*Article*

# Pan-Cancer Analysis Reveals Functional Similarity of Three lncRNAs across Multiple Tumors

Abir Khazaal [1,2], Seid Miad Zandavi [1,3], Andrei Smolnikov [1], Shadma Fatima [1,4] and Fatemeh Vafaee [1,2,*]

1 School of Biotechnology and Biomolecular Sciences, Faculty of Science, University of New South Wales, Sydney, NSW 2052, Australia
2 UNSW Data Science Hub, University of New South Wales, Sydney, NSW 2052, Australia
3 Harvard Medical School, Harvard University, Boston, MA 02115, USA
4 Ingham Institute of Applied Medical Research, Sydney, NSW 2170, Australia
* Correspondence: f.vafaee@unsw.edu.au

**Abstract:** Long non-coding RNAs (lncRNAs) are emerging as key regulators in many biological processes. The dysregulation of lncRNA expression has been associated with many diseases, including cancer. Mounting evidence suggests lncRNAs to be involved in cancer initiation, progression, and metastasis. Thus, understanding the functional implications of lncRNAs in tumorigenesis can aid in developing novel biomarkers and therapeutic targets. Rich cancer datasets, documenting genomic and transcriptomic alterations together with advancement in bioinformatics tools, have presented an opportunity to perform pan-cancer analyses across different cancer types. This study is aimed at conducting a pan-cancer analysis of lncRNAs by performing differential expression and functional analyses between tumor and non-neoplastic adjacent samples across eight cancer types. Among dysregulated lncRNAs, seven were shared across all cancer types. We focused on three lncRNAs, found to be consistently dysregulated among tumors. It has been observed that these three lncRNAs of interest are interacting with a wide range of genes across different tissues, yet enriching substantially similar biological processes, found to be implicated in cancer progression and proliferation.

**Keywords:** long non-coding RNA; cancer; mRNA; pan-cancer analysis; functional analysis; gene ontology

## 1. Introduction

Cancer is a complex disease that continues to be a health burden globally [1]. It is characterised by dynamic genomic alterations, including somatic mutations, epigenetic modifications, copy number variations, and changes in expression profiles [2–4]. The emergence of massively parallel sequencing technologies has enabled systematic documentation of the genetic changes in tumors and introduced the concept of the cancer genome [5–7]. Considered a landmark in cancer genomics, The Cancer Genome Atlas (TCGA) program has produced, to date, more than 2.5 PB of multi-layered omic data, along with clinical profiles for more than 11,000 patients across 33 cancer types [8–10]. TCGA has improved our understanding of cancer genomics, revolutionised cancer classification and identified therapeutic targets [8,11,12].

Although cancers have their own genetic identity with distinct, tissue-specific changes, many tumors share similar genetic alterations that disrupt common biological processes [13,14]. Emerging computational technologies and rich datasets present an opportunity to explore the differences and similarities of genetic and molecular changes across different tumor types using a set of techniques collectively referred to as pan-cancer analyses [14,15]. The importance of pan-cancer profiling lies in its ability to provide a comprehensive analysis of the genetic changes associated with multiple cancers. In addition, not only does it identify shared patterns that aid in the development of uniform treatments strategies, but also distinguishes those unique alterations and enhances personalised care [14].

With the decreasing cost of whole genome sequencing, there is a growing focus on performing pan-cancer analysis on non-coding regions of the genome. An increasing body of evidence suggests non-coding RNAs (ncRNAs) play an important role in biogenesis of cancer [16]. While some ncRNAs have been well studied, such as microRNAs [17], other types have been studied less extensively, including long ncRNAs (lncRNAs). LncRNAs are transcripts with a length greater than 200 nucleotides, exhibiting similar molecular characteristics to messenger RNAs (mRNAs) but lacking an appreciable potential to code for proteins [18,19]. Localised either in the nucleus or cytoplasm, lncRNAs form a complex network of interactions with DNA, RNA, and proteins [20]. Although it is still debatable whether the majority of lncRNAs are simply transcriptional noise, some have been found to be associated with important biological roles. For example, lncRNA *XIST* is known to initiate silencing of the inactive X chromosome during X inactivation [21,22]. More generally, studies have suggested lncRNAs as cis- and trans-acting regulators of gene expression via chromatin reprogramming [23,24]. They have also been implicated in post-transcriptional regulation, including mRNA translation [25], as well as cell differentiation and development [26]. Despite these findings, lncRNA functions remain poorly understood.

LncRNAs are engaged in many cellular functions, and their dysregulation has been linked to diseases, including cancer [16,27]. Aberrant expression of lncRNAs has been identified in different tumors including brain, breast, and colon cancer [28,29]. Many lncRNAs have also been shown to be regulated by oncogenes and tumor suppressors, suggesting a role in oncogenesis [30]. Furthermore, functional studies have revealed validated cancer roles for more than hundred lncRNAs in tumors [31]. A wide range of cancer treatments are currently available [32,33]; chemotherapy, however, continues to be preferred despite its diminishing effectiveness when cancer has advanced or metastasised [33,34]. Poor prognosis is probably due to late diagnoses of cancer, together with tumors having acquired drug resistance, and continues to be a major challenge in treating malignancies [34]. It is thus important to search for new biomarkers for early diagnosis and therapeutic targets for more effective treatments. A plethora of evidence has revealed dysregulation of lncRNAs to be associated with cell proliferation, apoptosis, invasion and drug resistance, processes found implicated in the pathogenesis of cancer [35,36]. These findings put forward lncRNAs as potential biomarkers and therapy agents.

At present, there is an increasing focus on identifying lncRNAs associated with tumorigenesis and elucidating their functional implications. Rich RNA-seq datasets are a promising tool for this purpose, but their use can be computationally challenging. To highlight the importance of lncRNA association with cancer and overcome these challenges, The Atlas of Noncoding RNAs in Cancer (TANRIC) was developed [37]. TANRIC is a free and interactive database which gives users access to genomic, proteomic, clinical and lncRNAs expression data of 8143 samples (tumorous and non-neoplastic) from TCGA and others.

In order to characterise common, aberrantly expressed lncRNAs we performed a pan-cancer analysis on lncRNA expression profiles from a TCGA derived dataset using TANRIC platform. We hypothesise that those found to be implicated in different cancer types may exhibit similar functional implications across cancers. To assess this hypothesis, we sought to identify dysregulated lncRNAs across eight TCGA cancer types and explored commonality among different malignancies. We then investigated their functional implications, by performing functional analysis and exploiting the similarity of enriched biological processes. Previous pan-cancer studies have focused on somatic mutations of whole genomes [15], tumor microenvironments [38] as well as proteomic profiles [39]. Putative functions have been studied for onco-lncRNAs dysregulated in multiple cancers [40] without focusing on common consistently dysregulated lncRNAs or exploring the similarity of gene ontologies across different cancer types as performed in our study.

