



Article

# Machine Learning Gene Signature to Metastatic ccRCC Based on ceRNA Network

Epitácio Farias <sup>1</sup>, Patrick Terrematte <sup>2,\*</sup> and Beatriz Stransky <sup>1,3</sup>

<sup>1</sup> Bioinformatics Multidisciplinary Environment (BioME), Federal University of Rio Grande do Norte (UFRN), Natal 59078-400, Brazil; epitacio.farias.101@ufrn.edu.br (E.F.); beatriz.stransky@ufrn.br (B.S.)

<sup>2</sup> Metropolis Digital Institute (IMD), Federal University of Rio Grande do Norte (UFRN), Natal 59078-400, Brazil

<sup>3</sup> Biomedical Engineering Department, Center of Technology, Federal University of Rio Grande do Norte (UFRN), Natal 59078-970, Brazil

\* Correspondence: patrick.terrematte@ufrn.br

**Abstract:** Clear-cell renal-cell carcinoma (ccRCC) is a silent-development pathology with a high rate of metastasis in patients. The activity of coding genes in metastatic progression is well known. New studies evaluate the association with non-coding genes, such as competitive endogenous RNA (ceRNA). This study aims to build a ceRNA network and a gene signature for ccRCC associated with metastatic development and analyze their biological functions. Using data from The Cancer Genome Atlas (TCGA), we constructed the ceRNA network with differentially expressed genes, assembled nine preliminary gene signatures from eight feature selection techniques, and evaluated the classification metrics to choose a final signature. After that, we performed a genomic analysis, a risk analysis, and a functional annotation analysis. We present an 11-gene signature: *SNHG15*, *AF117829.1*, *hsa-miR-130a-3p*, *hsa-mir-381-3p*, *BTBD11*, *INSR*, *HECW2*, *RFLNB*, *PTTG1*, *HMMR*, and *RASD1*. It was possible to assess the generalization of the signature using an external dataset from the International Cancer Genome Consortium (ICGC-RECA), which showed an Area Under the Curve of 81.5%. The genomic analysis identified the signature participants on chromosomes with highly mutated regions. The *hsa-miR-130a-3p*, *AF117829.1*, *hsa-miR-381-3p*, and *PTTG1* were significantly related to the patient's survival and metastatic development. Additionally, functional annotation resulted in relevant pathways for tumor development and cell cycle control, such as RNA polymerase II transcription regulation and cell control. The gene signature analysis within the ceRNA network, with literature evidence, suggests that the lncRNAs act as "sponges" upon the microRNAs (miRNAs). Therefore, this gene signature presents coding and non-coding genes and could act as potential biomarkers for a better understanding of ccRCC.

**Keywords:** renal carcinoma; ceRNA network; transcriptional signature; machine learning; metastasis



**Citation:** Farias, E.; Terrematte, P.; Stransky, B. Machine Learning Gene Signature to Metastatic ccRCC Based on ceRNA Network. *Int. J. Mol. Sci.* **2024**, *25*, 4214. <https://doi.org/10.3390/ijms25084214>

Academic Editors: Ramin Ranjbarzadeh and Andrzej Kloczkowski

Received: 13 October 2023

Revised: 5 January 2024

Accepted: 19 January 2024

Published: 11 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Renal cancer is a group of neoplasms originating in the renal tissues and classified by cell type or histologic characteristics, such as clear-cell renal-cell carcinoma (ccRCC), papillary renal carcinoma (pRCC), and chromophobe renal carcinoma (chRCC) [1–3]. Due to the silent characteristic of this disease [4], the diagnosis at the metastatic state occurs in approximately 30% of ccRCC patients [5,6].

In a study of 537 ccRCC patients, The Cancer Genome Atlas (TCGA) consortium [7] characterized significant alterations in the ccRCC cohort. The changes include mutations in genes such as *VHL*, *PBRM1*, *SETD2*, and *BAP1*; the deletion of the q arm of chromosome 3; and distinct arrangements involving messenger RNA (mRNA) and microRNA (miRNA). These alterations signify crucial mechanisms in ccRCC. More recently, other studies revealed important roles for non-coding RNAs (ncRNAs), a class of RNAs that comprise approximately 80% of the transcriptome [8–10].

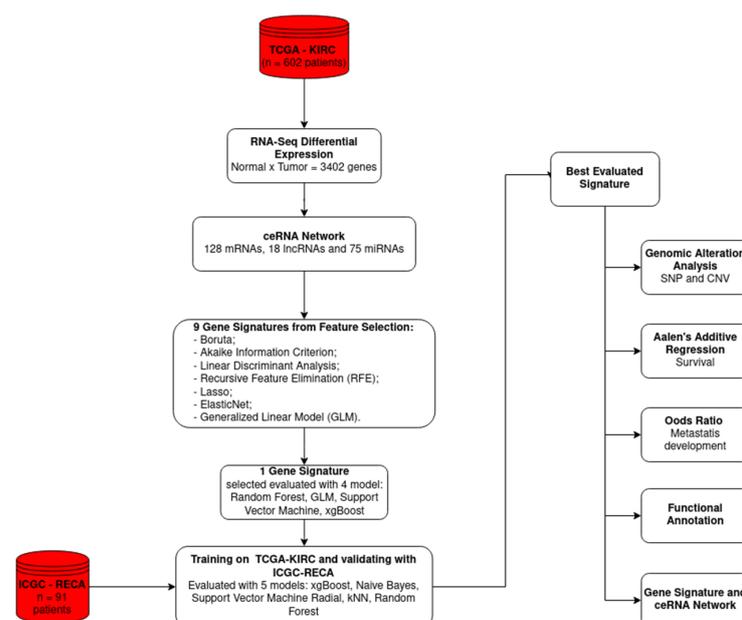
The functions of lncRNAs are determined by their interactions with DNA, proteins, or other RNAs and their cellular localization [9–13]. The lncRNAs can act as a (i) decoy or “sponge” modulating the effector of their targets, (ii) guide the enzyme modifiers of histones or chromatin, and (iii) respond to various stimuli [14,15]. In particular, the ligation of the lncRNA with the miRNA affects their targets, characterizing endogenous competition between the lncRNA and the mRNA target of the miRNA [9,10].

The proposed “Competing Endogenous RNA” (ceRNA) hypothesis was based on the idea of communication between miRNAs, mediated by the miRNA recognition elements (MREs), with mRNA, lncRNA, and other ncRNAs [16]. Alteration in the ceRNA networks is observed in cancer and other pathologies, associating them with biomarkers for metastasis and other clinical outcomes or therapeutic targets [10,17–20].

Research involving RNA expression generates extensive and intricate datasets. Integrating this data with clinical information through machine learning techniques could facilitate the extraction of patterns in gene expression, enriching our comprehension of gene functions within the biological context [21,22]. Among the vast applications of ML, methods of classification and prediction are commonly applied in health research [23,24]. However, the lack of feature selection associated with the outcome variable could influence the performances of the algorithms [25]. Feature selection involves the analysis of variables based on their impact on the outcome, eliminating irrelevant ones, and enhancing their consistency and relevance for the model [26].

Recently, studies investigated the ceRNA network and gene signature association in ccRCC. Most of them focus on the relationship between ceRNA, immune response regulation, and prognosis [27–36], and there is a lack of information about gene signatures involving the ceRNA network and metastasis in ccRCC.

In this study, we constructed a ceRNA network and generated a gene signature based on feature selection algorithms to classify the metastatic profiles of ccRCC patients. We achieved an Area Under the Curve (AUC) of 81.5% and an accuracy of 72% in the classification task. The signature was validated using an independent dataset, and the biological functions of its components were investigated in the ceRNA network. The flowchart shown in Figure 1 displays a summarized view of the discovery process for the novel Recursive Feature Elimination (RFE) gene signature of ccRCC.

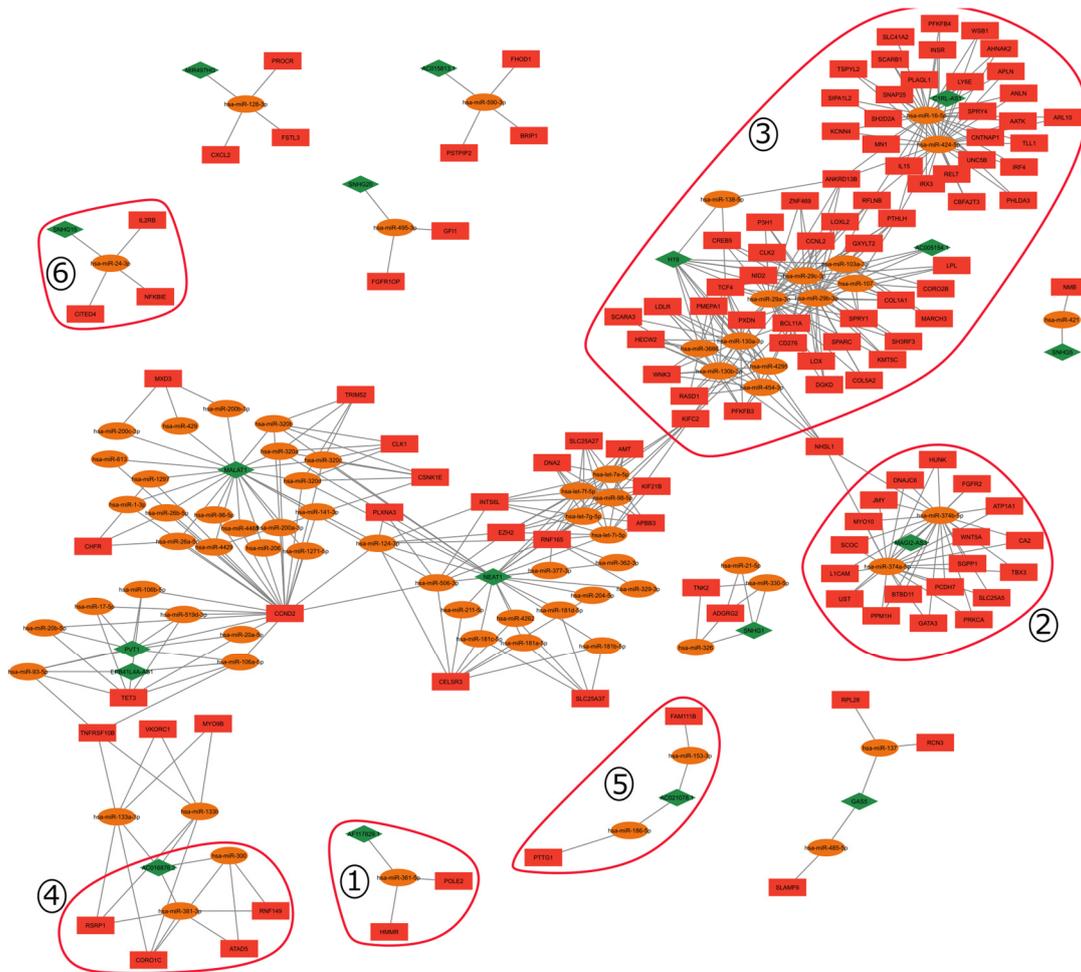


**Figure 1.** Flowchart of the current study to obtain a gene signature based on the Recursive Feature Elimination (RFE) approach. The datasets are indicated by the cylindrical shape; the white rectangles represent the steps of the study. TCGA-KIRC and ICGC-RECA are the ccRCC datasets.

