

## REVIEW ARTICLE OPEN



# Blood-derived lncRNAs as biomarkers for cancer diagnosis: the Good, the Bad and the Beauty

Cedric Badowski<sup>1</sup>✉, Bing He<sup>2</sup> and Lana X. Garmire<sup>1,2</sup>✉

Cancer ranks as one of the deadliest diseases worldwide. The high mortality rate associated with cancer is partially due to the lack of reliable early detection methods and/or inaccurate diagnostic tools such as certain protein biomarkers. Cell-free nucleic acids (cfNA) such as circulating long noncoding RNAs (lncRNAs) have been proposed as a new class of potential biomarkers for cancer diagnosis. The reported correlation between the presence of tumors and abnormal levels of lncRNAs in the blood of cancer patients has notably triggered a worldwide interest among clinicians and oncologists who have been actively investigating their potentials as reliable cancer biomarkers. In this report, we review the progress achieved (“the Good”) and challenges encountered (“the Bad”) in the development of circulating lncRNAs as potential biomarkers for early cancer diagnosis. We report and discuss the diagnostic performance of more than 50 different circulating lncRNAs and emphasize their numerous potential clinical applications (“the Beauty”) including therapeutic targets and agents, on top of diagnostic and prognostic capabilities. This review also summarizes the best methods of investigation and provides useful guidelines for clinicians and scientists who desire conducting their own clinical studies on circulating lncRNAs in cancer patients via RT-qPCR or Next Generation Sequencing (NGS).

*npj Precision Oncology* (2022)6:40; <https://doi.org/10.1038/s41698-022-00283-7>

## INTRODUCTION

Cancer ranks as one of the deadliest diseases worldwide. Despite ongoing efforts to develop new treatments and a better understanding of the mechanisms underlying tumorigenesis, it remains difficult to treat cancers, particularly when diagnosed at late stages with a poor prognosis. The high mortality rate associated with cancer is partially due to the lack of early detection methods and/or inaccurate diagnostic tools, such as certain protein biomarkers. Protein or peptide-based biomolecules such as glycoproteins constitute most of the currently available cancer biomarkers. Variations in their levels in tissues or blood may indicate the development of diseases such as cancer. Protein markers can be detected in tissue biopsy sections analyzed by immunohistochemistry (IHC) upon diagnostic notably to determine cancer molecular subtype. For instance, breast tumor tissues are commonly assessed for the presence of estrogen receptor (ER) to determine their ER-positive or ER-negative status. However, some protein biomarkers are reportedly unreliable as they generate a significant amount of false-positive and/or false-negative results. Plasma alpha fetoprotein (AFP), one of the most frequently used biomarkers for diagnosis of hepatocellular carcinoma (HCC)<sup>1</sup>, has been described by many as a marker with low sensitivity and/or specificity<sup>2–5</sup>. Conventional serological biomarkers such as carbohydrate antigen 153 (CA153), cancer antigen 125 (CA125), CA27.29 and carcinoembryonic antigen (CEA) remain controversial due to poor specificity and sensitivity<sup>6–11</sup>. The poor reliability of certain protein biomarkers is partially due to the nature of the biomarker itself. The detection of proteins and peptides indeed relies on the use of antibodies that may or may not be specific to the desired marker as the epitope recognized by the antibodies may be present on other tissue components<sup>12</sup>. Unreliable antibodies currently represent a major issue in biomedical research in general and can significantly comprise the outcome of a study or diagnosis.

Another issue with traditional histology analyses is the need for actual tissue biopsies. This invasive and inconvenient technique may discourage potential cancer patients to proceed with the entire diagnosis procedure. Thus, the development of noninvasive nonprotein biomarkers is currently needed.

Cell-free nucleic acids (cfNA) or circulating nucleic acids (CNAs) have recently been proposed as a new class of potential biomarkers that could improve cancer diagnosis<sup>13</sup>. CNA PCA3 (prostate cancer associated 3) has notably been approved by the FDA and is currently being sold as Progenisa by Hologic Gen Probe (Marlborough, MA, USA) for the diagnosis of prostate cancer<sup>14–16</sup>. Circulating long noncoding RNAs or lncRNAs (noncoding RNAs of 200 nucleotides or more), such as PCA3 seem more reliable than other CNAs due to their high stability in the bloodstream and poor sensitivity to nuclease-mediated degradation. Arita et al. especially showed that plasmatic lncRNAs are resistant to degradation induced by repetitive freeze-thaw cycles, as well as prolonged exposure to 45 C and room temperatures<sup>17</sup>. The stability of lncRNAs in the bloodstream appears to originate from the presence of extensive secondary structures<sup>18</sup>, the transport by protective exosomes<sup>19</sup>, as well as stabilizing posttranslational modifications. The reported prevalence of ncRNAs in the mammalian genome and the known association between aberrant lncRNA expressions and tumorigenesis undeniably highlight the crucial biological importance of ncRNAs in health and disease. ncRNAs are particularly known to be major regulators of cell proliferation and differentiation during development and in adult life through complex mechanisms which are still being investigated. In a pioneering study published in 2007, Rinn et al. notably reported that lncRNA HOTAIR (HOX transcript antisense RNA) was capable of repressing transcription in trans across the HOXD locus and interacting with Polycomb Repressive Complex 2 (PRC2) while being required for PRC2 occupancy and histone H3 lysine-27 trimethylation of HOXD locus<sup>20</sup>. Many more mechanisms

<sup>1</sup>University of Hawaii Cancer Center, Epidemiology, 701 Ilalo Street, Honolulu, HI 96813, USA. <sup>2</sup>Present address: Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48105, USA. ✉email: cedricbadowski@gmail.com; LGarmire@med.umich.edu

have been described and continue to be discovered, as scientists and clinicians actively investigate the mechanisms of action of lncRNAs as well as their potential as reliable cancer biomarkers.

The high stability and relative abundance of lncRNAs in the circulation may make them more reliable cancer biomarkers compared to other analytes such as circulating tumor cells (CTCs), cell-free DNA (cfDNA, which includes circulating tumor DNA ctDNA) and exosomes. CTCs and ctDNA are present in limited quantities in the fluids of cancer patients especially those with early-stage cancers, which may significantly hinder their quantification in clinic, while impairing the detection of low allelic frequency mutations<sup>21,22</sup>. Moreover, CTCs are very heterogeneous<sup>21</sup>, and the value of CTCs as diagnostic biomarkers remains currently unclear as early lesions may still be benign and devoid of CTCs<sup>21</sup>. ctDNA on the other hand, may not be sufficient to provide an accurate diagnosis and is often used in combination with other methods in diagnostic and prognostic studies. As for tumor-derived exosomes, the detection of glycoprotein biomarkers on their surface relies heavily on the specificity of antibodies. Lysed exosomes could be alternatives that release the nonprotein content including lncRNAs, which are easier to detect compared to proteins.

In this report, we review the progress achieved and challenges encountered in the development of circulating lncRNAs as potential biomarkers for early cancer diagnosis. We report and discuss the specificity and sensitivity of blood-based lncRNAs currently considered as promising biomarkers for various cancers such as hepatocellular carcinoma, colorectal cancer, gastric cancer and prostate cancer. We also highlight potential therapeutic applications for circulating lncRNAs both as therapeutic targets and agents, on top of diagnostic and prognostic purposes. Based on recommendations from different published works, we finally provide recommendations for investigators who seek to investigate and compare the levels of circulating lncRNAs in the blood of cancer patients compared to healthy subjects by RT-qPCR or Next Generation Sequencing.

## BLOOD-BASED LNCRNAs AS POTENTIAL CIRCULATING BIOMARKERS FOR CANCER DIAGNOSIS

### Changes in circulating lncRNA levels specifically correlate with cancer development

Most studies focusing on circulating lncRNAs have been initiated based on prior observations reporting changes in lncRNA levels in cancer tissue samples. For instance, MALAT-1 (metastasis-associated lung adenocarcinoma transcript 1) was first shown to be upregulated in various cancer tissues including lung and prostate tumors<sup>23,24</sup>. Using peripheral blood cells as a lncRNA source for their study, Weber et al. later showed that MALAT-1 levels could reflect the presence of nonsmall-cell lung cancer with a specificity of 96%<sup>25</sup> (Table 1). lncRNA MALAT-1 was also detected in significantly higher quantities in the plasma of patients with prostate cancer as compared to healthy subjects<sup>26</sup> and these changes in circulating MALAT-1 levels correlated with prostate cancer with relatively high specificity (84.8%)<sup>26</sup>. This study showed that tumors were at the origin of MALAT-1 variations, since the surgical removal of the cancerous tissues induced a dramatic reduction in circulating MALAT-1, while plasmatic levels of this lncRNA increased upon ectopic implantation of a tumoral xenograft in mice<sup>26</sup>. More studies support the concept that circulating lncRNAs are, directly or indirectly, correlated with the presence of tumors in vivo. For instance, the blood of patients with hepatocellular carcinoma was shown to contain elevated levels of lncRNA HULC (for “highly upregulated in liver cancer”)<sup>27,28</sup>. Moreover, HULC, H19, HOTAIR and GACAT2 (for “gastric cancer-associated transcript 2”) were found to be significantly increased in the plasma of gastric cancer

(GC) patients compared to healthy individuals<sup>29–32</sup>. Alike MALAT-1 which was primarily detected in tumoral tissue, lncRNA GIHCG (for “gradually increased during hepatocarcinogenesis”) was originally found to be upregulated in cancer tissue samples from HCC and RCC (renal cell carcinoma) tumors<sup>33,34</sup>. Higher levels of GIHCG as well as ARSR (for “activated in RCC with sunitinib resistance”) were also reported in the circulation of renal cell carcinoma patients<sup>34–36</sup>. Serum GIHCG levels were notably able to distinguish RCC patients from healthy individuals with a specificity of 84.8%. Levels of circulating lncRNAs GIHCG and ARSR significantly dropped after resection of RCC tumors, while plasma levels of H19, A174084 and GACAT2 markedly decreased in GC patients postoperatively, further supporting a direct correlation between abnormal levels of circulating lncRNAs and tumorigenesis<sup>29,32,34,35,37,38</sup>. In fact, some of these circulating lncRNAs have shown greater diagnostic performance than conventional glycoprotein markers. For instance, circulating H19 and RP11-445H22.4 have been reported as more reliable than carcinoembryonic antigen (CEA) and/or carbohydrate antigen 153 (CA153) for the diagnosis of breast cancer<sup>39,40</sup>. Likewise, a serum three-lncRNA signature consisting of PTENP1, LSINCT-5 and CUDR (also known as UCA1) significantly outperformed CEA and CA19-9 in gastric cancer diagnostic studies<sup>41</sup>.

Other lncRNAs have been reported to detect various cancer types with relatively high specificity. For instance, HOTAIR has shown high efficacy in identifying samples from colorectal cancer patients with a specificity of 92.5%<sup>42</sup>. Changes in plasmatic levels of lncRNA LINC00152 were found to correlate with gastric cancer with a specificity of 85.2%<sup>19</sup> (Table 1). LINC00152 has also been suggested as a reliable blood-based biomarker for hepatocellular carcinoma<sup>43,44</sup>. The high prevalence of HCC in certain parts of the world such as Asia or Africa is undeniably alarming, and it has become a major public health matter in many countries. Reliable biomarkers are desperately needed to detect this deadly cancer at an early stage. Many circulating lncRNAs have shown a significant correlation with HCC and represent promising candidates for HCC diagnostic applications (Table 1). Several studies from Egypt identified lncRNA-UCA1 as a potential serum-based biomarker for the detection of HCC. The specificities obtained were 82.1%<sup>45</sup> and 88.6%<sup>46</sup>. These studies also reported WRAP53 and CTBP as potential biomarkers for HCC with a specificity of 82.1%<sup>45</sup> and 88.5%<sup>46</sup>, respectively. In Asia, Jing et al. showed that lncRNA SPRY4-IT1 represents another promising blood-based biomarker for the diagnosis of hepatocellular carcinoma<sup>47</sup>.

Many more circulating lncRNAs have been proposed as potential blood-based biomarkers for cancer diagnosis, some with relatively high specificity (Table 1)<sup>48–50</sup>.

### Challenges and potential impacts on diagnosis using lncRNA as biomarkers

The diagnostic power of circulating biomarkers has yet to reach its maximum potential. Indeed, the diagnostic performance of many circulating lncRNAs remains relatively poor when taken individually. Several lncRNAs reportedly have either poor sensitivity or poor specificity towards a specific cancer type, affecting their potentials as diagnosis biomarkers. Below are some examples:

MALAT-1 has shown a sensitivity of only 58.6% when testing plasma samples from prostate cancer patients and healthy subjects. This moderate sensitivity implies that the use of MALAT-1 as a blood-based prostate cancer biomarker may result in a significant number of false-negative results, as actual cancer samples may not be detected. MALAT-1 has also been investigated as a potential biomarker for nonsmall-cell lung cancer<sup>25,51</sup>. However, with a sensitivity of only 56%, MALAT-1 may also face multiple challenges before becoming a reliable blood-based biomarker for lung cancer diagnosis (Table 1). One unsolved issue is notably the reported lack of correlation between