## 2. Results

### 2.1. Common Dysregulated lncRNAs

In total, 9616 lncRNAs manifested significant differential expression across cancers (Table S1). Whilst similarity between pairs of cancers, represented by the Jaccard index, appears low (Figure S1), collectively the number of shared lncRNAs of one cancer type with the remaining types is quite high, with overlap ranging from ~80% to ~97% (Table S2). Of the large number of lncRNAs found to be overlapping among different cancer pairs, seven were observed to be differentially expressed in all cancer types (Table S3). Often the same lncRNA deemed upregulated in one cancer type can be found downregulated in another, or the other way around. Nonetheless, three lncRNAs were found to be consistently dysregulated across all cancers: ENSG00000235904 (*RBMS3-AS3*) (hereafter, "Antisense") and ENSG00000261472 (Novel transcript) (hereafter, "Novel") are both upregulated, and ENSG00000272455 (*MRPL20-DT*) (hereafter, "Divergent") is downregulated (Table S4). We chose to focus in the present study on the three consistently dysregulated lncRNAs: Antisense, Novel and Divergent. Correlation analysis revealed in total, 3141 coding genes ($|rs| \geq 0.5$ and $p$-value $\leq 0.01$), with 2185 mRNAs found to be co-expressed with Antisense, 69 mRNAs with Novel and 1026 mRNAs with Divergent (Figure 1a,b). We found little overlap between correlated mRNAs across different cancers. It appears that for a given lncRNA, the group of co-expressed mRNAs is specific for each cancer type (Figure 1c). The full list of correlated mRNAs can be found in Table S5.
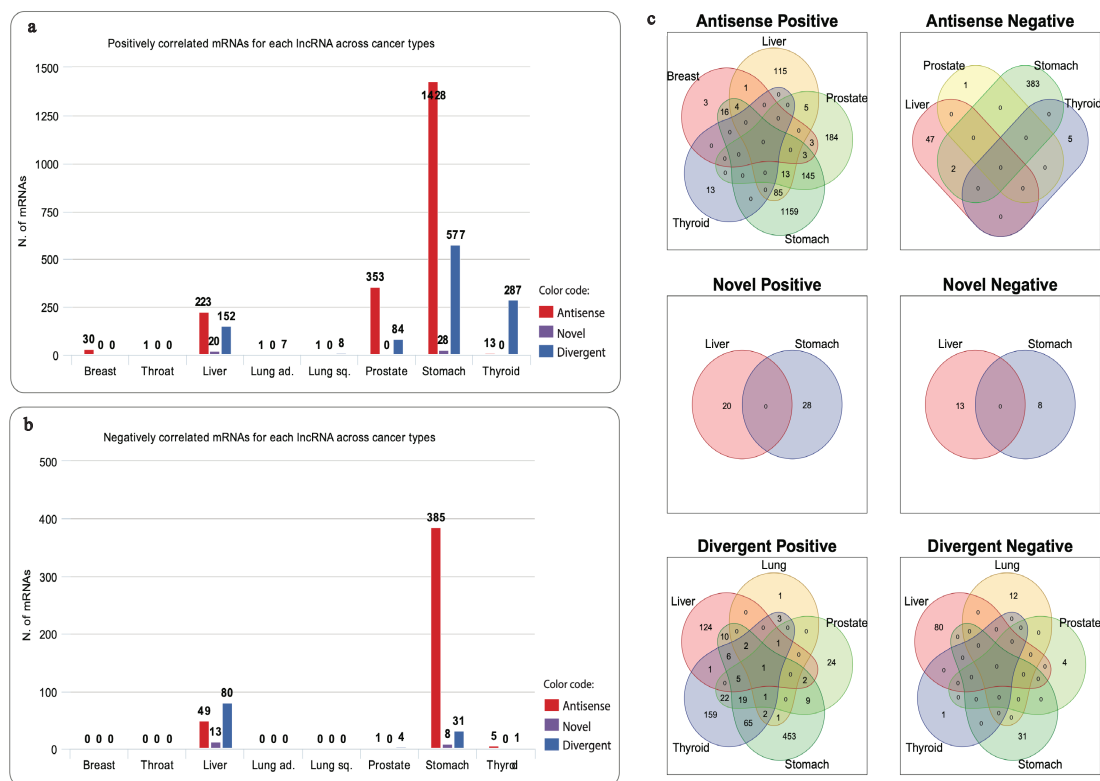


**Figure 1.** Bar plots showing the number of mRNAs related to the three lncRNAs, per cancer type. Red represents Antisense, purple represents Novel and blue represents Divergent. (**a**) number of positively correlated mRNAs. (**b**) number of negatively correlated mRNAs. (**c**) Venn diagram showing the overlap between lists of positively and negatively correlated mRNAs with lncRNAs: Antisense, Novel and Divergent, across different cancer types: breast (BRCA), thyroid (THCA), stomach (STAD), prostate (PRAD), lung (LUAD and LUSC) and liver (LIHC).

### 2.2. Gene Ontologies (GO) for Inference of Functional Similarity

After identifying and combining GO terms enriched by each lncRNA, additional filtering (FDR $\leq$ 0.05) resulted in the acquisition of three GO terms lists: Two lists of GO

terms associated with Antisense and one list with Divergent, with no records linked with lncRNA Novel (Figure S2). In order to provide functional elucidation of the remaining lncRNAs of interest, we explored similarity between GO term pairs using NaviGO and GO pairwise similarity networks were created (RSS ≥ 0.05).

Similarity Networks

Starting with GO list enriched by mRNAs positively correlated with Antisense, three clusters in the network of functionally similar GO terms were identified (Figure 2a). The first cluster contains GO terms predominantly involved in tissues and vessel morphogenesis, together with tissues and vasculature development (Figure 2b). The second cluster includes system and cellular processes such as actin-mediated cell contraction in addition to localisation and movement of cell and/or subcellular component (Figure 2c). Finally, GO terms in the third cluster found of extracellular matrix, structure organisation along with cell adhesion (Figure 2d). Network of mRNAs negatively correlated with Antisense, displayed GO terms appear to be mainly associated with ncRNA metabolic processes, particularly ribosomal RNA (rRNA), ribosome biogenesis, and ubiquitination (Figure 3). Lastly, mRNAs positively correlated with Divergent have enriched substantially similar GO terms, immersed with mRNA processing, splicing and metabolism, in addition to processes associated with cell cycle (Figure 4).
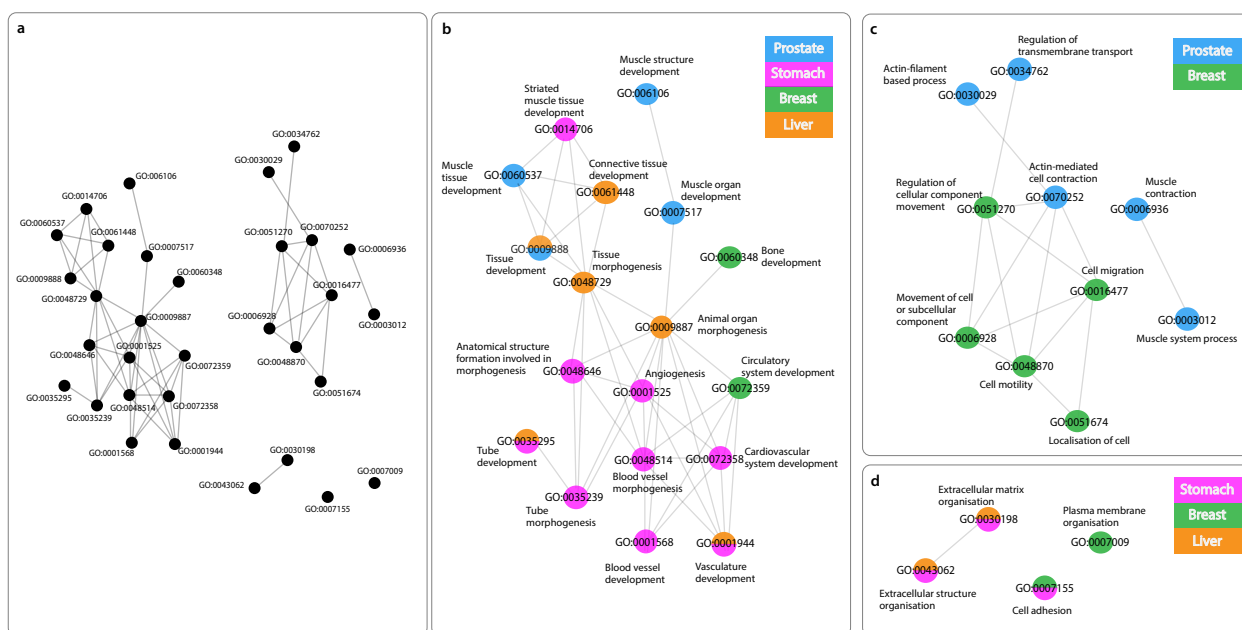