## 2. Results

### 2.1. ceRNA Network

To construct the ceRNA network, we used the differentially expressed (DE) genes of the TCGA-KIRC (n = 602) project. This analysis resulted in 2842 mRNAs and 271 lncRNAs DE based on the thresholds of  $|\log_2FC| > 2$  and  $p$ -value adjusted for the False Discovery Rate (FDR)  $< 0.01$  (Figure S1). With these DE genes, we constructed the ceRNA network composed of 18 lncRNAs, 128 mRNAs, and 75 miRNAs (Figure 2). The miRNAs were included in the ceRNA network as described (Section 4.2).



**Figure 2.** The ceRNA network was constructed based on the differentially expressed (DE) genes from ccRCC patients. The network is composed of 18 lncRNAs (green diamond), 75 miRNAs (orange ellipses), and 128 mRNAs (red rectangles). Individual clusters and clusters composed of gene signatures with their first neighbors are enumerated from 1 to 6, and highlighted by the red line circle.

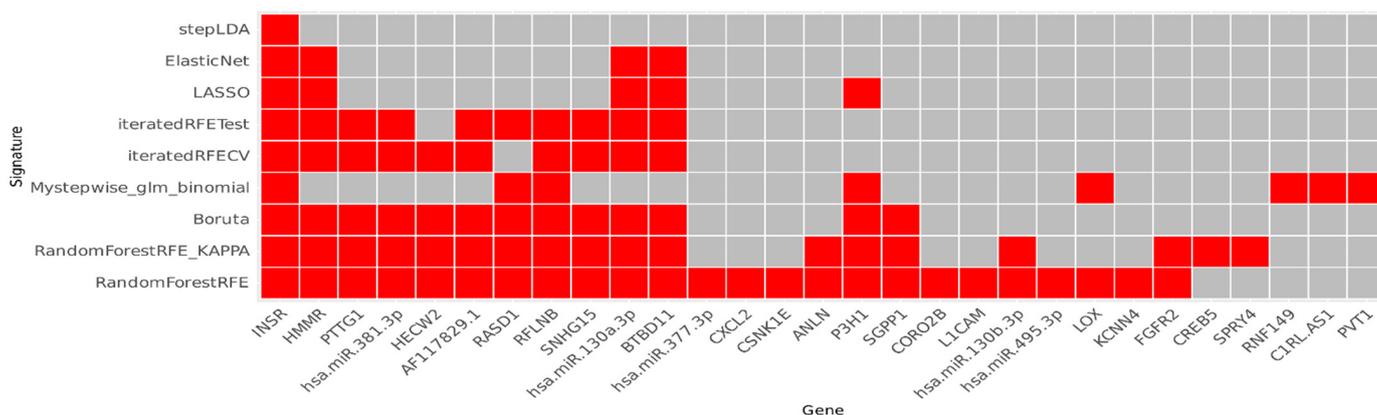
The network structure explicitly reveals the connections between miRNAs, lncRNAs, and mRNAs. Within this inferred network, we have observed the presence of a diverse group of genes that share miRNAs and some lncRNAs. This characteristic points to a clustered organization, where these genes are related to a common outcome. Upon closer analysis of this cluster organization, we noticed a pattern in which one cluster is fully connected while others are sparser.

In order to evaluate the topology of our ceRNA network, we tested the fitness of the degree distribution to a power-law model,  $P(k) = \alpha k^\alpha$ , resulting in  $\alpha = 2.163$ . We performed a Kolmogorov–Smirnov test with our ceRNA network, and the distribution of our data did not fit strictly to a power-law model. Nevertheless, using the likelihood

ratio test, our network fitted between a power-law and a log-normal positive distribution (Figure S2).

### 2.2. Feature Selection

With the expression data from the 221 ceRNA network genes and the metastatic classification from the 192 patients, we conducted balanced performance assessments. Subsequently, we executed a training process for feature selection and developed nine initial gene signatures (Figure 3). Among the feature selection techniques, only stepAIC did not converge. The Recursive Feature Elimination (RFE) shows an accuracy of 76.30% and a Kappa coefficient of 0.5663, showing a moderate level of agreement between the actual metastatic samples and the predicted ones.



**Figure 3.** Heat plot with the 29 unique genes reported based on the nine preliminary gene signatures constructed. On the Y-axis are the models applied to the signature construction, and on the X-axis are the genes (red squares) from each obtained signature.

We dismissed the outcomes from the stepAIC method and performed the first benchmark. The xgbTree presented the best result, with an accuracy of 80% during the training, 60% for the test, and 68.3% for validation. Employing the Youden statistics, we selected the top four signatures. The four signatures shared some genes, and we constructed the final signature through majority voting, composed of *INSR*, *PTTG1*, *BTBD11*, *RASD1*, *HECW2*, *HMMR*, *RFLNB*, *hsa-miR-130a-3p*, *hsa-miR-381-3p*, *SNHG15*, and *AF117829.1*. To evaluate gene importance in random forest, we present a multi-way importance plot (Figure S3).

We conducted the second benchmark to the constructed signature (Table 1), using the ICGC-RECA project as a test dataset. We observed an accuracy of 72%, an AUC (Figure S4) of 81.5%, and a Brier Score of 0.1955.

**Table 1.** Metrics evaluated for validation using an external dataset.

Method	Accuracy	AUC	Brier Score
<b>Random forest</b> <sup>1</sup>	<b>72.2%</b>	<b>81.48%</b>	<b>0.1955442</b>
SVM	50%	66.67%	0.2500714
xgBoost	61.1%	62.34%	0.2343498
kNN	50%	61.72%	0.4817816
Naïve Bayes	50%	54.32%	0.5000000

<sup>1</sup> Bold represents the best classification.

To highlight the separability of the data through the 11-signature genes, we applied k-means clustering to partition TCGA-KIRC samples into two groups (C1 and C2). Subsequently, for dimensionality reduction and visualization, we implemented principal component analysis (PCA). Predominantly, the metastatic samples (M1) are located within the positive range of the first dimension (C1), whereas the non-metastatic samples are

positioned on the opposite side (C2) (Figure S5a). The chi-squared test revealed a significant association ( $p$ -value = 0.007) between the metastatic samples and the cluster 1. The analysis of gene contribution with Cos2 to the sample characteristics shows that positively correlated genes are located in the first quadrant within the majority of metastatic samples (M1) and group C1 samples (Figure S5b).

### 2.3. Integrative Analysis of the Transcriptional Signature Components

#### 2.3.1. Genomic Alteration Analysis

Performing a genome-level alteration analysis enables us to evaluate their impact on the gene product. The alterations can include changes in the genetic structure, disruptions in protein synthesis, or variations in the quantity of the gene product. We used the Maftools package to investigate single-nucleotide polymorphisms (SNPs) and copy number variations (CNVs) in the TCGA-KIRC cohort.

The missense mutation is the most frequent alteration in SNP data, with approximately 44 variants per sample and a prevalence of cytosine and thymine transversions. Moreover, ten samples showed mutations in signature-coding genes (Figure S6). Specifically, we detected missense mutations in the genes HECW2, BTBD11, INSR, and PTTG1; a frameshift deletion in BTBD11; and a multi-hit mutation in HECW2. However, the genes HMMR, RASD1, and RFLNB did not present any mutations.

The analysis of copy number variation reveals substantial and frequent alterations in chromosomes 1, 4, 5, 6, 7, 12, 17, 18, and 20 across the samples. Upon investigating the chromosome locations of our gene signature in the National Center of Biotechnology Information, we found that while our genes were situated in the chromosomes undergoing significant alterations, they were not specifically located in the regions exhibiting notable modifications (Table S2).

#### 2.3.2. Risk Analysis

To evaluate the relationship between the expression level of signature genes, the metastatic development, and the survival status of the patients, we performed a risk analysis. Aalen's additive regression shows a significant relationship between some genes from the gene signature and patient survival, such as (i) *AF1117829.1* ( $p$ -value = 0.0001627), (ii) *hsa-miR-130a-3p* ( $p$ -value = 0.016), (iii) *hsa.miR.381.3p* ( $p$ -value = 0.027), and (iv) *PTTG1* ( $p$ -value = 0.020), see Figure S7.

The odds ratio analysis shows that the miRNA *hsa-miR-130a-3p* and the lncRNA *AF117829.1* are the only ones that had significant associations, with  $p$ -value = 0.011 and  $p$ -value = 0.029, respectively (Figure 4).

