

Review

Non-Coding RNAs in Lung Cancer: Contribution of Bioinformatics Analysis to the Development of Non-Invasive Diagnostic Tools

Meik Kunz ¹, Beat Wolf ^{1,2}, Harald Schulze ³, David Atlan ⁴, Thorsten Walles ⁵, Heike Walles ^{6,7} and Thomas Dandekar ^{1,8,*}

- ¹ Functional Genomics and Systems Biology Group, Department of Bioinformatics, Biocenter, University of Wuerzburg, 97074 Wuerzburg, Germany; meik.kunz@uni-wuerzburg.de (M.K.); beat.wolf@hefr.ch (B.W.)
 - ² University of Applied Sciences and Arts of Western Switzerland, Perolles 80, 1700 Fribourg, Switzerland
 - ³ Institute of Experimental Biomedicine, University Hospital Wuerzburg, 97080 Wuerzburg, Germany; schulze_h@ukw.de
 - ⁴ Phenosystems SA, 137 Rue de Tubize, 1440 Braine le Château, Belgium; phenosystems@gmail.com
 - ⁵ Department of Cardiothoracic Surgery, University Hospital of Wuerzburg, 97080 Wuerzburg, Germany; walles_t@ukw.de
 - ⁶ Department of Tissue Engineering and Regenerative Medicine, University Hospital Wuerzburg, Roentgenring 11, 97070 Wuerzburg, Germany; heike.walles@igb.fraunhofer.de
 - ⁷ Translational Center Wuerzburg “Regenerative therapies in oncology and musculoskeletal disease” Wuerzburg branch of the Fraunhofer Institute Interfacial Engineering and Biotechnology (IGB), Roentgenring 11, 97070 Wuerzburg, Germany
 - ⁸ BioComputing Unit, European Molecular Biology Laboratory (EMBL) Heidelberg, Meyerhofstraße 1, 69117 Heidelberg, Germany
- * Correspondence: dandekar@biozentrum.uni-wuerzburg.de; Tel.: +49-931-318-4551

Academic Editor: George A. Calin

Received: 16 October 2016; Accepted: 15 December 2016; Published: 26 December 2016

Abstract: Lung cancer is currently the leading cause of cancer related mortality due to late diagnosis and limited treatment intervention. Non-coding RNAs are not translated into proteins and have emerged as fundamental regulators of gene expression. Recent studies reported that microRNAs and long non-coding RNAs are involved in lung cancer development and progression. Moreover, they appear as new promising non-invasive biomarkers for early lung cancer diagnosis. Here, we highlight their potential as biomarker in lung cancer and present how bioinformatics can contribute to the development of non-invasive diagnostic tools. For this, we discuss several bioinformatics algorithms and software tools for a comprehensive understanding and functional characterization of microRNAs and long non-coding RNAs.

Keywords: lung cancer; non-invasive biomarkers; miRNAs; lncRNAs; early diagnosis; bioinformatics; algorithm

1. Introduction

Lung cancer is the main cause of cancer-associated mortality worldwide. Non-small cell lung cancer (NSCLC) with the two major pathologic types adenocarcinoma (AC) and squamous cell carcinoma (SQ) [1,2] represents with 85% the most often diagnosed subtype [3,4], whereas small-cell lung carcinoma (SCLC, 15%) is the most aggressive subtype but less observed [5]. Several studies reported differences in the gene expression characteristics between AC and SQ, but also between both lung cancer subtypes [1,5–8]. Most prominently altered molecular marker genes in SCLC are associated with neuronal differentiation and/or growth such as the human achaete-scute homolog 1

(ASCL1) and the glycine receptor alpha1 subunit gene (GLRA1) [8,9], whereas NSCLC is characterized by deregulation of the Epidermal growth factor receptor (EGFR) and tumor-suppressor p53 [10,11].

About 25%–30% of patients are at limited stage of disease at time of diagnosis warranting surgical therapy. The remaining majority of patients depends on systemic therapy. Lung cancer detection is often too late due to missing patient symptoms at early disease and the lack of accurate diagnostic tools [12]. Notably, diagnosis at early stages can increase patient survival rate up to 45%–53% [12]. Diagnosis currently includes computer tomography (CT) or PET-CT scan which have a high false-positive rate (fall-out rate) up to 28% (incorrect as lung cancer identified individuals which are healthy, resulting in unnecessary additional diagnostics), followed by invasive biopsy and bronchoscopy [12–14]. Systemic therapy is based on platinum-based chemotherapy that can be combined with targeted therapies (e.g., tyrosine kinase inhibitor gefitinib) according to the biomarker status of the tumor. Targeted therapies raise hope for treatment improvements, but unfortunately lead commonly to resistance development after several months [11]. The five-year survival rate in these patients is less than 5%.

Over the past decades, diagnostics have mainly focused on proteins which represent only a small part of the entire human genome [12,15]. For example, the serum markers Cytokeratin 19 Fragment, Cancer Antigen 125, Cancer Antigen 19-9, tissue polypeptide antigen and Neuron-specific Enolase were extensively investigated [12,16–19]. However, due to the high false positives and low sensitivity, none of them can currently serve as stand-alone biomarker for early lung cancer detection [12,20,21]. Thus, accurate and innovative diagnostic biomarker tests for early disease detection are required. Ideally, for potential clinical application such tests should have the following characteristics: they should be minimal invasive, easily accessible in body fluids and highly specific, sensitive and robust [15,22].

Transcriptome analysis revealed that non-coding RNAs (ncRNAs; no protein encoding) such as microRNAs (miRNAs) and long non-coding RNAs (lncRNAs) represent more than 90% of the transcribed human genome [15,23–25]. Recent studies demonstrated that miRNAs and lncRNAs regulate several targets and are associated with important biological processes and signaling pathways, also often show a tissue specific expression [26–28]. Moreover, several miRNAs and lncRNAs were found to play an important role in lung cancer pathogenesis and can function as molecular marker for cancer metastasis and prognosis [15,29–31]. Thus, investigating the functional role of miRNAs and lncRNAs can lead to a better understanding of lung cancer tumorigenesis which opens new windows for effective diagnostic strategies for a better clinical management of lung cancer.

In this review, we present how bioinformatics can contribute to the development of non-invasive diagnostic tools, in particular we focus on lung cancer as an example disease. We report the current state of research of miRNAs and lncRNAs as novel diagnostic biomarkers and discuss bioinformatics algorithm and software tools for understanding the functional role of miRNAs and lncRNAs. However, all presented bioinformatics analysis tools can also be applied to other tumors and, where appropriate, to other diseases.

2. miRNAs as Diagnostic Biomarkers

miRNAs are small ncRNAs (~22 nucleotides) which are highly conserved among different organism. They lie in intronic or intergenic regions of coding and non-coding genes. miRNAs regulate gene expression in the cytoplasm through binding with their conserved seed-region to messenger RNA (mRNA) in a sequence-specific manner, resulting in mRNA degradation or inhibition which leads to a reduced translational activity or inhibition of translation [30,32–36] (see Figure 1).

miRNAs such as miRNA-21, let-7, miRNA-145, miRNA-155 and miRNA-210 are known to have an implication in lung tumorigenesis [37–41]. They are reported to show altered expression levels in lung cancer and correlated with tumor stage and patient survival [42–44]. Moreover, several miRNAs show a co-expression in lung cancer patient, e.g., miRNA-21, miRNA-145 and miRNA-155 or let-7g and miRNA-21 [37,45]. Table 1 summarizes different clinical miRNA lung cancer studies including background information (number of patient, tumor stage and histological subtype; if available).

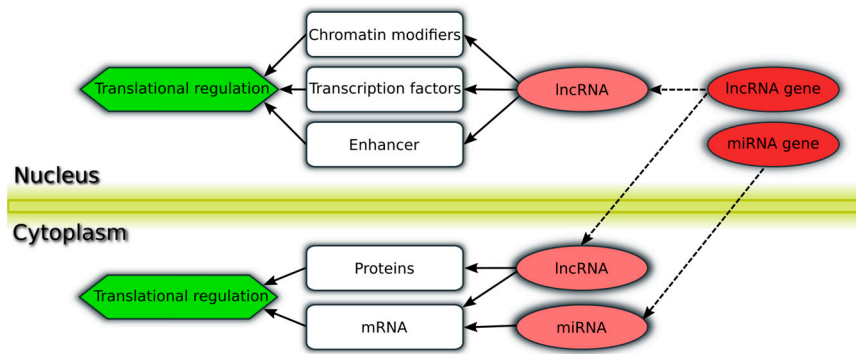


Figure 1. Schematic overview of miRNA and lncRNA regulation. miRNAs are transcribed as miRNA gene (light red circle) in the nucleus and further transported (dashed arrows) in the cytoplasm (darker red circle) where they regulate (continuous arrows) translation (green hexagon) through complementary mRNA binding (white rectangle). lncRNAs are transcribed in the nucleus (light red circle) and not only transported (dashed arrows) in the cytoplasm (darker red circle) and influence translation, e.g., through mRNA and protein binding (white rectangles), but can also regulate transcription (green hexagon), e.g., through chromatin modifier binding (white rectangles) in the nucleus.

Table 1. Table summarizing the different lung cancer miRNA studies.

Study (Ref.)	Patient Cohort	Important Reported Findings
Tang et al. (2013) [37]	Training set: 62 patients/60 healthy smokers Test set: 34 malignant tumor patients, 30 patients with benign pulmonary nodules, 32 healthy smokers	Plasma miRNA-21, miRNA-145 and miRNA-155 have strong potential as novel noninvasive biomarkers for early detection of lung cancer
Yu et al. (2008) [42]	112 NSCLC patients: AC (55), SQ (50), others (7)	five-miRNA signature (miRNA-221, let-7a, miRNA-137, miRNA-372, miRNA-182 *) for NSCLC treatment prediction outcome
Yanaihara et al. (2006) [43]	AC (65), AC normal (65); SQ (39), SQ normal (39) Stage classification: AC: stage I (41), stage II (8), stages III and IV (16); SQ: stage I (24), stage II (9), stages III and IV (6)	miRNA-155 and let-7a-2 correlates with poor survival; miRNA-205, miRNA-99b, miRNA-203, miRNA-202, miRNA-102 and miRNA-204-prec differentially expressed in AC and SQ; miRNA-21, miRNA-191, miRNA-155, miRNA-210, miRNA-126 * and miRNA-224 differentiate AC and SQ
Geng et al. (2014) [39]	Training set: 25 NSCLC patients: AC (8), SQ (13), others (4); stage: I (9), II (16); 25 healthy controls Test sample: 126 NSCLC patients: AC (45), SQ (64), others (17); stage: I (54), II (72); 42 NCPD (non-cancerous pulmonary disease) patients; 60 healthy controls	Plasma miRNA-20a, miRNA-145, miRNA-21, miRNA-223 and miRNA-221 as potential biomarkers in early-stage NSCLC
Zhu et al. (2016) [40]	112 NSCLC patients: AC (90), SQ (22); lymph node metastasis: negative (95), positive (17); stage: 0 (0), IA, IB (82), IIA, IIB (15), IIIA, IIIB (10); 104 controls (20 current healthy smokers, 23 pneumonia patients, 21 gastric cancer patients, 40 healthy controls)	Serum miRNA-182, miRNA-183, miRNA-210 and miRNA-126 levels serve as a diagnostic biomarker for NSCLC early detection; serum levels distinguish NSCLC or early-stage NSCLC from current smokers without lung cancer and pneumonia or gastric cancer patients
Bjaanaes et al. (2014) [41]	154 resected AC: stage: IA (45), IB (46), IIA (24), IIB (12), IIIA (26), IV (1); EGFR status: mutated (22), wt (130), not tested (2); KRAS status: mutated (52), wt (96), not tested (6); 20 normal tissue samples; independent cohort of 103 lung cancer patients	129 significantly differentially expressed miRNAs in AC compared with normal lung tissue; 17 differentially expressed miRNAs between EGFR-mutated and EGFR wildtype AC; miRNA-9, miRNA-21, miRNA-126, miRNA-133a and miRNA-500a deregulated in AC compared with normal lung tissue
Saito et al. (2011) [44]	317 AC patient tissues: stage: I (220), II (76), III (21)	miRNA-21 is associated with disease progression and survival in stage I AC
Capodanno et al. (2013) [45]	80 NSCLC patients: EGFR status: Wt: AC (32), SQ (20), LCC and other (3); stage: I (9), II (20), III (13), IV (3); mutated: AC (22), SQ (0), LCC and other (1); stage: I (3), II (11), III (1), IV (3) KRAS status: Wt: AC (39), SQ (20), LCC and other (3); stage: I (9), II (27), III (9), IV (4); mutated: AC (15), SQ (0), LCC and other (1); stage: I (3), II (4), III (5), IV (2)	let-7g and miRNA-21 combined with KRAS mutational status are useful biomarkers for NSCLC patients

Table 1. Cont.