**Table 1.** List of blood-based lncRNAs investigated as potential biomarkers for diagnosis of various cancers.

lncRNA	Cancer type	Source	Sensitivity (%)	Specificity (%)	AUC / QUADAS	Normalization	Reference
MALAT-1	Nonsmall-cell lung cancer	Blood cells	56	96	AUC 0.79	GAPDH	25
	Nonsmall-cell lung cancer	Whole blood	N.A.	N.A.	AUC 0.718	GAPDH	51
	Prostate cancer	Plasma	58.6	84.8	AUC 0.836	Standard curve	26
	Hepatocellular carcinoma	Plasma	51.1	89.3	AUC 0.66	MALAT-1	48
LINC00152	Gastric cancer	Plasma	48.1	85.2	AUC 0.675	GAPDH	19
	Hepatocellular carcinoma	Plasma	N.A.	N.A.	AUC 0.85	5 S	43
	Hepatocellular carcinoma	Serum	78.3	89.2	AUC 0.877	GAPDH	44
UCA1	Hepatocellular carcinoma	Serum	92.7	82.1	AUC 0.861	GAPDH	45
	Hepatocellular carcinoma	Serum	91.4	88.6	QUADAS 11	$\beta$ -actin	46
	Colorectal cancer	Plasma	N.A.	N.A.	N.A.	Cel-miR-39	178
	Gastric cancer	Plasma	N.A.	N.A.	AUC 0.928	GAPDH	104
	Osteosarcoma	Serum	N.A.	N.A.	AUC 0.831	GAPDH	179
H19	Gastric cancer	Plasma	74	58	AUC 0.64	LncRNA levels	17
	Gastric cancer	Plasma	82.9	72.9	AUC 0.838	Standard curve	38
	Gastric cancer	Plasma	68.75	56.67	AUC 0.724	GAPDH	60
	Breast cancer	Plasma	56.7	86.7	AUC 0.81	$\beta$ -actin	39
PVT1	Cervical cancer	Serum	71.6	98.8	AUC 0.932	GAPDH	176
	Melanoma	Serum	94.12	85.11	AUC 0.938	GAPDH	177
WRAP53	Hepatocellular carcinoma	Serum	85.4	82.1	AUC 0.896	GAPDH	45
HULC	Hepatocellular carcinoma	Blood cells	N.A.	N.A.	N.A.	$\beta$ -actin	27
	Hepatocellular carcinoma	Plasma	N.A.	N.A.	N.A.	GAPDH	28
	Hepatocellular carcinoma	Plasma	N.A.	N.A.	AUC 0.78	5 S	43
	Hepatocellular carcinoma	Plasma	58	80	AUC 0.65	GAPDH	30
HOTAIR	Gastric cancer	Blood cells	67	92.5	AUC 0.87	PPIA	42
	Colorectal cancer	Serum	N.A.	N.A.	N.A.	GAPDH	105
CTBP	Hepatocellular carcinoma	Serum	91	88.5	QUADAS 11	$\beta$ -actin	46
GIHCG	Renal cell carcinoma	Serum	87	84.8	AUC 0.920	N.A.	34
	Cervical cancer	Serum	88.7	87.5	AUC 0.940	$\beta$ -actin	90
PCA3	Prostate cancer	Periph. Blood	32	94	N.A.	N.A.	14
RP11-445H22.4	Breast cancer	Serum	92	74	AUC 0.904	U6	40
uc003wbd	Hepatocellular carcinoma	Serum	N.A.	N.A.	AUC 0.86	$\beta$ -actin	49
AF085935	Hepatocellular carcinoma	Serum	N.A.	N.A.	AUC 0.96	$\beta$ -actin	49
GACAT2	Gastric cancer	Plasma	87	28	AUC 0.622	GAPDH	29
SPRY4-IT1	Hepatocellular carcinoma	Plasma	87.3	50	QUADAS 12	18 S	47
uc001ncr	Hepatocellular carcinoma	Serum	N.A.	N.A.	AUC 0.885	GAPDH	50
AX800134	Hepatocellular carcinoma	Serum	N.A.	N.A.	AUC 0.925	GAPDH	50
ZNF1-AS1	Gastric cancer	Plasma	84	68	AUC 0.85	GAPDH	30
LINC00152 + AFP	Hepatocellular carcinoma	Serum	85.3	83.4	AUC 0.906	GAPDH	44
XIST + HIF1A-AS1	Nonsmall-cell lung cancer	Serum	N.A.	N.A.	AUC 0.931	GAPDH	62

Table 1 continued

LncRNA	Cancer type	Source	Sensitivity (%)	Specificity (%)	AUC / QUADAS	Normalization	Reference
PVT1 + uc002mbe.2	Hepatocellular carcinoma	Serum	60.5	90.6	QUADAS 11	GAPDH	64
GAS5 + SRA	Pancreatic cancer (IPMN)	Plasma	82	59	AUC 0.729	$\beta$ -actin PGK1 PPIB	69
SPRY4-IT1 + ANRIL + NEAT1	Nonsmall-cell lung cancer	Plasma	82.8	92.3	AUC 0.876	N.A.	61
LINC00152 + UCA1 + AFP	Hepatocellular carcinoma	Serum	82.9	88.2	AUC 0.912	GAPDH	44
CUDR (UCA1) + LSINCT-5 + PTENP1	Gastric cancer	Serum	81.8	85.2	AUC 0.829	$\beta$ -actin	41
SPRY4-IT1 + POU3F3 + HNF1A-AS1	Esophageal squamous cell carcinoma	Plasma	72.8	89.4	AUC 0.842	GAPDH	63
XLOC_006844 + LOC152578 + XLOC_000303	Colorectal cancer	Plasma	80	84	AUC 0.975	N.A.	65
RP11-160H22.5 + XLOC_014172 + LOC149086	Hepatocellular carcinoma	Plasma	82	73	AUC 0.896	$\beta$ -actin	3
UCA1 + POU3F3 ESCCAL-1 + PEG10	Esophageal squamous cell carcinoma	Serum	80.2	80.2	AUC 0.853	GAPDH	66
LET + PVT1 + PANDAR + PTENP1 + linc00963	Renal cell carcinoma	Serum	67.6	91.4	AUC 0.823	$\beta$ -actin	68
AOC4P + BANCR + CCAT2 + LINC00857 + TINCR	Gastric cancer	Plasma	0.82	0.87	AUC 0.91	GAPDH	67

N.A. not available / data presented in graphical format in original report. Information reported includes the name of lncRNA, cancer type, source of lncRNA, lncRNA specificity, lncRNA sensitivity, AUC (ROC) value (area under the ROC curve - receiver operating characteristic), QUADAS score, normalization method and literature reference.

the levels of circulating MALAT-1 in lung cancer patients and the levels of this lncRNA in lung cancer tissues. Indeed, the comparative analysis of whole blood samples from 105 lung cancer patients and 65 healthy subjects revealed a decrease in blood MALAT-1 levels in cancer patients, while lung cancer tissues showed higher MALAT1 expression<sup>51</sup>. The lack of strong sensitivity and the poor correlation between tissue and blood levels may arise from the fact that MALAT-1 is reportedly undergoing a certain degree of degradation in the bloodstream<sup>26</sup>. One of the resulting fragments has notably been referred to as MD-mini RNA (for metastasis associated in lung adenocarcinoma transcript 1 derived miniRNA)<sup>26</sup>.

The degradation of MALAT-1 in the bloodstream may not be an isolated case and, probably, many more lncRNAs are actively being degraded once they enter the circulation. Degradation of circulating lncRNAs may increase in cancer patients as several studies reported that tumorigenesis is often associated with higher RNase activity in the bloodstream<sup>52</sup>. In fact, long before circulating lncRNAs were considered as potential cancer biomarkers, increased RNase activity in the serum of cancer patients was suggested as a mean of early cancer detection<sup>53,54</sup>. In their study, Reddi and Holland notably reported that 90% of the patients with pancreatic cancer showed a dramatic increase in serum RNase levels (above 250 units/mL). They hence promoted the use of high serum RNase activity as a biomarker for pancreatic carcinoma. Other cancers such as chronic myeloid leukemia have also been reported to be associated with a higher level of plasmatic RNase activity<sup>55</sup>. RNases circulating in the bloodstream notably constitute cytotoxic agents secreted by immune cells as part of anti-cancer defense mechanisms that aim at lysing transformed cells by activating cell death pathways<sup>56</sup>. For instance, an RNase secreted by human eosinophils is known to induce the specific apoptosis of Kaposi's sarcoma cells without affecting normal human fibroblasts<sup>57</sup>. RNase L was shown to suppress prostate tumorigenesis by initiating a cellular stress response that leads to cancer cell apoptosis<sup>58,59</sup>. Tumors, on the other hand, reportedly display lower RNase activity to promote protein synthesis and cell proliferation<sup>52</sup>. The reported difference in RNase activity in tumors versus circulation may explain seemingly paradoxical data when comparing lncRNA levels in tissues and blood such as in the case of MALAT1. While many studies have shown positive correlations between tissue and blood lncRNAs, the reported increased RNase activity in the blood of some cancer patients may promote the degradation of circulating lncRNAs to a degree that would depend on the nature of cancer and/or lncRNA studied. This could represent a significant challenge for investigators as RT-qPCR analyses may not detect fragments of an investigated lncRNA possibly compromising the outcome of a study.

LINC00152 is another circulating lncRNA that has been actively investigated as a potential cancer biomarker. However, LINC00152 has shown a sensitivity of only 48.1% when analyzing plasma samples from gastric patients and healthy subjects, limiting its diagnostic performance as well (Table 1). It is currently not clear if LINC00152 is undergoing degradation in the bloodstream. Other circulating lncRNAs have shown poor specificity in the detection of specific cancers. For instance, GACAT2 reportedly has a specificity of only 28% when comparing plasma samples from gastric cancer patients and healthy subjects<sup>29</sup>, while several studies have shown that H19 is capable of detecting samples from gastric cancer patients with a specificity of only 58 %<sup>17</sup> or 56.67%<sup>60</sup> (Table 1). This implies that diagnosis based on the quantification of plasmatic levels of H19 or GACAT2 may potentially result in a significant number of false-positive results when testing for gastric cancer. It is also the case for lncRNA SPRY4-IT1 regarding the diagnosis of hepatocellular carcinoma (HCC) with a specificity of only 50%, and HULC for the detection of gastric cancer (with a specificity of only 58%)<sup>30</sup> (Table 1).

Therefore, significant improvements are required before most individual circulating lncRNAs become reliable blood-based cancer biomarkers.

### Combination of circulating lncRNAs for greater diagnostic performance and new technologies for improved lncRNA detection

To compensate for the moderate specificity/sensitivity of certain circulating lncRNAs and increase their diagnostic performance, several studies have combined the diagnostic values of several circulating lncRNAs. For instance, Hu et al., integrated lncRNAs SPRY4-IT1, ANRIL and NEAT1 in their studies on nonsmall-cell lung cancer and obtained a specificity of 92.3%, a sensitivity of 82.8%, and an AUC (ROC) (area under the ROC curve - receiver operating characteristic) of 0.876<sup>61</sup> (Table 1). The combination of serum XIST and HIF1A-AS1 was able to accurately detect nonsmall-cell lung cancer as well<sup>62</sup>. When combined with POU3F3 and HNF1AAS1, SPRY4-IT1 displayed a sensitivity of 72.8% and a specificity of 89.4% (AUC: 0.842) in the detection of esophageal squamous cell carcinoma<sup>63</sup>. Yu et al. reported that the combination of circulating lncRNAs PVT1 and uc002mbe.2 reflected the presence of hepatocellular carcinoma with a specificity of 90.6% and a sensitivity of 60.5%<sup>64</sup>. The integrated analysis of plasmatic levels of XLOC\_006844, LOC152578 and XLOC\_000303 allowed the detection of colorectal cancer with a specificity of 84%, a sensitivity of 80% and an AUC of 0.975<sup>65</sup>. Other examples include the combination of lncRNAs RP11-160H22.5, XLOC\_014172 and LOC149086 which produced a sensitivity of 82% and a specificity of 73% (AUC: 0.896) for the diagnosis of hepatocellular carcinoma<sup>3</sup> (Table 1). Some studies have investigated the diagnostic signature of more than 3 circulating lncRNAs. For instance, Yan et al., reported that a 4-lncRNA panel comprising UCA1, POU3F3, ESCCAL-1 and PEG10 constitutes a remarkable diagnostic tool for the accurate and reliable detection of esophageal squamous cell carcinoma (ESCC) since this multi-lncRNA panel was capable of distinguishing ESCC patients from healthy controls with a sensitivity of 80.20%, a specificity of 80.20% and an AUC of 0.853<sup>66</sup>. The authors emphasized that, in terms of diagnostic performance, the 4-lncRNA panel outperformed each individual lncRNA, further supporting the clinical value of such a combinatory approach. In a separate study, Zhang et al. identified a panel of five plasma lncRNAs (BANCR, AOC4P, TINCR, CCAT2 and LINC00857) that was able to discriminate GC patients from healthy controls with an AUC of 0.91, outperforming CEA biomarker<sup>67</sup>. Wu et al. have reported that a 5-lncRNA signature could accurately distinguish serum samples of patients with renal cell carcinoma (RCC) from those of healthy subjects<sup>68</sup>. The combination of lncRNA-LET, PVT1, PANDAR, PTENP1 and linc00963 identified RCC samples with an AUC of 0.823. Each of these 5 lncRNAs was not individually capable of performing as well as the 5-lncRNA signature. PVT1 and PANDAR have also been investigated as part of a 8-lncRNA signature in plasma samples of patients with pancreatic ductal adenocarcinoma<sup>69</sup>. The 8-lncRNA signature was identified by using a custom nCounter Expression Assay (Nanostring Technologies, USA) that allows multiplex qPCR analyses using TaqMan probes. A better diagnostic performance may also be obtained through the improved detection of lncRNAs in human samples and novel highly sensitive methods have been recently developed to achieve this purpose. In a remarkable study, Chen et al. recently developed a novel biocompatible electrochemical biosensor referred to as "SPCE Au NCs/MWCNT-NH2" for the ultrasensitive detection of lncRNA MALAT1 in non-small cell lung cancer<sup>70</sup>. Importantly, the authors highlighted that, compared to traditional RT-PCR, this new method presents several major advantages including faster detection and lower cost while being simpler to operate. In another outstanding study, Morlion et al. developed a unique custom lncRNA capture sequencing

approach that relies on a set of 565,878 capture probes for 49,372 human lncRNA genes and which is reportedly capable of enhancing detection sensitivity<sup>71</sup>. This custom enrichment approach achieved major advancements in lncRNA detection, since it enables the detection of a broad repertoire of lncRNAs with better reproducibility and higher coverage than classic total RNA-sequencing methods.

Overall, the signature generated by the combination of several blood-based lncRNAs reportedly provides better diagnostic performance than most individual circulating lncRNAs, while the emergence of new technologies paves the way for a better detection of lncRNAs in human biofluids.

### Circulating lncRNAs as potential blood-based biomarkers for cancer prognosis

Besides being potential blood-based biomarkers for early cancer diagnosis, circulating lncRNAs may also constitute valuable prognosis markers. Most studies assessing the ability of lncRNAs to predict disease evolution and eventual clinical outcome have been performed on cancer tissue samples<sup>72-74</sup>. However, a few studies based on the analysis of blood-derived samples indicate that circulating lncRNAs may also be able to reflect cancer prognosis. For instance, changes in plasmatic levels of lncRNAs XLOC\_014172 and LOC149086 can distinguish metastatic HCC from non-metastatic HCC with a specificity of 90%, a sensitivity of 91% and an AUC of 0.934 (combined)<sup>3</sup>. HOTAIR can also be used as a negative prognostic marker for colorectal cancer with a sensitivity of 92.5%, a specificity of 67% and an AUC of 0.87<sup>42</sup>. Moreover, lncRNA GIHCG has been proposed as a potential prognostic biomarker for renal cell carcinoma<sup>34</sup>. The 5-lncRNA signature reported by Wu et al., was also capable of discriminating benign renal tumors from metastatic renal cell carcinoma<sup>68</sup>. Similarly, the 8-lncRNA signature recently described by Permuth et al., reportedly distinguished indolent (benign) intraductal papillary mucinous neoplasms (IPMNs) from aggressive (malignant) IPMNs<sup>69</sup>. This 8-lncRNA-signature reportedly had greater accuracy than standard clinical and radiological features. It was further improved when combined with plasma miRNA data and quantitative radiomic imaging.