#### 2.3.3. Functional Annotation Analysis

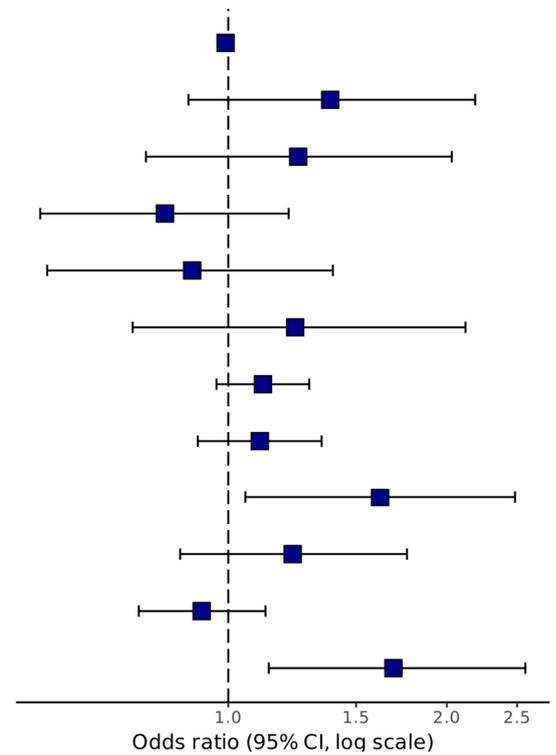
We performed a functional analysis using the signature-coding genes and the targets of the signature miRNAs against the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

When evaluating the targets from the miRNAs and their biological pathways, well-known oncology-related pathways, such as the PI3K-AKT signaling pathway, the p53 signaling pathway, the transforming growth factor-beta (TGF-beta) signaling pathway, renal cancer, and the HIF-alfa pathway, were also observed in a statistically significant manner (Figure S8a).

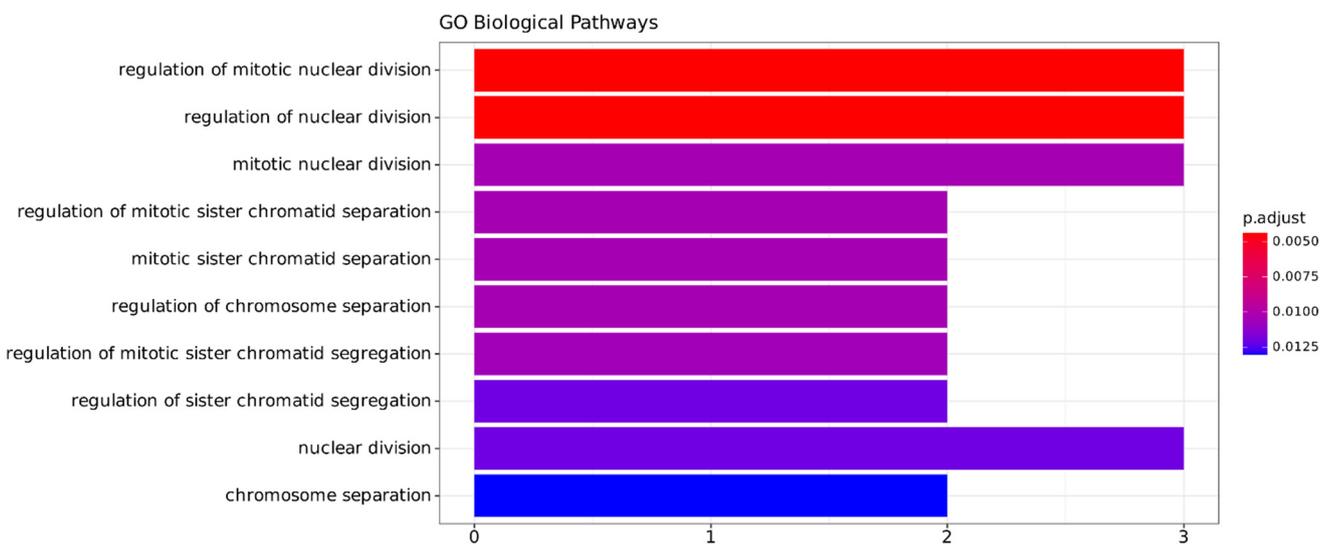
The annotated biological processes were associated with cellular division regulation, such as chromatid sister separation, and chromosome segregation (Figure 5). The pathways annotated for miRNA targets were related to the cellular division process. Also, other pathways related to signal transduction, growth factors, and DNA polymerase I regulation were significantly enriched ( $p$ -value < 0.05) (Figure S8b).

pathologic\_M: OR (95% CI, *p*-value)

age	[26.0,90.0]	0.99 (0.97–1.01, <i>p</i> =0.456)
HMMR	[-1.9,5.1]	1.38 (0.88–2.19, <i>p</i> =0.162)
PTTG1	[-1.1,6.1]	1.25 (0.77–2.03, <i>p</i> =0.369)
HECW2	[-0.9,7.6]	0.82 (0.55–1.21, <i>p</i> =0.316)
RFLNB	[0.9,9.8]	0.89 (0.56–1.39, <i>p</i> =0.620)
INSR	[3.5,9.7]	1.24 (0.74–2.12, <i>p</i> =0.428)
BTBD11	[-6.7,6.1]	1.12 (0.96–1.29, <i>p</i> =0.141)
RASD1	[-1.6,8.2]	1.11 (0.91–1.34, <i>p</i> =0.316)
AF117829.1	[-0.9,5.3]	1.62 (1.06–2.48, <i>p</i> =0.027)
SNHG15	[0.5,6.3]	1.23 (0.86–1.76, <i>p</i> =0.264)
hsa.miR.381.3p	[-3.1,8.5]	0.92 (0.75–1.13, <i>p</i> =0.409)
hsa.miR.130a.3p	[3.1,8.3]	1.69 (1.14–2.57, <i>p</i> =0.011)



**Figure 4.** The odds ratio of each gene in the signature regarding metastatic development and the 95% confidence interval. The miRNA hsa-miR-130a-3p and the lncRNA AF117829.1 were the only ones that were significantly associated (*p*-value < 0.05).



**Figure 5.** Functional annotation from the gene ontology, focusing on the biological process related to the seven coding genes from the signature (*HMMR*, *RASD1*, *RFLNB*, *PTTG1*, *INSR*, *HECW2*, *BTBD11*). On the Y-axis, the pathways annotated are listed, and the X-axis represents the gene count of the signature in each pathway.

#### 2.4. Gene Signature and ceRNA Network

As the signature construction was performed upon the genes from the ceRNA network, the evaluation of their location and their first neighbors could improve knowledge about their functions and the possible metastatic effects in the ccRCC.

The ceRNA network presents a cluster organization, and the signature genes are located in clusters with distinct properties. Two of the signature genes are located in cluster 1, while others resided in locations with dense interconnections, such as cluster 2. Additionally, certain areas contained only one gene, as observed in clusters 3, 4, 5, and 6. Table 2 presents the genes from the signature, their first ligands within the ceRNA network, and their cluster localization.

**Table 2.** Gene signature participants, their first ligands within the ceRNA network, and their cluster localization. Bold represents the gene from the signature's first ligands.

Cluster	Gene	First Ligands
1	<i>AF117829.1</i>	<i>hsa-miR-361-5p</i> , <i>POLE2</i> , <b><i>HMMR</i></b>
2	<i>BTBD11</i>	<i>hsa-miR-374a-5p</i> , <i>hsa-miR-374b-5p</i> , <i>MAGI2-AS3</i>
3	<i>HECW2</i>	<b><i>hsa-miR-130a-3p</i></b> , <i>hsa-miR-130b-3p</i> , <i>hsa-miR-454-3p</i> , <i>hsa-miR-4295</i> , <i>hsa-miR-3666</i> , <i>H19</i>
1	<i>HMMR</i>	<i>hsa-miR-361-5p</i> , <i>POLE2</i> , <b><i>AF117829.1</i></b>
3	<i>hsa-miR-130a-3p</i>	<b><i>HECW2</i></b> , <i>WNK3</i> , <i>RASD1</i> , <i>PFKFB3</i> , <i>SCARA3</i> , <i>LDLR</i> , <i>PMEPA1</i> , <i>TCF4</i> , <i>PXDB</i> , <i>BCL11A</i> , <i>NHSL1</i> , <i>H19</i>
4	<i>hsa-miR-381-3p</i>	<i>RSRP1</i> , <i>CORO1C</i> , <i>ATAD5</i> , <i>RNF149</i> , <i>AC016876.2</i>
3	<i>INSR</i>	<i>hsa-miR-16-5p</i> , <i>hsa-miR-424-5p</i> , <i>C1RL-AS1</i> .
5	<i>PTTG1</i>	<i>hsa-miR-186-5p</i> , <i>AC021078.1</i>
3	<i>RFLNB</i>	<i>hsa-miR-29a-3p</i> , <i>hsa-miR-29b-3p</i> , <i>hsa-miR-29c-3p</i> , <i>hsa-miR-16-5p</i> , <i>hsa-miR-424-5p</i> , <i>H19</i> , <i>AC005154.1</i>
3	<i>RASD1</i>	<b><i>hsa-miR-130a-3p</i></b> , <i>hsa-miR-130b-3p</i> , <i>hsa-miR-3666</i> , <i>hsa-miR-4295</i> , <i>hsa-miR-454-3p</i>
6	<i>SNHG15</i>	<i>hsa-miR-24-3p</i> , <i>IL2RB</i> , <i>NFKBIE</i> , <i>CITED4</i>

### 3. Discussion

In the present study, a transcriptional signature associated with metastatic development was constructed using feature selection techniques in conjunction with ceRNA network data for the cases of patients diagnosed with clear-cell renal-cell carcinoma. Additionally, the biological behavior of the genes comprising the signature was evaluated to understand their actions within the tumoral environment in ccRCC.

