of Response to Immunotherapy for Breast Cancer. *Medicine* **2023**, *102*, e34287. [CrossRef]
- 63. Zhou, Q.-N.; Lei, R.-E.; Liang, Y.-X.; Li, S.-Q.; Guo, X.-W.; Hu, B.-L. Oxaliplatin Related lncRNAs Prognostic Models Predict the Prognosis of Patients given Oxaliplatin-Based Chemotherapy. *Cancer Cell Int.* **2023**, *23*, 103. [CrossRef]
- Li, G.; Huo, D.; Guo, N.; Li, Y.; Ma, H.; Liu, L.; Xie, H.; Zhang, D.; Qu, B.; Chen, X. Integrating Multiple Machine Learning Algorithms for Prognostic Prediction of Gastric Cancer Based on Immune-Related lncRNAs. *Front. Genet.* 2023, 14, 1106724. [CrossRef] [PubMed]
- Yu, Y.; Ren, W.; He, Z.; Chen, Y.; Tan, Y.; Mao, L.; Ouyang, W.; Lu, N.; Ouyang, J.; Chen, K.; et al. Machine Learning Radiomics of Magnetic Resonance Imaging Predicts Recurrence-Free Survival after Surgery and Correlation of lncRNAs in Patients with Breast Cancer: A Multicenter Cohort Study. *Breast Cancer Res.* 2023, 25, 132. [CrossRef] [PubMed]
- 66. Maimaiti, A.; Aili, Y.; Turhon, M.; Kadeer, K.; Aikelamu, P.; Wang, Z.; Niu, W.; Aisha, M.; Kasimu, M.; Wang, Y.; et al. Modification Patterns of DNA Methylation-Related lncRNAs Regulating Genomic Instability for Improving the Clinical Outcomes and Tumour Microenvironment Characterisation of Lower-Grade Gliomas. *Front. Mol. Biosci.* 2022, 9, 844973. [CrossRef] [PubMed]
- 67. Wang, S.; Wang, R.; Gao, F.; Huang, J.; Zhao, X.; Li, D. Pan-Cancer Analysis of the DNA Methylation Patterns of Long Non-Coding RNA. *Genomics* 2022, 114, 110377. [CrossRef]
- Song, P.; Li, Y.; Wang, F.; Pu, L.; Bao, L.; Gao, H.; Zhu, C.; Wang, M.; Tao, L. Genome-Wide Screening for Differentially Methylated Long Noncoding RNAs Identifies LIFR-AS1 as an Epigenetically Regulated lncRNA That Inhibits the Progression of Colorectal Cancer. *Clin. Epigenetics* 2022, 14, 138. [CrossRef] [PubMed]
- 69. Gao, Y.; Li, X.; Zhi, H.; Zhang, Y.; Wang, P.; Wang, Y.; Shang, S.; Fang, Y.; Shen, W.; Ning, S.; et al. Comprehensive Characterization of Somatic Mutations Impacting IncRNA Expression for Pan-Cancer. *Mol. Ther. Nucleic Acids* **2019**, *18*, 66–79. [CrossRef]
- 70. Esposito, R.; Lanzós, A.; Uroda, T.; Ramnarayanan, S.; Büchi, I.; Polidori, T.; Guillen-Ramirez, H.; Mihaljevic, A.; Merlin, B.M.; Mela, L.; et al. Tumour Mutations in Long Noncoding RNAs Enhance Cell Fitness. *Nat. Commun.* 2023, 14, 3342. [CrossRef]
- 71. Zhang, Y.; Han, P.; Guo, Q.; Hao, Y.; Qi, Y.; Xin, M.; Zhang, Y.; Cui, B.; Wang, P. Oncogenic Landscape of Somatic Mutations Perturbing Pan-Cancer IncRNA-ceRNA Regulation. *Front. Cell Dev. Biol.* **2021**, *9*, 658346. [CrossRef]
- 72. Brady, S.W.; Gout, A.M.; Zhang, J. Therapeutic and Prognostic Insights from the Analysis of Cancer Mutational Signatures. *Trends Genet.* 2022, *38*, 194–208. [CrossRef] [PubMed]