While early studies suggest that the analysis of circulating lncRNA levels may contribute to the evaluation of disease progression, more investigations focusing on blood-based lncRNAs are needed to truly appreciate the prognosis power of circulating lncRNAs. The best diagnostic/prognostic performance may actually emerge from the integration of several analytic methods that combine circulating lncRNA data, miRNA data, clinical data, quantitative imaging features<sup>69</sup> and/or conventional glycoprotein antigens such as carcinoembryonic antigen (CEA)<sup>60</sup> or prostate-specific antigen (PSA)<sup>14</sup>.

### Circulating lncRNAs as potential therapeutic agents/targets for cancer treatment

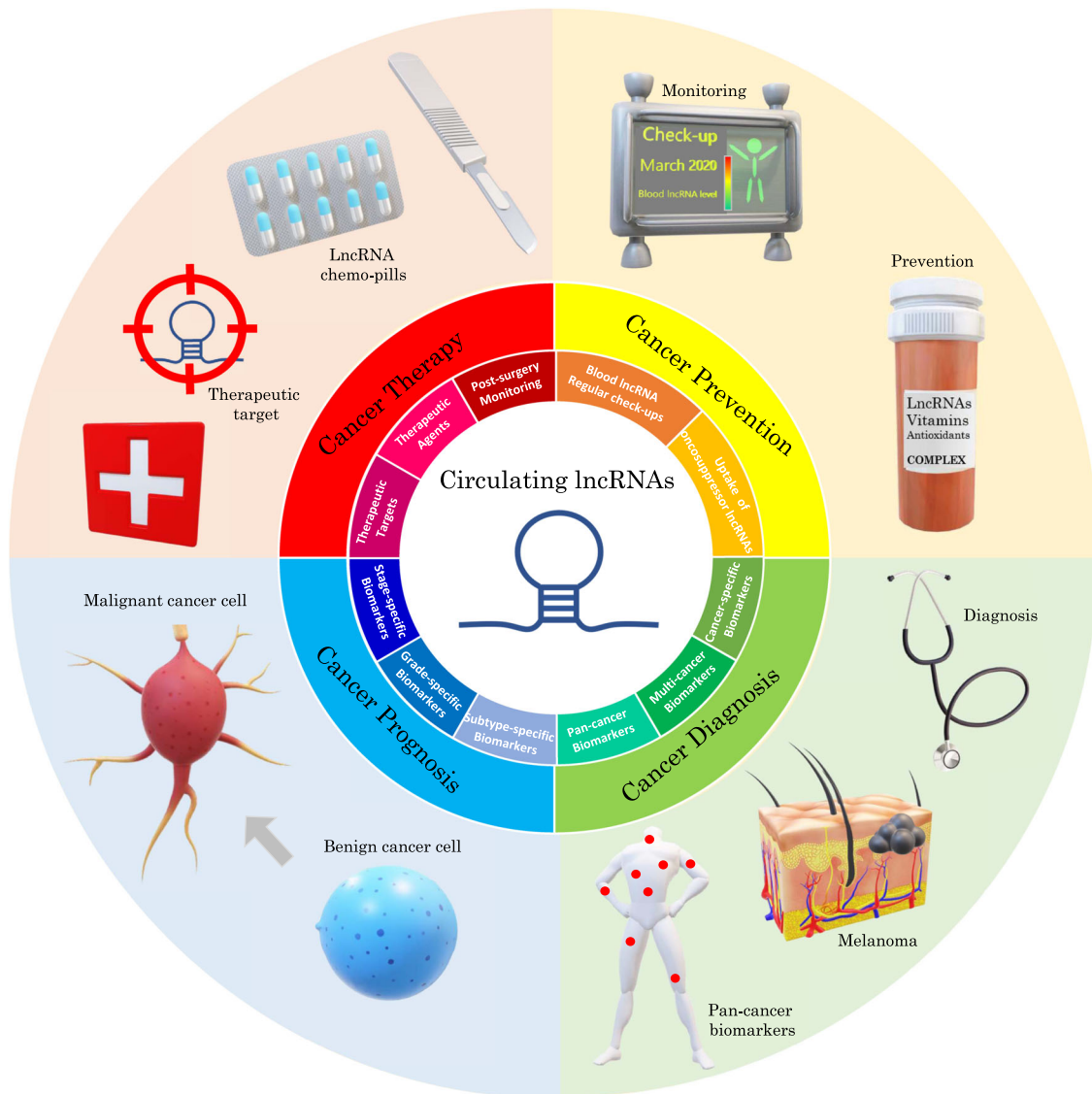
Circulating lncRNAs should not be considered only as passive biomedical tools that solely enable the detection and monitoring of various diseases. They may also constitute effective therapeutic agents and/or targets in innovative strategies that could treat various types of cancers including colorectal cancer and renal cell carcinoma<sup>34,75-77</sup>. Indeed, lncRNAs have been shown to trigger or contribute to tumorigenesis notably by interfering with tumor-suppressive signaling pathways or acting as oncogenic stimuli<sup>78-82</sup>. In a Genome-wide analysis of the human p53 transcriptional network, Sanchez et al. notably revealed the existence of a lncRNA tumor suppressor signature<sup>83</sup>. GAS5, CCND1, LET, PTENP1 and lincRNA-p21 have been described as tumor suppressors<sup>36,75,84-87</sup>, while MALAT-1, PANDAR, HOTAIR, H19, PVT1, GIHCG and ANRIL have been characterized as oncogenic lncRNAs<sup>36,75,88-90</sup>. At the molecular level, lncRNAs

can promote tumorigenesis by acting as chromatin structure regulators that modify gene expression<sup>91</sup>, scaffolds for oncogenic RNA-binding proteins<sup>92</sup> or RNA sponges for oncosuppressor microRNAs<sup>93,94</sup>. For instance, lncRNA HOTTIP (HOXA transcript at the distal tip) was shown to act as a sponge for the tumor-suppressive microRNA miR-615-3p and dysregulation of HOTTIP expression was shown to alter levels of miR-615-3p and its target IGF-2, promoting the formation of RCC tumors<sup>94</sup>. Many more mechanisms have been described and continue to be discovered. Through various pathways, dysregulation of lncRNAs levels eventually promotes cancer cell proliferation, migration, invasion and/or metastasis<sup>94–97</sup>. Therefore, lncRNAs do constitute legitimate therapeutic targets. However, most mechanistic studies have been done on cancer tissues or cells, so it is still unclear if targeting lncRNAs in blood would be sufficient to treat tumors located deep inside layers of tissues. A more fundamental question may be to determine whether circulating lncRNAs can actually penetrate cells and tissues. Nucleic acids are usually unable to cross the hydrophobic cellular plasma membrane due to their large size and negative charges carried by the phosphate groups of nucleotides. In vitro DNA transfection is usually achieved by using specific carriers such as lipofectamine. Answers may come from reports indicating that circulating lncRNAs are, at least for a part, transported in the blood via extracellular vesicles such as exosomes<sup>19</sup>. It has even been reported that 3.36 % of the total exosomal RNA content is represented by lncRNAs<sup>98</sup>. Circulating exosomes are lipid-based extracellular vesicles that promote the transport of various biomolecules across long distances within the human body. Microvesicles and exosomes have notably been characterized as potent messengers that enable cancer cells to communicate with each other (autocrine messengers) and also with non-cancerous cells (paracrine and endocrine messengers<sup>99</sup>). Because of their lipidic structure, exosomes can fuse with the plasma membrane of a targeted cell and release their content inside it, including lncRNAs. It is thus conceivable that exosome-borne lncRNAs may be used by cancer cells to spread within the human body. Therefore, circulating lncRNAs may constitute bonafide therapeutic targets as much as tissue lncRNAs do (Fig. 1). Besides exosomes, some circulating lncRNAs may be transported as complexes with circulatory proteins such as Argonaute (Ago) or nucleophosmin 1 (NPM1) similar to circulating miRNAs<sup>100,101</sup>. Others may be transported in blood without any binding partner or specific protective structure. These lncRNAs may constitute the easiest targets for lncRNA-interfering cancer therapy. While the circulatory system is devoid of cellular machinery that degrades RNA-RNA and RNA-DNA hybrids, targeting lncRNAs using ASOs (RNAseH-dependent antisense oligonucleotide) can effectively produce significant antitumoral effects in vivo. Arun et al. have notably shown that the systemic knockdown of Malat-1 by subcutaneous injections of ASOs in an MMTV-PyMT mouse mammary carcinoma model resulted in slower tumor growth and a reduction in metastasis<sup>102</sup>.

Other studies have highlighted the existence of lncRNAs that are downregulated in cancer tissues<sup>103</sup> and the circulation of cancer patients<sup>51</sup>. Such downregulated lncRNAs may be oncosuppressor lncRNAs of which expression is dysregulated during tumorigenesis. The ectopic delivery of synthetic or purified oncosuppressor lncRNAs may constitute a promising therapeutic strategy in the future (Fig. 1). These therapeutic oncosuppressor lncRNAs may be administrated as an exosome-based formula which could possibly treat primary and secondary tumors as it spreads throughout the body via the circulatory system. If some circulating lncRNAs are indeed shown to have oncosuppressive properties in vivo, they may also be uptaken prior to cancer formation for cancer prevention purposes, similar to anti-oxidants (Fig. 1).

### Cancer-specific, multicancer and pan-cancer circulating lncRNA biomarkers and therapeutic targets

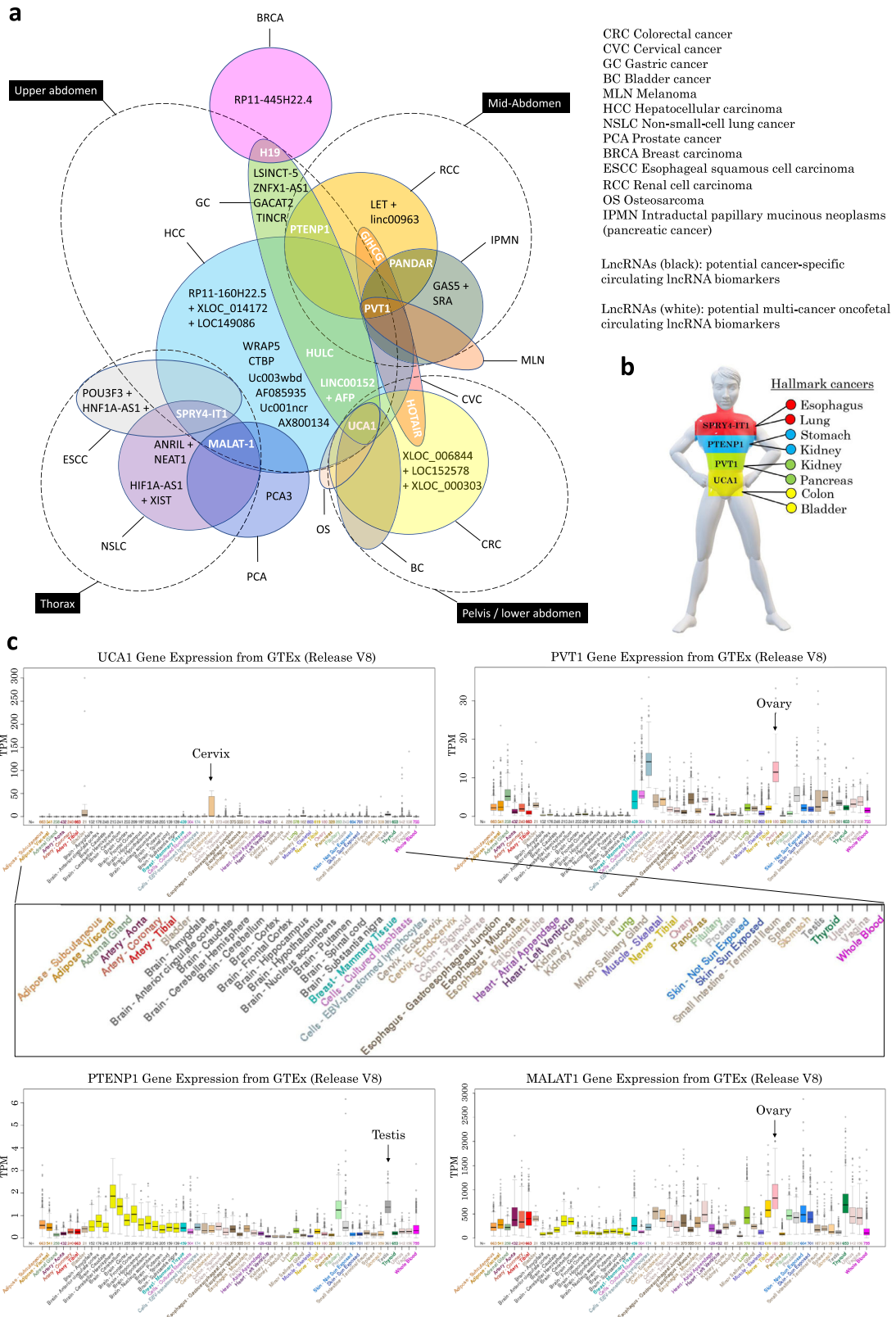
A significant number of circulating lncRNAs have been reported to be associated with only one cancer type so far (Table 1). While this could be due to a lack of studies on these lncRNAs in other cancer types, it could also imply that certain blood-based lncRNAs may really be specific to a unique type of cancer only, which has significant translational applications especially in cancer screening since the detection of abnormal levels of such lncRNAs in the circulation would not only be indicative of a cancer diagnosis but also pinpoint with accuracy the organ affected by the tumor. More studies need to be undertaken to evaluate the plausibility of these two scenarios. Interestingly, the integrated analysis of the most reported circulating lncRNAs and their specific association with certain cancers seems to reveal a pattern where some circulating lncRNAs are apparently able to reflect multiple cancers especially in organs that are close anatomically and/or embryologically (Fig. 2a, lncRNAs in white letters). For instance, circulating LINC00152, HULC and UCA1 have been associated with gastric and liver cancer, two organs that are in close proximity within the upper abdomen and which both originate from the foregut of the embryonic endoderm<sup>19,30,43,45,46,104</sup>. Lung and esophagus which are located in the thorax and share common embryological origins (before they split apart during development) also show a similar circulating lncRNA - SPRY4-IT1 - upon tumorigenesis<sup>61,63</sup>. Circulating HOTAIR has been detected in the blood of patients with cancers of the uterus and colon/rectum, organs that are located in the pelvis and sometimes fused in congenital diseases such as persistent cloaca<sup>42,105</sup>. Levels of circulating lncRNAs PVT-1 and PANDA reportedly reflect tumorigenesis or malignancy in the kidney and pancreas, two organs that are in close proximity and often grafted together<sup>68,69</sup>. Circulating PVT-1 also reflects tumor formation in the liver, an organ close anatomically and embryologically to the pancreas<sup>64</sup>. The fact that cancers from the same anatomical region or embryological origin display a similar circulating lncRNA molecular signature is consistent with the findings from an integrative study published in 2018 that analyzed the complete set of tumors in The Cancer Genome Atlas (TCGA), consisting of approximately 10,000 specimens and representing 33 cancer types<sup>106</sup>. In this study, the authors performed molecular clustering based on RNA expression levels and other key features and concluded that clustering is primarily organized by histology, tissue type, or anatomic origin<sup>106</sup>. Moreover, the embryological origin of human tumors has been largely discussed and is notably supported by evidence suggesting that adult somatic cells retain an embryonic program that can be reactivated in certain pathological conditions promoting the dedifferentiation into stem cells and eventually tumorigenesis<sup>107</sup>. In addition, machine learning has enabled the identification of key stemness features that are associated with oncogenic dedifferentiation<sup>108</sup> while embryonic stem cell-like gene expression signatures have been identified in human tumors<sup>109–111</sup>. Because of their involvement in both tumorigenesis and development, several genes including some coding for lncRNAs have been referred to as “oncofetal”<sup>112</sup>. They are reportedly upregulated in the embryo and downregulated in adults<sup>113</sup>. However, in some cancers, these oncofetal lncRNAs may be re-expressed contributing to tumorigenesis and malignancy<sup>114</sup>. In this context, cancer may arise due to loss of cellular differentiation and gain of pluri- or multipotency with the high proliferative potential characteristic of stem cells<sup>115</sup>. This concept notably led to the characterization of cancer stem cells. In fact, it is believed that, as somatic cells from different organs of the same anatomic region dedifferentiate into cancer stem cells, they may indirectly try to recreate the same embryonic organ that was originally responsible for their formation during embryogenesis (which they share in common). Based on this cumulative information, it is perhaps not surprising