Regarding the network topology, our ceRNA presents a characteristic topology that does not follow the power-law degree distribution or represent a scale-free network. As indicated by Broido-Clauset [37] and Clauset, Shalizi, and Newman [38], the existence of scale-free topologies in the real world is rare, and most of them follow a log-normal distribution, like ours, or an exponential distribution once they have a heavy-tailed pattern.

#### 3.1. Gene Signature

A set of nine feature selection methods produced preliminary gene signatures for the metastatic classification of ccRCC. The learning curves derived from RFE present a Kappa coefficient falling within the range of 0.41 to 0.6, which signifies a substantial agreement between the method's outcomes and the data [39]. The combined application of these two metrics enhances classification accuracy [40].

Some benchmark models displayed overfitting, and we used the Youden statistics to select the models with the best sensitivity and specificity performances. Among the top four signatures, their Youden coefficients ranged from 0.13 to 0.18, approaching values closer to 1, signifying the optimal classification results [41].

The use of majority voting with the top four signatures results in the final signature of our work, composed of seven mRNAs: *PTTG1*, *BTBD11*, *HECW2*, *INSR*, *RFLNB*, *HMMR*, *RASD1*; two lncRNAs, *SNHG15* and *AF117829.1*; and two miRNAs, *hsa-miR-381-3p* and *hsa-miR-130a-3p*.

Validation with an external dataset is a process in the ML field used to evaluate model generalization [42]. We performed external validation to classify metastatic tissue using the gene expression of our signature. The training was performed with TCGA-KIRC and

testing with ICGC-RECA, resulting in an accuracy of 72% and an AUC of 81.5%. Since the classification relied on health data, these evaluation metrics might not perfectly align with the objective of the study [43]. This limitation can be partially explained by the heterogeneity of the data, the sample size, and the inherent complexity of the biological process underlying the ccRCC. As far as we know, this is the first study that analyzes the relationship between a ceRNA network and a metastatic gene signature in ccRCC.

### 3.2. Validation and Biological Interpretation

#### 3.2.1. Genomic and Functional Alterations

The somatic alterations of the coding genes in the signature were more commonly associated with missense or frame\_shift\_del, except for HMMR and RFLNB. Regarding the copy number variations, the amplified or deleted regions were not in the exact location of the genes in the signature.

Analyzing the risk associated with survival or metastasis development showed a significant association between four genes in the gene signature. The lncRNA AF117829.1 and the miRNA hsa-miR-130a-3p were present in both analyses. The miRNA association is related to various cancers, such as bladder, breast, hepatocellular, glioma, and osteosarcoma [44–49]. Therefore, the presence of the PTTG1 and hsa.miR.130a.3p genes corroborate the literature, where in a situation of high expression, the prognosis is poor, and the hsa.miR.13p and hsa.miR.381.3p are associated with metastatic development. However, the lncRNA remains unknown, and these features could be added to its actions, which are still under study.

The functional annotation revealed diversified pathways. Both approaches, using the coding genes and miRNAs, highlighted biological processes associated with cell cycle regulation, controlling the separation and segregation of the sister chromatids, RNA polymerase II transcription, and the up-regulation and accommodation of the transcription activity of coding and non-coding genes [50], as well as processes related to cell–cell communication.

#### 3.2.2. Gene Cluster Analysis

The ceRNA network presents a cluster organization, showing dense regions with highly connected gene networks, others with more sparse networks, and some isolated small clusters. We perform a cluster-by-cluster analysis, using the signature genes and their first ligands (Table 2) to evaluate a possible role in metastasis development.

The first cluster is composed of two genes from the signature: the mRNA HMMR and the lncRNA AF117829.1, as well as the miRNA hsa.mir.361.5p and the mRNA POLE2. The Hyaluronan Mediated Motility Receptor (HMMR) is responsible for the regulation of tumor cell motility [51], and its knockdown reduced peritoneal metastasis in gastric cancer [52]. The role of lncRNA AF117829.1 remains unknown, but it was described as related to the proliferation, differentiation, and regulation of T-cell immunity [53,54], and its expression is implied with metastatic development and the worst prognosis of ccRCC patients. In this context, the lncRNA AF117829.1 could be acting as a sponge over the miRNA, impairing the degradation of POLE2 and HMMR, and promoting cell differentiation and metastasis development.

Cluster 2 presents the BTB Domain Containing 11 (BTBD11) gene from the signature. Its mechanism remains unknown, but it is described as a target in the TGF-beta pathway, responsible for cell cycle and apoptosis regulation [55]. The BTBD11 first ligands are the lncRNA MAGI2-AS3 and the miRNAs hsa.miR.374a.5p and hsa.miR.374b.5p isoforms. The lncRNA–miRNA interactions are related to tumor suppression in breast and hepatocellular cancers [56,57], suppressing proliferation, migration, and invasion. With the down-expression of the lncRNA, the miRNA can degrade BTBD11, negatively regulating the TGF-beta pathway and promoting tumor development.

The third cluster is located in the most dense region of the ceRNA network and presents five of the signature protein genes, as well as three lncRNAs and eight miRNAs directly linked to the former. The insulin receptor (INSR) regulates the insulin signaling pathway

and activates the oncogenic PI3K/Akt/mTOR pathway, and its high expression is inversely associated with patient survival in ccRCC and gastric cancer [58]. Refilin b (RFLNB) is responsible for the epithelial–mesenchymal transition (EMT) and inhibits tumoral growth in neuroblastoma and pleural malignant mesothelioma [59–61]. HECT-Type E3 Ubiquitin Transferase HECW2 (HECW2) acts in apoptosis regulation, and its high expression is related to a good prognosis in ccRCC [62,63]. Ras-Related Dexamethasone-Induced 1 (RASD1) inhibits the RAS superfamily of the short GTPases and, in high expression, induces a decrease in cell growth, leading to apoptosis [64]. The increased expression of H19, C1RL-AS1, and AC005154.1 lncRNAs presented in the cluster suggests their roles as miRNA sponges (Table 2). The impairment or attenuation of miRNAs could potentially stabilize RFLNB and INSR expression, promoting tumor growth. Furthermore, the decreased expression of RASD1 implies that the miRNAs remain stable, also favoring a tumorigenic environment.

Cluster 4 is composed of the miRNA hsa.miR.381.3p and its targets. Coronin 1C (CORO1C) regulates apoptosis and cell cycle progression [65], acting as an oncogene in ccRCC and non-small-cell lung cancers [66,67]. The ATPase Family AAA Domain Containing 5 (ATAD5) is responsible for DNA duplication [68] and cell cycle regulation in neuroendocrine hepatic tumors [69]. The Arginine and Serine Rich Protein 1 (RSRP1) is involved in spliceosome assembly and has a good prognosis in breast cancer, but its biological mechanism is still unknown [70]. Ring Finger Protein 149 (RNF149) regulates ubiquitination and proteasomal degradation and is associated with pancreatic cancer [71,72]. The high expression of lncRNA AC016876.2 can promote the capture of the hsa.miR.381.3p, hence stabilizing the miRNA targets and facilitating tumor development.

Cluster 5 presents the signature Pituitary Tumor-Transforming Gene 1 (PTTG1), an oncogene that regulates sister chromatid separation [73]. The interaction with the miRNA hsa.miR.186.5 regulates the TGF-beta and MAPK pathways in breast cancer and ccRCC [74]. These pathways are associated with essential processes, such as tissue development, proliferation, senescence, migration, apoptosis, and cell differentiation [75,76]. The lncRNA AC021078.1 is involved in cell differentiation and DNA repair [77], and its high expression can negatively regulate the miRNAs, giving the PTTG1 the possibility to act in tumoral and metastatic progression.

The signature lncRNA Small Nucleolar RNA Host Gene 15 (SNHG15) is located on the sixth cluster and regulates the NF-kappa-B pathway that represses cell proliferation and the epithelial–mesenchymal transition (EMT) in ccRCC [78]. In cases of high expression, SNHG15 correlates to metastatic progression in colorectal and non-small-cell lung cancers [79,80]. This cluster is also composed of the protein coding genes' NFKB Inhibitor Epsilon (NFKBIE) inhibitor of the NF-kappa-B signaling pathway associated with the inflammatory process in cancer [81,82], the Interleukin 2 Receptor Subunit Beta (IL2RB), and Cbp/P300 Interacting Transactivator With Glu/Asp Rich Carboxy-Terminal Domain 4 (CITED4), which are the regulators of the T-cell immune response and gene transcription, respectively [83,84], both, in cases of high expression, present a poor prognosis and are related to metastasis development. As the SNHG15 presented an elevated expression, it could indicate a sponge effect upon the miRNA on the cluster, promoting the normal activity of miRNAs-target.

Thus, the behavior observed in some genes from signature corroborates with the literature. SNHG15, hsa.miR.130a.3p, PTTG1, INSR and HMMR were described in a ccRCC environment, exhibiting the higher expression that induces metastasis and promotes cancer development. Conversely, lower expression of miRNA hsa.miR.381.3p is associated with a poor prognosis and linked to the development of metastasis. However, the remaining genes in the signature are reported in the literature across several other solid tumors, and play a crucial role in cancer and metastasis development.

## 4. Materials and Methods

### 4.1. Data

The RNA-seq and clinical datasets from the TCGA-KIRC project ( $n = 602$ ) were downloaded from Genomic Data Commons (<https://portal.gdc.cancer.gov/>, accessed on 1 May 2023) [85] and UCSC Xenabrowser (<https://xena.ucsc.edu/>, accessed on 1 May 2023) [86] (University of California, Santa Cruz, CA, USA). For external validation, we used the dataset of ccRCC ( $n = 91$  patients) from the International Cancer Genome Consortium (ICGC-RECA, accessed on 1 June 2023) [87] (Ontario Institute of Cancer Research in Toronto, Canada).

### 4.2. ceRNA Network Construction

The ceRNA network was constructed from the differentially expressed genes' mRNAs and lncRNAs. We use the DESeq2 (v1.36.0) [88] package for the differential expression analysis between the normal ( $n = 72$ ) and tumor tissues ( $n = 530$ ) from the TCGA-KIRC cohort, with an absolute  $|\log\text{-fold change (LFC)}| > 2$  and an adjusted  $p$ -value (FDR)  $< 0.01$ .