**Figure 2.** Functional similarity of GO terms associated with mRNAs of positive correlation with lncRNA Antisense. Network visualisation of functional similarity of GO terms enriched by positively correlated mRNAs with lncRNA "Antisense", across cancers (when applicable, RSS ≥ 0.05). (**a**) Three clusters can be visually identified. (**b**) Close up of cluster 1. (**c**) Close up of cluster 2. (**d**) Close up of cluster 3. (**b–d**): Each node indicates a GO term and the edges (lines in between) represent a functional similarity, representing relationships, such as "is a" or "part of". The color(s) of nodes represent the cancer type(s) in which GO term was found to be enriched, green for breast cancer (BRCA), orange for liver cancer (LIHC), blue for prostate cancer (PRAD) and purple for stomach cancer (STAD).
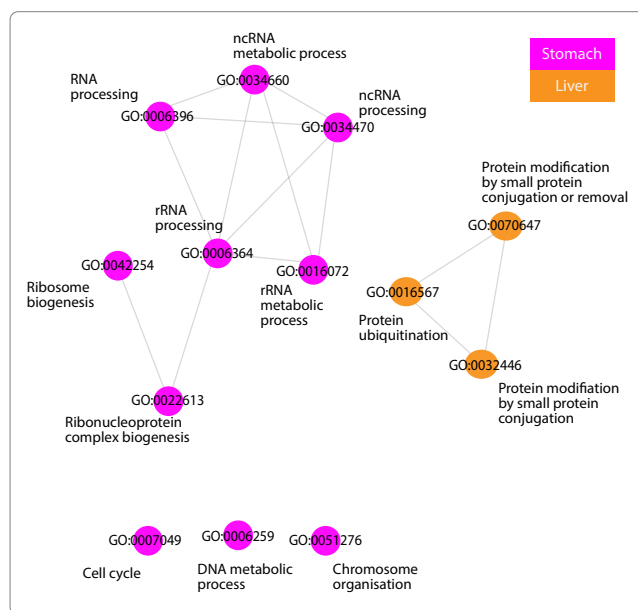
**Figure 3.** Functional similarity of GO terms associated with mRNAs of negative correlation with lncRNA Antisense. Network visualisation of functional similarity of GO terms enriched by negatively correlated mRNAs with lncRNA "Antisense", across cancers (when applicable, RSS ≥ 0.05). The color(s) of nodes represent the cancer type(s) in which GO term was found to be enriched, orange for liver cancer (LIHC) and purple for stomach cancer (STAD).
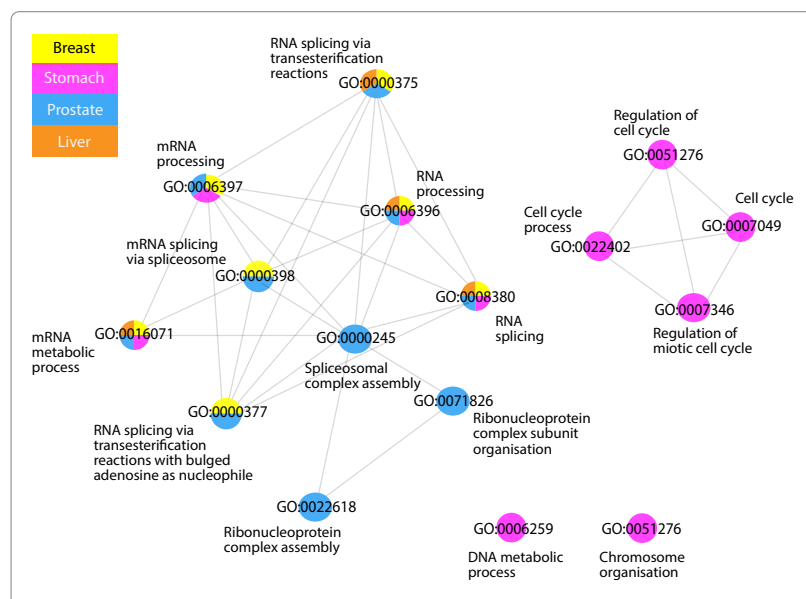


**Figure 4.** Functional similarity of GO terms associated with mRNAs of positive correlation with lncRNA Divergent. Network visualisation of functional similarity of GO terms enriched by negatively correlated mRNAs with lncRNA "Antisense", across cancers (when applicable, RSS ≥ 0.05). The color(s) of nodes represent the cancer type(s) in which GO term was found to be enriched, orange for liver cancer (LIHC), blue for prostate cancer (PRAD), purple for stomach cancer (STAD) and yellow for thyroid cancer (THCA).

Finally, when GO terms enriched by both lncRNAs were compared, high similarity was found between GO terms enriched by mRNAs negatively correlated with Antisense and those enriched by mRNAs positively correlated with Divergent (Figure S3). Interestingly, after further scrutinisation of the different networks, four GO terms were found to be

shared between the two: DNA metabolic process, chromosome organisation, cell cycle and RNA processing (Figure 5).
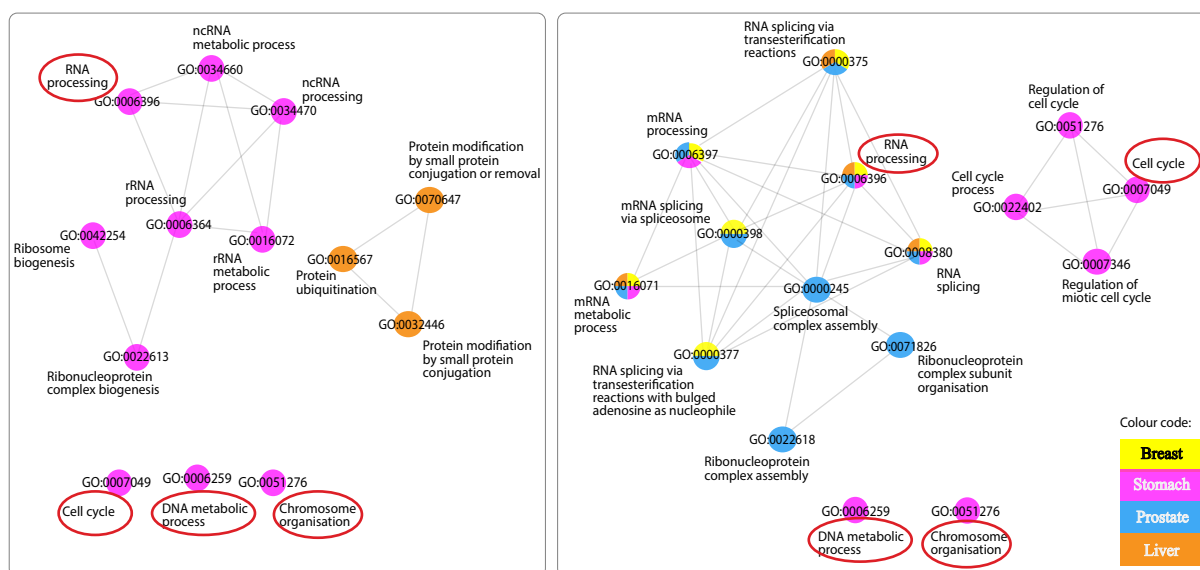


**Figure 5.** Comparison of Antisense and Divergent GO terms similarity networks. Similarity networks among GO terms enriched by mRNAs related negatively to Antisense (left) and mRNAs related positively to Divergent (right). Shared GO terms in both networks are encircled in red.

## 3. Discussion

### 3.1. Dysregulation of lncRNAs among Cancers

The identification of 9616 dysregulated lncRNAs suggests pervasive variation of lncRNA expression in cancers, consistent with previous studies [17,41]. Therefore, understanding functional implications of lncRNAs in malignancies is of high importance, as not only can this serve in developing diagnostic tools, but it can also lead to new treatment strategies. Upon exploring commonality of dysregulated lncRNAs, it was observed that tumors share a substantial number of genes (average overlap ~90%), suggesting that potentially the same lncRNAs could be associated with different tumors across different tissues. Whilst it has been suggested that lncRNAs are cancer specific, with tumors of different types and subtypes exhibiting different expression patterns [41,42], some evidence show otherwise. For instance, *MALAT1* was suggested to be involved in multiple tumors; inhibition of the well-studied lncRNA was found to prevent lung cancer metastasis [43]. Conversely, a more recent study showed that knocking out *MALAT1* actually promotes metastasis in breast cancer, suggesting its role as a metastasis suppressant [44]. Additionally, an oncogenic role has also been proposed for *MALAT1* in colorectal carcinoma [45]. Nonetheless, we observed 2855 lncRNAs to be dysregulated uniquely in one cancer type, denoting some level of specificity.

### 3.2. Three Consistently Dysregulated lncRNAs

Intriguingly, three lncRNAs showed striking consistent dysregulation: ENSG00000235904, known as *RBMS3-AS3* gene and ENSG00000261472, a novel transcript, both exhibited upregulation in all tumor samples whilst ENSG00000272455, known as *MRPL20-DT* gene, manifested downregulation.

#### 3.2.1. ENSG00000235904, RBMS3-AS3 "Antisense"

As with the majority of lncRNAs, little is known about the functional implication of *RBMS3-AS3* or its association with tumors. According to lncATLAS [46], *RBMS3-AS3* is found to be expressed mainly in the cytoplasm. Generally, cytoplasmic lncRNAs are believed, through formation of complexes with RNA binding proteins, to be involved

in different mechanisms, such as mRNA translation and stability [47–49] and protein localisation [50,51]. With regard to cancer, *RBMS3-AS3* has been proposed as a competing endogenous RNA (ceRNA), targeted by several miRNAs in breast cancer [52]. In addition, *RBMS3-AS3* was shown to be serving as miRNA sponge, acting as a tumor suppressor in prostate cancer [53]. We explored TANRIC's survival analysis for both of these cancer types and found the survival rate across patients to be higher in those who have lower expression of *RBMS3-AS3* (Kaplan–Meier analysis and log-rank test, *p*-value < 0.05). Furthermore, both gastric and colorectal cancers exhibited dysregulation of *RBMS3-AS3* and involvement in ceRNA network [54,55]. Taken together, *RBMS3-AS3* seems to display aberrant expression patterns across tumors; further studies are required to investigate its function and potential involvement in cancer.

### 3.2.2. ENSG00000261472 "Novel"

Although this lncRNA was also found to be consistently upregulated across tumors, little is known about its association with cancer. ENSG00002614172 is small in size (<500 bp), and unlike *RBMS3-AS3*, no localisation information was found in lncATLAS [46]. Expression across tissues was almost negligible, with body fat having the highest value of only 1.7 TPM, according to the Genotype Tissue Expression project GTEX [56]. Lack of information on this transcript is possibly due to it being newly annotated, and most importantly minimally expressed across tissues. In the literature, we came across a breast cancer analysis where ENSG00000261472 was listed among other enriched lncRNAs [57]. However, to the best of our knowledge, no other studies have been published with reference to cancer. Concisely, ENSG00000261472 is a novel lncRNA whose cellular function is yet to be discovered.