- Rizzo, A.; Ricci, A.D.; Di Federico, A.; Frega, G.; Palloni, A.; Tavolari, S.; Brandi, G. Predictive Biomarkers for Checkpoint Inhibitor-Based Immunotherapy in Hepatocellular Carcinoma: Where Do We Stand? *Front. Oncol.* 2021, *11*, 803133. [CrossRef] [PubMed]
- 74. Ding, C.; Shan, Z.; Li, M.; Xia, Y.; Jin, Z. Exploration of the Associations of lncRNA Expression Patterns with Tumor Mutation Burden and Prognosis in Colon Cancer. *OncoTargets Ther.* **2021**, *14*, 2893–2909. [CrossRef] [PubMed]
- 75. Lee, J.J.-K.; Park, S.; Park, H.; Kim, S.; Lee, J.; Lee, J.; Youk, J.; Yi, K.; An, Y.; Park, I.K.; et al. Tracing Oncogene Rearrangements in the Mutational History of Lung Adenocarcinoma. *Cell* **2019**, 177, 1842–1857.e21. [CrossRef] [PubMed]
- 76. Bae, K.; Kim, J.H.; Jung, H.; Kong, S.-Y.; Kim, Y.-H.; Kim, S.; Lee, G.K.; Lee, J.S.; Lee, J.J.-K.; Ju, Y.S.; et al. A Fusion of CD63-BCAR4 Identified in Lung Adenocarcinoma Promotes Tumorigenicity and Metastasis. *Br. J. Cancer* 2021, 124, 290–298. [CrossRef] [PubMed]
- 77. Jordan, E.J.; Kim, H.R.; Arcila, M.E.; Barron, D.; Chakravarty, D.; Gao, J.; Chang, M.T.; Ni, A.; Kundra, R.; Jonsson, P.; et al. Prospective Comprehensive Molecular Characterization of Lung Adenocarcinomas for Efficient Patient Matching to Approved and Emerging Therapies. *Cancer Discov.* 2017, 7, 596–609. [CrossRef] [PubMed]
- 78. Koivunen, J.P.; Mermel, C.; Zejnullahu, K.; Murphy, C.; Lifshits, E.; Holmes, A.J.; Choi, H.G.; Kim, J.; Chiang, D.; Thomas, R.; et al. EML4-ALK Fusion Gene and Efficacy of an ALK Kinase Inhibitor in Lung Cancer. *Clin. Cancer Res.* 2008, 14, 4275–4283. [CrossRef]
- Wang, C.; Yin, R.; Dai, J.; Gu, Y.; Cui, S.; Ma, H.; Zhang, Z.; Huang, J.; Qin, N.; Jiang, T.; et al. Whole-Genome Sequencing Reveals Genomic Signatures Associated with the Inflammatory Microenvironments in Chinese NSCLC Patients. *Nat. Commun.* 2018, 9, 2054. [CrossRef]
- 80. Godinho, M.F.E.; Sieuwerts, A.M.; Look, M.P.; Meijer, D.; Foekens, J.A.; Dorssers, L.C.J.; van Agthoven, T. Relevance of BCAR4 in Tamoxifen Resistance and Tumour Aggressiveness of Human Breast Cancer. *Br. J. Cancer* **2010**, *103*, 1284–1291. [CrossRef]
- Guichard, C.; Amaddeo, G.; Imbeaud, S.; Ladeiro, Y.; Pelletier, L.; Maad, I.B.; Calderaro, J.; Bioulac-Sage, P.; Letexier, M.; Degos, F.; et al. Integrated Analysis of Somatic Mutations and Focal Copy-Number Changes Identifies Key Genes and Pathways in Hepatocellular Carcinoma. *Nat. Genet.* 2012, 44, 694–698. [CrossRef]
- 82. Shao, X.; Lv, N.; Liao, J.; Long, J.; Xue, R.; Ai, N.; Xu, D.; Fan, X. Copy Number Variation Is Highly Correlated with Differential Gene Expression: A Pan-Cancer Study. *BMC Med. Genet.* **2019**, *20*, 175. [CrossRef] [PubMed]
- 83. Zack, T.I.; Schumacher, S.E.; Carter, S.L.; Cherniack, A.D.; Saksena, G.; Tabak, B.; Lawrence, M.S.; Zhsng, C.-Z.; Wala, J.; Mermel, C.H.; et al. Pan-Cancer Patterns of Somatic Copy Number Alteration. *Nat. Genet.* **2013**, *45*, 1134–1140. [CrossRef] [PubMed]