With the differentially expressed genes, the ceRNA was constructed using the GDCR-NATools package (v. 1.16.6) [89] associated with the starBase [90]. This database provides the interaction networks through numerous RBPs and RNAs, to supply the miRNAs shared by the differentially expressed lncRNAs and mRNAs from our KIRC dataset. The pair were selected using the following statistical analyses: (i) the hypergeometric test, where the probability of miRNAs shared by the lncRNA-mRNA pair was evaluated, observing the success of finding an association between the lncRNA-mRNA pair with the same miRNA; (ii) the Pearson correlation, used to measure the expression correlation between the lncRNA and the mRNA to understand the relation between them; and (iii) regulatory similarity, which will count the Pearson correlation and the total of miRNAs shared by the lncRNA-mRNA pair.

This analysis used a  $p$ -value threshold of 0.01 for the Pearson correlation and hypergeometric test and a value different from 0 for the regulatory similarity. The ceRNA network visualization was implemented using Cytoscape software (v 3.10.1) [91].

Once the network was constructed, its topology was evaluated following the Barabási-Oltvai [92] concepts of network biology, associating it with the likelihood ratio test based on the method of Broido-Clauset [37], using the package powerlaw [93] in Python (v 3.12).

### 4.3. Dataset Construction, Feature Selection, and Gene Signature Construction

The signature construction used the ceRNA network genes following the methodology of Terrematte and colleagues [94]. The gene signatures were produced using the feature selection techniques in Table S1 and the OmicSelector package (v1.0.0) [95].

Within the gene expression dataset from the TCGA-KIRC ( $n = 602$ ), a missing metastatic classification in 30 patients was observed, causing their removal, and due to the unbalanced characteristic of the metastasis classification of presence (M1) or absence (M0), a propensity matching score balance was performed, maintaining 190 patients, with 95 from each class.

This new dataset was split randomly into three new datasets, following the ratio of 60% for training ( $n = 114$ ), 20% for testing ( $n = 38$ ), and 20% for validation ( $n = 38$ ). For the signature construction process, we used the following feature selection techniques: Recursive Feature Elimination (RFE) and two iterated versions, Boruta, the Generalized Linear Model (GLM), the Akaike Information Criterion (AIC), Linear Discriminant Analysis (LDA), Lasso, and ElasticNet.

To improve the construction of the signature and optimize computational efficiency, we performed hyperparameter adjustments to the feature selection. The RFE techniques used cross-validation with ten folds, using a window frame of 50 genes in each iteration, and the iterated RFE versions used a window frame of ten genes for the signature.

With the nine signatures constructed, a first benchmarking stage was performed to select the signature with the best metrics for metastatic classification using the test and validation datasets. The first benchmark compared the signatures using the following mod-

els: random forest (rf), the Generalized Linear Model (GLM), eXtreme Gradient Boosting (xgbTree), and the Support Vector Machine with a Radial Kernel (svmRadial), performed ten times to seek the best parameter adjustment for each of them. The metrics used to evaluate this benchmark were accuracy, specificity, sensitivity, and the Youden statistics.

To evaluate the signature generalization, the external dataset from the ICGC-RECA project (n = 91) was used with the mlr3verse package (v0.2.7) [96] to perform the second benchmark, applying the following classification techniques: random forest, naive Bayes, kNN, svmradial, and XGBoost. The evaluation metrics were accuracy, balanced accuracy, the Brier score, and the AUC. The validation process used the TCGA-KIRC for training and the ICGC-RECA for testing.

#### 4.4. Somatic and Copy Number Alteration Analysis

The somatic alterations analysis was conducted using the Mutation Annotation Format (MAF) datafile, using the Maftools package (v2.16.0) [97], extracting information about (a) the types and classification of variations, (b) the variation quantity by sample, and (c) the top 10 genes altered.

The copy number variation analysis requires the construction of the GISTIC file. The Genomic Identification of Significant Targets in Cancer (GISTIC) pipeline [98] resulted in information about amplification and deletions within the data, analyzed via the Maftools package. To perform GISTIC analysis, it was necessary to obtain the segmentation file obtained from the GDC Data Portal [85] and the reference genome, version 41, from the GENCODE [99].

#### 4.5. Risk Analysis

To evaluate the relationship between the expression level of the signature genes, the metastatic development, and the survival statuses of the patients, we performed a risk analysis. With the survival (v3.5.7) [100] and finalfit (v1.0.6) [101] packages, we executed Aalen's additive regression and an odds ratio analysis, respectively. Aalen's regression acts as a complementary, or alternative, form of the Cox model. In this method (Aalen Regression), the covariables associations and their effects are determined, taking into account the gene set association to the death event [102]. The odds ratio quantifies the strength of association between each of the analyzed covariables separately and the outcome (metastasis) [103].

#### 4.6. Functional Annotation Analysis

The identification of the pathways enriched by the genes of the signature was performed against the gene ontology [104], focusing on the biological processes and molecular functions, using the clusterProfiler package (v4.8.2) [105] and the mirPath platform (v3.0) [106] for the functional characterization of miRNAs from the signature. Enriched terms with  $p$ -values < 0.05 were considered statistically significant.

#### 4.7. Development

This study was constructed using the R programming language (v4.2.0) with the RStudio platform (v4) hosted on the servers of the Multiuser Bioinformatics Center of the Metropolis Digital Institute at UFRN. The constructed codes are available in the GitHub repository ([https://github.com/epfarias/transcriptional\\_sig\\_ceRNA\\_KIRC](https://github.com/epfarias/transcriptional_sig_ceRNA_KIRC), accessed on 27 July 2023).

## 5. Conclusions

This study aimed to build a transcriptional signature of clear-cell renal-cell carcinoma from differentially expressed genes that act as a competitor endogenous RNA network.

Using feature selection techniques for signature construction represents a promising application in this vast area of pattern recognition and machine learning. By integrating expression data with clinical information, we successfully constructed transcriptional

signatures comprising multiple genes. The incorporation of evaluative metrics allowed us to gain valuable insights into the signature, assessing the metrics of the accuracy, sensitivity, and specificity of the signature in order to classify metastatic tissue expression. Using the external dataset permitted the examination of the signature generalization, thus validating its action as a metastatic classifier in clear-cell renal cancer.

With the cluster analysis, it was possible to know the actions performed by the signature genes within the cellular environment of clear-cell renal-cell carcinoma and how the effects of this regulatory process occur, indicating new roles for the lncRNA AF117829.1 and the mRNA RASD1. As research in the realm of lncRNA actions on cancer development undergoes constant evolution, our latest findings provide novel insights that illuminate promising avenues for future exploration. The dynamic nature of this field underscores the importance of our study, pointing toward potential directions for further investigation.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms25084214/s1>.

**Author Contributions:** Conceptualization, E.F., P.T. and B.S.; Data curation, E.F. and P.T.; Formal analysis, E.F. and P.T.; Investigation, E.F.; Methodology, E.F. and P.T.; Project administration, P.T. and B.S.; Resources, E.F.; Software, E.F.; Supervision, B.S. and P.T.; Validation, E.F., P.T. and B.S.; Visualization, E.F.; Writing original draft, E.F. and B.S.; Writing review and editing, P.T. and B.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** Eptácio Farias was funded by grant number 88887.702717/2022-00 of the Brazilian funding agency CAPES—National Coordination of High Education Personnel Formation Program.

**Institutional Review Board Statement:** This study did not require ethical review and approval because it performed a secondary analysis of publicly available data.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** This study utilized openly accessible datasets for analysis. The findings presented in this paper stem from information gathered by the TCGA Research Network. The TCGA-KIRC dataset (version 07-19-2019) can be accessed through the UCSC Xena Browser [86], while the ICGC-RECA dataset is available via the ICGC Data Portal [87].

**Acknowledgments:** The authors express their gratitude to Rodrigo Dalmolin, Alexandre Paschoal, Dhiego Souto Andrade, Iara Souza, and Rafaella Ferraz for their valuable input and suggestions while drafting the manuscript. The authors also thank the Center for High Performance Computing (Núcleo de Processamento de Alto Desempenho-NPAD/UFRN) available at <https://npad.ufrn.br> (accessed on 5 June 2023) and the Multidisciplinary Bioinformatics Environment (BioME) at UFRN for providing computing resources for data processing.

**Conflicts of Interest:** The writers assert that they possess no conflicting concerns and this study was carried out without any affiliations or monetary connections that might be perceived as possible sources of bias. The sponsors played no part in shaping this study's structure, acquiring and analyzing data, deciding to publish, or preparing the manuscript.