**Fig. 1** Diagram summarizing the full panel of possible clinical applications that can be derived from the analysis of blood-based lncRNAs. Information indicated includes four main domains of applications (cancer prevention, cancer diagnosis, cancer prognosis, cancer treatment) and smaller subdomains referring to the domain of the same color.

to observe similar patterns of blood lncRNA levels in cancers with the same embryological or anatomical origin as shown in Fig. 2a, b. However, there are some exceptions and circulating lncRNAs may not necessarily change upon tumorigenesis according to organ location or its embryological origin (e.g. endoderm, mesoderm, ectoderm). For instance, circulating lncRNAs associated with cancer from organs related to reproduction (e.g. prostate, breast) may not follow such an anatomic/embryonic pattern as sexual organs are usually not developed during embryogenesis. Although, in healthy adults, sexual organs appear to be the main sources of some of the most widely reported cancer-associated lncRNAs such as PVT1 and MALAT1 that are mostly expressed in the ovaries of healthy women, while PTENP1 is largely expressed in the testis of healthy men (Fig. 2c). Those lncRNAs mostly remain poorly expressed in other tissues of healthy individuals. The fact that many of these lncRNAs are suppressed in most adult tissues but remain extensively expressed in sexual organs (either ovaries or testis, exclusively) suggests the likely involvement of so-called “genomic imprinting”. It essentially consists in the reprogramming of the epigenetic make-up of certain key genes according to the sex of the individual during

gametogenesis, which results in the fetus in a parent-of-origin type of gene expression with transcription occurring only on one allele while being suppressed on the other (notably through DNA methylation and histone modification). *H19* for instance is an imprinted gene that is known to be transcribed exclusively from the maternal allele and silenced on the paternal allele<sup>116</sup>. *H19* is in fact the first imprinted lncRNA-encoding gene ever identified<sup>113</sup> and its product, the lncRNA *H19* (*H19* Imprinted Maternally Expressed Transcript), has since been the object of numerous studies to understand its implications in health and disease. *H19* lncRNA has notably been reported to play critical roles in both developments<sup>117–119</sup> and tumorigenesis<sup>120–127</sup> and therefore legitimately belongs to the class of oncofetal lncRNAs<sup>112,128,129</sup>. A major mechanism by which imprinted lncRNAs such as *H19* induce or contribute to tumorigenesis likely involves a still poorly understood event known as “loss-of-imprinting” or LOI that abnormally restores gene expression on both alleles (i.e. “biallelic expression”) in adult somatic cells potentially promoting cancer formation. The reasons for sporadic LOI are not fully understood but likely involve the partial or complete loss of the imprinted epigenetic code of certain key regulatory regions within the DNA



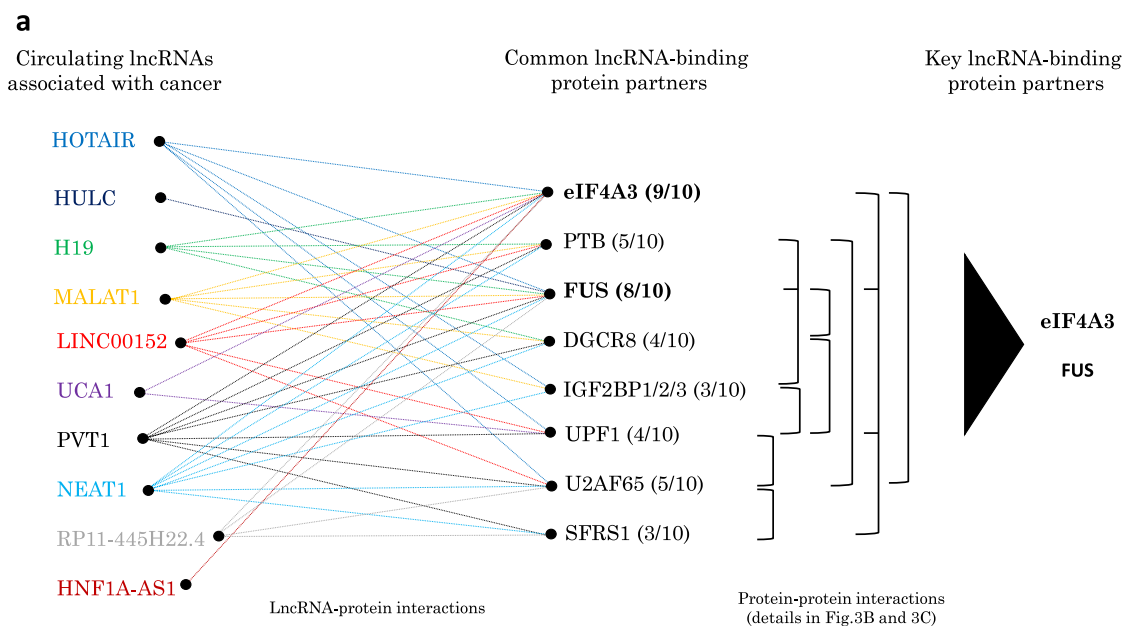
**Fig. 2 Cancer-specific and multicancer blood-derived lncRNA biomarkers.** **a** Diagram showing circulating lncRNAs reported in the literature regrouped by cancer type. Some lncRNAs (in black letters) are cancer-specific. Other circulating lncRNAs (in white letters) such as MALAT1, SPRY4-IT1, PVT1, UCA1 and LINC00152 reflect tumorigenesis in multiple organs. **b** Simplified cartoon representing the specificity of certain circulating lncRNAs towards cancers of organs located in designated anatomic segments of the human body. **c** Gene tissue expression of some of the most widely reported circulating lncRNAs with high multicancer diagnosis potential (GTEx, obtained from UCSC genome browser<sup>188–197</sup>, <https://genome.ucsc.edu/>).



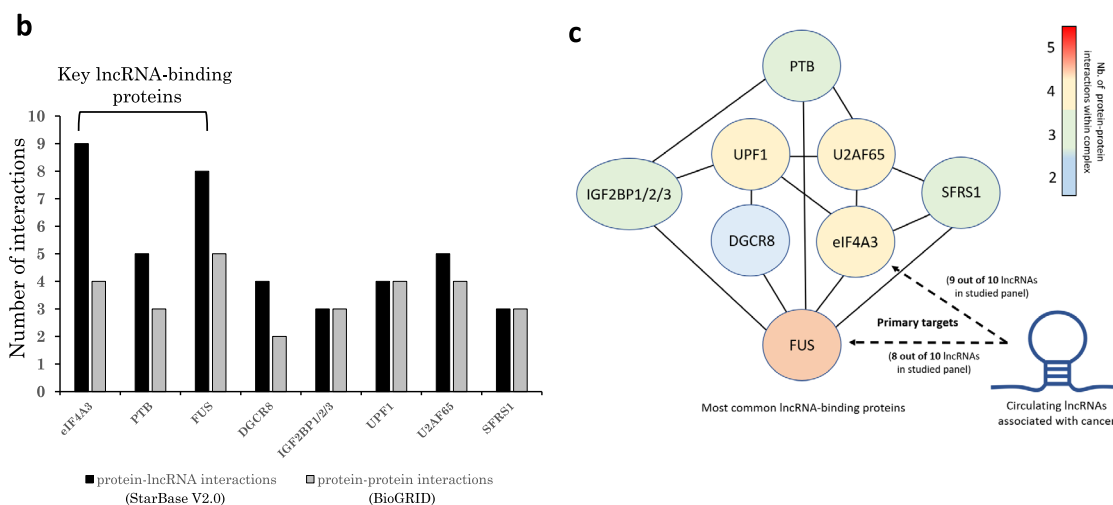
sequence notably due to major changes in methylation patterns (e.g. hypomethylation or hypermethylation) that can reportedly be induced by exposure to cigarette smoke for instance. This may affect the ability to recruit insulating proteins such as CTCF resulting in changes in the chromatic structure including de-condensation potentially promoting gene expression on the allele that should otherwise be suppressed. Eventually, it is undeniably clear that circulating imprinted lincRNAs that are expressed during development and which reflect, in adults, tumors from organs with a same embryonic origin could constitute potential “oncofetal imprinted lincRNA biomarkers” as well as promising therapeutic targets. These embryo-derived lincRNAs do represent promising multicancer biomarkers that would not only enable the detection of various types of cancers but also determine the likely

location of the tumor in the adult body as well as the organ(s) affected by tumorigenesis. Embryo-related biomarkers such as the carcinoembryonic antigen (CEA) are already in use for the diagnosis of many cancers.

The existence of potential pan-cancer circulating lincRNA biomarkers has also been investigated, including by our lab. Indeed, in a leading study based on the rigorous and systematic statistical analysis of gene expression profiles of twelve different cancer types extracted from multiple publicly available databases, our lab identified 6 promising pan-cancer lincRNA biomarkers subsequently termed “PCAN” lincRNAs that are systematically dysregulated in cancer<sup>103</sup>. Active efforts are currently undertaken to explore the full potential of these PCAN lincRNAs by extending the study to cancers beyond the original 12 cancer types. Upon



Inspired by Mohamadkhani, 2014

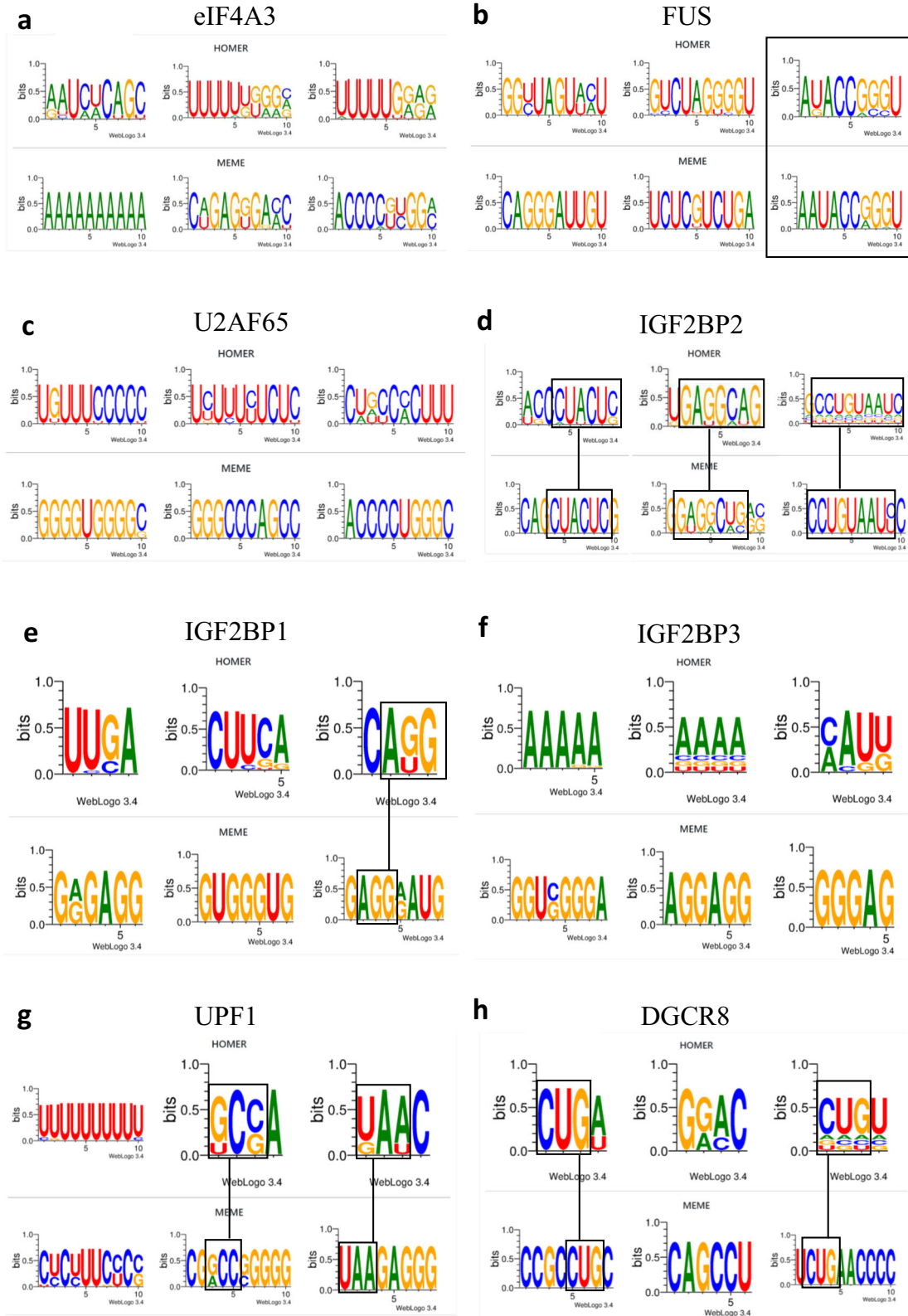


**Fig. 3** Circulating lincRNAs and a common set of protein partners. **a** Data extracted from starBase V2.0 and lincRNome databases reporting lincRNA-protein interactions occurring in tissues. Indicated lincRNAs share the same set of interacting proteins that are also known to be involved in tumorigenesis. These main proteins may constitute an oncogenic pan-lincRNA core protein interactome. Displayed protein-protein interactions are based on data from BioGRID database. **b** Graph bars representing the number of interactions with lincRNAs and proteins for each RNA-binding protein shown in (a). **c** Putative pan-cancer multimeric RNA-binding protein complex showing the different interactions between the proteins that are the most commonly recruited by cancer-related lincRNAs as shown in (a).