- Zhang, Y.; Yang, L.; Kucherlapati, M.; Chen, F.; Hadjipanayis, A.; Pantazi, A.; Bristow, C.A.; Lee, E.A.; Mahadeshwar, H.S.; Tang, J.; et al. A Pan-Cancer Compendium of Genes Deregulated by Somatic Genomic Rearrangement across More Than 1400 Cases. *Cell Rep.* 2018, 24, 515–527. [CrossRef] [PubMed]
- 85. Beroukhim, R.; Mermel, C.H.; Porter, D.; Wei, G.; Raychaudhuri, S.; Donovan, J.; Barretina, J.; Boehm, J.S.; Dobson, J.; Urashima, M.; et al. The Landscape of Somatic Copy-Number Alteration across Human Cancers. *Nature* **2010**, *463*, 899–905. [CrossRef]
- 86. Ping, Y.; Zhou, Y.; Hu, J.; Pang, L.; Xu, C.; Xiao, Y. Dissecting the Functional Mechanisms of Somatic Copy-Number Alterations Based on Dysregulated ceRNA Networks across Cancers. *Mol. Ther. Nucleic Acids* **2020**, *21*, 464–479. [CrossRef] [PubMed]
- Hu, X.; Feng, Y.; Zhang, D.; Zhao, S.D.; Hu, Z.; Greshock, J.; Zhang, Y.; Yang, L.; Zhong, X.; Wang, L.-P.; et al. A Functional Genomic Approach Identifies FAL1 as an Oncogenic Long Noncoding RNA That Associates with BMI1 and Represses p21 Expression in Cancer. *Cancer Cell* 2014, 26, 344–357. [CrossRef]
- Wang, Z.-L.; Li, B.; Piccolo, S.R.; Zhang, X.-Q.; Li, J.-H.; Zhou, H.; Yang, J.-H.; Qu, L.-H. Integrative Analysis Reveals Clinical Phenotypes and Oncogenic Potentials of Long Non-Coding RNAs across 15 Cancer Types. *Oncotarget* 2016, *7*, 35044–35055. [CrossRef]
- 89. Zheng, X.; Li, F.; Zhao, H.; Tang, Y.; Xue, K.; Zhang, X.; Liang, W.; Zhao, R.; Lv, X.; Song, X.; et al. A Novel Method to Identify and Characterize Personalized Functional Driver lncRNAs in Cancer Samples. *Comput. Struct. Biotechnol. J.* **2023**, *21*, 2471–2482. [CrossRef]
- 90. Djebali, S.; Davis, C.A.; Merkel, A.; Dobin, A.; Lassmann, T.; Mortazavi, A.; Tanzer, A.; Lagarde, J.; Lin, W.; Schlesinger, F.; et al. Landscape of Transcription in Human Cells. *Nature* **2012**, *489*, 101–108. [CrossRef]
- 91. Wei, R.; Zhao, M.; Zheng, C.-H.; Zhao, M.; Xia, J. Concordance between Somatic Copy Number Loss and down-Regulated Expression: A Pan-Cancer Study of Cancer Predisposition Genes. *Sci. Rep.* **2016**, *6*, 37358. [CrossRef]
- 92. Dagogo-Jack, I.; Shaw, A.T. Tumour Heterogeneity and Resistance to Cancer Therapies. *Nat. Rev. Clin. Oncol.* 2018, 15, 81–94. [CrossRef]
- 93. Fisher, R.; Pusztai, L.; Swanton, C. Cancer Heterogeneity: Implications for Targeted Therapeutics. *Br. J. Cancer* 2013, *108*, 479–485. [CrossRef] [PubMed]
- 94. Chen, H.; Xu, J.; Hong, J.; Tang, R.; Zhang, X.; Fang, J.-Y. Long Noncoding RNA Profiles Identify Five Distinct Molecular Subtypes of Colorectal Cancer with Clinical Relevance. *Mol. Oncol.* **2014**, *8*, 1393–1403. [CrossRef] [PubMed]