## References

1. Dall'Oglio, M.; Srougi, M.; Nesrallah, L. Câncer de Rim. In *Tratado de Clínica Médica*, 2nd ed.; Roca: Rio de Janeiro, Brazil, 2006; pp. 3264–3273.
2. Vinay, K.; Aster, J.C.; Abbas, A.K. *Robbins & Cotran: Patologia: Bases Patológicas das Doenças*; Elsevier: Amsterdam, The Netherlands, 2011.
3. Muglia, V.F.; Prando, A. Renal cell carcinoma: Histological classification and correlation with imaging findings. *Radiol. Bras.* **2015**, *48*, 166–174. [[CrossRef](#)] [[PubMed](#)]
4. NFK. *Renal Carcinoma Guidelines*; NFK—National Kidney Foundation: New York, NY, USA, 2017.
5. Wang, Y.; Li, Z.; Li, W.; Zhou, L.; Jiang, Y. Prognostic significance of long non-coding RNAs in clear cell renal cell carcinoma: A meta-analysis. *Medicine* **2019**, *98*, e17276. [[CrossRef](#)] [[PubMed](#)]
6. Cui, H.; Shan, H.; Miao, M.Z.; Jiang, Z.; Meng, Y.; Chen, R.; Zhang, L.; Liu, Y. Identification of the key genes and pathways involved in the tumorigenesis and prognosis of kidney renal clear cell carcinoma. *Sci. Rep.* **2020**, *10*, 4271. [[CrossRef](#)] [[PubMed](#)]
7. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **2013**, *499*, 43–49. [[CrossRef](#)] [[PubMed](#)]

8. Klinge, C.M. Non-coding RNAs: Long non-coding RNAs and microRNAs in endocrine-related cancers. *Endocr. Relat. Cancer* **2018**, *25*, R259–R282. [[CrossRef](#)] [[PubMed](#)]
9. Kazimierczyk, M.; Kasprowicz, M.K.; Kasprzyk, M.E.; Wrzesinski, J. Human Long Noncoding RNA Interactome: Detection, Characterization and Function. *Int. J. Mol. Sci.* **2020**, *21*, 1027. [[CrossRef](#)] [[PubMed](#)]
10. Statello, L.; Guo, C.-J.; Chen, L.-L.; Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 96–118. [[CrossRef](#)]
11. Morris, K.V.; Mattick, J.S. The rise of regulatory RNA. *Nat. Rev. Genet.* **2014**, *15*, 423–437. [[CrossRef](#)]
12. Yao, R.-W.; Wang, Y.; Chen, L.-L. Cellular functions of long noncoding RNAs. *Nat. Cell Biol.* **2019**, *21*, 542–551. [[CrossRef](#)]
13. Schmitz, S.U.; Grote, P.; Herrmann, B.G. Mechanisms of long noncoding RNA function in development and disease. *Cell. Mol. Life Sci.* **2016**, *73*, 2491–2509. [[CrossRef](#)]
14. Wang, P.-S.; Wang, Z.; Yang, C. Dysregulations of long non-coding RNAs—The emerging “lnc” in environmental carcinogenesis. *Semin. Cancer Biol.* **2021**, *76*, 163–172. [[CrossRef](#)] [[PubMed](#)]
15. Chiu, H.-S.; Somvanshi, S.; Patel, E.; Chen, T.-W.; Singh, V.P.; Zorman, B.; Patil, S.L.; Pan, Y.; Chatterjee, S.S.; Sood, A.K.; et al. Pan-Cancer Analysis of lncRNA Regulation Supports Their Targeting of Cancer Genes in Each Tumor Context. *Cell Rep.* **2018**, *23*, 297–312.e12. [[CrossRef](#)] [[PubMed](#)]
16. 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](#)] [[PubMed](#)]
17. Qi, X.; Lin, Y.; Chen, J.; Shen, B. Decoding competing endogenous RNA networks for cancer biomarker discovery. *Brief. Bioinform.* **2020**, *21*, 441–457. [[CrossRef](#)] [[PubMed](#)]
18. Chan, J.; Tay, Y. Noncoding RNA: RNA Regulatory Networks in Cancer. *Int. J. Mol. Sci.* **2018**, *19*, 1310. [[CrossRef](#)] [[PubMed](#)]
19. Bhan, A.; Soleimani, M.; Mandal, S.S. Long Noncoding RNA and Cancer: A New Paradigm. *Cancer Res.* **2017**, *77*, 3965–3981. [[CrossRef](#)] [[PubMed](#)]
20. Liu, S.J.; Dang, H.X.; Lim, D.A.; Feng, F.Y.; Maher, C.A. Long noncoding RNAs in cancer metastasis. *Nat. Rev. Cancer* **2021**, *21*, 446–460. [[CrossRef](#)] [[PubMed](#)]
21. Subramanian, I.; Verma, S.; Kumar, S.; Jere, A.; Anamika, K. Multi-omics Data Integration, Interpretation, and Its Application. *Bioinform. Biol. Insights* **2020**, *14*, 117793221989905. [[CrossRef](#)]
22. Reel, P.S.; Reel, S.; Pearson, E.; Trucco, E.; Jefferson, E. Using machine learning approaches for multi-omics data analysis: A review. *Biotechnol. Adv.* **2021**, *49*, 107739. [[CrossRef](#)]
23. Black, J.E.; Kueper, J.K.; Williamson, T.S. An introduction to machine learning for classification and prediction. *Fam. Pract.* **2023**, *40*, 200–204. [[CrossRef](#)]
24. Andrade, D.S.; Terrematte, P.; Rennó-Costa, C.; Zilberberg, A.; Efroni, S. GENTLE: A novel bioinformatics tool for generating features and building classifiers from T cell repertoire cancer data. *BMC Bioinform.* **2023**, *24*, 32. [[CrossRef](#)] [[PubMed](#)]
25. Kann, B.H.; Hosny, A.; Aerts, H.J.W.L. Artificial intelligence for clinical oncology. *Cancer Cell.* **2021**, *39*, 916–927. [[CrossRef](#)] [[PubMed](#)]
26. Liu, H.; Motoda, H. (Eds.) *Computational Methods of Feature Selection*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2008.
27. Zhou, L.; Ye, J.; Wen, F.; Yu, H. Identification of Novel Prognostic Signatures for Clear Cell Renal Cell Carcinoma Based on ceRNA Network Construction and Immune Infiltration Analysis. *Dis. Markers* **2022**, *2022*, 4033583. [[CrossRef](#)] [[PubMed](#)]
28. Zhang, Y.; Dai, J.; Huang, W.; Chen, Q.; Chen, W.; He, Q.; Chen, F.; Zhang, P. Identification of a competing endogenous RNA network related to immune signature in clear cell renal cell carcinoma. *Aging* **2021**, *13*, 25980–26002. [[CrossRef](#)] [[PubMed](#)]
29. Yu, J.; Mao, W.; Sun, S.; Hu, Q.; Wang, C.; Xu, Z.; Liu, R.; Chen, S.; Xu, B.; Chen, M. Identification of an m6A-Related lncRNA Signature for Predicting the Prognosis in Patients with Kidney Renal Clear Cell Carcinoma. *Front. Oncol.* **2021**, *11*, 663263. [[CrossRef](#)] [[PubMed](#)]
30. Yin, H.; Wang, X.; Zhang, X.; Wang, Y.; Zeng, Y.; Xiong, Y.; Li, T.; Lin, R.; Zhou, Q.; Ling, H.; et al. Integrated analysis of long noncoding RNA associated-competing endogenous RNA as prognostic biomarkers in clear cell renal carcinoma. *Cancer Sci.* **2018**, *109*, 3336–3349. [[CrossRef](#)] [[PubMed](#)]
31. Wang, Y.; Sun, Z.; Lu, S.; Zhang, X.; Xiao, C.; Li, T.; Wu, J. Identification of PLAUR-related ceRNA and immune prognostic signature for kidney renal clear cell carcinoma. *Front. Oncol.* **2022**, *12*, 834524. [[CrossRef](#)]
32. Sun, P.; Xu, H.; Zhu, K.; Li, M.; Han, R.; Shen, J.; Xia, X.; Chen, X.; Fei, G.; Zhou, S.; et al. The cuproptosis related genes signature predicts the prognosis and correlates with the immune status of clear cell renal cell carcinoma. *Front. Genet.* **2022**, *13*, 1061382. [[CrossRef](#)]
33. Song, J.; Peng, J.; Zhu, C.; Bai, G.; Liu, Y.; Zhu, J.; Liu, J. Identification and Validation of Two Novel Prognostic lncRNAs in Kidney Renal Clear Cell Carcinoma. *Cell Physiol. Biochem.* **2018**, *48*, 2549–2562. [[CrossRef](#)]
34. Quan, J.; Huang, B. Identification and validation of the molecular subtype and prognostic signature for clear cell renal cell carcinoma based on neutrophil extracellular traps. *Front. Cell Dev. Biol.* **2022**, *10*, 1021690. [[CrossRef](#)]
35. Peng, Y.; Wu, S.; Xu, Z.; Hou, D.; Li, N.; Zhang, Z.; Wang, L.; Wang, H. A prognostic nomogram based on competing endogenous RNA network for clear-cell renal cell carcinoma. *Cancer Med.* **2021**, *10*, 5499–5512. [[CrossRef](#)]
36. Lin, G.; Wang, H.; Wu, Y.; Wang, K.; Li, G. Hub Long Noncoding RNAs with m6A Modification for Signatures and Prognostic Values in Kidney Renal Clear Cell Carcinoma. *Front. Mol. Biosci.* **2021**, *8*, 682471. [[CrossRef](#)] [[PubMed](#)]
37. Broido, A.D.; Clauset, A. Scale-free networks are rare. *Nat. Commun.* **2019**, *10*, 1017. [[CrossRef](#)] [[PubMed](#)]