### 3.2.3. ENSG00000272455, MRPL20-DT "Divergent"

Similar to Novel, no localisation information was detected in lncATLAS for *MRPL20-DT* [46]. A recent study, however, found that *MRPL20-DT* promoter is consistently upregulated across 13 tumors [58]. Likewise, its upregulation amid other lncRNAs was reported in muscle invasive bladder cancer [59]. In contradiction, TANRIC's survival analysis displayed better survival probability for those with higher expression of *MRPL20-DT* in bladder cancer (Kaplan–Meier analysis and log-rank test, *p*-value < 0.05), which comes in conformity with our results of it being downregulated in malignancy. In essence, *MRPL20-DT* role is still undetermined, but evidence suggest its possible association with cancer. Future research is needed to investigate its dysregulation in tumors and better understand the molecular mechanisms involved.

### 3.3. Specificity of Correlated mRNAs

Co-expressed mRNAs lists, classified between positively and negatively correlated, ranged in size across different cancers and different lncRNAs, proposing a wide and diverse network of gene interactions across tumors. In addition, the number of correlated mRNAs of a given lncRNA was dependent on the tissue type. For instance, 1428 mRNAs were positively correlated with Antisense in stomach cancer sample set, compared to only 353 and 30 mRNAs in prostate and breast cancer, respectively (Figure 1a), suggesting some level of tissue specificity, which comes in accordance with previous findings [60]. Although there is little intersect between malignancies, mRNAs were noted to be predominantly different (Figure 1c), suggesting that, despite the commonality of the three lncRNAs of interest, they appear to be interacting with different mRNAs in different tissues. Taken together, lncRNAs seem to manifest both tissue-specific and ubiquitous relations, interacting with a broad range of genes across tumors.

### 3.4. Enriched GO Terms

GO enrichment analysis differentiated three GO term lists; no GO terms were found to be notably associated with Novel post-filtration, possibly because the number of correlated mRNAs was the lowest compared to Antisense and Divergent, hence, as a consequence

no significant ontology enrichment was detected. The absence of GO terms is somewhat surprising but does not undermine the possible involvement of this lncRNA with tumors. It is worth noting that there is evidence of enrichment of Novel in breast cancer [57], in addition to the present study where consistent upregulation was outlined across all eight cancer types. These initial findings are promising; further studies would make a worthwhile contribution, to better understand the underlying mechanisms related to cancer.

### 3.5. Networks of Functionally Similar GO Terms

In order to understand the functional similarities of the two remaining lncRNAs, Antisense and Divergent, we identified three GO similarity networks. It is worth noting that the scoring scheme (RSS) we adopted in creating these networks showed minimal variation when compared with Resnik and Lin's semantic similarity, another two widely used and known measures [61,62]. Whilst this increases our confidence with results presented here, a caveat with this approach is that network representation differs slightly based on the cut-off used with these scoring schemes. With that said, this is generally the case with many analyses and statistical tests which rely on arbitrary cut-off values, and in our analysis, cut-off value does not change the number or nature of biological processes involved, rather the way they are represented in a network.

### 3.5.1. GO Terms of Positively Related mRNAs with "Antisense"

Three clusters can be identified in this network (Figure 2a). The first cluster comprised of biological processes linked to angiogenesis, blood vessel and tissue morphogenesis as well as vasculature development, all known to be critical for cancer growth (Figure 2b). For instance, it is well-established that angiogenesis is one of the hallmarks of cancer [3]; tumor cells recruit new blood vessels to allow for nutrients and oxygen delivery, as well to be able to metastasise to other tissues [63,64], and this also involves the development of new blood vessels and tissues. This is of importance as it implies that Antisense might be involved with pivotal mechanisms of tumorigenesis. Simultaneously, the second cluster encompassed GO terms of cell motility and migration along with actin filament-based processes (Figure 2c). These processes are also linked to those seen in the first cluster; taking cell motility for example, this is essential in allowing tumor cells to enter the vasculature, transport through blood vessels and invade other sites [65]. Moreover, networks of actin protein filaments form actin cytoskeleton, involved primarily in cell migration and motility in cancer, leading to metastasis [66,67]. Finally, the third cluster involved extracellular matrix and structure organisation together with cell adhesion, processes also associated with cancer progression (Figure 2d). Indeed, there has been a focus on understanding the dysregulation of the extracellular matrix in complex diseases such as cancer. Being the major component of the tumor 'microenvironment', it has been suggested to modulate cell behavior and influence adhesion and migration of cells [68,69]. Collectively, the GO terms presented in this network appear to be closely related, describing vital processes for the proliferation and progression of malignancies. The question remains though, whether Antisense is exerting a regulatory role in this network or is simply a by-pass product. Further investigation is required to understand its potential role.

### 3.5.2. GO Terms of Negatively Related mRNAs with "Antisense"

The smaller list of negatively correlated mRNAs with Antisense enriched important functions relating to ncRNA metabolic processes, particularly rRNA, ribosome biogenesis, and ubiquitination (Figure 3). Mounting evidence suggests that impaired ribosomal activity drives tumorigenesis [70,71]. Ribosome biogenesis is an important regulator of cellular activities, including cell growth and cell cycle progression [72,73]; an increase in rRNA processing is observed during G1 of interphase, in preparation for protein translation, whilst during mitosis, downregulation of ribosomal activity is needed to signal the ending of cell cycle. Uncontrolled cell proliferation, a common feature in cancer, is a consequence of impaired ribosomal activity. Furthermore, it is now believed that perturbation of ribosomal

biogenesis is sufficient to lead to malignant transformation [74]. Another process found is ubiquitination (also known as ubiquitylation), a post-translational mechanism in which proteins are tagged by the conjugation of ubiquitin, for modification. Ubiquitin is a small regulatory protein that is highly conserved in eukaryotes, most commonly found to initiate proteins degradation, apart from also altering protein–protein interactions and modulating cellular processes such as cell cycle, apoptosis cell signaling and DNA repair [75,76]. It has been shown that cytoplasmic lncRNAs interfere with protein expression by either obstructing or promoting ubiquitination [77]. With a balanced ribosomal genesis for instance, tumor suppressor protein p53 is usually post-translationally downregulated through ubiquitination. However, studies have revealed that disruption of ribosomal biogenesis primarily causes activation of p53, and consequently disrupts its degradation through ubiquitination [70]. Hence, it comes as no surprise that these processes are interconnected, and their perturbation is associated with carcinogenesis. lncRNA Antisense, identified in this study, has been shown to be associated with a range of important biochemical processes, explicitly ubiquitination. Currently, novel strategies are being developed to target certain pathways in which ubiquitin is primarily involved, resulting in potentiation of drug efficacy, and overcoming multi drug resistance [78].

### 3.5.3. GO Terms of Positively Related mRNAs with "Divergent"

Finally, the network similar GO terms related with "Divergent" displayed processes, enriched in more than one cancer type, relating to mRNA processing, in particular, splicing pathways in addition to cell cycle regulation, specifically mitosis (Figure 4). Accounting primarily for protein diversity, splicing is a fundamental step of mRNA processing where the same coding gene can have different, even at times, opposing functional transcripts called isoforms. It has been revealed that aberrant splicing is linked with cancer, yielding cancer-specific isoforms favorable for tumor growth [79,80]. In addition, defective splicing also perturbs the cell cycle, which is also enriched in this network. Consistent with the literature, lncRNAs have been suggested to influence splicing in diseases such as cancer [81]. Moreover, new mechanism of cell death modulation has also been linked to lncRNA through interaction with protein factors, leading to apoptosis resistance [82]. However, the downfall of these findings is that the underlying mechanisms still remain largely unknown. Nonetheless, Divergent has been presented in this study to be consistently downregulated across tumors and seems to be associated with pivotal processes, involved in cancer growth and development; investigating its possible role in splicing events and cell growth may be of benefit.

### 3.5.4. GO Terms Shared across Networks

We reported four processes to be shared by both lncRNAs (Figure 5). Interestingly, these processes were found in the network of negatively correlated mRNAs with Antisense, reported to be upregulated, and that of positively correlated mRNAs with Divergent, found to be downregulated in cancers. Taken together, it appears that although these two lncRNAs are interacting with different sets of mRNAs across different cancers, and possibly through different mechanisms with one being upregulated and the other shown to be downregulated; they are both enriching substantially similar processes, found to be fundamental in cancer proliferation and progression.