Study (Ref.)	Patient Cohort	Important Reported Findings
Du et al. (2010) [5]	19 lung cancer cell lines: 7 NSCLC (2 AC, 3 SQ, 2 other); 9 SCLC; 3 immortalized normal	41 out of 136 differentially expressed miRNAs distinguish NSCLC and SCLC (e.g., miRNA-17-5p, miRNA-135, miRNA-103, miRNA-107, miRNA-301 and miRNA-338 altered in SCLC relative to NSCLC); miRNA-29a/b/c, miRNA-24, miRNA-21 and miRNA-221/222 down-regulated in SCLC
Lee et al. (2011) [46]	26 NSCLC cell lines, 14 SCLC cell lines, 31 SCLC tumors	miRNA-21, miRNA-29b, miRNA-34a/b/c, miRNA-155 and let-7a not related to SCLC patients
Landi et al. (2010) [30]	290 tissue samples: AC (165): stage: I (65), II (43), III (46), IV (11); SQ (125): stage: I (52), II (42), III (30), IV (1)	34 miRNAs differentiate AC from SQ in male smoker patients; five-miRNA signature predicts survival for SQ
Lebanony et al. (2009) [47]	Training set: 122 AC and SQ; 47 NSCLC FFPE samples Test set: 79 blinded cohort of NSCLC FFPE samples	miRNA-205 expression distinguishes SQ from AC
Bishop et al. (2010) [48]	102 resected NSCLC: AC (50): grades: well (9), moderate (24), poor (17); SQ (52): grades: well (2), moderate (35), poor (15); 21 preoperative biopsies/aspirates	miRNAs such as miRNA-205 are reliable to classify NSCLC
Montani et al. (2015) [14]	COSMOS study with high-risk individuals (n = 1115): Calibration Set: lung cancer (12): stage: I (11), II-III (1); lung cancer deaths (1); no lung cancer (12) Validation Set: lung cancer (36): stage: I (31), II-III (5); lung cancer deaths (3); no lung cancer (972); PN+COPD (81) Specificity Set: no lung cancer (83); PN + COPD + NOD (78); Benign (5) All (CT screening): lung cancer (48): stage: I (42), II-III (6); lung cancer deaths (4); no lung cancer (1067); PN + COPD + NOD (159); Benign (5)	miR-Test using 13 miRNAs (miRNA-92a-3p, miRNA-30b-5p, miRNA-191-5p, miRNA-484, miRNA-328-3p, miRNA-30c-5p, miRNA-374a-5p, let-7d-5p, miRNA-331-3p, miRNA-29a-3p, miRNA-148a-3p, miRNA-223-3p, miRNA-140-5p) represent a useful tool for lung cancer screening in high-risk individuals
Sozzi et al. (2014) [13]	MILD trial study: 939 plasma samples: no lung cancer (870); lung cancer (69): stage: I (37); II to III (12); IV (19)	Plasma-based miRNA signatures from patients in two independent LDCT screening studies of 24 circulating miRNAs has diagnostic and prognostic performance
Hennessey et al. (2012) [49]	Phase I/II serum biomarker study: Training set: 30 NSCLC patients: AC (20), SQ (10); stage: I (10), II (9), III (10), IV (0); 20 healthy controls Test set: 55 NSCLC patients: AC (30), SQ (25); stage: I (33), II (13), III (7), IV (2); 75 healthy controls	Combination of miRNA-15b and miRNA-27b discriminate NSCLC from healthy controls
Markou et al. (2013) [50]	59 resected NSCLC and adjacent normal tissue: Training set: 19 tumor/normal tissues Test set: 40 tumor: AC (16), SCC (21), Other (3); stage: I (11), II-IV (29); lymph node: negative (23), positive (17); plasma samples from 37 NSCLC patients and 28 healthy donors	8 circulating plasma miRNAs (miRNA-21, miRNA-30d, miRNA-451, miRNA-10a, miRNA-30e-5p, miRNA-126*, miRNA-126, miRNA-145) were differential expressed in NSCLC; miRNA-520d, miRNA-489, miRNA-181b, miRNA-513, miRNA-26b, miRNA-189 and miRNA-520e were differentially expressed between AC and SQ
Wang et al. (2011) [51]	88 NSCLC patients: AC (37), SQ (21), other (30); stage: I-II (47), III (41); lymph node metastasis: No (53), Yes (35); 17 healthy individual	Serum miRNA-21 expression useful as a prognostic marker for NSCLC patients

* Indicates the antisense miRNA product.

Interestingly, miRNAs can also distinguish between the two lung cancer subtypes NSCLC and SCLC. For example, Du et al. (2010) [5] identified 41 differentially expressed miRNAs that effectively distinguish both subtypes, for instance miRNA-17-5p, miRNA-135, miRNA-103, miRNA-107, miRNA-301 and miRNA-338 were over-expressed in SCLC relative to NSCLC cells and have a potential to serve as potential biomarker in SCLC diagnosis. More interestingly, they observed that the known NSCLC associated miRNAs miRNA-29a/b/c, miRNA-24 and the oncogenic miRNA-21 and miRNA-221/222 are significantly down-regulated in SCLC. Similarly, Lee et al. (2011) reported that the expression of the seven miRNAs miRNA-21, miRNA-29b, miRNA-34a/b/c, miRNA-155 and let-7a are not related to SCLC patients [46].

Beside them, studies reported that miRNA signatures can differentiate the histological types of NSCLC. Lebanony et al. (2009) [47] identified miRNA-205 as highly specific marker for SQ to clearly distinguish it from AC which was independently validated by Bishop et al. (2010) [48] using resected NSCLC biopsies. Moreover, Landi et al. (2010) [30] identified that miRNA expression profiles significantly differentiated AC from SQ in which an altered expression of five miRNAs (miRNA-25, miRNA-34c-5p, miRNA-191, let-7e and miRNA-34a) strongly predicted survival for SQ.

As altered miRNA expression levels do not only correlate with patient relapse and survival but can also clearly distinguish between lung cancers, thus they might function as novel diagnostic biomarkers.

Several studies highlight miRNA expression signatures as non-invasive diagnostic tools for early detection and classification of lung cancer including first promising clinical results. For example, Hennessey et al. (2012) [49] reported from a phase I/II biomarker study the potential of a serum miRNA-15b/miRNA-27b pair as sensitive and effective biomarkers for the early NSCLC detection. Moreover, Sozzi et al. (2014) [13] described the clinical utility of blood based miRNA signatures which have a high potential to reduce false-positive rates of low dose CT (LDCT) screenings (3.7% compared to 19.7% for LDCT alone) as an additional diagnostic tool (MILD-trial). Similarly, Montani et al. (2015) developed the blood miRNA-test for identification of optimal patients for LDCT screening (COSMOS-trial; accuracy, sensitivity and specificity approximately 75%–78%) [14].

In the context of having been mainly developed for smokers as a first-line tool for LDCT screenings with *per se* high false-positive rates, these miRNA blood tests reduce unnecessary LDCT screenings in disease-free individuals of such high-risk cohorts. In case of a general population such tests appear probably not yet very beneficial and need further improvements and validation [13,14,49].

On the other hand, studies investigating the potential of miRNAs to differentiate between NSCLC patients and smokers have been started. For example, Zhu et al. (2016) [40] validated the diagnostic value of a four serum miRNA signature (miRNA-182, miRNA-183, miRNA-210 and miRNA-126) for early diagnosis which can also distinguish NSCLC patients from current smokers. This points to the usage of a combined biomarker signature for a better diagnostic value and an effective molecular characterization of lung cancer patients [37,45].

Nevertheless, miRNA based signatures have drawbacks and limitations: They depend on technical accuracy and proper validation, standardization and normalization schemes. miRNAs were detected either in tumor tissue or blood samples. However, miRNA levels are different in plasma and serum, and the overlap of miRNAs present both in blood and tissue is small [14,50]. This points out the challenge of miRNA signatures as a clinically reliable blood test. However, given the low overlap of miRNAs between different samples, there exists a common consensus of known miRNAs expressed in lung cancer, e.g., miRNA-21 and/or miRNA-30d, in which one of them is at least found across different studies and/or known to be expressed in tissue, plasma and serum [13,44,49–51]. Such statistically significantly expressed miRNAs were also independently validated by bioinformatics meta-analyses with miRNA-21 and miRNA-210 as most consistently reported miRNAs in lung cancer [52,53]. Thus, circulating blood signatures based on those miRNAs represent a high potential as an effective diagnostic tool for early lung cancer detection.

3. lncRNAs as Diagnostic Biomarkers

lncRNAs represent a large and diverse class of ncRNAs that play an important biological role. They are multi-exonic transcripts greater than 200 nucleotides and often less conserved among species [7,28,54,55]. lncRNAs show altered expression and regulate important biological processes and pathways associated with lung cancer pathogenesis and progression [15]. They have a complex regulatory effect ranging from transcriptional regulation (enhancer, chromatin modifier and transcription factor (TF) binding) and regulation of epigenetic processes (nuclear lncRNAs) to translational regulation through mRNA and protein binding as well as miRNA sponging (cytoplasm lncRNAs) [54–58] (see Figure 1). Thus, functional characterization is challenging and most lncRNAs are not well understood: currently, several thousand lncRNA transcripts have been annotated, however the majority is functionally uncharacterized [59–62].

As lncRNAs show a tissue- and disease-specific expression they emerged as potential diagnostic biomarkers [2,7]. lncRNAs are already known as promising biomarkers for different diseases, e.g., LIPCAR as biomarker for cardiac remodeling [63] or HULC for liver cancer [15,64]. Moreover, lncRNA based diagnostic tests such as the lncRNA prostate cancer antigen 3 (PCA3) as urinary biomarker for prostate cancer are already available for clinical use [65].

Therefore, lncRNAs are currently also under investigation for lung cancer, showing high potential for developing effective screening tests for diagnosis [66,67]. Table 2 summarizes different clinical lncRNA lung cancer studies including background information (number of patient, tumor stage, histological subtype; if available).

Table 2. Table summarizing the different lung cancer lncRNA studies.

Study (Ref.)	Patient Cohort	Important Reported Findings
Zhu et al. (2015) [29]	Meta-Analysis: 8 studies with 845 patients: NSCLC (2), colorectal cancer (1), gastric cancer (1), pancreatic cancer (2), clear cell renal cell carcinoma (1), osteosarcoma (1)	MALAT-1 serve as a molecular marker for cancer metastasis and prognosis
Ji et al. (2003) [68]	NSCLC patients (70): AC (26), SQ (34), LCC (10); stage: I (37), II (13), IIIA (20)	MALAT-1 and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer
Weber et al. (2013) [22]	45 NSCLC patients: AC (21): Stage: I/II (0),III/IV (21); SQ (24): I/II (3), III/IV (21); 25 controls	MALAT1 as complementary diagnostic biomarker in NSCLC
Yao et al. (2012) [12]	65 NSCLC patient serum: biopanning stage: tissue and serum samples (25): AC (12), SQ (12), LCC (1); stage: I (6), II (5), III (9), IV (5) identification stage: NSCLC serum samples (40): AC (17), SQ (22), LCC (1); stage: I (11), II (8), III (17), IV (4); 41 normal controls	A four serum biomarker SMOX, NOLC1, MALAT1 and HMMR show a high diagnostic accuracy for detecting early stage NSCLC
Qiu et al. (2014) [69]	NSCLC tissues/paired adjacent normal tissues	CCAT2 is an AC-specific lncRNA and promotes invasion of NSCLC; biomarker for lymph node metastasis.
Chen et al. (2016) [70]	SCLC tissues and cell lines	CCAT2 serves as an oncogenic lncRNA, and an independent unfavorable prognostic factor in SCLC patients
Liu et al. (2013) [71]	Tissues from 42 NSCLC/adjacent non-tumor lung patients: stage: I/II (25), III/IV (17); 4 NSCLC cell lines: AC (3; A549, SPC-A1, NCI-H1975); SQ (1; SK-MES-1); normal human bronchial epithelial cell line (1; 16HBE)	HOTAIR represent diagnostic biomarker of poor prognosis in NSCLC
Yang et al. (2013) [72]	A549 cells and cisplatin resistant A549/CDDP cells (microarray profiling of mRNAs, lncRNAs and miRNAs)	8 mRNAs (BMP4, CTSB, NKD2, BAG1, TGFBI, EGFR, JUN, CUL2), 8 lncRNAs (AK123263, CES1P1-001, RP3-508115.14, AK126698, TP53TG1, AC090952.4.1, uc003bgl.1, NCRNA00210) and 5 miRNAs (miRNA-17, miRNA-21, let-7i, miRNA-138, miRNA-194) potentially play a key role in cisplatin resistance; AK126698 appears to confer cisplatin resistance by targeting the Wnt pathway
Sui et al. (2016) [73]	465 AC patient RNA sequencing profiles (from TCGA); 53 AC patients	Correlation of AFAP1-AS1 and LINC00472 as potential biomarkers for diagnosis and prognosis
Tantai et al. (2015) [74]	64 NSCLC tissues/matched adjacent non-tumor patient tissues; stage: I (15), II/III (17)	Combination of XIST and HIF1A-AS1 had a higher positive diagnostic efficiency of NSCLC patient screening
Gong et al. (2016) [75]	498 lung cancer patients (467 patients at least two cycles of platinum-based chemotherapy); 213 healthy controls	HOTTIP, CCAT2, H19, HOTAIR, MALAT1 and ANRIL potential clinical biomarkers to predict lung cancer risk and platinum-based chemotherapy response
Yuan et al. (2016) [76]	Meta-analysis of eight published GWAS datasets with 17,153 cases and 239,337 controls	SNP rs114020893 of NEXN-AS1 at 1p31.1 may contribute to lung cancer susceptibility
Yang et al. (2014) [6]	5 NSCLC gene expression datasets from GEO: Training set: GSE27262, GSE19804, GSE19188 and GSE30219 Test set: GSE18842	47 lncRNAs differentially expressed in NSCLC; 19 lncRNAs differed in expression between SCC and AC
White et al. (2014) [7]	Three lung RNA-Seq datasets: 72 AC/adjacent normal pairs 55 AC/adjacent normal pairs + 243 unmatched tumors from TCGA 34 SQ/adjacent normal pairs + 163 unmatched tumors from TCGA	463 and 315 up- and down-regulated lncRNA in AC tumors relative to SQ; 27 lncRNAs differentially expressed between AC and SQ
Zhang et al. (2015) [2]	AC and SQ microarray	1646 differentially expressed lncRNA LINC01133 showed the largest up-regulation in SQ but not in AC
Wei et al. (2016) [1]	Paired tissue samples of RNA sequencing or microarray data from TCGA and GEO	lncRNA expression is different in AC and SQ knockdown of the up-regulated lncRNA AFAP1-AS1 and LINC00511 impaired AC cell proliferation; knockdown of PVT1 inhibited SQ cell growth four 6-lncRNAs signature expression patterns were found to be significantly associated with AC and SQ patient overall and progression-free survival

The most promising lncRNA candidate is the metastasis-associated lung AC transcript 1 (MALAT1) which is highly expressed e.g., in lung and pancreas. MALAT1 is functionally well-characterized and known to be a prognostic marker for early-stage NSCLC as well as cancer metastasis [29,68]. Moreover, MALAT1 has been shown to serve as blood-based biomarker for the early detection of NSCLC. For example, Yao et al. (2012) identified a four serum biomarker signature containing MALAT1 and three proteins that show a high diagnostic accuracy for detecting early stage NSCLC [12]. Similarly, Weber et al. (2013) reported MALAT1 as non-invasive and effective diagnostic biomarker for NSCLC diagnosis [22]. As the authors observed a low sensitivity, they suggest that MALAT1 might not be a single biomarker but applicable as a complementary biomarker [22]. The lung AC-specific lncRNA colon cancer-associated transcript 2 (CCAT2) displays altered expression, promotes invasion of NSCLC and can serve as a biomarker for lymph node metastasis [69]. The lncRNA HOX transcript antisense intergenic RNA (HOTAIR) functions as gene expression repressor through recruitment of chromatin modifiers and correlates with metastasis and poor prognosis in NSCLC [71,77].