38. Clauset, A.; Shalizi, C.R.; Newman, M.E.J. Power-law distributions in empirical data. *arXiv* **2007**, arXiv:0706.1062. [[CrossRef](#)]
39. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159–174. [[CrossRef](#)] [[PubMed](#)]
40. Bendavid, A. Comparison of classification accuracy using Cohen’s Weighted Kappa. *Expert Syst. Appl.* **2008**, *34*, 825–832. [[CrossRef](#)]
41. Youden, W.J. Index for rating diagnostic tests. *Cancer* **1950**, *3*, 32–35. [[CrossRef](#)] [[PubMed](#)]
42. Ho, S.Y.; Phua, K.; Wong, L.; Bin Goh, W.W. Extensions of the External Validation for Checking Learned Model Interpretability and Generalizability. *Patterns* **2020**, *1*, 100129. [[CrossRef](#)]
43. Hicks, S.A.; Strümke, I.; Thambawita, V.; Hammou, M.; Riegler, M.A.; Halvorsen, P.; Parasa, S. On evaluation metrics for medical applications of artificial intelligence. *Sci. Rep.* **2022**, *12*, 5979. [[CrossRef](#)]
44. Zhu, J.; Luo, Y.; Zhao, Y.; Kong, Y.; Zheng, H.; Li, Y.; Gao, B.; Ai, L.; Huang, H.; Huang, J.; et al. circEHBP1 promotes lymphangiogenesis and lymphatic metastasis of bladder cancer via miR-130a-3p/TGFβR1/VEGF-D signaling. *Mol. Ther.* **2021**, *29*, 1838–1852. [[CrossRef](#)]
45. Chen, J.; Yan, D.; Wu, W.; Zhu, J.; Ye, W.; Shu, Q. MicroRNA-130a promotes the metastasis and epithelial-mesenchymal transition of osteosarcoma by targeting PTEN. *Oncol. Rep.* **2016**, *35*, 3285–3292. [[CrossRef](#)] [[PubMed](#)]
46. Li, B.; Huang, P.; Qiu, J.; Liao, Y.; Hong, J.; Yuan, Y. MicroRNA-130a is down-regulated in hepatocellular carcinoma and associates with poor prognosis. *Med. Oncol.* **2014**, *31*, 230. [[CrossRef](#)] [[PubMed](#)]
47. Stückerath, I.; Rack, B.; Janni, W.; Jäger, B.; Pantel, K.; Schwarzenbach, H. Aberrant plasma levels of circulating *miR-16*, *miR-107*, *miR-130a* and *miR-146a* are associated with lymph node metastasis and receptor status of breast cancer patients. *Oncotarget* **2015**, *6*, 13387–13401. [[CrossRef](#)] [[PubMed](#)]
48. Ma, F.; Xie, Y.; Lei, Y.; Kuang, Z.; Liu, X. The microRNA-130a-5p/RUNX2/STK32A network modulates tumor invasive and metastatic potential in non-small cell lung cancer. *BMC Cancer* **2020**, *20*, 580. [[CrossRef](#)] [[PubMed](#)]
49. Xu, C.-H.; Xiao, L.-M.; Liu, Y.; Chen, L.-K.; Zheng, S.-Y.; Zeng, E.-M.; Li, D.-. H. The lncRNA HOXA11-AS promotes glioma cell growth and metastasis by targeting miR-130a-5p/HMGB2. *Eur. Rev. Med. Pharmacol. Sci.* **2019**, *23*, 241–252. [[PubMed](#)]
50. Schier, A.C.; Taatjes, D.J. Structure and mechanism of the RNA polymerase II transcription machinery. *Genes Dev.* **2020**, *34*, 465–488. [[CrossRef](#)] [[PubMed](#)]
51. Hardwick, C.; Hoare, K.; Owens, R.; Hohn, H.; Hook, M.; Moore, D.; Cripps, V.; Austen, L.; Nance, D.; Turley, E. Molecular cloning of a novel hyaluronan receptor that mediates tumor cell motility. *J. Cell Biol.* **1992**, *117*, 1343–1350. [[CrossRef](#)]
52. Yang, M.; Chen, B.; Kong, L.; Chen, X.; Ouyang, Y.; Bai, J.; Yu, D.; Zhang, H.; Li, X.; Zhang, D. HMMR promotes peritoneal implantation of gastric cancer by increasing cell-cell interactions. *Discov. Oncol.* **2022**, *13*, 81. [[CrossRef](#)]
53. Li, Y.; Deng, L.; Pan, X.; Liu, C.; Fu, R. The Role of lncRNA AF117829.1 in the Immunological Pathogenesis of Severe Aplastic Anaemia. *Oxidative Med. Cell. Longev.* **2021**, *2021*, 5587921. [[CrossRef](#)]
54. Xia, F.; Yan, Y.; Shen, C. A Prognostic Pyroptosis-Related lncRNAs Risk Model Correlates With the Immune Microenvironment in Colon Adenocarcinoma. *Front. Cell Dev. Biol.* **2021**, *9*, 811734. [[CrossRef](#)]
55. Filho, G.S.; Caballé-Serrano, J.; Sawada, K.; Bosshardt, D.D.; Bianchini, M.A.; Buser, D.; Gruber, R. Conditioned Medium of Demineralized Freeze-Dried Bone Activates Gene Expression in Periodontal Fibroblasts In Vitro. *J. Periodontol.* **2015**, *86*, 827–834. [[CrossRef](#)]
56. Du, S.; Hu, W.; Zhao, Y.; Zhou, H.; Wen, W.; Xu, M.; Zhao, P.; Liu, K. Long non-coding RNA MAGI2-AS3 inhibits breast cancer cell migration and invasion via sponging microRNA-374a. *Cancer Biomark.* **2019**, *24*, 269–277. [[CrossRef](#)] [[PubMed](#)]
57. Yin, Z.; Ma, T.; Yan, J.; Shi, N.; Zhang, C.; Lu, X.; Hou, B.; Jian, Z. LncRNA MAGI2-AS3 inhibits hepatocellular carcinoma cell proliferation and migration by targeting the miR-374b-5p/SMG1 signaling pathway. *J. Cell. Physiol.* **2019**, *234*, 18825–18836. [[CrossRef](#)] [[PubMed](#)]
58. Takahashi, M.; Inoue, T.; Huang, M.; Numakura, K.; Tsuruta, H.; Saito, M.; Maeno, A.; Nakamura, E.; Narita, S.; Tsuchiya, N.; et al. Inverse relationship between insulin receptor expression and progression in renal cell carcinoma. *Oncol. Rep.* **2017**, *37*, 2929–2941. [[CrossRef](#)] [[PubMed](#)]
59. Pothapragada, S.P.; Gupta, P.; Mukherjee, S.; Das, T. Matrix mechanics regulates epithelial defence against cancer by tuning dynamic localization of filamin. *Nat. Commun.* **2022**, *13*, 218. [[CrossRef](#)] [[PubMed](#)]
60. Jamal, S.; Cheriyan, V.T.; Muthu, M.; Munie, S.; Levi, E.; Ashour, A.E.; Pass, H.I.; Wali, A.; Singh, M.; Rishi, A.K. CARP-1 Functional Mimetics Are a Novel Class of Small Molecule Inhibitors of Malignant Pleural Mesothelioma Cells. *PLoS ONE* **2014**, *9*, e89146. [[CrossRef](#)] [[PubMed](#)]
61. Muthu, M.; Cheriyan, V.T.; Munie, S.; Levi, E.; Frank, J.; Ashour, A.E.; Singh, M.; Rishi, A.K. Mechanisms of Neuroblastoma Cell Growth Inhibition by CARP-1 Functional Mimetics. *PLoS ONE* **2014**, *9*, e102567. [[CrossRef](#)] [[PubMed](#)]
62. Wang, Y.; Argiles-Castillo, D.; Kane, E.I.; Zhou, A.; Spratt, D.E. HECT E3 ubiquitin ligases—Emerging insights into their biological roles and disease relevance. *J. Cell Sci.* **2020**, *133*, jcs228072. [[CrossRef](#)] [[PubMed](#)]
63. Xie, S.; Xia, L.; Song, Y.; Liu, H.; Wang, Z.; Zhu, X. Insights into the Biological Role of NEDD4L E3 Ubiquitin Ligase in Human Cancers. *Front. Oncol.* **2021**, *11*, 774648. [[CrossRef](#)]
64. Vaidyanathan, G.; Cismowski, M.J.; Wang, G.; Vincent, T.S.; Brown, K.D.; Lanier, S.M. The Ras-related protein AGS1/RASD1 suppresses cell growth. *Oncogene* **2004**, *23*, 5858–5863. [[CrossRef](#)]