### 3.6. Evaluation of lncRNA DE on Independent Datasets

We were interested to evaluate if our findings are generalisable to other cancer types not included in the discovery phase of the three lncRNAs. Accordingly, we assessed the dysregulation of Antisense, Novel, and Divergent on independent TCGA and non-TCGA datasets obtained from TANRIC repository. These include the total of 1047 tumor samples and their related adjacent 218 non-neoplastic tissue samples across five cancer subtypes, four of which are sourced from TCGA including urothelial bladder carcinoma, BLCA (19 non-neoplastic and 252 tumor samples), kidney chromophobe, KICH (25 non-neoplastic and 66 tumor), kidney renal clear cell carcinoma, KIRC (67 non-neoplastic and 448 tumor), and cervical kidney renal papillary cell carcinoma, KIRP (30 non-neoplastic and 198 tumor samples), and one is a non-TCGA dataset collected in Seoul, South Korea comprising 77 non-neoplastic and 83 tumor samples of lung adenocarcinoma (LUAD_KOREA) [83]. We investigated the DE of the three lncRNAs of interest across each dataset. In addition to the *t*-test, we estimated *p*-values using a non-parametric test (i.e., the two-sample Wilcoxon test or the Mann–Whitney test) to mitigate the biases associated with the underlying assumptions of hypothesis tests. The *p*-values were adjusted for multiple hypothesis testing using the Bonferroni correction. In agreement with our results, we observed the three lncRNAs to be consistently dysregulated across all five cancer types, with adjusted *p*-value $\leq 0.01$ (Figure 6 and Table S6). While the direction of dysregulation can be different in the validation datasets compared to the discovery datasets, the significance of DE (independent of the hypothesis test) supports the potential role of these lncRNAs across diverse cancer types and provides more evidence that lncRNAs identified in this study could serve in potentially developing new pan-cancer biomarkers or therapeutic agents.

### 3.7. Future Direction

The foundational processes identified in this study, such as angiogenesis, underly all tumors regardless of tissue type. Understanding the role of lncRNAs in enriching these perturbations is thus very important and can be advantageous, particularly when malignant cells spread to other tissues, leading to current treatment strategies in becoming somewhat ineffective. A possible future direction would be to use other computational approaches and databases to decipher putative functions of lncRNAs. For instance, taking an integrative approach considering transcriptomic and epigenomic data (e.g., methylation profile [84]), or genomic changes such as copy number alterations (CNAs) [85]. In addition, using UCSC genome browser [86], we could also examine the genomic location of lncRNAs along with co-expressed coding genes in order to explore cis and trans-relationships, which can be an initial step in discovering potential regulatory roles. Further analysis of the dysregulation of the identified lncRNAs can also be performed across other datasets, allowing further evidence to be provided for their dysregulation across other cancer types. Moreover, we have focused in the present study on commonly dysregulated lncRNAs, the complimentary future approach could be to explore lncRNAs that are specific to each cancer type/subtype, with the aim of investigating cancer-specific functions, which can also be of benefit.
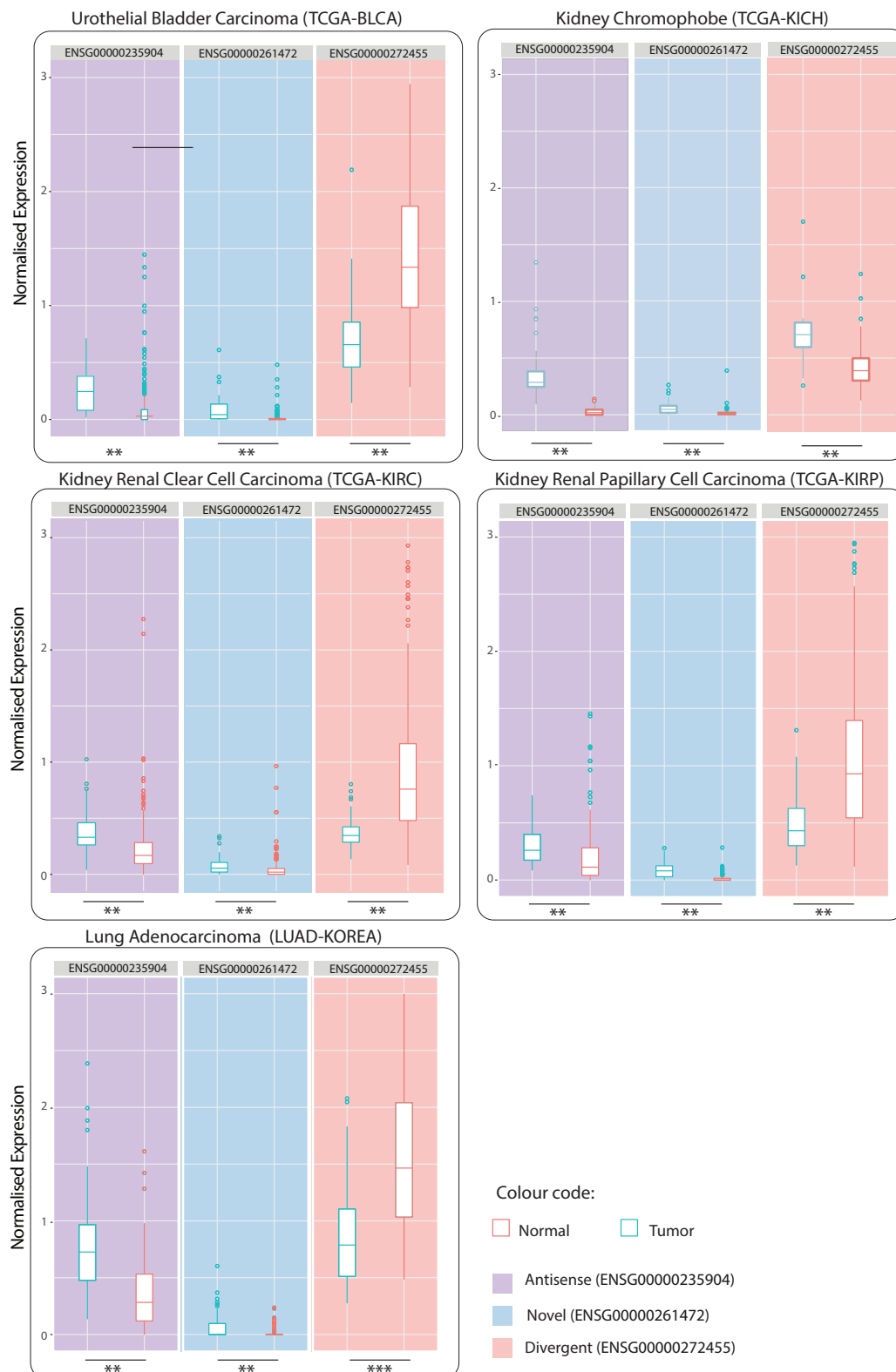
**Figure 6.** Boxplots of normalised expression values of lncRNAs Antisense, Novel and Divergent, in tumor versus non-neoplastic samples across five other cancer types which have not been used during the discovery phase of these lncRNAs. These include TCGA datasets, namely urothelial bladder carcinoma (BLCA), kidney chromophobe (KICH), kidney renal clear cell carcinoma (KIRC), and cervical kidney renal papillary cell carcinoma (KIRP), and one non-TCGA lung adenocarcinoma dataset collected in Seoul, South Korea (LUAD_KOREA). Significant adjusted *p*-values (adj $p \leq 0.01$) are represented by a star, from left to right: the *t*-test and the Wilcoxon test.

## 4. Materials and Methods

### 4.1. Data Source

We used TANRIC to access expression data of lncRNAs of 3326 tumor samples and their related adjacent 416 non-neoplastic tissue samples for eight cancer subtypes as categorised in TCGA. Included were 105 non-neoplastic and 837 tumor samples of breast invasive carcinoma (BRCA), 58 non-neoplastic and 488 tumor samples of lung adenocarcinoma (LUAD); 17 non-neoplastic and 220 tumor samples of lung squamous cell carcinoma (LUSC); 52 non-neoplastic and 374 tumor samples of prostate adenocarcinoma (PRAD); 59 non-neoplastic and 497 tumor samples of thyroid carcinoma (THCA); 33 non-neoplastic and 285 tumor samples of stomach adenocarcinoma (STAD), 50 non-neoplastic and 200 tumor samples of liver hepatocellular carcinoma (LIHC); and 42 non-neoplastic and 425 tumor samples of head and neck squamous cell carcinoma (HNSC). An overview of the workflow of this study is shown in Figure S4.

### 4.2. Differential Exprssion Analysis and Commonality Exploration

We carried out differential expression analysis by comparing lncRNA expression levels between tumor and related adjacent non-neoplastic samples of a given set. Expression data for a total of 12,727 lncRNAs was downloaded from TANRIC v2.0 for each of the eight TCGA cancer types. Differentially expressed lncRNAs (up- and downregulated) were identified ($|\log_2(FC)| > 1$) and Student's $t$-test was applied to calculate $p$-values followed by the Benjamini and Hochberg method to control the false discovery rate (FDR) [87]. Differentially expressed lncRNAs with adjusted $p$-value $\leq 0.01$ were considered statistically significant. We then identified common dysregulated lncRNAs (found in 2+ cancers) and unique dysregulated lncRNAs (specific to a given cancer type). Commonality was evaluated between each pair of cancers and represented by the Jaccard index (*J*) [88]. These analyses were implemented in MATLAB with the code available on the GitHub repository (https://github.com/VafaeeLab/PanCancer-lncRNAs).