Several lncRNAs show a co-expression in lung cancer patients. For example, Yang et al. (2013) validated the differential expression of 8 mRNAs, 8 lncRNAs and 5 miRNAs in NSCLC cells, in which co-expression between genes (e.g., FN1, CTSB, EGFR and NKD2), lncRNAs (e.g., BX648420, ENST00000366408 and AK126698) and miRNAs (e.g., miRNA-26a and let-7i) were identified, also playing a potential key role in cisplatin resistance [72]. More recently, Sui et al. (2016) confirmed a correlation between AFAP1-AS1 and LINC00472 and clinical features in AC patients [73].

Further lncRNAs known to be associated with different cancer types are also under investigation as diagnostic biomarker in lung cancer. For instance, Tantai et al. (2015) showed that the lncRNAs XIST and HIF1A-AS1 have a significantly increased level in tumor tissues or serum from NSCLC patients, highlighting a clinical significance as effective diagnostic screening for NSCLC when combining both lncRNAs [74].

lncRNAs were also investigated regarding their potential as clinical biomarkers to predict lung cancer risk and treatment response. In this context, studies of Gong et al. (2016) found that genetic polymorphisms of well-characterized lncRNAs such as CCAT2, HOTAIR and MALAT1 were significantly associated with lung cancer susceptibility and platinum-based chemotherapy response, indicating that they might function as clinical biomarkers [75]. Furthermore, Yuan et al. (2016) found in a large-scale meta-analysis of 690,564 single-nucleotide polymorphism (SNPs) in 15,531 autosomal lncRNAs a genetic SNP risk locus (1p31.1) in the lncRNA NEXN-AS1 which influence the secondary structure and is statistically associated with lung cancer risk [76]. It could thus serve as potential risk biomarker for lung cancer diagnosis. The potential of lncRNAs as diagnostic biomarkers was also confirmed by several meta-analyses with MALAT1 and the human urothelial carcinoma associated 1 (UCA1) as most promising candidates in lung cancer patient [29,78–80].

In addition, altered lncRNA expression levels might also accurately distinguish between AC and SQ and predict the clinical outcome for both NSCLC subtypes [1,2,6,7]. White et al. (2014) characterized 567 RNA-Seq datasets from AC and SQ tumors and found 463 and 315 up- and down-regulated lncRNA, respectively, in AC tumors relative to SQ. Moreover, they reported that 27 lncRNAs were differentially expressed between AC and SQ that can potentially serve as important biomarkers for lung cancer diagnosis [7]. Furthermore, studies of Zhang et al. (2015) identified 1646 differentially expressed lncRNA transcripts, in which the lncRNA LINC01133 showed the largest up-regulation in SQ but not in the AC samples and correlates with shorter survival time [2]. Recently, Wei et al. (2016) reported four six-lncRNA signature patterns that are significantly associated with AC and SQ patient survival [1]. More interestingly, the authors also demonstrated that knockdown of the up-regulated lncRNA AFAP1-AS1 and LINC00511 impaired AC cell proliferation, while knockdown of PVT1 inhibited SQ cell growth [1].

However, to the best of our knowledge, currently no study has been directly carried out to identify lncRNA expression signatures that can differentiate NSCLC and SCLC lung cancer subtypes.

An explanation might be that investigating lncRNAs in lung cancer is a relatively new research field and therefore only very few lncRNAs are well-characterized. Moreover, lncRNA studies focused primarily on NSCLC as the more commonly diagnosed subtype, but also mainly on lncRNAs that were already reported from other cancer types, e.g., the colon cancer lncRNA CCAT2. However, there exist reports on lncRNA characterization in different lung cancer subtypes. For example, Qiu et al. (2014) identified that the lncRNA CCAT2 shows a specific expression in AC and might function as biomarker for lymph node metastasis [69], whereas Chen et al. (2016) recently found that CCAT2 serves also as an independent unfavorable prognostic factor in SCLC patients [70]. In this regard, it would be of high importance for the future to investigate differentially expressed lncRNAs between NSCLC and SCLC in order to develop a diagnostic biomarker signature that can accurately distinguish between both lung cancer subtypes.

4. Bioinformatics Databases, Algorithms and Analysis Tools

The number of annotated miRNAs and lncRNAs is growing dramatically, however experimental characterization is challenging [15,59,81,82]. Bioinformatics analysis can contribute to a comprehensive functional understanding of miRNAs and lncRNAs. Instead of discussing experimental methods (several experimental methods are reviewed in [83]), in this section we will discuss several databases and bioinformatics algorithms. A summary of them is given in Tables 3–5. It is worth mentioning that all the presented databases and tools are also applicable to other tumors and diseases.

Table 3. Databases and software packages for ncRNA research.

Tool (Ref.)	Purpose	Website
Rfam [84]	ncRNA database	http://rfam.xfam.org/
Ensembl	genome browser	http://www.ensembl.org/
UCSC	genome browser	https://genome.ucsc.edu/
BLAST [85]	Sequence analysis	http://blast.ncbi.nlm.nih.gov/Blast.cgi
RNAfold [86]	Folding prediction	http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi
Mfold [87]	Folding prediction	http://unafold.rna.albany.edu/?q=mfold
RNAalifold [88]	Folding prediction	http://rna.tbi.univie.ac.at/cgi-bin/RNAalifold.cgi
FOLDALIGN [89]	Folding prediction	http://rth.dk/resources/foldalign/server/index.html
LocARNA [90]	Folding prediction	http://rna.informatik.uni-freiburg.de/LocARNA/Input.jsp
RNAshapes [91]	Folding prediction	https://bibiserv2.cebitec.uni-bielefeld.de/rnashapes
4SALE [92]	Folding prediction	http://4sale.bioapps.biozentrum.uni-wuerzburg.de/
GO database [93]	Functional classification	http://geneontology.org/page/go-database
AmiGO [94]	Functional analysis	http://amigo.geneontology.org/amigo
Panther [95]	Functional analysis	http://www.pantherdb.org/
Reactome [96]	Interactions/pathways	http://www.reactome.org/
KEGG [97]	Interactions/pathways	http://www.genome.jp/kegg/
WikiPathways [98]	Interactions/pathways	http://wikipathways.org/index.php/WikiPathways
Cytoscape [99]	Visualization/Functional analysis	http://www.cytoscape.org/
starBase v2.0 [100]	Functions/Interactions/networks	http://starbase.sysu.edu.cn/
TRANSFAC [101]	Promotor analysis	http://www.gene-regulation.com/pub/databases.html
JASPAR [102]	Promotor analysis	http://jaspar.genereg.net/
Allgen PROMO [103]	Promotor analysis	http://algggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3
MatInspector [104]	Promotor analysis	https://www.genomatix.de/online_help/help_matinspector/matinspector_help.html
miRanda [105]	Target prediction	http://www.microrna.org/microrna/home.do
RNAup [106]	Target prediction	http://rna.tbi.univie.ac.at/cgi-bin/RNAup.cgi
IntaRNA [107]	Target prediction	http://rna.informatik.uni-freiburg.de/IntaRNA/Input.jsp
RNAcentral [108]	ncRNA sequence database	http://rnacentral.org/

Table 4. Databases and software packages for miRNA research.

Tool (Ref.)	Purpose	Website
miRBase [109]	miRNA database	http://www.mirbase.org/
MiR2Disease [110]	Interactions/pathways	http://www.mir2disease.org/
TargetScan [111]	Target prediction	http://www.targetscan.org/vert_71/
PicTar [112]	Target prediction	http://www.pictar.org/
PITA [113]	Target prediction	https://genie.weizmann.ac.il/pubs/mir07/mir07_data.html

Table 5. Databases and software packages for lncRNA research.

Tool (Ref.)	Purpose	Website
LncRBase [60]	lncRNA database	http://bicesources.jcbose.ac.in/zhumur/lncbase/
LNCipedia [59]	lncRNA database	http://www.lncipedia.org/
LncRNADisease [114]	Interactions/pathways	http://www.cuilab.cn/lncrnadisease
Rtools [115]	Interactions/pathways	http://rtools.cbrc.jp/cgi-bin/RNARNA/index.pl
LncTar [116]	Interactions	http://www.cuilab.cn/lncstar
NPInter [117]	Interactions	http://www.bioinfo.org/NPInter/
PDB [118]	Interactions	http://www.rcsb.org/pdb/home/home.do
NDB [119]	Interactions	http://ndbserver.rutgers.edu/
BioGRID [120]	Interactions	https://thebiogrid.org/
IntAct [121]	Interactions	http://www.ebi.ac.uk/intact/
PRD [122]	Interactions	http://pri.hgc.jp/
RPIntDB [82]	Interactions	http://pridb.gdcb.iastate.edu/RPISeq/RPIntDB.html
iMEX [123]	Interactions	http://www.imexconsortium.org/
UniProt [124]	Interactions	http://www.uniprot.org/
PRIDB [125]	Interactions	http://pridb.gdcb.iastate.edu/
DrumPID [126]	Interactions/pathways	http://drumpid.bioapps.biozentrum.uni-wuerzburg.de/compounds/index.php
catRAPID [127]	Interaction prediction	http://s.tartagliab.com/page/catrapid_group
RPISeq [82]	Interaction prediction	http://pridb.gdcb.iastate.edu/RPISeq/
Pprint [128]	Interaction prediction	http://www.imtech.res.in/raghava/pprint/
KYG [129]	Interaction prediction	http://cib.cf.ocha.ac.jp/KYG/
Struct-NB [130]	Interaction prediction	http://www.public.iastate.edu/~ftowfic
PRINTR [131]	Interaction prediction	http://210.42.106.80/printr/
lncRNator [132]	Functions/Interactions/networks	http://lncrnator.ewha.ac.kr/index.htm

4.1. Bioinformatics Databases

Popular databases such as Rfam [84] represent information about several RNA families including sequence and consensus secondary structure information, whereas LNCipedia [59], LncRBase [60] and miRBase [109] provide information about specific families including further information, e.g., about experimental data, tissue expression and targets. On the other hand, databases such as MiR2Disease [110] and LncRNADisease [114] focus on the disease and interaction specific context based on the literature and/or experimentally disease data. All these databases provide information about miRNAs and lncRNAs and allow a fast overview regarding their sequence, structure and functional role. However, most miRNAs and lncRNAs are newly detected without concrete knowledge about their functional role which requires integrated bioinformatics analysis for comprehensive understanding. In this context, they should combine phylogenetic sequence-structure conservation analysis with functional interaction partner, biological process and pathway as well as promoter analysis (see Figure 2).

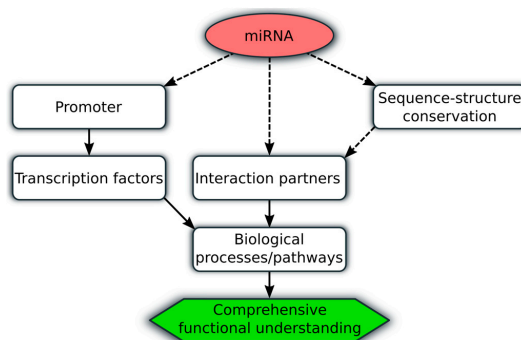


Figure 2. Workflow for integrated bioinformatics functional analysis of miRNAs and lncRNAs. Illustration of integrated bioinformatics analysis of ncRNAs (miRNAs, lncRNAs; red circle) which should focus on the sequence, structure, promoter and interaction partner prediction combined with functional analysis (rectangles). Dashed arrows represent the three main analysis steps (e.g., promoter analysis), whereas continuous arrows show the subsequently functional analysis step using the obtained results from the previous steps (e.g., transcription factors) to get a comprehensive functional understanding of ncRNAs (green hexagon).

4.2. Phylogenetic Sequence-Structure Conservation

Sequence data are available from the Ensembl (<http://www.ensembl.org/>) and UCSC genome browser. The sequence can first be analyzed using the BLAST algorithm [85] to find homologous sequences among mammalian species, e.g., human, chimpanzee and mouse. Resulting sequences can be further analyzed for sequence and structure conservation using bioinformatics secondary prediction algorithms. Dynamic programming algorithms such as the Zuker algorithm which are implemented in RNAfold and Mfold calculate for a sequence the thermodynamic optimal secondary structure based on a minimum free energy. These algorithms calculate accurately the optimal secondary structure, they are however not useful for a large-scale application with several sequences due to the high calculation time [86,87,133,134]. A more effective approach represents the Sankoff algorithm which can simultaneously align and fold multiple sequences [88,133,135–137]. Programs using the Sankoff algorithm are for instance RNAalifold [88], FOLDALIGN [89] and LocARNA [90], in which a pairwise sequence alignment is generated and subsequently aligned and folded to calculate the optimal conserved secondary structure [135]. However, further extensions of the Sankoff algorithm are more efficient, allowing a faster calculation, e.g., the RNASHAPES program has a linear calculation time based on a non-heuristic approach instead of being exponential (Sankoff algorithm) [91,138]. Another useful program for large sequence and structure data sets is the 4SALE algorithm which allows a fast sequence and synchronous secondary analysis including further analyses and manual editing [92].