- 95. Akrami, R.; Jacobsen, A.; Hoell, J.; Schultz, N.; Sander, C.; Larsson, E. Comprehensive Analysis of Long Non-Coding RNAs in Ovarian Cancer Reveals Global Patterns and Targeted DNA Amplification. *PLoS ONE* **2013**, *8*, e80306. [CrossRef]
- 96. Volders, P.-J.; Anckaert, J.; Verheggen, K.; Nuytens, J.; Martens, L.; Mestdagh, P.; Vandesompele, J. LNCipedia 5: Towards a Reference Set of Human Long Non-Coding RNAs. *Nucleic Acids Res.* **2019**, *47*, D135–D139. [CrossRef]
- 97. Sun, Y.; Ma, L. New Insights into Long Non-Coding RNA in Cancer and Metastasis. Cancers 2019, 11, 216. [CrossRef] [PubMed]

- 98. Bhan, A.; Mandal, S.S. IncRNA HOTAIR: A Master Regulator of Chromatin Dynamics and Cancer. *Biochim. Biophys. Acta* 2015, 1856, 151–164. [CrossRef]
- Ferreira, L.B.; Palumbo, A.; de Mello, K.D.; Sternberg, C.; Caetano, M.S.; de Oliveira, F.L.; Neves, A.F.; Nasciutti, L.E.; Goulart, L.R.; Gimba, E.R.P. PCA3 Noncoding RNA Is Involved in the Control of Prostate-Cancer Cell Survival and Modulates Androgen Receptor Signaling. *BMC Cancer* 2012, 12, 507. [CrossRef]
- 100. Ji, P.; Diederichs, S.; Wang, W.; Böing, S.; Metzger, R.; Schneider, P.M.; Tidow, N.; Brandt, B.; Buerger, H.; Bulk, E.; et al. MALAT-1, a Novel Noncoding RNA, and Thymosin beta4 Predict Metastasis and Survival in Early-Stage Non-Small Cell Lung Cancer. Oncogene 2003, 22, 8031–8041. [CrossRef]
- Okugawa, Y.; Toiyama, Y.; Hur, K.; Toden, S.; Saigusa, S.; Tanaka, K.; Inoue, Y.; Mohri, Y.; Kusunoki, M.; Boland, C.R.; et al. Metastasis-Associated Long Non-Coding RNA Drives Gastric Cancer Development and Promotes Peritoneal Metastasis. *Carcino-genesis* 2014, 35, 2731–2739. [CrossRef] [PubMed]
- 102. Zheng, H.-T.; Shi, D.-B.; Wang, Y.-W.; Li, X.-X.; Xu, Y.; Tripathi, P.; Gu, W.-L.; Cai, G.-X.; Cai, S.-J. High Expression of IncRNA MALAT1 Suggests a Biomarker of Poor Prognosis in Colorectal Cancer. Int. J. Clin. Exp. Pathol. 2014, 7, 3174–3181. [PubMed]
- Lu, R.; Zhang, J.; Zhang, W.; Huang, Y.; Wang, N.; Zhang, Q.; Qu, S. Circulating HOTAIR Expression Predicts the Clinical Response to Neoadjuvant Chemotherapy in Patients with Breast Cancer. *Cancer Biomark.* 2018, 22, 249–256. [CrossRef]
- 104. Yang, Z.; Zhou, L.; Wu, L.-M.; Lai, M.-C.; Xie, H.-Y.; Zhang, F.; Zheng, S.-S. Overexpression of Long Non-Coding RNA HOTAIR Predicts Tumor Recurrence in Hepatocellular Carcinoma Patients Following Liver Transplantation. Ann. Surg. Oncol. 2011, 18, 1243–1250. [CrossRef]
- 105. Lemos, A.E.G.; da Rocha Matos, A.; Ferreira, L.B.; Gimba, E.R.P. The Long Non-Coding RNA: An Update of Its Functions and Clinical Applications as a Biomarker in Prostate Cancer. *Oncotarget* **2019**, *10*, 6589–6603. [CrossRef] [PubMed]
- 106. Pal, R.P.; Maitra, N.U.; Mellon, J.K.; Khan, M.A. Defining Prostate Cancer Risk before Prostate Biopsy. Urol. Oncol. 2013, 31, 1408–1418. [CrossRef]