#### 4.2.1. mRNA Correlation Analysis

Guided by the "guilt by association" principle [89], we utilised TANRIC to explore mRNAs correlated with common dysregulated lncRNAs across each cancer type. Spearman rank correlation coefficient (rs) were calculated to examine correlation relationships between lncRNAs of interest and mRNAs expression. Lists of correlated mRNAs were extracted based on cut-offs of rs $\geq 0.5$ or rs $\leq -0.5$, for positive and negative correlation, respectively, with correlation $p$-value $\leq 0.01$.

#### 4.2.2. Functional Enrichment Analysis

Exploration of large sets of genes can be achieved by organising them based on common functional features and one of the most widely used ways to understand genes and their products is to explore gene ontologies (GO) [90,91]. Thus, to investigate functional implications of lncRNAs of interest, we performed GO enrichment analysis with particular focus on biological processes. We exploited WebGestalt (WEB-based gene set analysis toolkit), to identify GO terms enriched by mRNAs lists, found to be correlated with common differentially expressed lncRNAs, with the aim of perusing their functional role [92,93]. Statistical analysis of GO enrichment was performed using a Fisher's exact test with a hypergeometric null distribution [94,95]. Significantly enriched GO terms were determined as FDR $\leq 0.05$.

#### 4.2.3. GO Similarity Analysis

GO, considered as a universal vocabulary, is structured as a hierarchical directed acyclic graph where each node represents a class of gene function (GO term), and the connection between two GO terms indicates different relationships. This hierarchy allows exploring semantic similarities among enriched GO terms, which could imply functional similarities between the associated genes [61,96]. Following the annotation of mRNAs lists by ontology, we merged GO terms enriched by each lncRNA across different cancers

and distinguished two separate groups: GO terms enriched by positively and negatively correlated genes. After additional filtering, based on FDR ≤0.05, GO terms were then further investigated. In order to measure closeness of GO terms using NaviGO, an interactive software that allows the retrieval of functional similarity scores and visualisation as networks [97]. From the six different scoring schemes offered by NaviGO, we relied on Relevance Semantic Similarity (RSS), which measures the relative depth and rareness of the biological processes involved [61]. RSS ranges from 0 to 1, with 0 representing zero similarity and 1 indicating very high similarity. Functional similarity networks were created using the NaviGO visualiser based on threshold RSS ≥ 0.5.

## 5. Conclusions

Once seen as transcriptional by-products, lncRNAs are now emerging as key players in cellular function, regulating a wide range of biological processes, and involved in their disruption. Our study provides evidence that lncRNAs may be contributing to the hallmarks of cancer, regardless of cancer type. These results are promising, suggesting that lncRNAs can serve as potential therapeutic targets to be applied in multiple cancer subtypes; future treatment strategies may potentially include non-coding genes in addition to specific protein targets. Indeed, lncRNAs have brought a promising new era to cancer biology, especially in terms of diagnosis and therapy.

## References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef]
2. Esteller, M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet.* **2007**, *8*, 286–298. [CrossRef]
3. You, J.S.; Jones, P.A. Cancer genetics and epigenetics: Two sides of the same coin? *Cancer Cell* **2012**, *22*, 9–20. [CrossRef] [PubMed]
4. Weinberg, R.; Hanahan, D. The hallmarks of cancer. *Cell* **2000**, *100*, 57–70.
5. Stratton, M.R.; Campbell, P.J.; Futreal, P.A. The cancer genome. *Nature* **2009**, *458*, 719–724. [CrossRef] [PubMed]
6. Pleasance, E.D.; Cheetham, R.K.; Stephens, P.J.; McBride, D.J.; Humphray, S.J.; Greenman, C.D.; Varela, I.; Lin, M.-L.; Ordóñez, G.R.; Bignell, G.R. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **2010**, *463*, 191–196. [CrossRef] [PubMed]
7. Ley, T.J.; Mardis, E.R.; Ding, L.; Fulton, B.; McLellan, M.D.; Chen, K.; Dooling, D.; Dunford-Shore, B.H.; McGrath, S.; Hickenbotham, M. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **2008**, *456*, 66–72. [CrossRef] [PubMed]

8.    Hutter, C.; Zenklusen, J.C. The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. *Cell* **2018**, *173*, 283–285. [CrossRef]
9.    Wei, L.; Jin, Z.; Yang, S.; Xu, Y.; Zhu, Y.; Ji, Y. TCGA-assembler 2: Software pipeline for retrieval and processing of TCGA/CPTAC data. *Bioinformatics* **2018**, *34*, 1615–1617. [CrossRef]
10.   Boehm, J.S.; Garnett, M.J.; Adams, D.J.; Francies, H.E.; Golub, T.R.; Hahn, W.C.; Iorio, F.; McFarland, J.M.; Parts, L.; Vazquez, F. Cancer research needs a better map. *Nature* **2021**, *589*, 514–516. [CrossRef]
11.   Tomczak, K.; Czerwińska, P.; Wiznerowicz, M. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Contemp. Oncol.* **2015**, *19*, A68. [CrossRef]
12.   Colaprico, A.; Silva, T.C.; Olsen, C.; Garofano, L.; Cava, C.; Garolini, D.; Sabedot, T.S.; Malta, T.M.; Pagnotta, S.M.; Castiglioni, I. TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **2016**, *44*, e71. [CrossRef]
13.   Ciriello, G.; Miller, M.L.; Aksoy, B.A.; Senbabaoglu, Y.; Schultz, N.; Sander, C. Emerging landscape of oncogenic signatures across human cancers. *Nat. Genet.* **2013**, *45*, 1127–1133. [CrossRef]
14.   Weinstein, J.N.; Collisson, E.A.; Mills, G.B.; Shaw, K.R.M.; Ozenberger, B.A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J.M. The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* **2013**, *45*, 1113–1120. [CrossRef]
15.   The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* **2020**, *578*, 82.
16.   Adams, B.D.; Parsons, C.; Walker, L.; Zhang, W.C.; Slack, F.J. Targeting noncoding RNAs in disease. *J. Clin. Investig.* **2017**, *127*, 761–771. [CrossRef]
17.   Schmitt, A.M.; Chang, H.Y. Long noncoding RNAs in cancer pathways. *Cancer Cell* **2016**, *29*, 452–463. [CrossRef]
18.   Zampetaki, A.; Albrecht, A.; Steinhofel, K. Long non-coding RNA structure and function: Is there a link? *Front. Physiol.* **2018**, *9*, 1201. [CrossRef]
19.   Ulitsky, I. Evolution to the rescue: Using comparative genomics to understand long non-coding RNAs. *Nat. Rev. Genet.* **2016**, *17*, 601–614. [CrossRef] [PubMed]
20.   Rinn, J.L.; Chang, H.Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **2012**, *81*, 145–166. [CrossRef] [PubMed]
21.   Penny, G.D.; Kay, G.F.; Sheardown, S.A.; Rastan, S.; Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **1996**, *379*, 131–137. [CrossRef]
22.   Plath, K.; Mlynarczyk-Evans, S.; Nusinow, D.A.; Panning, B. Xist RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.* **2002**, *36*, 233–278. [CrossRef]
23.   Geisler, S.; Coller, J. RNA in unexpected places: Long non-coding RNA functions in diverse cellular contexts. *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 699–712. [CrossRef] [PubMed]
24.   Tsai, M.-C.; Manor, O.; Wan, Y.; Mosammaparast, N.; Wang, J.K.; Lan, F.; Shi, Y.; Segal, E.; Chang, H.Y. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **2010**, *329*, 689–693. [CrossRef]
25.   Prensner, J.R.; Chinnaiyan, A.M. The emergence of lncRNAs in cancer biology. *Cancer Discov.* **2011**, *1*, 391–407. [CrossRef] [PubMed]
26.   Fatica, A.; Bozzoni, I. Long non-coding RNAs: New players in cell differentiation and development. *Nat. Rev. Genet.* **2014**, *15*, 7–21. [CrossRef] [PubMed]
27.   Anastasiadou, E.; Jacob, L.S.; Slack, F.J. Non-coding RNA networks in cancer. *Nat. Rev. Cancer* **2018**, *18*, 5–18. [CrossRef]
28.   Ellis, B.C.; Molloy, P.L.; Graham, L.D. CRNDE: A long non-coding RNA involved in cancer, neurobiology, and development. *Front. Genet.* **2012**, *3*, 270. [CrossRef]
29.   Gupta, R.A.; Shah, N.; Wang, K.C.; Kim, J.; Horlings, H.M.; Wong, D.J.; Tsai, M.-C.; Hung, T.; Argani, P.; Rinn, J.L. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **2010**, *464*, 1071–1076. [CrossRef]
30.   Huarte, M.; Guttman, M.; Feldser, D.; Garber, M.; Koziol, M.J.; Kenzelmann-Broz, D.; Khalil, A.M.; Zuk, O.; Amit, I.; Rabani, M. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **2010**, *142*, 409–419. [CrossRef]
31.   Carlevaro-Fita, J.; Lanzós, A.; Feuerbach, L.; Hong, C.; Mas-Ponte, D.; Pedersen, J.S.; Johnson, R. Cancer LncRNA Census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis. *Commun. Biol.* **2020**, *3*, 56. [CrossRef] [PubMed]
32.   Schrama, D.; Reisfeld, R.A.; Becker, J.C. Antibody targeted drugs as cancer therapeutics. *Nat. Rev. Drug Discov.* **2006**, *5*, 147–159. [CrossRef] [PubMed]
33.   Mansoori, B.; Mohammadi, A.; Davudian, S.; Shirjang, S.; Baradaran, B. The different mechanisms of cancer drug resistance: A brief review. *Adv. Pharm. Bull.* **2017**, *7*, 339. [CrossRef] [PubMed]
34.   Bukowski, K.; Kciuk, M.; Kontek, R. Mechanisms of multidrug resistance in cancer chemotherapy. *Int. J. Mol. Sci.* **2020**, *21*, 3233. [CrossRef]
35.   Peng, K.; Liu, R.; Yu, Y.; Liang, L.; Yu, S.; Xu, X.; Liu, T. Identification and validation of cetuximab resistance associated long noncoding RNA biomarkers in metastatic colorectal cancer. *Biomed. Pharmacother.* **2018**, *97*, 1138–1146. [CrossRef] [PubMed]
36.   Peng, F.; Wang, R.; Zhang, Y.; Zhao, Z.; Zhou, W.; Chang, Z.; Liang, H.; Zhao, W.; Qi, L.; Guo, Z. Differential expression analysis at the individual level reveals a lncRNA prognostic signature for lung adenocarcinoma. *Mol. Cancer* **2017**, *16*, 1–12. [CrossRef]
37.   Li, J.; Han, L.; Roebuck, P.; Diao, L.; Liu, L.; Yuan, Y.; Weinstein, J.N.; Liang, H. TANRIC: An Interactive Open Platform to Explore the Function of lncRNAs in CancerThe Atlas of Noncoding RNAs in Cancer. *Cancer Res.* **2015**, *75*, 3728–3737. [CrossRef]
38.   Aran, D.; Sirota, M.; Butte, A.J. Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* **2015**, *6*, 8971. [CrossRef] [PubMed]
39.   Akbani, R.; Ng, P.K.S.; Werner, H.M.; Shahmoradgoli, M.; Zhang, F.; Ju, Z.; Liu, W.; Yang, J.-Y.; Yoshihara, K.; Li, J. A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nat. Commun.* **2014**, *5*, 1–15. [CrossRef]