**Table 2.** Experimental data supporting interactions between lncRNAs and RNA-binding proteins (RBPs) that are commonly associated with cancer.

		lncRNAs commonly associated with cancer								
		HOTAIR	HULC	H19	MALAT1	LINC00152 (CYTOR)	UCA1	PVT1	NEAT1	HNF1A-AS1
Most common RNA-binding proteins (RBPs)										
eIF4A3				CLIP-seq / HITS-CLIP (NPinter, POSTAR3) <sup>143</sup>	CLIP-Seq / HITS-CLIP; iCLIP (NPinter, POSTAR3) <sup>143</sup>			CLIP-seq / HITS-CLIP; iCLIP (NPinter, POSTAR3) <sup>143</sup>	CLIP-Seq / HITS-CLIP (NPinter, POSTAR3) <sup>143</sup>	
PTB (PTBP1)	iCLIP (POSTAR3)	Affinity Capture Mass Spectrometry (BioGRID) <sup>142</sup>	Affinity Capture Mass Spectrometry (BioGRID) <sup>142</sup>	PAR-CLIP; HITS-CLIP (NPinter, lncRNome, POSTAR3)	PAR-CLIP; HITS-CLIP; iCLIP (NPinter, lncRNome, POSTAR3)	PAR-CLIP; HITS-CLIP; iCLIP; eCLIP (NPinter, lncRNome, POSTAR3)	eCLIP (POSTAR3)	PAR-CLIP; HITS-CLIP; iCLIP (NPinter, lncRNome, POSTAR3)	PAR-CLIP; HITS-CLIP; iCLIP (NPinter, lncRNome, POSTAR3)	
FUS		CLIP-seq (NPinter) <sup>144</sup>	CLIP-seq (NPinter) <sup>144</sup>	CLIP-Seq (NPinter) <sup>144</sup>	CLIP-Seq; PAR-CLIP (NPinter, POSTAR3) <sup>150,144,153</sup>	CLIP-seq / HITS-CLIP; eCLIP (NPinter, POSTAR3) <sup>144</sup>	CLIP-seq / HITS-CLIP; eCLIP (NPinter, POSTAR3) <sup>144</sup>	CLIP-Seq; PAR-CLIP (NPinter, POSTAR3) <sup>150,144</sup>	RIP; CHART-seq CLIP-Seq; PAR-CLIP (NPinter, POSTAR3) <sup>150,144,145,156</sup>	PAR-CLIP (POSTAR3)
DGCR8				eCLIP (NPinter, POSTAR3) <sup>154</sup>	eCLIP; HITS-CLIP (NPinter, POSTAR3) <sup>154</sup>	eCLIP (NPinter) <sup>154</sup>	eCLIP (POSTAR3)	eCLIP (NPinter) <sup>154</sup>	eCLIP; CHART-seq; HITS-CLIP (NPinter, POSTAR3) <sup>154,156</sup>	eCLIP (NPinter) <sup>154</sup>
IGF2BP1/2/3	PAR-CLIP (NPinter, lncRNome, POSTAR3) <sup>151</sup>	Affinity Chromatography (NPinter) <sup>157</sup>	Affinity Chromatography (NPinter) <sup>157</sup>	iCLIP; PAR-CLIP; RT-PCR in situ Hybridization Northern Blot (NPinter, POSTAR3) <sup>165,166</sup>	eCLIP; PAR-CLIP; iCLIP (NPinter, lncRNome, POSTAR3) <sup>151,154,155</sup>	PAR-CLIP (POSTAR3)	iCLIP; eCLIP (POSTAR3)	PAR-CLIP; eCLIP; iCLIP (NPinter, lncRNome, POSTAR3) <sup>154</sup>	eCLIP; PAR-CLIP; iCLIP; CHART-Seq (NPinter, POSTAR3) <sup>151,154-156</sup>	eCLIP (NPinter) <sup>154</sup>
UPF1	PAR-CLIP; HITS-CLIP (NPinter, POSTAR3) <sup>152</sup>			HITS-CLIP (POSTAR3)	eCLIP; PAR-CLIP; HITS-CLIP; iCLIP (NPinter, POSTAR3) <sup>154,146,152</sup>	eCLIP; HITS-CLIP; iCLIP (NPinter, POSTAR3) <sup>154</sup>	iCLIP (POSTAR3)	eCLIP; PAR-CLIP; iCLIP; HITS-CLIP (NPinter, POSTAR3) <sup>154,146,152</sup>	eCLIP; PAR-CLIP; HITS-CLIP (NPinter, POSTAR3) <sup>152</sup>	
U2AF65	iCLIP (POSTAR3)			iCLIP (POSTAR3)	iCLIP (POSTAR3)	iCLIP (POSTAR3)		iCLIP (POSTAR3)	iCLIP (POSTAR3)	
SFRS1	iCLIP (POSTAR3)			PAR-CLIP; iCLIP (POSTAR3)	iCLIP; Microarray eCLIP; CLIP-Seq CHART-seq; PAR-CLIP (NPinter, POSTAR3) <sup>147-149,154,156</sup>	eCLIP; iCLIP (NPinter, POSTAR3) <sup>154</sup>	eCLIP; iCLIP (POSTAR3)	eCLIP; CLIP-Seq; PAR-CLIP; iCLIP (NPinter, POSTAR3) <sup>147,154</sup>	eCLIP; CLIP; PAR-CLIP CLIP-Seq; CHART-seq (NPinter, POSTAR3) <sup>147,154,156</sup>	eCLIP (NPinter) <sup>154</sup>
Other RBPs of interest	AGO2, ELAVL1, EZH2. Affinity Capture-RNA; Protein-RNA (BioGRID, POSTAR3) <sup>160-164</sup>	AGO2. PAR-CLIP; HITS-CLIP Affinity Capture - RNA (BioGRID) <sup>158</sup>	AGO2. PAR-CLIP; HITS-CLIP Affinity Capture - RNA (BioGRID) <sup>158</sup>	AGO2, ELAVL1, EZH2. PAR-CLIP; HITS-CLIP Affinity Capture - RNA (BioGRID) <sup>158</sup>	AGO2, ELAVL1, EZH2. PAR-CLIP; HITS-CLIP Affinity Capture - RNA (POSTAR3, BioGRID) <sup>158</sup>	AGO2, ELAVL1. PAR-CLIP; HITS-CLIP (POSTAR3)	AGO2, ELAVL1. PAR-CLIP; HITS-CLIP (POSTAR3)	AGO2, ELAVL1. PAR-CLIP; HITS-CLIP (POSTAR3) <sup>147,154,156</sup>	AGO2, ELAVL1. PAR-CLIP; HITS-CLIP (POSTAR3)	AGO2, ELAVL1. PAR-CLIP Affinity Capture - RNA (POSTAR3, BioGRID) <sup>159</sup>

Information extracted from several databases including NPinter<sup>136-139</sup>, BioGRID<sup>140</sup>, lncRNome<sup>134</sup> and POSTAR3<sup>141</sup>. CLIP UV Cross-Linking and Immunoprecipitation, PAR-CLIP Photoactivatable Ribonucleoside-enhanced Crosslinking and Immunoprecipitation, eCLIP Enhanced CLIP, iCLIP Individual-nucleotide resolution UV Crosslinking and Immunoprecipitation, CHART-seq Capture Hybridization Analysis of RNA Targets, RIP RNA Immunoprecipitation.



**Fig. 4 Putative consensus motifs in lncRNAs for the specific binding of key RNA-binding proteins.** Data extracted from POSTAR3 database (CLIPseq-based)<sup>141</sup> and processed by HOMER and MEME algorithms that are commonly used for motif discovery and next-generation sequencing (NGS) data analysis. Square boxes highlight similar patterns identified in the motifs provided by both algorithms. **a** Consensus motif for binding of RNA-binding protein eIF4A3 (eukaryotic initiation factor 4A-III). **b** Consensus motif for binding of RNA-binding protein FUS (fused in sarcoma). **c** Consensus motif for binding of RNA-binding protein U2AF65 (splicing factor U2AF 65kDa subunit). **d** Consensus motif for binding of RNA-binding protein IGF2BP2 (insulin-like growth factor 2 mRNA-binding protein 2). **e** Consensus motif for binding of RNA-binding protein IGF2BP1 (insulin-like growth factor 2 mRNA-binding protein 1). **f** Consensus motif for binding of RNA-binding protein IGF2BP3 (insulin-like growth factor 2 mRNA-binding protein 3). **g** Consensus motif for binding of RNA-binding protein UPF1 (regulator of nonsense transcripts 1). **h** Consensus motif for binding of RNA-binding protein DGCR8 (microprocessor complex subunit DGCR8, DiGeorge syndrome critical region 8).

validation in blood-based samples, this panel of PCAN biomarkers could potentially constitute the first set of circulating lincRNAs capable of detecting any kind of cancer in the human body. Further investigations would also be required to better understand the molecular mechanisms associated with the upregulation of these PCAN lincRNAs in cancer and to assess whether they could constitute potential pan-cancer therapeutic targets as well as imprinted oncofetal genes similar to *H19*.

### Circulating lincRNAs and association with RNA-binding proteins

While RNA-binding proteins may not interact with circulating lincRNAs once they reach the bloodstream, they may bind lincRNAs inside the tumor cells prior to secretion and may actively contribute to the tumorigenic process. Indeed, many RNA-binding proteins that interact with lincRNAs have also been characterized as oncofetal<sup>130,131</sup>. This suggests that lincRNA-related tumorigenesis is likely the result of a complex and diversified molecular mechanism that involves the upregulation of several oncofetal genes, including genes coding for oncofetal lincRNAs and oncofetal lincRNA-binding proteins. Investigators can find information of lincRNA-binding partners by screening databases such as lincRNome, lincRNAMap, starBase V2.0 and UCSC genome browser<sup>132–135</sup>. Further information on the experimental data which support the lincRNA-protein interactions described in Fig. 3 can be found in Table 2. This table provides substantial scientific

information that has been extracted from other highly valuable databases such as NPInter<sup>136–139</sup>, BioGRID<sup>140</sup> and POSTAR3<sup>141</sup> which rigorously report data from Affinity Capture-Mass Spectrometry (BioGRID terminology)<sup>142</sup>, UV Cross-Linking and Immunoprecipitation (CLIP) / CLIP-seq / HITS-CLIP<sup>143–149</sup>, Photoactivatable Ribonucleoside-enhanced Crosslinking and Immunoprecipitation (PAR-CLIP)<sup>150–153</sup>, Enhanced CLIP (eCLIP)<sup>154,155</sup>, Individual-nucleotide resolution UV Crosslinking and Immunoprecipitation (iCLIP), Capture Hybridization Analysis of RNA Targets (CHART-seq)<sup>156</sup>, Affinity Chromatography<sup>157</sup>, as well as other methods such as RNA Immunoprecipitation (RIP), Affinity Capture-RNA (BioGRID terminology)<sup>158–161</sup> and other “Protein-RNA” methods (BioGRID terminology)<sup>162–164</sup> which may also include a combination of Immunocytochemistry (ICC), In Situ Hybridization, Northern Blot and/or RT-PCR<sup>165,166</sup>.

Systematic analysis of these databases actually revealed a common set of proteins that consistently interacts with the most reported cancer-related lincRNAs (Fig. 3a)<sup>167</sup>. Most of these proteins are associated with cancer formation upon dysregulation, especially IGF2BP3<sup>168,169</sup>, FUS<sup>170,171</sup> and eIF4A3<sup>172</sup>. This suggests the likely existence of a pan-lincRNA core protein interactome that may, by itself, be sufficient to promote tumorigenesis. However, some of these proteins appear to be more frequently involved in lincRNA interactions than others and may play a more central role in cancer formation. For instance, eIF4A3 was found to interact with 9 of out 10 lincRNAs in the

**Table 3.** Guidelines recommended for the study of circulating lincRNAs as biomarkers for cancer diagnosis, based on troubleshooting performed by previous works.

Step	Recommended	Reason	Reference
Patient selection	Exclude patients with inflammation	Higher / different levels of white blood cells associated with inflammation may impact levels of circulating RNAs upon cytolysis	198,199
	Recruit patients with same gender, age and race	Minimize variation in lincRNA levels due to possible inter-individual confounding factors (such as SNPs, CNV, etc.)	63
	May include questionnaire about diet and lifestyle	Diet and lifestyle (alcohol consumption, smoking) can affect lincRNA levels	200,201
Blood sample preparation	Prepare serum or plasma. Discard cellular fraction	Cellular fraction of blood may contain different levels of blood cells which in return may impact levels of circulating RNAs upon cytolysis	199,202
	Strict standard operating procedures when preparing serum/plasma	Minimize variations in circulating RNAs due to sample preparation. Avoid hemolysis.	202
	Measure $A_{414}$ , $A_{541}$ , $A_{576}$	Assess for hemolyzed samples	69
RNA extraction	Use kits compatible with liquid samples	Enable extraction of circulating lincRNAs from plasma or serum samples	Kit manufacturers
	Use kits combining both solid (filter) and liquid phase (organic) extraction	Maximize extraction of circulating lincRNAs from plasma or serum samples	17,24,42
	Use as much plasma/serum as possible	Maximize RNA yield after extraction	Our recommendation
Reverse Transcription	Use same volume of RNA extracts	Allow maximum RNA input for Reverse Transcription	Our recommendation
qPCR (relative quantification with $\Delta\Delta Ct$ method)	Test several reference genes. Carefully choose best reference gene(s) using NormFinder, RefFinder or Genorm algorithms. Most popular: GAPDH, beta-actin, 18 S To avoid: RPLPO, GUSB, HPRT	The right reference gene is needed for accurate relative quantification using $\Delta\Delta Ct$ method. GAPDH, beta-actin, 18 S present in large quantities in blood. RPLPO levels inconsistent in blood GUSB, HPRT levels too low in blood	3,25,41,42,47,203
	Careful in interpretation of data when using spike-in controls	Spike-in controls do not account for variations in lincRNA concentrations in blood-derived samples prior to RNA extraction step	180
	Measure transcript levels of <i>MB</i> , <i>NGB</i> , <i>CYGB</i> genes	Assess for contamination from red blood cells	69
	Measure transcript levels of <i>APOE</i> , <i>CD68</i> , <i>CD2</i> , <i>CD3</i> genes	Assess for contamination from white blood cells	69

Information reported includes step of the analysis, actual recommendation, reason for the recommendation and related literature reference.

lncRNA panel reported here (Fig. 3a, b), while FUS was recruited by 8 out of 10 lncRNAs. Therefore, eIF4A3 and FUS may constitute key lncRNA-binding proteins that could be part of a pan-cancer molecular mechanism that mediates the tumorigenic properties of most oncogenic lncRNAs and/or generally promotes lncRNA secretion into the systemic circulation from the tumor site. Thus, eIF4A3 and FUS may represent major pan-cancer therapeutic targets. While other RNA-binding proteins appear to be less frequently recruited by cancer-related lncRNAs, they may still exert pan-tumorigenic properties since all RNA-binding proteins reported here in Fig. 3a are part of a very same multimeric protein complex based on data from an extensive search of protein-protein interactions using BioGRID database (Fig. 3c). Interestingly, eIF4A3 and FUS showed the highest ability to interact with other RNA-binding proteins (respectively binding 4 and 5 other protein partners within the complex), which may explain why they are often associated with lncRNAs since the more lncRNA-binding proteins they bind, the more lncRNAs they collect. Given the relatively high frequency of recruitment of eIF4A3, FUS and related RNA-binding proteins (RBPs) by cancer-associated lncRNAs and their known roles in tumorigenesis, we here provide in Fig. 4 the putative consensus motifs that enable lncRNAs to specifically bind these RBPs, as this may help investigators to identify novel interactions between their lncRNA of interest and these tumorigenic RBPs (consensus motifs extracted from POSTAR3 database which reports CLIP-seq data<sup>141</sup>).