- 107. Li, Z.; Liu, L.; Feng, C.; Qin, Y.; Xiao, J.; Zhang, Z.; Ma, L. LncBook 2.0: Integrating Human Long Non-Coding RNAs with Multi-Omics Annotations. *Nucleic Acids Res.* 2023, 51, D186–D191. [CrossRef]
- 108. Carlevaro-Fita, J.; Lanzós, A.; Feuerbach, L.; Hong, C.; Mas-Ponte, D.; Pedersen, J.S.; Johnson, R.; PCAWG Drivers and Functional Interpretation Group. PCAWG Consortium Author Correction: Cancer IncRNA Census Reveals Evidence for Deep Functional Conservation of Long Noncoding RNAs in Tumorigenesis. *Commun. Biol.* 2022, *5*, 1347. [CrossRef]
- 109. Mularoni, L.; Sabarinathan, R.; Deu-Pons, J.; Gonzalez-Perez, A.; López-Bigas, N. OncodriveFML: A General Framework to Identify Coding and Non-Coding Regions with Cancer Driver Mutations. *Genome Biol.* **2016**, *17*, 128. [CrossRef]
- Lanzós, A.; Carlevaro-Fita, J.; Mularoni, L.; Reverter, F.; Palumbo, E.; Guigó, R.; Johnson, R. Discovery of Cancer Driver Long Noncoding RNAs across 1112 Tumour Genomes: New Candidates and Distinguishing Features. *Sci. Rep.* 2017, 7, 41544. [CrossRef]
- 111. Chiu, H.-S.; Somvanshi, S.; Patel, E.; Chen, T.-W.; Singh, V.P.; Zorman, B.; Patil, S.L.; Pan, Y.; Chatterjee, S.S.; Cancer Genome Atlas Research Network; et al. Pan-Cancer Analysis of IncRNA Regulation Supports Their Targeting of Cancer Genes in Each Tumor Context. Cell Rep. 2018, 23, 297–312.e12. [CrossRef] [PubMed]
- Mitra, R.; Adams, C.M.; Eischen, C.M. Systematic lncRNA Mapping to Genome-Wide Co-Essential Modules Uncovers Cancer Dependency on Uncharacterized lncRNAs. *eLife* 2022, *11*, e77357. [CrossRef] [PubMed]
- 113. Yang, Z.; Xu, F.; Wang, H.; Teschendorff, A.E.; Xie, F.; He, Y. Pan-Cancer Characterization of Long Non-Coding RNA and DNA Methylation Mediated Transcriptional Dysregulation. *eBioMedicine* **2021**, *68*, 103399. [CrossRef]
- 114. Salmena, L.; Poliseno, L.; Tay, Y.; Kats, L.; Pandolfi, P.P. A ceRNA Hypothesis: The Rosetta Stone of a Hidden RNA Language? *Cell* 2011, 146, 353–358. [CrossRef]
- 115. Arunkumar, G.; Baek, S.; Sturgill, D.; Bui, M.; Dalal, Y. Oncogenic lncRNAs Alter Epigenetic Memory at a Fragile Chromosomal Site in Human Cancer Cells. *Sci. Adv.* **2022**, *8*, eabl5621. [CrossRef]
- 116. Galon, J.; Bruni, D. Tumor Immunology and Tumor Evolution: Intertwined Histories. Immunity 2020, 52, 55–81. [CrossRef]
- 117. Li, Y.; Jiang, T.; Zhou, W.; Li, J.; Li, X.; Wang, Q.; Jin, X.; Yin, J.; Chen, L.; Zhang, Y.; et al. Pan-Cancer Characterization of Immune-Related IncRNAs Identifies Potential Oncogenic Biomarkers. *Nat. Commun.* **2020**, *11*, 1000. [CrossRef]
- 118. Hur, K.; Kim, S.-H.; Kim, J.-M. Potential Implications of Long Noncoding RNAs in Autoimmune Diseases. *Immune Netw.* 2019, 19, e4. [CrossRef] [PubMed]