40. Cabanski, C.R.; White, N.M.; Dang, H.X.; Silva-Fisher, J.M.; Rauck, C.E.; Cicka, D.; Maher, C.A. Pan-cancer transcriptome analysis reveals long noncoding RNAs with conserved function. *RNA Biol.* **2015**, *12*, 628–642. [CrossRef]

41. Yan, X.; Hu, Z.; Feng, Y.; Hu, X.; Yuan, J.; Zhao, S.D.; Zhang, Y.; Yang, L.; Shan, W.; He, Q. Comprehensive genomic characterization of long non-coding RNAs across human cancers. *Cancer Cell* **2015**, *28*, 529–540. [CrossRef]

42. Gutschner, T.; Diederichs, S. The hallmarks of cancer: A long non-coding RNA point of view. *RNA Biol.* **2012**, *9*, 703–719. [CrossRef] [PubMed]

43. Gutschner, T.; Hämmerle, M.; Eißmann, M.; Hsu, J.; Kim, Y.; Hung, G.; Revenko, A.; Arun, G.; Stentrup, M.; Groß, M. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res.* **2013**, *73*, 1180–1189. [CrossRef] [PubMed]

44. Kim, J.; Piao, H.-L.; Kim, B.-J.; Yao, F.; Han, Z.; Wang, Y.; Xiao, Z.; Siverly, A.N.; Lawhon, S.E.; Ton, B.N. Long noncoding RNA MALAT1 suppresses breast cancer metastasis. *Nat. Genet.* **2018**, *50*, 1705–1715. [CrossRef] [PubMed]

45. Sun, Z.; Ou, C.; Liu, J.; Chen, C.; Zhou, Q.; Yang, S.; Li, G.; Wang, G.; Song, J.; Li, Z. YAP1-induced MALAT1 promotes epithelial–mesenchymal transition and angiogenesis by sponging miR-126-5p in colorectal cancer. *Oncogene* **2019**, *38*, 2627–2644. [CrossRef]

46. Mas-Ponte, D.; Carlevaro-Fita, J.; Palumbo, E.; Pulido, T.H.; Guigo, R.; Johnson, R. LncATLAS database for subcellular localization of long noncoding RNAs. *Rna* **2017**, *23*, 1080–1087. [CrossRef]

47. Gong, C.; Maquat, L.E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature* **2011**, *470*, 284–288. [CrossRef]

48. Kretz, M. TINCR, staufen1, and cellular differentiation. *RNA Biol.* **2013**, *10*, 1597–1601. [CrossRef]

49. Yoon, J.-H.; Abdelmohsen, K.; Srikantan, S.; Yang, X.; Martindale, J.L.; De, S.; Huarte, M.; Zhan, M.; Becker, K.G.; Gorospe, M. LincRNA-p21 suppresses target mRNA translation. *Mol. Cell* **2012**, *47*, 648–655. [CrossRef]

50. Wang, P.; Xue, Y.; Han, Y.; Lin, L.; Wu, C.; Xu, S.; Jiang, Z.; Xu, J.; Liu, Q.; Cao, X. The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation. *Science* **2014**, *344*, 310–313. [CrossRef]

51. Liu, B.; Sun, L.; Liu, Q.; Gong, C.; Yao, Y.; Lv, X.; Lin, L.; Yao, H.; Su, F.; Li, D. A cytoplasmic NF-κB interacting long noncoding RNA blocks IκB phosphorylation and suppresses breast cancer metastasis. *Cancer Cell* **2015**, *27*, 370–381. [CrossRef] [PubMed]

52. Wang, X.; Wan, J.; Xu, Z.; Jiang, S.; Ji, L.; Liu, Y.; Zhai, S.; Cui, R. Identification of competitive endogenous RNAs network in breast cancer. *Cancer Med.* **2019**, *8*, 2392–2403. [CrossRef] [PubMed]

53. Jiang, Z.; Zhang, Y.; Chen, X.; Wu, P.; Chen, D. Long noncoding RNA RBMS3-AS3 acts as a microRNA-4534 sponge to inhibit the progression of prostate cancer by upregulating VASH1. *Gene Ther.* **2020**, *27*, 143–156. [CrossRef] [PubMed]

54. Fan, Q.; Liu, B. Comprehensive analysis of a long noncoding RNA-associated competing endogenous RNA network in colorectal cancer. *OncoTargets Ther.* **2018**, *11*, 2453–2466. [CrossRef]

55. Peng, J.; Zhu, Y.; Dong, X.; Mao, X.; Lou, Y.; Mu, Y.; Xue, D.; Zhou, H. Construction and analysis of lncRNA-associated ceRNA network identified potential prognostic biomarker in gastric cancer. *Transl. Cancer Res.* **2019**, *8*, 1116–1128. [CrossRef] [PubMed]

56. Lonsdale, J.; Thomas, J.; Salvatore, M.; Phillips, R.; Lo, E.; Shad, S.; Hasz, R.; Walters, G.; Garcia, F.; Young, N. The genotype-tissue expression (GTEx) project. *Nat. Genet.* **2013**, *45*, 580–585. [CrossRef]

57. Vishnubalaji, R.; Shaath, H.; Elkord, E.; Alajez, N.M. Long non-coding RNA (lncRNA) transcriptional landscape in breast cancer identifies LINC01614 as non-favorable prognostic biomarker regulated by TGFβ and focal adhesion kinase (FAK) signaling. *Cell Death Discov.* **2019**, *5*, 109. [CrossRef]

58. Imada, E.L.; Sanchez, D.F.; Collado-Torres, L.; Wilks, C.; Matam, T.; Dinalankara, W.; Stupnikov, A.; Lobo-Pereira, F.; Yip, C.-W.; Yasuzawa, K. Recounting the FANTOM CAGE-associated transcriptome. *Genome Res.* **2020**, *30*, 1073–1081. [CrossRef]

59. Sheng, H.; Zhang, G.; Huang, Y.; Sun, L.; Shi, G.; Ye, D. A 5-lncRNA Signature Associated with Smoking Predicts the Overall Survival of Patients with Muscle-Invasive Bladder Cancer. *Dis. Markers* **2021**, *2021*, 8839747. [CrossRef]

60. Jiang, C.; Li, Y.; Zhao, Z.; Lu, J.; Chen, H.; Ding, N.; Wang, G.; Xu, J.; Li, X. Identifying and functionally characterizing tissue-specific and ubiquitously expressed human lncRNAs. *Oncotarget* **2016**, *7*, 7120. [CrossRef]

61. Schlicker, A.; Domingues, F.S.; Rahnenführer, J.; Lengauer, T. A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinform.* **2006**, *7*, 302. [CrossRef] [PubMed]

62. Resnik, P. Using information content to evaluate semantic similarity in a taxonomy. *arXiv* **1995**, arXiv:cmp-lg/9511007.