4.3. Functional Interaction Partner Analysis

miRNAs and lncRNAs show enormous clinical potential but currently have some limitations, as exact knowledge about the functional interaction context is necessary. Experimental methods for RNA-target detection range from quantitative proteomic analysis and high throughput experiments such as tissue-specific microarray and RNA-Seq analysis to UV cross-linking immunoprecipitation-high-throughput sequencing (CLIP-Seq) [33,34,83,139]. Such methods are essential for correct identification of interaction partners, but technically challenging, time and cost intensive, e.g., interaction analysis considering the whole interactome is laborious and methods such as CLIP-Seq require a specific target RNA or protein [82,83,115,140]. Over the past decades, several bioinformatics interaction prediction algorithms have been developed which are helpful e.g., for large-scale application and for filtering and pre-selection of candidates for further experimental testing. However, RNA–RNA interaction prediction is challenging due to high combinatorial number of RNA pairs, time complexity of calculating the joint secondary structures of both RNA molecules as well as knowledge of the intra-molecular and inter-molecular base-pairs interactions between both RNA molecules [115]. In the following we will discuss computational databases and prediction algorithm for RNA interactions analysis (extensively reviewed in [33,115,139,141–143]).

4.4. miRNA Target Prediction Algorithms

miRNAs bind mRNA by seed region matching which can be predicted by bioinformatics algorithms. Numerous prediction algorithms were developed which are mainly based on seed region similarity, but new approaches also include sequence matching combined with structure and/or thermodynamic parameters (folding energy) or target site accessibility [33,34,107,144]. The popular TargetScan algorithm [111] predicts target interactions based on conserved seed region matching, whereas miRanda [105] and PicTar [112] algorithms allow seed region mismatches but include free folding energy, and the PITA algorithm [113] includes target site accessibility for seed interaction prediction [139,145–147]. In addition, several other interaction prediction algorithm such as RNAup (target site accessibility [106]) and IntaRNA (seed region and target site accessibility [107]) exist which focus mainly on specific RNA–RNA interactions such as miRNA–mRNA or bacterial small RNAs, but can also be used for other RNA types.

It is worth mentioning that such mRNA target prediction algorithms have some drawbacks and limitations which requires a carefully compliance by users. For instance, most algorithms are not validated by experimental data, they do not include tissue-specific miRNA expression, and are based on different parameters, resulting in a less target overlap between them [33,34,139,144]. Moreover, target predictions often shows large number of false positives, e.g., miRanda has a approximated false-positive rate between 24%–39%, whereas TargetScan shows 22%–31%, PicTar 30% and PITA 20% [113,148,149]. The false positive rate reflects in this case all erroneously predicted target interactions that have not been experimentally validated. Thus, this parameter is of high interest as it gives an approximation of how specific and usable the algorithm is as this is important for computational target interaction prediction especially without available experimental supported results [139].

Beside these facts, several studies highlighted that bioinformatics prediction algorithms are beneficial as an additional tool for experiments [150–152]. In this context, studies demonstrated that algorithms using high seed region similarity represent the highest accuracy and overlap of predicted and experimentally validated targets [33,34,150], e.g., TargetScanS (requires perfect complementarity), miRanda and PicTar seem to be the effective methods with a sensitivity nearly 65% with experimentally validated interactions, but miRanda algorithm predicts much more mRNA targets [139,148].

Similarly, Busch et al. (2008) compared several algorithm and demonstrated that IntaRNA shows the highest accuracy, whereas RNAup shows an overall accuracy closest to IntaRNA (average sensitivity of IntaRNA is 0.783, RNAup 0.752) [107]. Therefore, users should use different algorithms and their overlapping targets in combination with experiments to find the best mRNA interaction partners in a functional and tissue-specific manner [33,34,139].

4.5. lncRNA–RNA Interaction Prediction Algorithms

Due to the high computational costs and the fact that lncRNAs are a new research field, methods for predicting lncRNA–RNA interactions are limited [115]. To the best of our knowledge, there are currently only two tools available which allow direct lncRNA–RNA interaction prediction. The Rtools pipeline calculates interactions between lncRNA–lncRNA and lncRNA–mRNA considering seed matching and target site accessibility by combining different existing algorithms, e.g., IntaRNA, to reduce the computation complexity. The interactions between lncRNAs and the human transcriptome were calculated, but were also validated using different experimentally RNA–RNA interaction datasets [115]. According to the authors, the pipeline is comparable with existing algorithms but significantly faster, however it has room for further improvements to reach higher accuracy especially for lncRNA interaction predictions [115]. On the other hand, the LncTar software allows calculation of lncRNA–RNA interactions in large-scale datasets [116]. It assumes that the primer-dimer detection is not only important for real-time polymerase chain reaction (PCR) design, but also an important process of base pairing in nature, thus enabling prediction of lncRNA–RNA interactions. Therefore, the software uses a modified exact melting-temperature and primer-dimer prediction algorithms of the PerlPrimer [153] code, a developed platform for real-time PCR primer design [116]. It was demonstrated by the authors that LncTar efficiently predicts lncRNA–RNA interaction partners with highly accuracy, which was further validated by experimentally lncRNA–mRNA interactions curated from the LncRNADisease and NPInter databases [116]. However, it has currently some limitations which need further improvements, e.g., it did not take the stacking base pairs and loop energy for searching the stable joint secondary structures and the RNA tertiary structure into account which are known to play a role in RNA–RNA interactions [116].

miRNAs potentially sponged by lncRNAs can be predicted e.g., using the above described miRanda algorithm [105]. As the RNAup [106] and IntaRNA [107] algorithms can be used for several RNA types, they can also be applied for lncRNA–mRNA interaction prediction. However, they are not efficient for large-scale prediction of lncRNA targets and/or for larger lncRNAs and mRNA molecules, e.g., due to sequence length limitation (IntaRNA ≤ 2 kb, RNAup ≤ 5 kb) [107,116] which requires carefully consideration of their specific features and further steps such as locally software use by users.

For more information and a detailed overview about several *in silico* approaches for functional prediction and mechanistic characterization of lncRNAs please see to the recent review by Signal et al. (2016) [62].

4.6. RNA-Protein Interaction Prediction Algorithms

lncRNAs regulate not only RNAs, but also interact with proteins, pointing out the importance of an integrated bioinformatics analysis of potential interaction partners for a tissue and disease-specific functional characterization of lncRNAs. Popular databases such as NPInter [117], the Protein Data Bank (PDB) [118] and Nucleic Acid Database (NDB) [119] include information about the experimentally determined structure of proteins and nucleic acids as well as RNA-protein complex assemblies, whereas BioGRID [120] and IntAct [121] contain protein-protein and RNA-protein interactions from several organisms. Furthermore, several approaches combine different databases, thus allowing a comprehensive overview of interactions and further individual analysis. For example, the RPIIntDB [82], PRD [122], iMEX [123] and UniProt [124] databases provide functional information and experimentally interactions curated e.g., from BioGRID [120] and IntAct [121], whereas the Protein-RNA Interface Database (PRIDB) [125] collects interactions based on PDB [141]. In addition, the DrumPID database [126] which was developed by our group, focusses especially on the drug-target interaction context combining interaction and pathway data, but also allows organism- or tissue-specific analyses. Additional features include structural and sequence domain analyses of proteins and RNAs which help in detecting functional interaction and recognition binding-sites such as the RNA recognition motif and RNA-binding domain in proteins [126,141,154].

As these databases mainly contain experimentally validated RNA-protein interactions, they are not applicable especially for newly annotated lncRNAs. To support this, several bioinformatics prediction algorithms were developed that focus on the sequence and/or structure by using different machine learning algorithm. For example, the software “fast predictions of RNA and protein interactions and domains at the Center for Genomic Regulation, Barcelona, Catalonia” (catRAPID) [127] performs predictions based on a sequence HMMscan (probabilistic statistical profile Hidden Markov Model (HMM)) using propensities of individual residues from PDB [127,141,142]. The RPISeq software calculates interactions using different machine learning classifiers [82,141,142], whereas Pprint predicts RNA binding sites using evolutionary position-specific scoring matrix (PSSM) information combined with a support vector machine (SVM) [128,155]. The KYG algorithm focuses on the structure and calculates binding regions by applying a position-specific multiple sequence profile on protein-RNA structures from PDB, also without biochemical or functional data [129,155]. Other algorithm like Struct-NB combine sequence and structural-based information of protein-RNA complex interfaces from PDB using a Gaussian Naive Bayes classifiers machine learning algorithm [130], whereas PRINTR calculates interactions using a SVM and PSSM [131,142].

All these algorithms perform high-confidence lncRNA-protein interaction predictions and are helpful to find potential RNA-protein interaction partners [82,141]. Moreover, they reveal a high accuracy, which was validated by independent training datasets including experimentally validated physical interactions [127,129,142,155], e.g., from the NPInter server [117]. Most prominently, catRAPID correctly calculates the interaction of the human lncRNAs XIST and HOTAIR with the Polycomb Repressive Complex 2 (PRC2) but also the interaction between HOTAIR and Suz12 predicted by RPISeq and catRAPID were in agreement with experimental data [82,127,141,142,156].

Nevertheless, these prediction algorithms also have some limitations. For instance, most of them do not consider the tissue and functional specific context of lncRNA-protein interactions and often show large number of interaction predictions [127,141,157]. Moreover, most of them were not systematically validated on general benchmark datasets, depend on different approaches and the prediction outcome is affected by the distance threshold for the interface residue definition [141,155,157]. In addition, several groups evaluated the influence of different machine learning classifiers and found out that the accuracy of sequence-based methods can be increased by using PSSM profiles [128,131,155,158].

For example, Walia et al. (2012) analyzed several sequence and simple structure-based with complex structure-based algorithms and reported that results are comparable between these approaches [155]. However, sequence-based methods using PSSM classifiers achieve comparable results to state-of-the-art structure-based methods, but the latter ones reach higher specificity compared to exclusively sequence-based approaches [155]. Thus, sequence-based approaches can effectively predict RNA–protein interactions, but a higher accuracy can be reached when using PSSM profiles and/or structure-based methods [128,155]. Furthermore, structure-based features also allow identifying substrate-binding clefts and how the RNAs and proteins specifically recognize each other but often show a higher degree of irregularity at the surface compared to non-interface residues. However, they require information on protein–RNA complexes as training templates, which are often limited [131,141,142,155,159].

As parameters and outputs differ between sequence and structure-based approaches, for a large-scale benchmark application and prediction of unknown RNA–protein interactions it is therefore of importance to compare different methods but also how to use them and to interpret the results. In this context, it is essential to have a detailed knowledge about the used features and datasets of each interaction prediction algorithm, the evaluation process and the definition of interface residues, e.g., background data and validation and usage of PSSM profiles [155].

4.7. Functional GO and Pathway Enrichment Analysis

Functional enrichment analysis of interaction partners and related regulatory networks is essential for understanding its tissue-specific and functional role and can boost the selection of best candidates for further experimental validation. There exist several databases and software tools for data mining and analysis of large gene list. The most prominent annotation is the Gene Ontology (GO) Consortium project which functionally specifies genes/proteins and their relationships in categories, so-called GO terms, regarding molecular function, cellular component and biological processes (including pathways) in a species-independent manner [93,94]. All annotation data can be downloaded or accessed online from the GO database or through the web-based application database AmiGO which also support a term enrichment analysis for user input lists using Panther [93,94]. The Protein ANalysis THrough Evolutionary Relationships (Panther) classification is a large database collection of protein families that are divided in functional categories using statistical HMMs, allowing functional analysis and classification of large gene lists and/or experimental datasets in significantly enriched ontology and pathway terms [95]. Other popular databases are Reactome, Kyoto Encyclopedia of Genes and Genomes (KEGG) and WikiPathways, enabling also analysis of signaling pathways including tissue analysis, e.g., also from gene lists and/or large-scale expression data sets [96–98,160]. Recently, Herwig et al. (2016) developed ConsensusPathDB which allows a functional and network-based characterization of biomolecules from a user input list and/or experimental high-throughput datasets such as RNA-seq [161]. For this, molecular interaction data from 32 different online available repositories from human, mouse and yeast were integrated, in which calculations of statistical significant over-represented and enriched interaction network modules and biochemical pathways are based on different computational analyses.

Limitations and drawbacks are, for instance, large output lists and over-prediction especially from large-scale gene lists. Moreover, results are not filtered for a specific biological process and pathway, and the programs do not consider interactions between genes/proteins and the interaction network in a cell-line and tissue-specific context [33,162,163]. Therefore, an individual collection by users of genes, proteins, processes and signaling pathways associated with lung cancer as well as tissue-specific information will specify the results from the functional enrichment analysis which allows a better functional interpretation in a biological context.

4.8. Bioinformatics Tools for Integrated Functional Analysis

There are several bioinformatics tools for integrated functional analysis and interpretation, e.g., co-expression, disease and tissue-specific analysis, enabling to comprehensively understand the functional role of miRNAs and lncRNAs from large input data lists. For example, the starBase web tool deciphers lncRNAs and miRNAs from experimental large-scale CLIP-Seq datasets and tumor samples and provides RNA-protein, miRNA-lncRNA and miRNA-mRNA interactions including further analysis, e.g., biological processes and signaling pathways [100]. Similarly, the lncRNAtor tool is especially designed to investigate and functional understand lncRNAs combining for instance lncRNA expression profiles, RNA-protein interactions and functional enrichment analysis by using RNA-Seq and CLIP-Seq data sets from publicly available databases such as The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO) and ENCODE [132]. Moreover, the Cytoscape software tool allows to visualize and analyze regulatory networks, e.g., regarding functional GO terms and the network topology, co-expression or identification of functional clusters of highly connected interaction partners [99,164,165]. As an example, the Cytoscape plugin ClueGO calculates statistically enriched processes and pathways from a user gene list using GO terms and information from KEGG, WikiPathways and Reactome [166]. Moreover, the ncINetView plugin integrates data from the ncRNA-DB, a database collection of ncRNA interactions from several sources, allowing to search for associated interaction partners and regulatory networks including related biological functions and diseases as well as filtering for a specific diseases [167].