- 119. Guo, W.; Wang, Y.; Yang, M.; Wang, Z.; Wang, Y.; Chaurasia, S.; Wu, Z.; Zhang, M.; Yadav, G.S.; Rathod, S.; et al. LincRNA-Immunity Landscape Analysis Identifies EPIC1 as a Regulator of Tumor Immune Evasion and Immunotherapy Resistance. *Sci. Adv.* 2021, 7, eabb3555. [CrossRef]
- 120. Li, J.; Zhang, X.; Liu, C. The Computational Approaches of lncRNA Identification Based on Coding Potential: And Challenges. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 3666–3677. [CrossRef]
- 121. Chi, Y.; Wang, D.; Wang, J.; Yu, W.; Yang, J. Long Non-Coding RNA in the Pathogenesis of Cancers. *Cells* **2019**, *8*, 1015. [CrossRef] [PubMed]
- 122. Zhong, L.; Zhen, M.; Sun, J.; Zhao, Q. Recent Advances on the Machine Learning Methods in Predicting ncRNA-Protein Interactions. *Mol. Genet. Genom.* 2021, 296, 243–258. [CrossRef] [PubMed]
- Li, G.; Luo, J.; Liang, C.; Xiao, Q.; Ding, P.; Zhang, Y. Prediction of lncRNA-Disease Associations Based on Network Consistency Projection. *IEEE Access* 2019, 7, 58849–58856. [CrossRef]

- Fan, X.-N.; Zhang, S.-W.; Zhang, S.-Y.; Zhu, K.; Lu, S. Prediction of IncRNA-Disease Associations by Integrating Diverse Heterogeneous Information Sources with RWR Algorithm and Positive Pointwise Mutual Information. *BMC Bioinform.* 2019, 20, 87. [CrossRef]
- 125. Zhou, J.-R.; You, Z.-H.; Cheng, L.; Ji, B.-Y. Prediction of lncRNA-Disease Associations via an Embedding Learning HOPE in Heterogeneous Information Networks. *Mol. Ther. Nucleic Acids* **2021**, *23*, 277–285. [CrossRef]
- 126. Wang, Y.; Juan, L.; Peng, J.; Zang, T.; Wang, Y. LncDisAP: A Computation Model for lncRNA-Disease Association Prediction Based on Multiple Biological Datasets. *BMC Bioinform.* **2019**, *20*, 582. [CrossRef]
- 127. Yu, J.; Ping, P.; Wang, L.; Kuang, L.; Li, X.; Wu, Z. A Novel Probability Model for lncRNA–Disease Association Prediction Based on the Naïve Bayesian Classifier. *Genes* 2018, *9*, 345. [CrossRef]
- Zhou, M.; Zhang, H.; Bai, Z.; Mann-Krzisnik, D.; Wang, F.; Li, Y. Single-Cell Multi-Omics Topic Embedding Reveals Cell-Type-Specific and COVID-19 Severity-Related Immune Signatures. *Cell Rep. Methods* 2023, *3*, 100563. [CrossRef]
- 129. Treppner, M.; Binder, H.; Hess, M. Interpretable Generative Deep Learning: An Illustration with Single Cell Gene Expression Data. *Hum. Genet.* 2022, 141, 1481–1498. [CrossRef]
- Nath, A.; Lau, E.Y.T.; Lee, A.M.; Geeleher, P.; Cho, W.C.S.; Huang, R.S. Discovering Long Noncoding RNA Predictors of Anticancer Drug Sensitivity beyond Protein-Coding Genes. *Proc. Natl. Acad. Sci. USA* 2019, 116, 22020–22029. [CrossRef]
- 131. Ahmad, A.; Poltronieri, P.; Uddin, S. *lncRNAs in Cancer Metastasis and Therapy Resistance*; Frontiers Media SA: Lausanne, Switzerland, 2023; ISBN 9782832526125.
- 132. Ding, Y.; Lei, X.; Liao, B.; Wu, F.-X. Machine Learning Approaches for Predicting Biomolecule-Disease Associations. *Brief. Funct. Genom.* **2021**, *20*, 273–287.