63. Hanahan, D.; Weinberg, R.A. Hallmarks of Cancer: The Next Generation. *Cell* **2011**, *144*, 646–674. [CrossRef] [PubMed]

64. Carmeliet, P.; Jain, R. Angiogenesis in cancer and other diseases. *Nature* **2000**, *407*, 249–257. [CrossRef] [PubMed]

65. Stuelten, C.H.; Parent, C.A.; Montell, D.J. Cell motility in cancer invasion and metastasis: Insights from simple model organisms. *Nat. Rev. Cancer* **2018**, *18*, 296–312. [CrossRef]

66. Yamazaki, D.; Kurisu, S.; Takenawa, T. Regulation of cancer cell motility through actin reorganization. *Cancer Sci.* **2005**, *96*, 379–386. [CrossRef]

67. Olson, M.F.; Sahai, E. The actin cytoskeleton in cancer cell motility. *Clin. Exp. Metastasis* **2009**, *26*, 273. [CrossRef] [PubMed]

68. Lu, P.; Weaver, V.M.; Werb, Z. The extracellular matrix: A dynamic niche in cancer progression. *J. Cell Biol.* **2012**, *196*, 395–406. [CrossRef]

69. Multhaupt, H.A.; Leitinger, B.; Gullberg, D.; Couchman, J.R. Extracellular matrix component signaling in cancer. *Adv. Drug Deliv. Rev.* **2016**, *97*, 28–40. [CrossRef]

70. Turi, Z.; Lacey, M.; Mistrik, M.; Moudry, P. Impaired ribosome biogenesis: Mechanisms and relevance to cancer and aging. *Aging* **2019**, *11*, 2512. [CrossRef]

71.  Narla, A.; Ebert, B.L. Ribosomopathies: Human disorders of ribosome dysfunction. *Blood* **2010**, *115*, 3196–3205. [CrossRef]
72.  Mayer, C.; Grummt, I. Ribosome biogenesis and cell growth: mTOR coordinates transcription by all three classes of nuclear RNA polymerases. *Oncogene* **2006**, *25*, 6384–6391. [CrossRef] [PubMed]
73.  Prakash, V.; Carson, B.B.; Feenstra, J.M.; Dass, R.A.; Sekyrova, P.; Hoshino, A.; Petersen, J.; Guo, Y.; Parks, M.M.; Kurylo, C.M. Ribosome biogenesis during cell cycle arrest fuels EMT in development and disease. *Nat. Commun.* **2019**, *10*, 1–16. [CrossRef] [PubMed]
74.  Pelletier, J.; Thomas, G.; Volarević, S. Ribosome biogenesis in cancer: New players and therapeutic avenues. *Nat. Rev. Cancer* **2018**, *18*, 51–63. [CrossRef]
75.  Hoeller, D.; Hecker, C.-M.; Dikic, I. Ubiquitin and ubiquitin-like proteins in cancer pathogenesis. *Nat. Rev. Cancer* **2006**, *6*, 776–788. [CrossRef] [PubMed]
76.  Hu, H.; Sun, S.-C. Ubiquitin signaling in immune responses. *Cell Res.* **2016**, *26*, 457–483. [CrossRef] [PubMed]
77.  Noh, J.H.; Kim, K.M.; McClusky, W.; Abdelmohsen, K.; Gorospe, M. Cytoplasmic functions of lncRNAs. *Wiley Interdiscip. Rev. RNA* **2018**, *9*, e1471. [CrossRef] [PubMed]
78.  Narayanan, S.; Cai, C.-Y.; Assaraf, Y.G.; Guo, H.-Q.; Cui, Q.; Wei, L.; Huang, J.-J.; Ashby, C.R., Jr.; Chen, Z.-S. Targeting the ubiquitin-proteasome pathway to overcome anti-cancer drug resistance. *Drug Resist. Updates* **2020**, *48*, 100663. [CrossRef]
79.  Wang, B.-D.; Lee, N.H. Aberrant RNA splicing in cancer and drug resistance. *Cancers* **2018**, *10*, 458. [CrossRef]
80.  Oltean, S.; Bates, D.O. Hallmarks of alternative splicing in cancer. *Oncogene* **2014**, *33*, 5311–5318. [CrossRef]
81.  Romero-Barrios, N.; Legascue, M.F.; Benhamed, M.; Ariel, F.; Crespi, M. Splicing regulation by long noncoding RNAs. *Nucleic Acids Res.* **2018**, *46*, 2169–2184. [CrossRef]
82.  Villamizar, O.; Chambers, C.B.; Riberdy, J.M.; Persons, D.A.; Wilber, A. Long noncoding RNA Saf and splicing factor 45 increase soluble Fas and resistance to apoptosis. *Oncotarget* **2016**, *7*, 13810. [CrossRef] [PubMed]
83.  Seo, J.-S.; Ju, Y.S.; Lee, W.-C.; Shin, J.-Y.; Lee, J.K.; Bleazard, T.; Lee, J.; Jung, Y.J.; Kim, J.-O.; Shin, J.-Y. The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res.* **2012**, *22*, 2109–2119. [CrossRef]
84.  Yang, Z.; Xu, F.; Wang, H.; Teschendorff, A.; Xie, F.; He, Y. Pan-cancer characterization of long non-coding RNA and DNA methylation mediated transcriptional dysregulation. *Ebiomedicine* **2021**, *68*, 103399. [CrossRef]
85.  Beroukhim, R.; Mermel, C.H.; Porter, D.; Wei, G.; Raychaudhuri, S.; Donovan, J.; Barretina, J.; Boehm, J.S.; Dobson, J.; Urashima, M. The landscape of somatic copy-number alteration across human cancers. *Nature* **2010**, *463*, 899–905. [CrossRef]
86.  Lee, C.M.; Barber, G.P.; Casper, J.; Clawson, H.; Diekhans, M.; Gonzalez, J.N.; Hinrichs, A.S.; Lee, B.T.; Nassar, L.R.; Powell, C.C. UCSC Genome Browser enters 20th year. *Nucleic Acids Res.* **2020**, *48*, D756–D761. [CrossRef]
87.  Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **1995**, *57*, 289–300. [CrossRef]
88.  Jaccard, P. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull. Soc. Vaud. Sci. Nat.* **1901**, *37*, 547–579.
89.  Oliver, S. Guilt-by-association goes global. *Nature* **2000**, *403*, 601–602. [CrossRef]
90.  Consortium, G.O. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* **2019**, *47*, D330–D338.
91.  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. Gene ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef] [PubMed]
92.  Wang, J.; Duncan, D.; Shi, Z.; Zhang, B. WEB-based gene set analysis toolkit (WebGestalt): Update 2013. *Nucleic Acids Res.* **2013**, *41*, W77–W83. [CrossRef] [PubMed]
93.  Zhang, B.; Kirov, S.; Snoddy, J. WebGestalt: An integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.* **2005**, *33*, W741–W748. [CrossRef]
94.  Adams, W.T.; Skopek, T.R. Statistical test for the comparison of samples from mutational spectra. *J. Mol. Biol.* **1987**, *194*, 391–396. [CrossRef] [PubMed]
95.  Falcon, S.; Gentleman, R. Hypergeometric testing used for gene set enrichment analysis. In *Bioconductor Case Studies*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 207–220.
96.  Liu, W.; Liu, J.; Rajapakse, J.C. Gene ontology enrichment improves performances of functional similarity of genes. *Sci. Rep.* **2018**, *8*, 12100. [CrossRef] [PubMed]
97.  Wei, Q.; Khan, I.K.; Ding, Z.; Yerneni, S.; Kihara, D. NaviGO: Interactive tool for visualization and functional similarity and coherence analysis with gene ontology. *BMC Bioinform.* **2017**, *18*, 177. [CrossRef]