Overall, it is clear that lncRNAs and their interacting partners will constitute innovative therapeutic targets and/or agents in future cancer therapy strategies.

## DISCUSSION AND FUTURE PERSPECTIVES

Circulating lncRNAs have been shown to constitute reliable biomarkers for both cancer diagnosis and prognosis. They have also been suggested as potential therapeutic targets, notably due to the fact that they are reportedly transported in the bloodstream by exosomes which are known to contribute to cancer progression and metastasis by enabling communication between cancer cells that produce those exosomes and non-cancerous “target” cells which may be incited to transform into new cancer cells under exposure to exosome-borne oncogenic lncRNAs<sup>99</sup>. Interestingly, those tumor-derived exosomes (or TD-exosomes) appear to display a unique molecular signature that differs from that of non-cancerous exosomes potentially providing a window of opportunity for future antitumoral therapies aiming to stop the formation of secondary tumors by specifically targeting TD-exosomes. In terms of diagnostic performance, while it can be improved by combining multiple lncRNAs, it is important to note that the “specificity” determined in the reported studies refers to the comparative analysis of samples from healthy volunteers and patients with specific cancer. In this particular context, “specificity” does not describe the ability to distinguish a certain cancer type from other cancers. This is particularly relevant since several circulating lncRNAs have been proposed as potential biomarkers for a large variety of different cancers. For instance, MALAT-1 could be used to diagnose prostate cancer<sup>26</sup> and non-small-cell lung cancer<sup>25,51</sup>. Similarly, HOTAIR has the potential to detect both colorectal<sup>42</sup> and cervical cancer<sup>105</sup>. LINC00152 could lead to the diagnosis of both hepatocellular carcinoma<sup>43</sup> and gastric cancer<sup>19</sup>. lncRNA GIHCG has been shown to be involved in the pathogenesis of many types of different cancers including liver, cervical, gastric, renal and colorectal cancer for which it may constitute a promising biomarker<sup>33,34,90,173–175</sup>. PVT1 has been reported as a potential circulating biomarker (alone or in combination with other lncRNAs) for at least five different types of cancers including RCC (kidney), IPMN (pancreas), HCC (liver), MLN (skin), and CVC (cervix)<sup>64,68,176,177</sup>. UCA1 constitutes another lncRNA with significant multicancer diagnostic potential since it has been reported to

### Box 1 Advice on patient recruitment and sample selection when studying circulating RNAs as biomarkers for early cancer diagnosis

- While whole blood has been successfully used in circulating lncRNA studies<sup>51</sup>, it is usually not recommended for accurate quantification of circulating RNAs due to variability associated with red and white blood cells<sup>202</sup>. Indeed, levels of white blood cells (and thus circulating RNAs) are likely to change if patients are experiencing chronic or acute inflammation which may not necessarily be related to the disease investigated<sup>198,199</sup>. Cell-free samples such as plasma (blood fraction obtained with anti-coagulants) and serum (blood fraction obtained after coagulation) are more reliable sources of circulating lncRNAs and have been largely used in studies comparing circulating lncRNA levels in cancer patients and healthy subjects (Table 1).
- Levels of circulating RNAs may also vary within the same group of individuals (e.g. healthy volunteers) due to internal factors such as patient hydration level or diet<sup>200,201</sup> as well as age, gender and race. Copy number variations (CNVs) and single nucleotide polymorphisms (SNPs) have also been proposed as possible sources of variations in levels of circulating lncRNAs. Consequently, investigators usually collect relevant patient information and compare individuals with similar records.
- Equal volumes of plasma from different patients may not contain the same RNA concentration. Inconsistent serum or plasma preparation across samples may add another level of variability in RNA content especially if hemolysis could not be avoided. To account for hemolysis, Permut et al. visually inspected their samples and measured absorbance at three different wavelengths. An absorbance exceeding 0.2 for any of these wavelengths indicated hemolyzed samples<sup>69</sup>. They further assessed for blood-cell contaminants by measuring levels of transcripts from *MB*, *NGB* and *CYGB* genes (for erythrocytes) as well as *APOE*, *CD68*, *CD2* and *CD3* (for leucocytes).

### Box 2 Extraction of circulating lncRNAs from liquid biopsy: Pitfalls and Recommendations

- Investigators that wish to use column-based kits should be aware that most commercially available kits are optimized for non-liquid samples such as cells or tissues, and not for plasma or serum. Some kits such as the miRNeasy Serum/ Plasma kit do allow RNA extraction from serum and plasma, but it is mostly designed for purification of microRNAs (miRNAs) and other small RNAs.
- Since lncRNAs are naturally scarce in circulation, investigators may wish to use large volumes of plasma or serum to increase the RNA yield upon extraction. However, most kits are provided with columns of limited size which may introduce variability in RNA yields, as investigators often have to perform successive column-based purifications with small volumes of the same sample. If different kit formats are available (for instance mini, midi and maxi), investigators should proceed with the kit that is the most suitable for their study based on the volume of samples that is available to them. Note that if the volume of the original plasma sample is too small, the RNA yield might be too low for RNA quantification and qPCR detection.
- Despite the relatively low RNA yield generated from blood-based samples, most kits provide RNA samples of high purity due to solid-phase extraction and multiple washing steps. Improved RNA extraction may come from the addition of an organic extraction based on liquid phase separation using phenol/chloroform. For instance, the mirVana kit which combines both solid phase (filter) and liquid phase (chloroform) RNA extraction has been largely used in cancer studies focusing on circulating lncRNAs<sup>17,26,42</sup>. This kit appears popular among investigators because it allows total RNA extraction from liquid samples (plasma/serum) as well as purification of small RNAs and lncRNAs.

effectively detect (alone or in combination with other lncRNAs) at least five distinct cancers such as HCC (liver), GC (stomach), BC (bladder), CRC (colon) and osteosarcoma (bone)<sup>41,45,46,104,178,179</sup>.

The increasing number of studies on circulating lncRNAs may eventually indicate that all circulating lncRNAs reflect more than one cancer and that there is no unique biomarker for each cancer type or subtype. It has especially been suggested that changes in lncRNA level in the circulation of cancer patients could be due to a general pathophysiological response from the body to the presence of tumors and not due to direct secretions from the tumors themselves<sup>180</sup>. This represents a strong argument as

### Box 3 Potential issues when performing RT-qPCR with circulating lncRNAs from liquid biopsy

- Investigators have the choice to use equal amounts of total RNAs or equal volumes of RNA extracts.
- Using equal quantities of total RNAs has drawbacks since the quantity of total RNAs obtained after extraction from plasma or serum samples is usually very low and may be undetectable using spectrophotometers, while the adjustment of all samples to the sample with the lowest RNA concentration may reduce the output of the subsequent qPCR reaction.
- When using equal volumes of RNA extracts, normalization using reference genes becomes indispensable for the analysis of qPCR data by relative quantification ( $\Delta\Delta C_t$  method). However, there is no consensus on the best reference genes. They are case-sensitive and must be evaluated for each study, which may be challenging due to inherent variability associated with cancer.<sup>199,202</sup> Ultimately, one must determine whether these variations are statistically significant or not.<sup>25</sup> Usually, the least variable candidate is selected by using algorithms such as NormFinder, RefFinder or Genorm.<sup>25,41,42</sup>
- Due to poor abundance or recurrent variability, several genes should be avoided. *RPLPO* which is a commonly used for tissue sample analysis is not recommended for the quantification of blood-based biomarkers<sup>25,203</sup>, while *HPRT* and *GUSB* transcripts are not abundant enough in normal human serum<sup>41</sup>.
- Exogenous spike-in controls may be used to account for variability introduced during RNA extraction, however they do not reflect inherent variations in RNA concentrations prior to RNA extraction<sup>180</sup>. We recommend using both spike-in controls and reference genes to better account for variations in circulating lncRNA levels. Investigators may also evaluate lncRNA levels by absolute qPCR quantification using standard curves made with reference standards.<sup>38</sup>
- All guidelines from Boxes 1, 2 and 3 are summarized in Table 3.

### Box 4 Tips for analyzing circulating lncRNAs with next-generation sequencing-based technologies

- There are two paradigms in next-generation sequencing technology: short-read sequencing (35–700 bp)<sup>204</sup> and long-read sequencing (>5000 bp)<sup>205</sup>. Short-read sequencing provide lower-cost, higher-accuracy data that are useful for population-level research and clinical variant discovery<sup>204</sup>, while long-read approaches provide read lengths that are well suited for de novo genome assembly applications and full-length isoform sequencing<sup>205</sup>. In practice, short-read sequencing is usually used for cancer early diagnosis<sup>206</sup>.
- It is necessary to enhance lncRNA concentration by building lncRNA-specific cDNA library using oligonucleotide capture technology<sup>207</sup>. With complementary oligonucleotide probes, this technology increases the concentration of lncRNA sequences by at least 25%<sup>207</sup>. The optimization of probe sequences is similar to that for microarray technologies<sup>208</sup>.
- The computational preprocessing of lncRNA sequencing requires using Quality-Control (QC) tools such as FastQC<sup>209,210</sup> or AfterQC<sup>211</sup>. The processed reads are then aligned to a noncoding transcriptome reference such as LNCipedia<sup>212</sup>, using tools such as HISAT2<sup>213</sup> or BowTie2<sup>214</sup>. The mapped reads are assembled to lncRNAs using tools such as StringTie2<sup>215</sup>. The quantity of lncRNAs is normalized by the sequencing depth and transcript length using RPKM, FPKM or TPM method<sup>216</sup>. The differentially expressed lncRNAs are detected by tools such as DESeq2<sup>217</sup> or edgeR<sup>218</sup>. Risk score analysis can determine the association between cancer and differentially expressed lncRNAs<sup>219</sup>. By associating lncRNAs with the nearest protein coding genes, their biological functions may be explored through gene ontology (GO)<sup>220</sup> and KEGG pathway<sup>221</sup> analyses. Finally, classification models can be used to identify potential lncRNA biomarkers, with individual training, testing and validation datasets<sup>222,223</sup>. Metrics such as AUC in ROC are used to measure the effectiveness of the lncRNAs in predicting cancer.<sup>103</sup>

significant levels of lncRNAs have been detected in the blood of cancer-free healthy subjects. This would also explain why there is sometimes a lack of correlation between circulating lncRNA levels and cancer tissue lncRNA levels. Thus, circulating lncRNAs may actually reflect the presence of tumors in general. In this context, it is likely that in the near future pan-cancer circulating biomarkers could be identified. On the other hand, the findings from recent studies suggest that the detection of a specific cancer type may be achieved by using multi-analyte liquid biopsy and multi-modal strategies, including lncRNA detection<sup>181,182</sup>. For instance, to

better predict specific lncRNA-cancer associations, Yan et al. developed an original method termed DRACA (for “detecting lncRNA-cancer association”), based on the analysis of five different types of features including lncRNAs, miRNAs, genes, cancer types and cancer prognosis (3)<sup>181</sup>. We here provide the name of the databases used by the authors, as these may be useful to other investigators. StarBase v2.0 was used for lncRNA-miRNA relationships<sup>135</sup>, lncReg for lncRNA-gene interactions<sup>183</sup>, lncRNADisease for lncRNA-cancer associations<sup>184</sup>, miRTarbase for miRNA-gene relationships<sup>185</sup>, MNDR v2.0 for miRNA-cancer relationships<sup>186</sup> and DisGeNet for gene-cancer relationships<sup>187</sup>. DRACA eventually outperformed other methods in predicting specific lncRNA-cancer associations<sup>181</sup>. In another outstanding study, Sanchez-Salcedo et al. reported that the specific detection of prostate cancer can be performed by using a dual electrochemical hybridization-based biosensor with enzymatic signal amplification for the detection of both PCA3 lncRNA and PSA mRNA (prostate-specific antigen, non-lncRNA)<sup>182</sup>. One major advantage of this technique compared to commercial tests, is that it reportedly enables the detection of PCA3 lncRNA in urine samples of prostate cancer patients without prior RNA amplification. Because the study of circulating lncRNAs via traditional RT-qPCR or next-generation sequencing methods can sometimes be quite challenging, we here provide relevant guidelines that may be useful to investigators who are new to the field (boxes 1–4, and Table 3).

Overall, while the study of circulating lncRNAs is still at an early stage, the worldwide growing interest in lncRNAs and the emergence of new technologies to improve their detection, specificity, and potential in clinical applications undeniably increases the chance of discovering one day reliable blood-based biomarkers that will allow the early and accurate detection of any type of cancer.

### Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### DATA AVAILABILITY

Weblinks to publicly available databases mentioned in this manuscript are provided below.

lncRNA-protein interactions:

NPInter: <http://bigdata.ibp.ac.cn/npinter4/>

POSTAR3: <http://POSTAR.ncrnalab.org/>

StarBase: <https://starbase.sysu.edu.cn/starbase2/>

lncRNome: <http://genome.igib.res.in/lncRNome/>

Protein-protein interactions:

BioGRID: <https://thebiogrid.org/>

lncRNA expression (GTEx):

UCSC Genome Browser<sup>188–197</sup>: <https://genome.ucsc.edu/>

3D models obtained from Paint 3D (Microsoft, <https://www.microsoft.com/en-us/>) and used with the permission from Microsoft.

The data generated and/or analyzed in the current study are available from the corresponding author on reasonable request.

All data generated or analyzed during this study are included in this published article.