- 133. Liu, Z.; Mi, M.; Li, X.; Zheng, X.; Wu, G.; Zhang, L. A lncRNA Prognostic Signature Associated with Immune Infiltration and Tumour Mutation Burden in Breast Cancer. J. Cell. Mol. Med. 2020, 24, 12444–12456. [CrossRef]
- 134. Zaheed, O.; Kiniry, S.J.; Baranov, P.V.; Dean, K. Exploring Evidence of Non-Coding RNA Translation with Trips-Viz and GWIPS-Viz Browsers. *Front. Cell Dev. Biol.* **2021**, *9*, 703374.
- 135. Leong, A.Z.-X.; Lee, P.Y.; Mohtar, M.A.; Syafruddin, S.E.; Pung, Y.-F.; Low, T.Y. Short Open Reading Frames (sORFs) and Microproteins: An Update on Their Identification and Validation Measures. *J. Biomed. Sci.* **2022**, *29*, 19.
- 136. Zheng, C.; Wei, Y.; Zhang, P.; Lin, K.; He, D.; Teng, H.; Manyam, G.; Zhang, Z.; Liu, W.; Lee, H.R.L.; et al. CRISPR-Cas9-Based Functional Interrogation of Unconventional Translatome Reveals Human Cancer Dependency on Cryptic Non-Canonical Open Reading Frames. *Nat. Struct. Mol. Biol.* 2023. [CrossRef]
- 137. Schlesinger, D.; Elsässer, S.J. Revisiting sORFs: Overcoming Challenges to Identify and Characterize Functional Microproteins. *FEBS J.* **2022**, *289*, 53–74.
- Patraquim, P.; Mumtaz, M.A.S.; Pueyo, J.I.; Aspden, J.L.; Couso, J.-P. Developmental Regulation of Canonical and Small ORF Translation from mRNAs. *Genome Biol.* 2020, 21, 128.
- Ingolia, N.T.; Ghaemmaghami, S.; Newman, J.R.S.; Weissman, J.S. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* 2009, 324, 218–223.
- 140. Patraquim, P.; Magny, E.G.; Pueyo, J.I.; Platero, A.I.; Couso, J.P. Translation and Natural Selection of Micropeptides from Long Non-Canonical RNAs. *Nat. Commun.* **2022**, *13*, 6515.
- 141. Li, M.; Liu, G.; Jin, X.; Guo, H.; Setrerrahmane, S.; Xu, X.; Li, T.; Lin, Y.; Xu, H. Micropeptide MIAC Inhibits the Tumor Progression by Interacting with AQP2 and Inhibiting EREG/EGFR Signaling in Renal Cell Carcinoma. *Mol. Cancer* **2022**, *21*, 181.
- 142. Zhang, M.; Zhao, K.; Xu, X.; Yang, Y.; Yan, S.; Wei, P.; Liu, H.; Xu, J.; Xiao, F.; Zhou, H.; et al. A Peptide Encoded by Circular Form of LINC-PINT Suppresses Oncogenic Transcriptional Elongation in Glioblastoma. *Nat. Commun.* **2018**, *9*, 4475. [PubMed]
- Jiang, L.; Yang, J.; Xu, Q.; Lv, K.; Cao, Y. Machine Learning for the Micropeptide Encoded by LINC02381 Regulates Ferroptosis through the Glucose Transporter SLC2A10 in Glioblastoma. *BMC Cancer* 2022, 22, 882. [CrossRef]
- 144. Polycarpou-Schwarz, M.; Groß, M.; Mestdagh, P.; Schott, J.; Grund, S.E.; Hildenbrand, C.; Rom, J.; Aulmann, S.; Sinn, H.-P.; Vandesompele, J.; et al. The Cancer-Associated Microprotein CASIMO1 Controls Cell Proliferation and Interacts with Squalene Epoxidase Modulating Lipid Droplet Formation. Oncogene 2018, 37, 4750–4768. [CrossRef]