65. Stelzer, G.; Rosen, N.; Plaschkes, I.; Zimmerman, S.; Twik, M.; Fishilevich, S.; Stein, T.I.; Nudel, R.; Lieder, I.; Mazon, Y.; et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr. Protoc. Bioinform.* **2016**, *54*, 1.30.1–1.30.33. [[CrossRef](#)] [[PubMed](#)]
66. Wang, X.J.; Yan, Z.J.; Luo, G.C.; Chen, Y.Y.; Bai, P.M. miR-26 suppresses renal cell cancer via down-regulating coronin-3. *Mol. Cell Biochem.* **2020**, *463*, 137–146. [[CrossRef](#)] [[PubMed](#)]
67. Liao, M.; Peng, L. MiR-206 may suppress non-small lung cancer metastasis by targeting CORO1C. *Cell. Mol. Biol. Lett.* **2020**, *25*, 22. [[CrossRef](#)] [[PubMed](#)]
68. Bell, D.W.; Sikdar, N.; Lee, K.; Price, J.C.; Chatterjee, R.; Park, H.-D.; Fox, J.; Ishiai, M.; Rudd, M.L.; Pollock, L.M.; et al. Predisposition to Cancer Caused by Genetic and Functional Defects of Mammalian Atad5. *PLoS Genet.* **2011**, *7*, e1002245. [[CrossRef](#)] [[PubMed](#)]
69. Yang, P.; Huang, X.; Lai, C.; Li, L.; Li, T.; Huang, P.; Ouyang, S.; Yan, J.; Cheng, S.; Lei, G.; et al. SET domain containing 1B gene is mutated in primary hepatic neuroendocrine tumors. *Int. J. Cancer* **2019**, *145*, 2986–2995. [[CrossRef](#)] [[PubMed](#)]
70. Hong, C.-Q.; Zhang, F.; You, Y.-J.; Qiu, W.-L.; Giuliano, A.E.; Cui, X.-J.; Zhang, G.-J.; Cui, Y.-K. Elevated C1orf63 expression is correlated with CDK10 and predicts better outcome for advanced breast cancers: A retrospective study. *BMC Cancer* **2015**, *15*, 548. [[CrossRef](#)] [[PubMed](#)]
71. Hong, S.-W.; Jin, D.-H.; Shin, J.-S.; Moon, J.-H.; Na, Y.-S.; Jung, K.-A.; Kim, S.M.; Kim, J.C.; Kim, K.P.; Hong, Y.S.; et al. Ring Finger Protein 149 Is an E3 Ubiquitin Ligase Active on Wild-type v-Raf Murine Sarcoma Viral Oncogene Homolog B1 (BRAF). *J. Biol. Chem.* **2012**, *287*, 24017–24025. [[CrossRef](#)] [[PubMed](#)]
72. Low, S.-K.; Kuchiba, A.; Zembutsu, H.; Saito, A.; Takahashi, A.; Kubo, M.; Daigo, Y.; Kamatani, N.; Chiku, S.; Totsuka, H.; et al. Genome-Wide Association Study of Pancreatic Cancer in Japanese Population. *PLoS ONE* **2010**, *5*, e11824. [[CrossRef](#)]
73. Zhang, X.; Horwitz, G.A.; Prezant, T.R.; Valentini, A.; Nakashima, M.; Bronstein, M.D.; Melmed, S. Structure, Expression, and Function of Human Pituitary Tumor-Transforming Gene (PTTG). *Mol. Endocrinol.* **1999**, *13*, 156–166. [[CrossRef](#)]
74. Mei, L. Multiple types of noncoding RNA are involved in potential modulation of PTTG1's expression and function in breast cancer. *Genomics* **2022**, *114*, 110352. [[CrossRef](#)]
75. Zi, Z. Molecular Engineering of the TGF- $\beta$  Signaling Pathway. *J. Mol. Biol.* **2019**, *431*, 2644–2654. [[CrossRef](#)] [[PubMed](#)]
76. Sun, Y.; Liu, W.-Z.; Liu, T.; Feng, X.; Yang, N.; Zhou, H.-F. Signaling pathway of MAPK/ERK in cell proliferation, differentiation, migration, senescence and apoptosis. *J. Recept. Signal Transduct.* **2015**, *35*, 600–604. [[CrossRef](#)]
77. Xiong, L.; He, X.; Wang, L.; Dai, P.; Zhao, J.; Zhou, X.; Tang, H. Hypoxia-associated prognostic markers and competing endogenous RNA coexpression networks in lung adenocarcinoma. *Sci. Rep.* **2022**, *12*, 21340. [[CrossRef](#)] [[PubMed](#)]
78. Du, Y.; Kong, C.; Zhu, Y.; Yu, M.; Li, Z.; Bi, J.; Li, Z.; Liu, X.; Zhang, Z.; Yu, X. Knockdown of SNHG15 suppresses renal cell carcinoma proliferation and EMT by regulating the NF- $\kappa$ B signaling pathway. *Int. J. Oncol.* **2018**, *53*, 384–394. [[CrossRef](#)] [[PubMed](#)]
79. Jin, B.; Jin, H.; Wu, H.-B.; Xu, J.-J.; Li, B. Long non-coding RNA SNHG15 promotes CDK14 expression via miR-486 to accelerate non-small cell lung cancer cells progression and metastasis. *J. Cell. Physiol.* **2018**, *233*, 7164–7172. [[CrossRef](#)] [[PubMed](#)]
80. Huang, L.; Lin, H.; Kang, L.; Huang, P.; Huang, J.; Cai, J.; Xian, Z.; Zhu, P.; Huang, M.; Wang, L.; et al. Aberrant expression of long noncoding RNA SNHG15 correlates with liver metastasis and poor survival in colorectal cancer. *J. Cell. Physiol.* **2019**, *234*, 7032–7039. [[CrossRef](#)] [[PubMed](#)]
81. Mitchell, S.; Vargas, J.; Hoffmann, A. Signaling via the NF $\kappa$ B system. *WIREs Mech. Dis.* **2016**, *8*, 227–241. [[CrossRef](#)]
82. Huttlin, E.L.; Bruckner, R.J.; Paulo, J.A.; Cannon, J.R.; Ting, L.; Baltier, K.; Colby, G.; Gebreab, F.; Gygi, M.P.; Parzen, H.; et al. Architecture of the human interactome defines protein communities and disease networks. *Nature* **2017**, *545*, 505–509. [[CrossRef](#)]
83. Li, G.; Wang, Y.; Cheng, Y. IL2RB Is a Prognostic Biomarker Associated with Immune Infiltrates in Pan-Cancer. *J. Oncol.* **2022**, *2022*, 2043880. [[CrossRef](#)]
84. Fox, S.B.; Bragança, J.; Turley, H.; Campo, L.; Han, C.; Gatter, K.C.; Bhattacharya, S.; Harris, A.L. CITED4 Inhibits Hypoxia-Activated Transcription in Cancer Cells, and Its Cytoplasmic Location in Breast Cancer Is Associated with Elevated Expression of Tumor Cell Hypoxia-Inducible Factor 1 $\alpha$ . *Cancer Res.* **2004**, *64*, 6075–6081. [[CrossRef](#)]
85. Grossman, R.L.; Heath, A.P.; Ferretti, V.; Varmus, H.E.; Lowy, D.R.; Kibbe, W.A.; Staudt, L.M. Toward a Shared Vision for Cancer Genomic Data. *N. Engl. J. Med.* **2016**, *375*, 1109–1112. [[CrossRef](#)] [[PubMed](#)]
86. Goldman, M.J.; Craft, B.; Hastie, M.; Repecka, K.; McDade, F.; Kamath, A.; Banerjee, A.; Luo, Y.; Rogers, D.; Brooks, A.N.; et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* **2020**, *38*, 675–678. [[CrossRef](#)] [[PubMed](#)]
87. Zhang, J.; Bajari, R.; Andric, D.; Gerthoffert, F.; Lepsa, A.; Nahal-Bose, H.; Stein, L.D.; Ferretti, V. The International Cancer Genome Consortium Data Portal. *Nat. Biotechnol.* **2019**, *37*, 367–369. [[CrossRef](#)] [[PubMed](#)]
88. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)] [[PubMed](#)]
89. Li, R.; Qu, H.; Wang, S.; Wei, J.; Zhang, L.; Ma, R.; Lu, J.; Zhu, J.; Zhong, W.-D.; Jia, Z. GDCRNATools: An R/Bioconductor package for integrative analysis of lncRNA, miRNA and mRNA data in GDC. *Bioinformatics* **2018**, *34*, 2515–2517. [[CrossRef](#)]
90. Li, J.-H.; Liu, S.; Zhou, H.; Qu, L.-H.; Yang, J.-H. starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucl. Acids Res.* **2014**, *42*, D92–D97. [[CrossRef](#)] [[PubMed](#)]
91. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)]

92. Barabási, A.-L.; Oltvai, Z.N. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **2004**, *5*, 101–113. [[CrossRef](#)]
93. Alstott, J.; Bullmore, E.; Plenz, D. powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions. *PLoS ONE* **2014**, *9*, e85777. [[CrossRef](#)]
94. Terrematte, P.; Andrade, D.; Justino, J.; Stransky, B.; De Araújo, D.; Dória Neto, A. A Novel Machine Learning 13-Gene Signature: Improving Risk Analysis and Survival Prediction for Clear Cell Renal Cell Carcinoma Patients. *Cancers* **2022**, *14*, 2111. [[CrossRef](#)]
95. Stawiski, K.; Kaszkowiak, M.; Mikulski, D.; Hogendorf, P.; Durczyński, A.; Strzelczyk, J.; Chowdhury, D.; Fendler, W. OmicSelector: Automatic feature selection and deep learning modeling for omic experiments. preprint. *Bioinformatics* **2022**. [[CrossRef](#)]
96. Lang, M.; Schratz, P. mlr3verse: Easily Install and Load the “mlr3” Package Family. Available online: <https://mlr3verse.ml-org.com> (accessed on 15 May 2023).
97. Mayakonda, A.; Lin, D.-C.; Assenov, Y.; Plass, C.; Koeffler, H.P. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **2018**, *28*, 1747–1756. [[CrossRef](#)] [[PubMed](#)]
98. Mermel, C.H.; Schumacher, S.E.; Hill, B.; Meyerson, M.L.; Beroukhi, R.; Getz, G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **2011**, *12*, R41. [[CrossRef](#)] [[PubMed](#)]
99. Frankish, A.; Diekhans, M.; Jungreis, I.; Lagarde, J.; Loveland, J.E.; Mudge, J.M.; Sisu, C.; Wright, J.C.; Armstrong, J.; Barnes, I.; et al. GENCODE 2021. *Nucleic Acids Res.* **2021**, *49*, D916–D923. [[CrossRef](#)]
100. Therneau, T.M.; Grambsch, P.M. *Modeling Survival Data: Extending the Cox Model*; Springer: New York, NY, USA, 2000.
101. Harrison, E.; Drake, T.; Pius, R. finalfit: Quickly Create Elegant Regression Results Tables and Plots when Modelling. Available online: <https://github.com/ewenharrison/finalfit> (accessed on 15 June 2023).
102. Aalen, O.O. A linear regression model for the analysis of life times. *Stat. Med.* **1989**, *8*, 907–925. [[CrossRef](#)] [[PubMed](#)]
103. Morris, J.A.; Gardner, M.J. Statistics in Medicine: Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. *BMJ* **1988**, *296*, 1313–1316. [[CrossRef](#)] [[PubMed](#)]
104. The Gene Ontology Consortium; Carbon, S.; Douglass, E.; Good, B.M.; Unni, D.R.; Harris, N.L.; Mungall, C.J.; Basu, S.; Chisholm, R.L.; Dodson, R.J.; et al. The Gene Ontology resource: Enriching a gold mine. *Nucleic Acids Res.* **2021**, *49*, D325–D334.
105. Wu, T.; Hu, E.; Xu, S.; Chen, M.; Guo, P.; Dai, Z.; Feng, T.; Zhou, L.; Tang, W.; Zhan, L.; et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2021**, *2*, 100141. [[CrossRef](#)]
106. Vlachos, I.S.; Zagganas, K.; Paraskevopoulou, M.D.; Georgakilas, G.; Karagkouni, D.; Vergoulis, T.; Dalamagas, T.; Hatzigeorgiou, A.G. DIANA-miRPath v3.0: Deciphering microRNA function with experimental support. *Nucleic Acids Res.* **2015**, *43*, W460–W466. [[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.