However, current bioinformatics tools only provide information about already known ncRNAs as they mainly analyzed public available large-scale datasets or focus on a specific disease and/or ncRNA class. Nevertheless, for specific analyses, e.g., interaction partners and functional enrichment analysis, there are powerful tools available such as ConsensusPathDB [161] on which further analyses (e.g., promoter, structure as discussed here) can build.

4.9. Promoter Analysis

Promoter analysis is an essential step in understanding the complex regulatory effects of ncRNAs, for instance transcriptional regulation of miRNA and lncRNA expression or in case of lncRNAs, also cooperatively working with a TF to guide them to the promoter e.g., to regulate its own transcription in a feedback loop, both helpful in terms of posttranscriptional therapeutically usage. TFs bind to specific transcription factor binding sites (TFBSs) in the promoter that can be bioinformatically represented by a IUPAC (International Union of Pure and Applied Chemistry) consensus nucleotide sequence or a position weight matrices (PWMs), whereas the latter one reflects a better representation by displaying the whole nucleotide distribution for each binding site position (extensively reviewed in [168]). Databases using PWMs are TRANSFAC, JASPAR, Allgen PROMO and MatInspector (implemented in the Genomatix software (Genomatix GmbH, Munich, Germany)). These tools allow not only prediction of putative TFBSs for a given sequence but also additional analysis, e.g., genome-wide and comparative regulatory region analysis [101–104].

Limitations and drawbacks are for instance, most methods are using different parameters for TFBS detection, they are mostly not based on experimental TFBS profiles and do not consider the tissue and functional context of the TF. Moreover, they are using different output parameters, e.g., dissimilarity threshold which controls how similar a sequence must be to the matrix to be reported as a hit, and do not include multiple statistical testing, indicating a high number of over-predictions. Thus, knowledge about the parameters is essential but also the combination of different prediction software to find overlapping hits. Moreover, analysis regarding the tissue and functional context can improve the detection of potentially functional TFBSs and minimize potential testing candidates which at least need to be further proven by experiments [104].

4.10. Automation

As discussed in the previous sections, analyzing ncRNAs is a complicated process involving a lot of different resources, e.g., various databases and prediction tools with different specializations. Especially considering the goal of using circulating miRNA and lncRNA biomarkers in the clinic, solutions for a more automated sample analysis are important. Partial automation has been achieved [115], where multiple prediction algorithms are combined. This already reduces the complexity of the analysis and improves its quality. A similar approach integrates databases, e.g., the online database RNAcentral [108] which provides a unified way to access various previously discussed databases not only using a unified API, but including a comprehensive versioning system which makes the data analysis reproducible through the use of stable identifiers. Nevertheless, complete solutions do not yet exist, but have to be developed using custom scripts or using flexible pipelines like Galaxy [169], Ruffus [170] or Snakemake [171]. Considering the rapidly evolving miRNAs and lncRNA annotations and analysis tools, reproducibility and versioning of the analysis pipeline is important. Pipeline specific [172] as well as generic approaches using Docker [173] have been proposed.

5. Conclusions and Future Directions

miRNAs and lncRNAs have a high potential working as non-invasive biomarker in lung cancer diagnosis (see Figure 3 for a summary of key points presented in this review). First large-scale blood biomarker signature studies based on miRNAs and lncRNAs show promising clinical results for NSCLC early diagnosis. However, studies to understand the implication and potential of such circulating biomarker signatures in lung cancer have just begun. There are major challenges for the transfer to the clinic, e.g., accuracy, reliability, and well established validity. Thus, further investigations and validation studies are required, identifying highly sensitive tests for efficient RNA-based lung cancer diagnostics. Regarding their functional characterization, experimental methods are technical challenging and laborious, but can be supported by integrated bioinformatics analysis for filtering and pre-selection of experimental candidates considering the tissue and functional interaction context. Especially lncRNA characterization is challenging and functional understanding of their role is limited which requires further studies regarding the diagnostic potential. Blood-based tests must be highly accurate and face substantial hurdles [15,22,174], however, such biomarker signatures have enormous potential as a first-line diagnostic tool. Despite its proven potential to reduce false-positive rates of LDCT screenings as an additional diagnostic tool, further refinement and evaluation of blood signatures are necessary. In this context, the use of tissue (e.g., from surgical NSCLC biopsies and/or GEO) and blood samples, examination of several cohorts (e.g., low- and high-risk lung cancer patients), expression normalization and analysis using different normalization strategies as well as bioinformatics meta-analysis will be essential to improve lung cancer diagnostics and application in the clinic in future. This will not only refine the blood biomarker signature but can also reduce the complexity of such blood tests, e.g., due to the identification of a combination of two or three differentially expressed miRNAs and lncRNAs. All this will contribute to better accuracy and less costs, resulting in higher number of correct and early diagnosed individuals as well as the reduction of unnecessary invasive diagnostics. The vision for the future is that circulating miRNAs and lncRNAs based blood tests can be transferred to the clinic for a better clinical management of lung cancer resulting in a better patient survival.

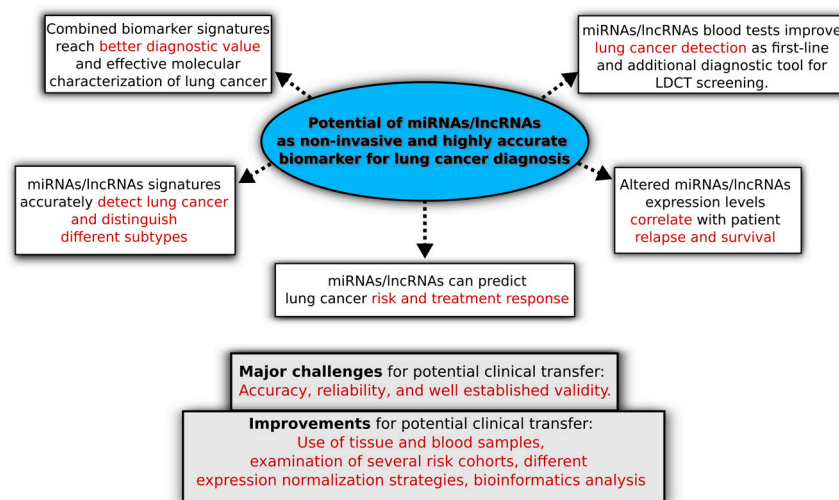


Figure 3. Schematic overview of key points of the potential of miRNAs and lncRNAs as non-invasive biomarker in lung cancer diagnosis. The Figure summarizes key points of the potential of miRNAs and lncRNAs as non-invasive biomarker in lung cancer diagnosis, current challenges and useful improvements for a clinical transfer in the future.

Acknowledgments: We thank IZKF (grant number BD-247; old project, preparatory work) and BMBF (grant number 031L0129B; consecutive new project on the topic) for funding and George Leigh for stylistic corrections.

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

References

- Wei, Y.; Zhang, X. Transcriptome analysis of distinct long non-coding RNA transcriptional fingerprints in lung adenocarcinoma and squamous cell carcinoma. *Tumour Biol.* **2016**. [CrossRef] [PubMed]
- Zhang, J.; Zhu, N.; Chen, X. A novel long noncoding RNA LINC01133 is upregulated in lung squamous cell cancer and predicts survival. *Tumour Biol.* **2015**, *36*, 7465–7471. [CrossRef] [PubMed]
- Herbst, R.S.; Heymach, J.V.; Lippman, S.M. Lung Cancer. *N. Engl. J. Med.* **2008**, *359*, 1367–1380. [CrossRef] [PubMed]
- GLOBOCAN project. Available online: <http://globocan.iarc.fr> (accessed on 13 August 2015).
- Du, L.; Schageman, J.J.; Irnov, G.L.; Hammond, S.M.; Minna, J.D.; Gazdar, A.F.; Pertsemlidis, A. MicroRNA expression distinguishes SCLC from NSCLC lung tumor cells and suggests a possible pathological relationship between SCLCs and NSCLCs. *J. Exp. Clin. Cancer Res.* **2010**. [CrossRef] [PubMed]
- Yang, J.; Lin, J.; Liu, T.; Chen, T.; Pan, S.; Huang, W.; Li, S. Analysis of lncRNA expression profiles in non-small cell lung cancers (NSCLC) and their clinical subtypes. *Lung Cancer* **2014**, *85*, 110–115. [CrossRef] [PubMed]
- White, N.M.; Cabanski, C.R.; Silva-Fisher, J.M.; Dang, H.X.; Govindan, R.; Maher, C.A. Transcriptome sequencing reveals altered long intergenic non-coding RNAs in lung cancer. *Genome Biol.* **2014**. [CrossRef] [PubMed]
- Taniwaki, M.; Daigo, Y.; Ishikawa, N.; Takano, A.; Tsunoda, T.; Yasui, W.; Inai, K.; Kohno, N.; Nakamura, Y. Gene expression profiles of small-cell lung cancers: molecular signatures of lung cancer. *Int. J. Oncol.* **2006**, *29*, 567–575. [CrossRef] [PubMed]
- Taneja, T.K.; Sharma, S.K. Markers of small cell lung cancer. *World J. Surg. Oncol.* **2004**. [CrossRef] [PubMed]
- Bangur, C.S.; Switzer, A.; Fan, L.; Marton, M.J.; Meyer, M.R.; Wang, T. Identification of genes over-expressed in small cell lung carcinoma using suppression subtractive hybridization and cDNA microarray expression analysis. *Oncogene* **2002**, *21*, 3814–3825. [CrossRef] [PubMed]
- Coate, L.E.; John, T.; Tsao, M.S.; Shepherd, F.A. Molecular predictive and prognostic markers in non-small-cell lung cancer. *Lancet Oncol.* **2009**, *10*, 1001–1010. [CrossRef]

12. Yao, Y.; Fan, Y.; Wu, J.; Wan, H.; Wang, J.; Lam, S.; Lam, W.L.; Girard, L.; Gazdar, A.F.; Wu, Z.; et al. Potential application of non-small cell lung cancer-associated autoantibodies to early cancer diagnosis. *Biochem. Biophys. Res. Commun.* **2012**, *423*, 613–619. [[CrossRef](#)] [[PubMed](#)]
13. Sozzi, G.; Boeri, M.; Rossi, M.; Verri, C.; Suatoni, P.; Bravi, F.; Roz, L.; Conte, D.; Grassi, M.; Sverzellati, N.; et al. Clinical utility of a plasma-based miRNA signature classifier within computed tomography lung cancer screening: A correlative MILD trial study. *J. Clin. Oncol.* **2014**, *32*, 768–773. [[CrossRef](#)] [[PubMed](#)]
14. Montani, F.; Marzi, M.J.; Dezi, F.; Dama, E.; Carletti, R.M.; Bonizzi, G.; Bertolotti, R.; Bellomi, M.; Rampinelli, C.; Maisonneuve, P.; et al. miR-Test: A blood test for lung cancer early detection. *J. Nat. Cancer Inst.* **2015**. [[CrossRef](#)] [[PubMed](#)]
15. Chen, J.; Wang, R.; Zhang, K.; Chen, L.-B. Long non-coding RNAs in non-small cell lung cancer as biomarkers and therapeutic targets. *J. Cell. Mol. Med.* **2014**, *18*, 2425–2436. [[CrossRef](#)] [[PubMed](#)]
16. Foa, P.; Fournier, M.; Miceli, R.; Seregni, E.; Santambrogio, L.; Nosotti, M.; Cataldo, I.; Sala, M.; Caldiera, S.; Bombardieri, E. Tumour markers CEA, NSE, SCC, TPA and CYFRA 21.1 in resectable non-small cell lung cancer. *Anticancer Res.* **1999**, *19*, 3613–3618. [[PubMed](#)]
17. Niklinski, J.; Furman, M. Clinical tumour markers in lung cancer. *Eur. J. Cancer Prev.* **1995**, *4*, 129–138. [[CrossRef](#)] [[PubMed](#)]
18. Kasimir-Bauer, S.; Schleucher, N.; Weber, R.; Neumann, R.; Seeber, S. Evaluation of different markers in non-small cell lung cancer: Prognostic value of clinical staging, tumour cell detection and tumour marker analysis for tumour progression and overall survival. *Oncol. Rep.* **2003**, *10*, 475–482. [[CrossRef](#)] [[PubMed](#)]
19. Tarro, G.; Perna, A.; Esposito, C. Early diagnosis of lung cancer by detection of tumor liberated protein. *J. Cell. Physiol.* **2005**, *203*, 1–5. [[CrossRef](#)] [[PubMed](#)]
20. Jacot, W.; Lhermitte, L.; Dossat, N.; Pujol, J.L.; Molinari, N.; Daurès, J.P.; Maudelonde, T.; Mangé, A.; Solassol, J. Serum proteomic profiling of lung cancer in high-risk groups and determination of clinical outcomes. *J. Thoracic Oncol.* **2008**, *3*, 840–850. [[CrossRef](#)] [[PubMed](#)]
21. Solassol, J.; Maudelonde, T.; Mange, A.; Pujol, J.L. Clinical relevance of autoantibody detection in lung cancer. *J. Thoracic Oncol.* **2011**, *6*, 955–962. [[CrossRef](#)] [[PubMed](#)]
22. Weber, D.G.; Johnen, G.; Casjens, S.; Bryk, O.; Pesch, B.; Jockel, K.H.; Kollmeier, J.; Brüning, T. Evaluation of long noncoding RNA MALAT1 as a candidate blood-based biomarker for the diagnosis of non-small cell lung cancer. *BMC Res. Notes* **2013**. [[CrossRef](#)] [[PubMed](#)]
23. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74.
24. Bartel, D.P. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **2004**, *116*, 281–297. [[CrossRef](#)]
25. Harrow, J.; Frankish, A.; Gonzalez, J.M.; Tapanari, E.; Diekhans, M.; Kokocinski, F.; Aken, B.L.; Barrell, D.; Zadissa, A.; Searle, S. GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res.* **2012**, *22*, 1760–1774. [[CrossRef](#)] [[PubMed](#)]
26. Stahlhut, C.; Slack, F.J. Combinatorial Action of MicroRNAs let-7 and miR-34 Effectively Synergizes with Erlotinib to Suppress Non-small Cell Lung Cancer Cell Proliferation. *Cell Cycle* **2015**, *14*, 2171–2180. [[CrossRef](#)] [[PubMed](#)]
27. Kasinski, A.L.; Kelnar, K.; Stahlhut, C.; Orellana, E.; Zhao, J.; Shimer, E.; Dysart, S.; Chen, X.; Bader, A.G.; Slack, F.J. A combinatorial microRNA therapeutics approach to suppressing non-small cell lung cancer. *Oncogene* **2015**, *34*, 3547–3555. [[CrossRef](#)] [[PubMed](#)]
28. Mercer, T.R.; Dinger, M.E.; Mattick, J.S. Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* **2009**, *10*, 155–159. [[CrossRef](#)] [[PubMed](#)]
29. Zhu, L.; Liu, J.; Ma, S.; Zhang, S. Long Noncoding RNA MALAT-1 Can Predict Metastasis and a Poor Prognosis: A Meta-Analysis. *Pathol. Oncol. Res.* **2015**, *21*, 1259–1264. [[CrossRef](#)] [[PubMed](#)]
30. Landi, M.T.; Zhao, Y.; Rotunno, M.; Koshiol, J.; Liu, H.; Bergen, A.W.; Rubagotti, M.; Goldstein, A.M.; Linnoila, I.; Marincola, F.M.; et al. MicroRNA expression differentiates histology and predicts survival of lung cancer. *Clin. Cancer Res.* **2010**, *16*, 430–441. [[CrossRef](#)] [[PubMed](#)]
31. Garofalo, M.; Romano, G.; Di Leva, G.; Nuovo, G.; Jeon, Y.J.; Ngankeu, A.; Sun, J.; Lovat, F.; Alder, H.; Condorelli, G.; et al. EGFR and MET receptor tyrosine kinase-altered microRNA expression induces tumorigenesis and gefitinib resistance in lung cancers. *Nat. Med.* **2012**, *18*, 74–82. [[CrossRef](#)] [[PubMed](#)]
32. Hou, J.; Meng, F.; Chan, L.W.; Cho, W.C.; Wong, S.C. Circulating Plasma MicroRNAs As Diagnostic Markers for NSCLC. *Front Genet.* **2016**. [[CrossRef](#)] [[PubMed](#)]