- 145. Zhang, C.; Zhou, B.; Gu, F.; Liu, H.; Wu, H.; Yao, F.; Zheng, H.; Fu, H.; Chong, W.; Cai, S.; et al. Micropeptide PACMP Inhibition Elicits Synthetic Lethal Effects by Decreasing CtIP and poly(ADP-Ribosyl)ation. *Mol. Cell* **2022**, *82*, 1297–1312.e8. [PubMed]
- 146. Huang, J.-Z.; Chen, M.; Chen, D.; Gao, X.-C.; Zhu, S.; Huang, H.; Hu, M.; Zhu, H.; Yan, G.-R. A Peptide Encoded by a Putative IncRNA HOXB-AS3 Suppresses Colon Cancer Growth. *Mol. Cell* **2017**, *68*, 171–184.e6. [CrossRef] [PubMed]
- 147. Guo, B.; Wu, S.; Zhu, X.; Zhang, L.; Deng, J.; Li, F.; Wang, Y.; Zhang, S.; Wu, R.; Lu, J.; et al. Micropeptide CIP2A-BP Encoded by LINC00665 Inhibits Triple-Negative Breast Cancer Progression. *EMBO J.* **2020**, *39*, e102190. [CrossRef] [PubMed]
- 148. Tsai, M.-C.; Manor, O.; Wan, Y.; Mosammaparast, N.; Wang, J.K.; Lan, F.; Shi, Y.; Segal, E.; Chang, H.Y. Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science* **2010**, *329*, 689–693. [CrossRef]
- 149. Li, Q.; Shao, Y.; Zhang, X.; Zheng, T.; Miao, M.; Qin, L.; Wang, B.; Ye, G.; Xiao, B.; Guo, J. Plasma Long Noncoding RNA Protected by Exosomes as a Potential Stable Biomarker for Gastric Cancer. *Tumour Biol.* **2015**, *36*, 2007–2012. [CrossRef]
- 150. Badowski, C.; He, B.; Garmire, L.X. Blood-Derived lncRNAs as Biomarkers for Cancer Diagnosis: The Good, the Bad and the Beauty. *npj Precis. Oncol.* **2022**, *6*, 40.

- 151. Profumo, V.; Forte, B.; Percio, S.; Rotundo, F.; Doldi, V.; Ferrari, E.; Fenderico, N.; Dugo, M.; Romagnoli, D.; Benelli, M.; et al. LEADeR Role of miR-205 Host Gene as Long Noncoding RNA in Prostate Basal Cell Differentiation. *Nat. Commun.* 2019, 10, 307. [CrossRef]
- Garofoli, M.; Volpicella, M.; Guida, M.; Porcelli, L.; Azzariti, A. The Role of Non-Coding RNAs as Prognostic Factor, Predictor of Drug Response or Resistance and Pharmacological Targets, in the Cutaneous Squamous Cell Carcinoma. *Cancers* 2020, 12, 2552. [CrossRef] [PubMed]
- 153. Smallegan, M.J.; Rinn, J.L. Linking Long Noncoding RNA to Drug Resistance. *Proc. Natl. Acad. Sci. USA* 2019, 116, 21963–21965. [CrossRef] [PubMed]
- 154. Cui, H.; Kong, H.; Peng, F.; Wang, C.; Zhang, D.; Tian, J.; Zhang, L. Inferences of Individual Drug Response-Related Long Non-Coding RNAs Based on Integrating Multi-Omics Data in Breast Cancer. *Mol. Ther. Nucleic Acids* **2020**, *20*, 128–139. [CrossRef]
- 155. Liu, H.; Wang, S.; Zhou, S.; Meng, Q.; Ma, X.; Song, X.; Wang, L.; Jiang, W. Drug Resistance-Related Competing Interactions of IncRNA and mRNA across 19 Cancer Types. *Mol. Ther. Nucleic Acids* 2019, 16, 442–451. [CrossRef] [PubMed]
- 156. Winkle, M.; El-Daly, S.M.; Fabbri, M.; Calin, G.A. Noncoding RNA Therapeutics—Challenges and Potential Solutions. *Nat. Rev. Drug Discov.* **2021**, *20*, 629–651.
- 157. Arun, G.; Diermeier, S.D.; Spector, D.L. Therapeutic Targeting of Long Non-Coding RNAs in Cancer. *Trends Mol. Med.* 2018, 24, 257–277. [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.