Received: 20 October 2021; Accepted: 13 May 2022;

Published online: 21 June 2022

### REFERENCES

1. Arieta, O. et al. The progressive elevation of alpha fetoprotein for the diagnosis of hepatocellular carcinoma in patients with liver cirrhosis. *BMC Cancer* **7**, 28 (2007).
2. Zhou, L., Liu, J. & Luo, F. Serum tumor markers for detection of hepatocellular carcinoma. *World J. Gastroenterol.* **12**, 1175–1181 (2006).
3. Tang, J. et al. Circulation long non-coding RNAs act as biomarkers for predicting tumorigenesis and metastasis in hepatocellular carcinoma. *Oncotarget* **6**, 4505–4515 (2015).

4. Di Carlo, I. et al. Persistent increase in alpha-fetoprotein level in a patient without underlying liver disease who underwent curative resection of hepatocellular carcinoma. A case report and review of the literature. *World J. Surg. Oncol.* **10**, 79 (2012).
5. Hsieh, C. B. et al. Is inconsistency of alpha-fetoprotein level a good prognosticator for hepatocellular carcinoma recurrence? *World J. Gastroenterol.* **16**, 3049–3055 (2010).
6. Duffy, M. J., Evoy, D. & McDermott, E. W. CA 15-3: uses and limitation as a biomarker for breast cancer. *Clin. Chim. Acta* **411**, 1869–1874 (2010).
7. Maric, P. et al. Tumor markers in breast cancer-evaluation of their clinical usefulness. *Coll. Antropol.* **35**, 241–247 (2011).
8. Patani, N., Martin, L. A. & Dowsett, M. Biomarkers for the clinical management of breast cancer: international perspective. *Int J. Cancer* **133**, 1–13 (2013).
9. Molina, R. et al. Tumor markers in breast cancer- European Group on Tumor Markers recommendations. *Tumour Biol.* **26**, 281–293 (2005).
10. Harris, L. et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J. Clin. Oncol.* **25**, 5287–5312 (2007).
11. Wu, S. G. et al. Serum levels of CEA and CA15-3 in different molecular subtypes and prognostic value in Chinese breast cancer. *Breast* **23**, 88–93 (2014).
12. Inamura, K. Update on Immunohistochemistry for the Diagnosis of Lung Cancer. *Cancers (Basel)* **10**, 30072 (2018).
13. Schwarzenbach, H., Hoon, D. S. & Pantel, K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat. Rev. Cancer* **11**, 426–437 (2011).
14. Neves, A. F., Dias-Oliveira, J. D., Araujo, T. G., Marangoni, K. & Goulart, L. R. Prostate cancer antigen 3 (PCA3) RNA detection in blood and tissue samples for prostate cancer diagnosis. *Clin. Chem. Lab Med* **51**, 881–887 (2013).
15. Sartori, D. A. & Chan, D. W. Biomarkers in prostate cancer: what's new? *Curr. Opin. Oncol.* **26**, 259–264 (2014).
16. Cui, Y. et al. Evaluation of prostate cancer antigen 3 for detecting prostate cancer: a systematic review and meta-analysis. *Sci. Rep.* **6**, 25776 (2016).
17. Arita, T. et al. Circulating long non-coding RNAs in plasma of patients with gastric cancer. *Anticancer Res* **33**, 3185–3193 (2013).
18. Tsai, M. C. et al. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
19. Li, Q. et al. Plasma long noncoding RNA protected by exosomes as a potential stable biomarker for gastric cancer. *Tumour Biol.* **36**, 2007–2012 (2015).
20. Rinn, J. L. et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
21. Yu, W. et al. Exosome-based liquid biopsies in cancer: opportunities and challenges. *Ann. Oncol.* **32**, 466–477 (2021).
22. Liebs, S. et al. Applicability of liquid biopsies to represent the mutational profile of tumor tissue from different cancer entities. *Oncogene* **40**, 5204–5212 (2021).
23. Ji, P. et al. MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* **22**, 8031–8041 (2003).
24. Ren, S. et al. RNA-seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and aberrant alternative splicings. *Cell Res* **22**, 806–821 (2012).
25. Weber, D. G. et al. Evaluation of long noncoding RNA MALAT1 as a candidate blood-based biomarker for the diagnosis of non-small cell lung cancer. *BMC Res Notes* **6**, 518 (2013).
26. Ren, S. et al. Long non-coding RNA metastasis associated in lung adenocarcinoma transcript 1 derived miniRNA as a novel plasma-based biomarker for diagnosing prostate cancer. *Eur. J. Cancer* **49**, 2949–2959 (2013).
27. Panzitt, K. et al. Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA. *Gastroenterology* **132**, 330–342 (2007).
28. Xie, H., Ma, H. & Zhou, D. Plasma HULC as a promising novel biomarker for the detection of hepatocellular carcinoma. *Biomed. Res Int* **2013**, 136106 (2013).
29. Tan, L., Yang, Y., Shao, Y., Zhang, H. & Guo, J. Plasma lncRNA-GACAT2 is a valuable marker for the screening of gastric cancer. *Oncol. Lett.* **12**, 4845–4849 (2016).
30. Xian, H. P., Zhuo, Z. L., Sun, Y. J., Liang, B. & Zhao, X. T. Circulating long non-coding RNAs HULC and ZNF1-AS1 are potential biomarkers in patients with gastric cancer. *Oncol. Lett.* **16**, 4689–4698 (2018).
31. Elsayed, E. T., Salem, P. E., Darwish, A. M. & Fayed, H. M. Plasma long non-coding RNA HOTAIR as a potential biomarker for gastric cancer. *Int J Biol Markers* **1**, 1724600818760244 (2018).
32. Yamamoto, H., Watanabe, Y., Sato, Y., Maehata, T. & Itoh, F. Non-Invasive Early Molecular Detection of Gastric Cancers. *Cancers (Basel)* **12**, 102880 (2020).
33. Sui, C. J. et al. Long noncoding RNA GIHCG promotes hepatocellular carcinoma progression through epigenetically regulating miR-200b/a/429. *J. Mol. Med (Berl.)* **94**, 1281–1296 (2016).
34. He, Z. H., Qin, X. H., Zhang, X. L., Yi, J. W. & Han, J. Y. Long noncoding RNA GIHCG is a potential diagnostic and prognostic biomarker and therapeutic target for renal cell carcinoma. *Eur. Rev. Med Pharm. Sci.* **22**, 46–54 (2018).
35. Qu, L. et al. Exosome-Transmitted lncARSR Promotes Sunitinib Resistance in Renal Cancer by Acting as a Competing Endogenous RNA. *Cancer Cell* **29**, 653–668 (2016).
36. Barth, D. A. et al. Circulating Non-coding RNAs in Renal Cell Carcinoma-Pathogenesis and Potential Implications as Clinical Biomarkers. *Front Cell Dev. Biol.* **8**, 828 (2020).
37. Shao, Y. et al. Gastric juice long noncoding RNA used as a tumor marker for screening gastric. *Cancer Cancer* **120**, 3320–3328 (2014).
38. Zhou, X., Yin, C., Dang, Y., Ye, F. & Zhang, G. Identification of the long non-coding RNA H19 in plasma as a novel biomarker for diagnosis of gastric cancer. *Sci. Rep.* **5**, 11516 (2015).
39. Zhang, K. et al. Circulating lncRNA H19 in plasma as a novel biomarker for breast cancer. *Cancer Biomark.* **17**, 187–194 (2016).
40. Xu, N. et al. Clinical significance of high expression of circulating serum lncRNA RP11-445H22.4 in breast cancer patients: a Chinese population-based study. *Tumour Biol.* **36**, 7659–7665 (2015).
41. Dong, L. et al. Circulating CUDR, LSINCT-5 and PTENP1 long noncoding RNAs in sera distinguish patients with gastric cancer from healthy controls. *Int J. Cancer* **137**, 1128–1135 (2015).
42. Svoboda, M. et al. HOTAIR long non-coding RNA is a negative prognostic factor not only in primary tumors, but also in the blood of colorectal cancer patients. *Carcinogenesis* **35**, 1510–1515 (2014).
43. Li, J. et al. HULC and linc00152 Act as Novel Biomarkers in Predicting Diagnosis of Hepatocellular Carcinoma. *Cell Physiol. Biochem* **37**, 687–696 (2015).
44. Huang, J. et al. A Circulating Long Noncoding RNA Panel Serves as a Diagnostic Marker for Hepatocellular Carcinoma. *Dis. Markers* **2020**, 5417598 (2020).
45. Kamel, M. M. et al. Investigation of long noncoding RNAs expression profile as potential serum biomarkers in patients with hepatocellular carcinoma. *Transl. Res* **168**, 134–145 (2016).
46. El-Tawdi, A. H. et al. Evaluation of Circulatory RNA-Based Biomarker Panel in Hepatocellular Carcinoma. *Mol. Diagn. Ther.* **20**, 265–277 (2016).
47. Jing, W. et al. Potential diagnostic value of lncRNA SPRY4-IT1 in hepatocellular carcinoma. *Oncol. Rep.* **36**, 1085–1092 (2016).
48. Konishi, H. et al. Plasma level of metastasis-associated lung adenocarcinoma transcript 1 is associated with liver damage and predicts development of hepatocellular carcinoma. *Cancer Sci.* **107**, 149–154 (2016).
49. Lu, J. et al. Investigation of serum lncRNA-uc003wbd and lncRNA-AF085935 expression profile in patients with hepatocellular carcinoma and HBV. *Tumour Biol.* **36**, 3231–3236 (2015).
50. Wang, K. et al. Serum lncRNAs Profiles Serve as Novel Potential Biomarkers for the Diagnosis of HBV-Positive Hepatocellular Carcinoma. *PLoS One* **10**, e0144934 (2015).
51. Guo, F. et al. Expression of MALAT1 in the peripheral whole blood of patients with lung cancer. *Biomed. Rep.* **3**, 309–312 (2015).
52. Shlyakhovenko, V. A. Ribonucleases in tumor growth. *Exp. Oncol.* **31**, 127–133 (2009).
53. Reddi, K. K. & Holland, J. F. Elevated serum ribonuclease in patients with pancreatic cancer. *Proc. Natl Acad. Sci. USA* **73**, 2308–2310 (1976).
54. Doran, G., Allen-Mersh, T. G. & Reynolds, K. W. Ribonuclease as a tumour marker for pancreatic carcinoma. *J. Clin. Pathol.* **33**, 1212–1213 (1980).
55. Naskalski, J. W. & Celinski, A. Determining of actual activities of acid and alkaline ribonuclease in human serum and urine. *Mater. Med Pol.* **23**, 107–110 (1991).
56. Youle, R. J., Newton, D., Wu, Y. N., Gadina, M. & Rybak, S. M. Cytotoxic ribonucleases and chimeras in cancer therapy. *Crit. Rev. Ther. Drug Carr. Syst.* **10**, 1–28 (1993).
57. Dricu, A., Sergiu-Bogdan, C., Brismar, K., Biberfeld, P. & Andersson, L. C. A synthetic peptide derived from the human eosinophil-derived neurotoxin induces apoptosis in Kaposi's sarcoma cells. *Anticancer Res* **24**, 1427–1432 (2004).
58. Xiang, Y. et al. Effects of RNase L mutations associated with prostate cancer on apoptosis induced by 2',5'-oligoadenylates. *Cancer Res* **63**, 6795–6801 (2003).
59. Silverman, R. H. Implications for RNase L in prostate cancer biology. *Biochemistry* **42**, 1805–1812 (2003).
60. Hashad, D., Elbanna, A., Ibrahim, A. & Khedr, G. Evaluation of the Role of Circulating Long Non-Coding RNA H19 as a Promising Novel Biomarker in Plasma of Patients with Gastric Cancer. *J. Clin. Lab Anal.* **30**, 1100–1105 (2016).
61. Hu, X. et al. The plasma lncRNA acting as fingerprint in non-small-cell lung cancer. *Tumour Biol.* **37**, 3497–3504 (2016).
62. 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.* **8**, 7887–7895 (2015).