33. Kunz, M.; Xiao, K.; Liang, C.; Viereck, J.; Pachel, C.; Frantz, S.; Thum, T.; Dandekar, T. Bioinformatics of cardiovascular miRNA biology. *J. Mol. Cell. Cardiol.* **2014**. [[CrossRef](#)] [[PubMed](#)]
34. Erson, A.E.; Petty, E.M. miRNAs and cancer: New research developments and potential clinical applications. *Cancer Biol. Therapy* **2009**, *8*, 2317–2322. [[CrossRef](#)]
35. Wang, Z. MicroRNA: A matter of life or death. *World J. Biol. Chem.* **2010**, *1*, 41–54. [[CrossRef](#)] [[PubMed](#)]
36. Jansson, M.D.; Lund, A.H. MicroRNA and cancer. *Mol. Oncol.* **2012**, *6*, 590–610. [[CrossRef](#)] [[PubMed](#)]
37. Tang, D.; Shen, Y.; Wang, M.; Yang, R.; Wang, Z.; Sui, A.; Jiao, W.; Wang, Y. Identification of plasma microRNAs as novel noninvasive biomarkers for early detection of lung cancer. *Eur. J. Cancer Prev.* **2013**, *22*, 540–548. [[CrossRef](#)] [[PubMed](#)]
38. McKiernan, P.J.; Greene, C.M. High-throughput profiling for discovery of non-coding RNA biomarkers of lung disease. *Expert Rev. Mol. Diagn.* **2016**, *16*, 173–185. [[CrossRef](#)] [[PubMed](#)]
39. Geng, Q.; Fan, T.; Zhang, B.; Wang, W.; Xu, Y.; Hu, H. Five microRNAs in plasma as novel biomarkers for screening of early-stage non-small cell lung cancer. *Respir. Res.* **2014**. [[CrossRef](#)] [[PubMed](#)]
40. Zhu, W.; Zhou, K.; Zha, Y.; Chen, D.; He, J.; Ma, H.; Liu, X.; Le, H.; Zhang, Y. Diagnostic Value of Serum miR-182, miR-183, miR-210, and miR-126 Levels in Patients with Early-Stage Non-Small Cell Lung Cancer. *PLoS ONE* **2016**, *11*, e0153046. [[CrossRef](#)] [[PubMed](#)]
41. Bjaanaes, M.M.; Halvorsen, A.R.; Solberg, S.; Jorgensen, L.; Dragani, T.A.; Galvan, A.; Galvan, A.; Colombo, F.; Anderlini, M.; Pastorino, U.; Kure, E.; et al. Unique microRNA-profiles in EGFR-mutated lung adenocarcinomas. *Int. J. Cancer* **2014**, *135*, 1812–1821. [[CrossRef](#)] [[PubMed](#)]
42. Yu, S.L.; Chen, H.Y.; Chang, G.C.; Chen, C.Y.; Chen, H.W.; Singh, S.; Cheng, C.L.; Yu, C.J.; Lee, Y.C.; Chen, H.S.; et al. MicroRNA signature predicts survival and relapse in lung cancer. *Cancer Cell* **2008**, *13*, 48–57. [[CrossRef](#)] [[PubMed](#)]
43. Yanaihara, N.; Caplen, N.; Bowman, E.; Seike, M.; Kumamoto, K.; Yi, M.; Stephens, R.M.; Okamoto, A.; Yokota, J.; Tanaka, T.; et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* **2006**, *9*, 189–198. [[CrossRef](#)] [[PubMed](#)]
44. Saito, M.; Schetter, A.J.; Mollerup, S.; Kohno, T.; Skaug, V.; Bowman, E.D.; Mathé, E.A.; Takenoshita, S.; Yokota, J.; Haugen, A.; et al. The association of microRNA expression with prognosis and progression in early-stage, non-small cell lung adenocarcinoma: A retrospective analysis of three cohorts. *Clin. Cancer Res.* **2011**, *17*, 1875–1882. [[CrossRef](#)] [[PubMed](#)]
45. Capodanno, A.; Boldrini, L.; Proietti, A.; Ali, G.; Pelliccioni, S.; Niccoli, C.; D’Incecco, A.; Cappuzzo, F.; Chella, A.; Lucchi, M.; et al. Let-7g and miR-21 expression in non-small cell lung cancer: Correlation with clinicopathological and molecular features. *Int. J. Oncol.* **2013**, *43*, 765–774. [[PubMed](#)]
46. Lee, J.H.; Voortman, J.; Dingemans, A.M.; Voeller, D.M.; Pham, T.; Wang, Y.; Giaccone, G. MicroRNA expression and clinical outcome of small cell lung cancer. *PLoS ONE* **2011**, *6*, e21300. [[CrossRef](#)] [[PubMed](#)]
47. Lebanony, D.; Benjamin, H.; Gilad, S.; Ezagouri, M.; Dov, A.; Ashkenazi, K.; Gefen, N.; Izraeli, S.; Rechavi, G.; Pass, H.; et al. Diagnostic assay based on hsa-miR-205 expression distinguishes squamous from nonsquamous non-small-cell lung carcinoma. *J. Clin. Oncol.* **2009**, *27*, 2030–2037. [[CrossRef](#)] [[PubMed](#)]
48. Bishop, J.A.; Benjamin, H.; Cholak, H.; Chajut, A.; Clark, D.P.; Westra, W.H. Accurate classification of non-small cell lung carcinoma using a novel microRNA-based approach. *Clin. Cancer Res.* **2010**, *16*, 610–619. [[CrossRef](#)] [[PubMed](#)]
49. Hennessey, P.T.; Sanford, T.; Choudhary, A.; Mydlarz, W.W.; Brown, D.; Adai, A.T.; Ochs, M.F.; Ahrendt, S.A.; Mambo, E.; Califano, J.A. Serum microRNA biomarkers for detection of non-small cell lung cancer. *PLoS ONE* **2012**, *7*, e32307. [[CrossRef](#)]
50. Markou, A.; Sourvinou, I.; Vorkas, P.A.; Yousef, G.M.; Lianidou, E. Clinical evaluation of microRNA expression profiling in non small cell lung cancer. *Lung Cancer* **2013**, *81*, 388–396. [[CrossRef](#)] [[PubMed](#)]
51. Wang, Z.X.; Bian, H.B.; Wang, J.R.; Cheng, Z.X.; Wang, K.M.; De, W. Prognostic significance of serum miRNA-21 expression in human non-small cell lung cancer. *J. Surg. Oncol.* **2011**, *104*, 847–851. [[CrossRef](#)] [[PubMed](#)]
52. Vosa, U.; Vooder, T.; Kolde, R.; Vilo, J.; Metspalu, A.; Annilo, T. Meta-analysis of microRNA expression in lung cancer. *Int. J. Cancer* **2013**, *132*, 2884–2893. [[CrossRef](#)] [[PubMed](#)]
53. Guan, P.; Yin, Z.; Li, X.; Wu, W.; Zhou, B. Meta-analysis of human lung cancer microRNA expression profiling studies comparing cancer tissues with normal tissues. *J. Exp. Clin. Cancer Res.* **2012**, *31*. [[CrossRef](#)] [[PubMed](#)]

54. Nagano, T.; Fraser, P. No-Nonsense Functions for Long Noncoding RNAs. *Cell* **2011**, *145*, 178–181. [[CrossRef](#)] [[PubMed](#)]
55. Ponting, C.P.; Oliver, P.L.; Reik, W. Evolution and Functions of Long Noncoding RNAs. *Cell* **2009**, *136*, 629–641. [[CrossRef](#)] [[PubMed](#)]
56. Lee, J.T. Epigenetic regulation by long noncoding RNAs. *Science* **2012**, *338*, 1435–1439. [[CrossRef](#)] [[PubMed](#)]
57. Ulitsky, I.; Bartel, D.P. lincRNAs: Genomics, evolution, and mechanisms. *Cell* **2013**, *154*, 26–46. [[CrossRef](#)] [[PubMed](#)]
58. Prasanth, K.V.; Spector, D.L. Eukaryotic regulatory RNAs: An answer to the “genome complexity” conundrum. *Genes Dev.* **2007**, *21*, 11–42. [[CrossRef](#)] [[PubMed](#)]
59. Volders, P.J.; Helsens, K.; Wang, X.; Menten, B.; Martens, L.; Gevaert, K.; Vandesompele, J.; Mestdagh, P. LNCipedia: A database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res.* **2013**, *41*, D246–D251. [[CrossRef](#)] [[PubMed](#)]
60. Chakraborty, S.; Deb, A.; Maji, R.K.; Saha, S.; Ghosh, Z. LncRBase: An enriched resource for lncRNA Information. *PLoS ONE* **2014**, *9*, e108010. [[CrossRef](#)] [[PubMed](#)]
61. Cabili, M.N.; Trapnell, C.; Goff, L.; Koziol, M.; Tazon-Vega, B.; Regev, A.; Rinn, J.L. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **2011**, *25*, 1915–1927. [[CrossRef](#)] [[PubMed](#)]
62. Signal, B.; Gloss, B.S.; Dinger, M.E. Computational approaches for functional prediction and characterisation of long noncoding RNAs. *Trends Genet.* **2016**, *32*, 620–637. [[CrossRef](#)] [[PubMed](#)]
63. Kumarswamy, R.; Bauters, C.; Volkmann, I.; Maury, F.; Fetisch, J.; Holzmann, A.; Lemesle, G.; de Groote, P.; Pinet, F.; Thum, T. Circulating long noncoding RNA, LIPCAR, predicts survival in patients with heart failure. *Circ. Res.* **2014**, *114*, 1569–1575. [[CrossRef](#)] [[PubMed](#)]
64. Panzitt, K.; Tschernatsch, M.M.; Guelly, C.; Moustafa, T.; Stradner, M.; Strohmaier, H.M.; Buck, C.R.; Denk, H.; Schroeder, R.; Trauner, M.; et al. Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA. *Gastroenterology* **2007**, *132*, 330–342. [[CrossRef](#)] [[PubMed](#)]
65. Durand, X.; Moutereau, S.; Xylinas, E.; de la Taille, A. Progensis PCA3 test for prostate cancer. *Expert Rev. Mol. Diagn.* **2011**, *11*, 137–144. [[CrossRef](#)] [[PubMed](#)]
66. Reis, E.M.; Verjovski-Almeida, S. Perspectives of long non-coding RNAs in cancer diagnostics. *Front. Genetic.* **2012**. [[CrossRef](#)] [[PubMed](#)]
67. Vencken, S.F.; Greene, C.M.; McKiernan, P.J. Non-coding RNA as lung disease biomarkers. *Thorax* **2015**, *70*, 501–503. [[CrossRef](#)] [[PubMed](#)]
68. Ji, P.; Diederichs, S.; Wang, W.; Boing, 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](#)] [[PubMed](#)]
69. Qiu, M.; Xu, Y.; Yang, X.; Wang, J.; Hu, J.; Xu, L.; Yin, R. CCAT2 is a lung adenocarcinoma-specific long non-coding RNA and promotes invasion of non-small cell lung cancer. *Tumour Biol.* **2014**, *35*, 5375–5380. [[CrossRef](#)] [[PubMed](#)]
70. Chen, S.; Wu, H.; Lv, N.; Wang, H.; Wang, Y.; Tang, Q.; Shao, H.; Sun, C. LncRNA CCAT2 predicts poor prognosis and regulates growth and metastasis in small cell lung cancer. *Biomed. Pharmacother.* **2016**, *82*, 583–588. [[CrossRef](#)] [[PubMed](#)]
71. Liu, X.H.; Liu, Z.L.; Sun, M.; Liu, J.; Wang, Z.X.; De, W. The long non-coding RNA HOTAIR indicates a poor prognosis and promotes metastasis in non-small cell lung cancer. *BMC Cancer* **2013**. [[CrossRef](#)] [[PubMed](#)]
72. Yang, Y.; Li, H.; Hou, S.; Hu, B.; Liu, J.; Wang, J. The noncoding RNA expression profile and the effect of lncRNA AK126698 on cisplatin resistance in non-small-cell lung cancer cell. *PLoS ONE* **2013**, *8*, e65309. [[CrossRef](#)] [[PubMed](#)]
73. Sui, J.; Li, Y.H.; Zhang, Y.Q.; Li, C.Y.; Shen, X.; Yao, W.Z.; Peng, H.; Hong, W.W.; Yin, L.H.; Pu, Y.P.; et al. Integrated analysis of long non-coding RNA-associated ceRNA network reveals potential lncRNA biomarkers in human lung adenocarcinoma. *Int. J. Oncol.* **2016**, *49*, 2023–2036. [[PubMed](#)]
74. Tantai, J.; Hu, D.; Yang, Y.; Geng, J. Combined identification of long non-coding RNA XIST and HIF1A-AS1 in serum as an effective screening for non-small cell lung cancer. *Int. J. Clin. Exp. Pathol.* **2015**, *8*, 7887–7895. [[PubMed](#)]

75. Gong, W.J.; Yin, J.Y.; Li, X.P.; Fang, C.; Xiao, D.; Zhang, W.; Zhou, H.H.; Li, X.; Liu, Z.Q. Association of well-characterized lung cancer lncRNA polymorphisms with lung cancer susceptibility and platinum-based chemotherapy response. *Tumour Biol.* **2016**, *37*, 8349–8358. [[CrossRef](#)] [[PubMed](#)]
76. Yuan, H.; Liu, H.; Liu, Z.; Owzar, K.; Han, Y.; Su, L.; Wei, Y.; Hung, R.J.; McLaughlin, J.; Brhane, Y.; et al. A novel genetic variant in long non-coding RNA gene NEXN-AS1 is associated with risk of lung cancer. *Sci. Rep.* **2016**. [[CrossRef](#)] [[PubMed](#)]
77. Loewen, G.; Jayawickramarajah, J.; Zhuo, Y.; Shan, B. Functions of lncRNA HOTAIR in lung cancer. *J. Hematol. Oncol.* **2014**. [[CrossRef](#)] [[PubMed](#)]
78. Zhang, J.; Zhang, B.; Wang, T.; Wang, H. LncRNA MALAT1 overexpression is an unfavorable prognostic factor in human cancer: evidence from a meta-analysis. *Int. J. Clin. Exp. Med.* **2015**, *8*, 5499–5505. [[PubMed](#)]
79. Hong, H.H.; Hou, L.K.; Pan, X.; Wu, C.Y.; Huang, H.; Li, B.; Nie, W. Long non-coding RNA UCA1 is a predictive biomarker of cancer. *Oncotarget* **2016**, *7*, 44442–44447. [[CrossRef](#)] [[PubMed](#)]
80. He, A.; Hu, R.; Chen, Z.; Liao, X.; Li, J.; Wang, D.; Lv, Z.; Liu, Y.; Wang, F.; Mei, H. Role of long noncoding RNA UCA1 as a common molecular marker for lymph node metastasis and prognosis in various cancers: A meta-analysis. *Oncotarget* **2016**. [[CrossRef](#)] [[PubMed](#)]
81. Bell, R.D.; Long, X.; Lin, M.; Bergmann, J.H.; Nanda, V.; Cowan, S.L.; Zhou, Q.; Han, Y.; Spector, D.L.; Zheng, D.; et al. Identification and initial functional characterization of a human vascular cell enriched long non-coding RNA. *Arterioscler. Thromb. Vasc. Biol.* **2014**, *34*, 1249–1259. [[CrossRef](#)] [[PubMed](#)]
82. Muppirala, U.K.; Honavar, V.G.; Dobbs, D. Predicting RNA-protein interactions using only sequence information. *BMC Bioinform.* **2011**, *12*, 1–11. [[CrossRef](#)] [[PubMed](#)]
83. Cook, K.B.; Hughes, T.R.; Morris, Q.D. High-throughput characterization of protein-RNA interactions. *Brief. Funct. Genom.* **2015**, *14*, 74–89. [[CrossRef](#)] [[PubMed](#)]
84. Nawrocki, E.P.; Burge, S.W.; Bateman, A.; Daub, J.; Eberhardt, R.Y.; Eddy, S.R.; Floden, E.W.; Gardner, P.P.; Jones, T.A.; Tate, J.; et al. Rfam 12.0: Updates to the RNA families database. *Nucleic Acids Res.* **2015**, *43*, D130–D137. [[CrossRef](#)] [[PubMed](#)]
85. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
86. Hofacker, I.L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **2003**, *31*, 3429–3431. [[CrossRef](#)] [[PubMed](#)]
87. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003**, *31*, 3406–3415. [[CrossRef](#)] [[PubMed](#)]
88. Bernhart, S.; Hofacker, I.; Will, S.; Gruber, A.; Stadler, P. RNAalifold: Improved consensus structure prediction for RNA alignments. *BMC Bioinform.* **2008**. [[CrossRef](#)] [[PubMed](#)]
89. Havgaard, J.H.; Lyngsø, R.B.; Gorodkin, J. The foldalign web server for pairwise structural RNA alignment and mutual motif search. *Nucleic Acids Res.* **2005**, *33*, W650–W653. [[CrossRef](#)] [[PubMed](#)]
90. Smith, C.; Heyne, S.; Richter, A.S.; Will, S.; Backofen, R. Freiburg RNA Tools: A web server integrating INTARNA, EXPARNA and LOCARNA. *Nucleic Acids Res.* **2010**, *38*, W373–W377. [[CrossRef](#)] [[PubMed](#)]
91. Steffen, P.; Voss, B.; Rehmsmeier, M.; Reeder, J.; Giegerich, R. RNASHAPES: An integrated RNA analysis package based on abstract shapes. *Bioinformatics* **2006**, *22*, 500–503. [[CrossRef](#)] [[PubMed](#)]
92. Seibel, P.N.; Müller, T.; Dandekar, T.; Schultz, J.; Wolf, M. 4SALE—A tool for synchronous RNA sequence and secondary structure alignment and editing. *BMC Bioinform.* **2006**, *7*, 1–7.
93. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29. [[CrossRef](#)] [[PubMed](#)]
94. Carbon, S.; Ireland, A.; Mungall, C.J.; Shu, S.; Marshall, B.; Lewis, S.; Web Presence Working Group. AmiGO: Online access to ontology and annotation data. *Bioinformatics* **2009**, *25*, 288–289. [[CrossRef](#)] [[PubMed](#)]
95. Mi, H.; Poudel, S.; Muruganujan, A.; Casagrande, J.T.; Thomas, P.D. PANTHER version 10: Expanded protein families and functions, and analysis tools. *Nucleic Acids Res.* **2016**, *44*, D336–D342. [[CrossRef](#)] [[PubMed](#)]
96. Croft, D.; O’Kelly, G.; Wu, G.; Haw, R.; Gillespie, M.; Matthews, L.; Caudy, M.; Garapati, P.; Gopinath, G.; Jassal, B.; et al. Reactome: A database of reactions, pathways and biological processes. *Nucleic Acids Res.* **2011**, *39*, D691–D697. [[CrossRef](#)] [[PubMed](#)]

97. Kanehisa, M.; Goto, S.; Furumichi, M.; Tanabe, M.; Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **2010**, *38*, D355–D360. [[CrossRef](#)] [[PubMed](#)]
98. Kutmon, M.; Riutta, A.; Nunes, N.; Hanspers, K.; Willighagen Egon, L.; Bohler, A.; Mélius, J.; Waagmeester, A.; Sinha, S.R.; Miller, R.; et al. WikiPathways: Capturing the full diversity of pathway knowledge. *Nucleic Acids Res.* **2016**, *44*, D488–D494. [[CrossRef](#)] [[PubMed](#)]
99. Saito, R.; Smoot, M.E.; Ono, K.; Ruschinski, J.; Wang, P.L.; Lotia, S.; Pico, A.R.; Bader, G.D.; Ideker, T. A travel guide to Cytoscape plugins. *Nat. Methods* **2012**, *9*, 1069–1076. [[CrossRef](#)] [[PubMed](#)]
100. 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. *Nucleic Acids Res.* **2014**, *42*, D92–D97. [[CrossRef](#)] [[PubMed](#)]
101. Wingender, E.; Chen, X.; Hehl, R.; Karas, H.; Liebich, I.; Matys, V.; Meinhardt, T.; Prüß, M.; Reuter, I.; Schacherer, F. TRANSFAC: An integrated system for gene expression regulation. *Nucleic Acids Res.* **2000**, *28*, 316–319. [[CrossRef](#)] [[PubMed](#)]
102. Sandelin, A.; Alkema, W.; Engstrom, P.; Wasserman, W.W.; Lenhard, B. JASPAR: An open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* **2004**, *32*, D91–D94. [[CrossRef](#)] [[PubMed](#)]
103. Messeguer, X.; Escudero, R.; Farre, D.; Nunez, O.; Martinez, J.; Alba, M.M. PROMO: Detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* **2002**, *18*, 333–334. [[CrossRef](#)] [[PubMed](#)]
104. Cartharius, K.; Frech, K.; Grote, K.; Klocke, B.; Haltmeier, M.; Klingenhoff, A.; Frisch, M.; Bayerlein, M.; Werner, T. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* **2005**, *21*, 2933–2942. [[CrossRef](#)] [[PubMed](#)]
105. Enright, A.J.; John, B.; Gaul, U.; Tuschl, T.; Sander, C.; Marks, D.S. MicroRNA targets in Drosophila. *Genome Biol.* **2003**. [[CrossRef](#)]
106. Muckstein, U.; Tafer, H.; Hackermuller, J.; Bernhart, S.H.; Stadler, P.F.; Hofacker, I.L. Thermodynamics of RNA–RNA binding. *Bioinformatics* **2006**, *22*, 1177–1182. [[CrossRef](#)] [[PubMed](#)]
107. Busch, A.; Richter, A.S.; Backofen, R. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* **2008**, *24*, 2849–2856. [[CrossRef](#)] [[PubMed](#)]
108. RNAcentral Consortium. RNAcentral: An international database of ncRNA sequences. *Nucleic Acids Res.* **2015**, *43*, D123–D129.
109. Kozomara, A.; Griffiths-Jones, S. miRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **2014**, *42*, D68–D73. [[CrossRef](#)] [[PubMed](#)]
110. Jiang, Q.; Wang, Y.; Hao, Y.; Juan, L.; Teng, M.; Zhang, X.; Li, M.; Wang, G.; Liu, Y. miR2Disease: A manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.* **2009**, *37*, D98–D104. [[CrossRef](#)] [[PubMed](#)]
111. Lewis, B.P.; Shih, I.H.; Jones-Rhoades, M.W.; Bartel, D.P.; Burge, C.B. Prediction of mammalian microRNA targets. *Cell* **2003**, *115*, 787–798. [[CrossRef](#)]
112. Krek, A.; Grun, D.; Poy, M.N.; Wolf, R.; Rosenberg, L.; Epstein, E.J.; MacMenamin, P.; Da Piedade, I.; Gunsalus, K.C.; Stoffel, M.; et al. Combinatorial microRNA target predictions. *Nat. Genet.* **2005**, *37*, 495–500. [[CrossRef](#)] [[PubMed](#)]
113. Kertesz, M.; Iovino, N.; Unnerstall, U.; Gaul, U.; Segal, E. The role of site accessibility in microRNA target recognition. *Nat. Genet.* **2007**, *39*, 1278–1284. [[CrossRef](#)] [[PubMed](#)]
114. Chen, G.; Wang, Z.; Wang, D.; Qiu, C.; Liu, M.; Chen, X.; Zhang, Q.; Yan, G.; Cui, Q. LncRNADisease: A database for long-non-coding RNA-associated diseases. *Nucleic Acids Res.* **2013**, *41*, D983–D986. [[CrossRef](#)] [[PubMed](#)]
115. Terai, G.; Iwakiri, J.; Kameda, T.; Hamada, M.; Asai, K. Comprehensive prediction of lncRNA–RNA interactions in human transcriptome. *BMC Genom.* **2016**, *17*, 153–164. [[CrossRef](#)] [[PubMed](#)]
116. Li, J.; Ma, W.; Zeng, P.; Wang, J.; Geng, B.; Yang, J.; Cui, Q. LncTar: A tool for predicting the RNA targets of long noncoding RNAs. *Brief Bioinform.* **2015**, *16*, 806–812. [[CrossRef](#)] [[PubMed](#)]
117. Wu, T.; Wang, J.; Liu, C.; Zhang, Y.; Shi, B.; Zhu, X.; Zhang, Z.; Skogerbo, G.; Chen, L.; Lu, H.; et al. NPInter: The noncoding RNAs and protein related biomacromolecules interaction database. *Nucleic Acids Res.* **2006**, *34*, D150–D152. [[CrossRef](#)] [[PubMed](#)]

118. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [[CrossRef](#)] [[PubMed](#)]
119. Narayanan, B.C.; Westbrook, J.; Ghosh, S.; Petrov, A.I.; Sweeney, B.; Zirbel, C.L.; Leontis, N.B.; Berman, H.M. The Nucleic Acid Database: New features and capabilities. *Nucleic Acids Res.* **2014**, *42*, D114–D122. [[CrossRef](#)] [[PubMed](#)]
120. Stark, C.; Breitkreutz, B.J.; Reguly, T.; Boucher, L.; Breitkreutz, A.; Tyers, M. BioGRID: A general repository for interaction datasets. *Nucleic Acids Res.* **2006**, *34*, D535–D539. [[CrossRef](#)] [[PubMed](#)]
121. Hermjakob, H.; Montecchi-Palazzi, L.; Lewington, C.; Mudali, S.; Kerrien, S.; Orchard, S.; Vingron, M.; Roechert, B.; Roepstorff, P.; Valencia, A.; et al. IntAct: An open source molecular interaction database. *Nucleic Acids Res.* **2004**, *32*, D452–D455. [[CrossRef](#)] [[PubMed](#)]
122. Fujimori, S.; Hino, K.; Saito, A.; Miyano, S.; Miyamoto-Sato, E. PRD: A protein-RNA interaction database. *Bioinformatics* **2012**, *8*, 729–730. [[CrossRef](#)] [[PubMed](#)]
123. Orchard, S.; Kerrien, S.; Abbani, S.; Aranda, B.; Bhate, J.; Bidwell, S.; Bridge, A.; Briganti, L.; Brinkman, F.S.; Cesareni, G.; et al. Protein interaction data curation: The International Molecular Exchange (IMEx) consortium. *Nat. Methods* **2012**, *9*, 345–350. [[CrossRef](#)] [[PubMed](#)]
124. Consortium, T.U. UniProt: A hub for protein information. *Nucleic Acids Res.* **2015**, *43*, D204–D212. [[CrossRef](#)] [[PubMed](#)]
125. Lewis, B.A.; Walia, R.R.; Terribilini, M.; Ferguson, J.; Zheng, C.; Honavar, V.; et al. PRIDB: A protein–RNA interface database. *Nucleic Acids Res.* **2011**, *39*, D277–D282. [[CrossRef](#)] [[PubMed](#)]
126. Kunz, M.; Liang, C.; Nilla, S.; Cecil, A.; Dandekar, T. The drug-minded protein interaction database (DrumPID) for efficient target analysis and drug development. *Database* **2016**. [[CrossRef](#)] [[PubMed](#)]
127. Bellucci, M.; Agostini, F.; Masin, M.; Tartaglia, G.G. Predicting protein associations with long noncoding RNAs. *Nat. Methods* **2011**, *8*, 444–445. [[CrossRef](#)] [[PubMed](#)]
128. Kumar, M.; Gromiha, M.M.; Raghava, G.P. Prediction of RNA binding sites in a protein using SVM and PSSM profile. *Proteins* **2008**, *71*, 189–194. [[CrossRef](#)] [[PubMed](#)]
129. Kim, O.T.; Yura, K.; Go, N. Amino acid residue doublet propensity in the protein-RNA interface and its application to RNA interface prediction. *Nucleic Acids Res.* **2006**, *34*, 6450–6460. [[CrossRef](#)] [[PubMed](#)]
130. Towfic, F.; Caragea, C.; Gemperline, D.C.; Dobbs, D.; Honavar, V. Struct-NB: Predicting protein-RNA binding sites using structural features. *Int. J. Data Min. Bioinform.* **2010**, *4*, 21–43. [[CrossRef](#)]
131. Wang, Y.; Xue, Z.; Shen, G.; Xu, J. PRINTR: Prediction of RNA binding sites in proteins using SVM and profiles. *Amino Acids* **2008**, *35*, 295–302. [[CrossRef](#)] [[PubMed](#)]
132. Park, C.; Yu, N.; Choi, I.; Kim, W.; Lee, S. lncRNator: A comprehensive resource for functional investigation of long non-coding RNAs. *Bioinformatics* **2014**, *30*, 2480–2485. [[CrossRef](#)] [[PubMed](#)]
133. Dalli, D.; Wilm, A.; Mainz, I.; Steger, G. STRAL: Progressive alignment of non-coding RNA using base pairing probability vectors in quadratic time. *Bioinformatics* **2006**, *22*, 1593–1599. [[CrossRef](#)] [[PubMed](#)]
134. Gomes, C.P.; Cho, J.H.; Hood, L.; Franco, O.L.; Pereira, R.W.; Wang, K. A Review of computational tools in microRNA discovery. *Front. Genet.* **2013**. [[CrossRef](#)] [[PubMed](#)]
135. Sankoff, D. Simultaneous solution of the RNA folding, alignment and protosequence problems. *SIAM J. Appl. Math.* **1985**, *45*, 810–825. [[CrossRef](#)]
136. Hofacker, I.L.; Fekete, M.; Stadler, P.F. Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.* **2002**, *319*, 1059–1066. [[CrossRef](#)]
137. Havgaard, J.H.; Gorodkin, J. RNA structural alignments, part I: Sankoff-based approaches for structural alignments. *Methods Mol. Biol.* **2014**, *1097*, 275–290. [[PubMed](#)]
138. Reeder, J.; Giegerich, R. Consensus shapes: An alternative to the Sankoff algorithm for RNA consensus structure prediction. *Bioinformatics* **2005**, *21*, 3516–3523. [[CrossRef](#)] [[PubMed](#)]
139. Sethupathy, P.; Megraw, M.; Hatzigeorgiou, A.G. A guide through present computational approaches for the identification of mammalian microRNA targets. *Nat. Methods* **2006**, *3*, 881–886. [[CrossRef](#)] [[PubMed](#)]
140. Friedersdorf, M.B.; Keene, J.D. Advancing the functional utility of PAR-CLIP by quantifying background binding to mRNAs and lncRNAs. *Genome Biol.* **2014**. [[CrossRef](#)] [[PubMed](#)]
141. Muppirla, U.K.; Lewis, B.A.; Dobbs, D. Computational tools for investigating RNA-protein interaction partners. *J. Comput. Sci. Syst. Biol.* **2013**, *6*, 182–187.
142. Cirillo, D.; Agostini, F.; Tartaglia, G.G. Predictions of protein–RNA interactions. *Wiley Interdiscip. Rev.* **2013**, *3*, 161–175. [[CrossRef](#)]

143. Hammell, M. Computational methods to identify miRNA targets. *Seminars Cell Dev. Biol.* **2010**, *21*, 738–744. [[CrossRef](#)] [[PubMed](#)]
144. Liu, B.; Li, J.; Cairns, M.J. Identifying miRNAs, targets and functions. *Brief. Bioinform.* **2014**, *15*, 1–19. [[CrossRef](#)] [[PubMed](#)]
145. Friedman, R.C.; Farh, K.K.; Burge, C.B.; Bartel, D.P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **2009**, *19*, 92–105. [[CrossRef](#)] [[PubMed](#)]
146. Betel, D.; Wilson, M.; Gabow, A.; Marks, D.S.; Sander, C. The microRNA.org resource: Targets and expression. *Nucleic Acids Res.* **2008**, *36*, D149–D153. [[CrossRef](#)] [[PubMed](#)]
147. Kiriakidou, M.; Nelson, P.T.; Kouranov, A.; Fitziev, P.; Bouyioukos, C.; Mourelatos, Z.; Hatzigeorgiou, A. A combined computational-experimental approach predicts human microRNA targets. *Genes Dev.* **2004**, *18*, 1165–1178. [[CrossRef](#)] [[PubMed](#)]
148. Maziere, P.; Enright, A.J. Prediction of microRNA targets. *Drug Discov. Today* **2007**, *12*, 452–458. [[CrossRef](#)] [[PubMed](#)]
149. Yue, D.; Liu, H.; Huang, Y. Survey of computational algorithms for microRNA Target prediction. *Curr. Genom.* **2009**, *10*, 478–492. [[CrossRef](#)] [[PubMed](#)]
150. Rajewsky, N. microRNA target predictions in animals. *Nat. Genet.* **2006**, *38*, S8–S13. [[CrossRef](#)] [[PubMed](#)]
151. Gusev, Y.; Schmittgen, T.D.; Lerner, M.; Postier, R.; Brackett, D. Computational analysis of biological functions and pathways collectively targeted by co-expressed microRNAs in cancer. *BMC Bioinform.* **2007**. [[CrossRef](#)] [[PubMed](#)]
152. Stark, A.; Brennecke, J.; Bushati, N.; Russell, R.B.; Cohen, S.M. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell* **2005**, *123*, 1133–1146. [[CrossRef](#)] [[PubMed](#)]
153. Marshall, O.J. PerlPrimer: Cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics* **2004**, *20*, 2471–2472. [[CrossRef](#)] [[PubMed](#)]
154. Chen, Y.; Varani, G. Engineering RNA-binding proteins for biology. *FEBS J.* **2013**, *280*, 3734–3754. [[CrossRef](#)] [[PubMed](#)]
155. Walia, R.R.; Caragea, C.; Lewis, B.A.; Towfic, F.; Terribilini, M.; El-Manzalawy, Y.; Dobbs, D.; Honavar, V. Protein-RNA interface residue prediction using machine learning: An assessment of the state of the art. *BMC Bioinform.* **2012**. [[CrossRef](#)] [[PubMed](#)]
156. He, W.; Cai, Q.; Sun, F.; Zhong, G.; Wang, P.; Liu, H.; Luo, J.; Yu, H.; Huang, J.; Lin, T. linc-UBC1 physically associates with polycomb repressive complex 2 (PRC2) and acts as a negative prognostic factor for lymph node metastasis and survival in bladder cancer. *Biochim. Biophys. Acta* **2013**, *1832*, 1528–1537. [[CrossRef](#)] [[PubMed](#)]
157. Li, A.; Ge, M.; Zhang, Y.; Peng, C.; Wang, M. Predicting long noncoding RNA and protein interactions using heterogeneous network model. *Biomed. Res. Int.* **2015**. [[CrossRef](#)] [[PubMed](#)]
158. Tong, J.; Jiang, P.; Lu, Z.H. RISP: A web-based server for prediction of RNA-binding sites in proteins. *Comput. Methods Programs Biomed.* **2008**, *90*, 148–153. [[CrossRef](#)] [[PubMed](#)]
159. Zhao, H.; Yang, Y.; Zhou, Y. Highly accurate and high-resolution function prediction of RNA binding proteins by fold recognition and binding affinity prediction. *RNA Biol.* **2011**, *8*, 988–996. [[CrossRef](#)] [[PubMed](#)]
160. Fabregat, A.; Sidiropoulos, K.; Garapati, P.; Gillespie, M.; Hausmann, K.; Haw, R. The Reactome pathway Knowledgebase. *Nucleic Acids Res.* **2016**, *44*, D481–D487. [[CrossRef](#)] [[PubMed](#)]
161. Herwig, R.; Hardt, C.; Lienhard, M.; Kamburov, A. Analyzing and interpreting genome data at the network level with ConsensusPathDB. *Nat. Protoc.* **2016**, *11*, 1889–1907. [[CrossRef](#)] [[PubMed](#)]
162. Tipney, H.; Hunter, L. An introduction to effective use of enrichment analysis software. *Hum. Genom.* **2010**, *4*, 202–206. [[CrossRef](#)] [[PubMed](#)]
163. Khatri, P.; Sirota, M.; Butte, A.J. Ten Years of pathway analysis: Current Approaches and outstanding challenges. *PLoS Comput. Biol.* **2012**, *8*, e1002375. [[CrossRef](#)] [[PubMed](#)]
164. Maere, S.; Heymans, K.; Kuiper, M. BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **2005**, *21*, 3448–3449. [[CrossRef](#)] [[PubMed](#)]
165. Bader, G.D.; Hogue, C.W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinform.* **2003**. [[CrossRef](#)]

166. Bindea, G.; Mlecnik, B.; Hackl, H.; Charoentong, P.; Tosolini, M.; Kirilovsky, A.; et al. ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **2009**, *25*, 1091–1093. [[CrossRef](#)] [[PubMed](#)]
167. Bonnici, V.; Russo, F.; Bombieri, N.; Pulvirenti, A.; Giugno, R. Comprehensive reconstruction and visualization of non-coding regulatory networks in human. *Front. Bioeng. Biotechnol.* **2014**. [[CrossRef](#)] [[PubMed](#)]
168. Stormo, G.D. DNA binding sites: Representation and discovery. *Bioinformatics* **2000**, *16*, 16–23. [[CrossRef](#)] [[PubMed](#)]
169. Giardine, B.; Riemer, C.; Hardison, R.C.; Burhans, R.; Elnitski, L.; Shah, P.; Zhang, Y.; Blankenberg, D.; Albert, I.; Taylor, J. Galaxy: A platform for interactive large-scale genome analysis. *Genome Res.* **2005**, *15*, 1451–1455. [[CrossRef](#)] [[PubMed](#)]
170. Goodstadt, L. Ruffus: A lightweight Python library for computational pipelines. *Bioinformatics* **2010**, *26*, 2778–2779. [[CrossRef](#)] [[PubMed](#)]
171. Koster, J.; Rahmann, S. Snakemake—A scalable bioinformatics workflow engine. *Bioinformatics* **2012**, *28*, 2520–2522. [[CrossRef](#)] [[PubMed](#)]
172. Dooley, D.M.; Petkau, A.J.; Van Domselaar, G.; Hsiao, W.W. Sequence database versioning for command line and Galaxy bioinformatics servers. *Bioinformatics* **2016**, *32*, 1275–1277. [[CrossRef](#)] [[PubMed](#)]
173. Boettiger, C. An introduction to Docker for reproducible research, with examples from the R environment. *ACM SIGOPS Oper. Syst. Rev.* **2015**. [[CrossRef](#)]
174. Sundaresan, T.K.; Haber, D.A. Fantastic voyage: The future of cancer diagnostics. *Lancet Oncol.* **2015**, *16*, 1596–1598. [[CrossRef](#)]



© 2016 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 (<http://creativecommons.org/licenses/by/4.0/>).