63. Tong, Y. S. et al. Identification of the long non-coding RNA POU3F3 in plasma as a novel biomarker for diagnosis of esophageal squamous cell carcinoma. *Mol. Cancer* **14**, 3 (2015).
64. Yu, J. et al. The long noncoding RNAs PVT1 and uc002mbe.2 in sera provide a new supplementary method for hepatocellular carcinoma diagnosis. *Med. (Baltim.)* **95**, e4436 (2016).
65. Shi, J. et al. Circulating lncRNAs associated with occurrence of colorectal cancer progression. *Am. J. Cancer Res* **5**, 2258–2265 (2015).
66. Yan, S. et al. Evaluation of Serum Exosomal lncRNAs as Diagnostic and Prognostic Biomarkers for Esophageal Squamous Cell Carcinoma. *Cancer Manag Res* **12**, 9753–9763 (2020).
67. Zhang, K. et al. Genome-Wide lncRNA Microarray Profiling Identifies Novel Circulating lncRNAs for Detection of Gastric Cancer. *Theranostics* **7**, 213–227 (2017).
68. Wu, Y. et al. A serum-circulating long noncoding RNA signature can discriminate between patients with clear cell renal cell carcinoma and healthy controls. *Oncogenesis* **5**, e192 (2016).
69. Permut, J. B. et al. Linc-ing Circulating Long Non-coding RNAs to the Diagnosis and Malignant Prediction of Intraductal Papillary Mucinous Neoplasms of the Pancreas. *Sci. Rep.* **7**, 10484 (2017).
70. Chen, M. et al. A novel biosensor for the ultrasensitive detection of the lncRNA biomarker MALAT1 in non-small cell lung cancer. *Sci. Rep.* **11**, 3666 (2021).
71. Morlion, A. et al. Custom long non-coding RNA capture enhances detection sensitivity in different human sample types. *RNA Biol.* **18**, 215–222 (2021).
72. Li, J. et al. lncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. *Gut* **63**, 1700–1710 (2014).
73. Meng, J., Li, P., Zhang, Q., Yang, Z. & Fu, S. A four-long non-coding RNA signature in predicting breast cancer survival. *J. Exp. Clin. Cancer Res* **33**, 84 (2014).
74. Chen, H. et al. Long noncoding RNA profiles identify five distinct molecular subtypes of colorectal cancer with clinical relevance. *Mol. Oncol.* **8**, 1393–1403 (2014).
75. Qi, P. & Du, X. The long non-coding RNAs, a new cancer diagnostic and therapeutic gold mine. *Mod. Pathol.* **26**, 155–165 (2013).
76. Pan, X., Zheng, G. & Gao, C. lncRNA PVT1: a Novel Therapeutic Target for Cancers. *Clin. Lab* **64**, 655–662 (2018).
77. Pichler, M. et al. Therapeutic potential of FLANC, a novel primate-specific long non-coding RNA in colorectal cancer. *Gut* **69**, 1818–1831 (2020).
78. Huarte, M. & Rinn, J. L. Large non-coding RNAs: missing links in cancer? *Hum. Mol. Genet* **19**, R152–R161 (2010).
79. Xu, M. D., Qi, P. & Du, X. Long non-coding RNAs in colorectal cancer: implications for pathogenesis and clinical application. *Mod. Pathol.* **27**, 1310–1320 (2014).
80. Shen, X. H., Qi, P. & Du, X. Long non-coding RNAs in cancer invasion and metastasis. *Mod. Pathol.* **28**, 4–13 (2015).
81. Chang, Y. N. et al. Hypoxia-regulated lncRNAs in cancer. *Gene* **575**, 1–8 (2016).
82. Chi, Y., Wang, D., Wang, J., Yu, W. & Yang, J. Long Non-Coding RNA in the Pathogenesis of Cancers. *Cells* **8**, 1015 (2019).
83. Sanchez, Y. et al. Genome-wide analysis of the human p53 transcriptional network unveils a lncRNA tumour suppressor signature. *Nat. Commun.* **5**, 5812 (2014).
84. Song, M. S., Salmena, L. & Pandolfi, P. P. The functions and regulation of the PTEN tumour suppressor. *Nat. Rev. Mol. Cell Biol.* **13**, 283–296 (2012).
85. Yu, G. et al. Pseudogene PTENP1 functions as a competing endogenous RNA to suppress clear-cell renal cell carcinoma progression. *Mol. Cancer Ther.* **13**, 3086–3097 (2014).
86. Weidle, U. H., Birzele, F., Kollmorgen, G. & Ruger, R. Long Non-coding RNAs and their Role in Metastasis. *Cancer Genomics Proteom.* **14**, 143–160 (2017).
87. Ye, Z., Duan, J., Wang, L., Ji, Y. & Qiao, B. lncRNA-LET inhibits cell growth of clear cell renal cell carcinoma by regulating miR-373-3p. *Cancer Cell Int* **19**, 311 (2019).
88. Derderian, C., Orunmuyi, A. T., Olapade-Olaopa, E. O. & Ogunwobi, O. O. PVT1 Signaling Is a Mediator of Cancer Progression. *Front Oncol.* **9**, 502 (2019).
89. Han, L. et al. Prognostic and Clinicopathological Significance of Long Non-coding RNA PANDAR Expression in Cancer Patients: A Meta-Analysis. *Front Oncol.* **9**, 1337 (2019).
90. Zhang, X. et al. Long noncoding RNA GIHCG functions as an oncogene and serves as a serum diagnostic biomarker for cervical cancer. *J. Cancer* **10**, 672–681 (2019).
91. Wilusz, J. E., Sunwoo, H. & Spector, D. L. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* **23**, 1494–1504 (2009).
92. Parasramka, M. A., Maji, S., Matsuda, A., Yan, I. K. & Patel, T. Long non-coding RNAs as novel targets for therapy in hepatocellular carcinoma. *Pharm. Ther.* **161**, 67–78 (2016).
93. Wang, S. H. et al. The lncRNA MALAT1 functions as a competing endogenous RNA to regulate MCL-1 expression by sponging miR-363-3p in gallbladder cancer. *J. Cell Mol. Med* **20**, 2299–2308 (2016).
94. Wang, Q. et al. Long non-coding RNA HOTTIP promotes renal cell carcinoma progression through the regulation of the miR-615/IGF-2 pathway. *Int J. Oncol.* **53**, 2278–2288 (2018).
95. Zhang, H. et al. Long non-coding RNA: a new player in cancer. *J. Hematol. Oncol.* **6**, 37 (2013).
96. Qiu, M. T., Hu, J. W., Yin, R. & Xu, L. Long noncoding RNA: an emerging paradigm of cancer research. *Tumour Biol.* **34**, 613–620 (2013).
97. Zhao, Y. et al. Role of long non-coding RNA HULC in cell proliferation, apoptosis and tumor metastasis of gastric cancer: a clinical and in vitro investigation. *Oncol. Rep.* **31**, 358–364 (2014).
98. Huang, X. et al. Characterization of human plasma-derived exosomal RNAs by deep sequencing. *BMC Genomics* **14**, 319 (2013).
99. Zhang, H. G. & Grizzle, W. E. Exosomes: a novel pathway of local and distant intercellular communication that facilitates the growth and metastasis of neoplastic lesions. *Am. J. Pathol.* **184**, 28–41 (2014).
100. Arroyo, J. D. et al. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc. Natl Acad. Sci. USA* **108**, 5003–5008 (2011).
101. Wang, K., Zhang, S., Weber, J., Baxter, D. & Galas, D. J. Export of microRNAs and microRNA-protective protein by mammalian cells. *Nucleic Acids Res* **38**, 7248–7259 (2010).
102. Arun, G. et al. Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. *Genes Dev.* **30**, 34–51 (2016).
103. Ching, T. et al. Pan-Cancer Analyses Reveal Long Intergenic Non-Coding RNAs Relevant to Tumor Diagnosis, Subtyping and Prognosis. *EBioMedicine* **7**, 62–72 (2016).
104. Gao, J., Cao, R. & Mu, H. Long non-coding RNA UCA1 may be a novel diagnostic and predictive biomarker in plasma for early gastric cancer. *Int J. Clin. Exp. Pathol.* **8**, 12936–12942 (2015).
105. Li, J., Wang, Y., Yu, J., Dong, R. & Qiu, H. A high level of circulating HOTAIR is associated with progression and poor prognosis of cervical cancer. *Tumour Biol.* **36**, 1661–1665 (2015).
106. Hoadley, K. A. et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291–304 e296 (2018).
107. Liu, J. The dualistic origin of human tumors. *Semin Cancer Biol.* **53**, 1–16 (2018).
108. Malta, T. M. et al. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* **173**, 338–354 e315 (2018).
109. Ben-Porath, I. et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat. Genet* **40**, 499–507 (2008).
110. Kim, J. et al. A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell* **143**, 313–324 (2010).
111. Kim, J. & Orkin, S. H. Embryonic stem cell-specific signatures in cancer: insights into genomic regulatory networks and implications for medicine. *Genome Med* **3**, 75 (2011).
112. Ariel, I. et al. The product of the imprinted H19 gene is an oncofetal RNA. *Mol. Pathol.* **50**, 34–44 (1997).
113. Pachnis, V., Belayew, A. & Tilghman, S. M. Locus unlinked to alpha-fetoprotein under the control of the murine raf and Rif genes. *Proc. Natl Acad. Sci. USA* **81**, 5523–5527 (1984).
114. Ariel, I. et al. The imprinted H19 gene as a tumor marker in bladder carcinoma. *Urology* **45**, 335–338 (1995).
115. Cofre, J. & Abdelhay, E. Cancer Is to Embryology as Mutation Is to Genetics: Hypothesis of the Cancer as Embryological Phenomenon. *ScientificWorldJournal* **2017**, 3578090 (2017).
116. Wood, A. J. & Oakey, R. J. Genomic imprinting in mammals: emerging themes and established theories. *PLoS Genet* **2**, e147 (2006).
117. Goshen, R. et al. The expression of the H-19 and IGF-2 genes during human embryogenesis and placental development. *Mol. Reprod. Dev.* **34**, 374–379 (1993).
118. Lustig, O. et al. Expression of the imprinted gene H19 in the human fetus. *Mol. Reprod. Dev.* **38**, 239–246 (1994).
119. Gabory, A., Jammes, H. & Dandolo, L. The H19 locus: role of an imprinted non-coding RNA in growth and development. *Bioessays* **32**, 473–480 (2010).
120. Steenman, M. J. et al. Loss of imprinting of IGF2 is linked to reduced expression and abnormal methylation of H19 in Wilms' tumour. *Nat. Genet* **7**, 433–439 (1994).
121. Hibi, K. et al. Loss of H19 imprinting in esophageal cancer. *Cancer Res* **56**, 480–482 (1996).
122. Kim, H. T., Choi, B. H., Niikawa, N., Lee, T. S. & Chang, S. I. Frequent loss of imprinting of the H19 and IGF-II genes in ovarian tumors. *Am. J. Med Genet* **80**, 391–395 (1998).
123. Chen, C. L., Ip, S. M., Cheng, D., Wong, L. C. & Ngan, H. Y. Loss of imprinting of the IGF-II and H19 genes in epithelial ovarian cancer. *Clin. Cancer Res* **6**, 474–479 (2000).



124. Lottin, S. et al. Overexpression of an ectopic H19 gene enhances the tumorigenic properties of breast cancer cells. *Carcinogenesis* **23**, 1885–1895 (2002).
125. Cui, H. et al. Loss of imprinting in colorectal cancer linked to hypomethylation of H19 and IGF2. *Cancer Res* **62**, 6442–6446 (2002).
126. Si, H., Chen, P., Li, H. & Wang, X. Long non-coding RNA H19 regulates cell growth and metastasis via miR-138 in breast cancer. *Am. J. Transl. Res* **11**, 3213–3225 (2019).
127. Cao, T. et al. H19 lncRNA identified as a master regulator of genes that drive uterine leiomyomas. *Oncogene* **38**, 5356–5366 (2019).
128. Biran, H., Ariel, I., de Groot, N., Shani, A. & Hochberg, A. Human imprinted genes as oncogene developmental markers. *Tumour Biol.* **15**, 123–134 (1994).
129. Ariel, I., de Groot, N. & Hochberg, A. Imprinted H19 gene expression in embryogenesis and human cancer: the oncofetal connection. *Am. J. Med Genet* **91**, 46–50 (2000).
130. Nielsen, J. et al. A family of insulin-like growth factor II mRNA-binding proteins represses translation in late development. *Mol. Cell Biol.* **19**, 1262–1270 (1999).
131. Bell, J. L. et al. Insulin-like growth factor 2 mRNA-binding proteins (IGF2BPs): post-transcriptional drivers of cancer progression? *Cell Mol. Life Sci.* **70**, 2657–2675 (2013).
132. Ching, T., Masaki, J., Weirather, J. & Garmire, L. X. Non-coding yet non-trivial: a review on the computational genomics of lincRNAs. *BioData Min.* **8**, 44 (2015).
133. Bolha, L., Ravnik-Glavac, M. & Glavac, D. Long Noncoding RNAs as Biomarkers in Cancer. *Dis. Markers* **2017**, 7243968 (2017).
134. Bhartiya, D. et al. lncRNome: a comprehensive knowledgebase of human long noncoding RNAs. *Database (Oxf.)* **2013**, bat034 (2013).
135. 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* **42**, D92–D97 (2014).
136. Wu, T. et al. NPInter: the noncoding RNAs and protein-related biomacromolecules interaction database. *Nucleic Acids Res* **34**, D150–D152 (2006).
137. Yuan, J. et al. NPInter v2.0: an updated database of ncRNA interactions. *Nucleic Acids Res* **42**, D104–D108 (2014).
138. Hao, Y. et al. NPInter v3.0: an upgraded database of noncoding RNA-associated interactions. *Database (Oxford)* **2016**, baw057 (2016).
139. Teng, X. et al. NPInter v4.0: an integrated database of ncRNA interactions. *Nucleic Acids Res* **48**, D160–D165 (2020).
140. Oughtred, R. et al. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* **30**, 187–200 (2021).
141. Zhao, W. et al. POSTAR3: an updated platform for exploring post-transcriptional regulation coordinated by RNA-binding proteins. *Nucleic Acids Res* **50**, D287–D294 (2022).
142. Wang, C. et al. Interactome analysis reveals that lncRNA HULC promotes aerobic glycolysis through LDHA and PKM2. *Nat. Commun.* **11**, 3162 (2020).
143. Sauliere, J. et al. CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. *Nat. Struct. Mol. Biol.* **19**, 1124–1131 (2012).
144. Lagier-Tourenne, C. et al. Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nat. Neurosci.* **15**, 1488–1497 (2012).
145. Ke, H. et al. NEAT1 is Required for Survival of Breast Cancer Cells Through FUS and miR-548. *Gene Regul. Syst. Bio* **10**, 11–17 (2016).
146. Colombo, M., Karousis, E. D., Bourquin, J., Bruggmann, R. & Muhlemann, O. Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and SMG7-mediated degradation pathways. *RNA* **23**, 189–201 (2017).
147. Sanford, J. R. et al. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res* **19**, 381–394 (2009).
148. Anko, M. L. & Neugebauer, K. M. Long noncoding RNAs add another layer to pre-mRNA splicing regulation. *Mol. Cell* **39**, 833–834 (2010).
149. Zong, X., Tripathi, V. & Prasanth, K. V. RNA splicing control: yet another gene regulatory role for long nuclear noncoding RNAs. *RNA Biol.* **8**, 968–977 (2011).
150. Hoell, J. I. et al. RNA targets of wild-type and mutant FET family proteins. *Nat. Struct. Mol. Biol.* **18**, 1428–1431 (2011).
151. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
152. Gregersen, L. H. et al. MOV10 is a 5' to 3' RNA helicase contributing to UPF1 mRNA target degradation by translocation along 3'UTRs. *Mol. Cell* **54**, 573–585 (2014).
153. Yokoshi, M. et al. Direct binding of Ataxin-2 to distinct elements in 3' UTRs promotes mRNA stability and protein expression. *Mol. Cell* **55**, 186–198 (2014).
154. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
155. Conway, A. E. et al. Enhanced CLIP Uncovers IMP Protein-RNA Targets in Human Pluripotent Stem Cells Important for Cell Adhesion and Survival. *Cell Rep.* **15**, 666–679 (2016).
156. West, J. A. et al. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell* **55**, 791–802 (2014).
157. Hammerle, M. et al. Posttranscriptional destabilization of the liver-specific long noncoding RNA HULC by the IGF2 mRNA-binding protein 1 (IGF2BP1). *Hepatology* **58**, 1703–1712 (2013).
158. Weinmann, L. et al. Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs. *Cell* **136**, 496–507 (2009).
159. Liu, H. T., Liu, S., Liu, L., Ma, R. R. & Gao, P. EGR1-Mediated Transcription of lncRNA-HNF1A-AS1 Promotes Cell-Cycle Progression in Gastric Cancer. *Cancer Res* **78**, 5877–5890 (2018).
160. Yoon, J. H. et al. Scaffold function of long non-coding RNA HOTAIR in protein ubiquitination. *Nat. Commun.* **4**, 2939 (2013).
161. Wang, L. et al. BRCA1 is a negative modulator of the PRC2 complex. *EMBO J.* **32**, 1584–1597 (2013).
162. Guil, S. et al. Intronic RNAs mediate EZH2 regulation of epigenetic targets. *Nat. Struct. Mol. Biol.* **19**, 664–670 (2012).
163. Kaneko, S. et al. Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. *Genes Dev.* **24**, 2615–2620 (2010).
164. Wu, L., Murat, P., Matak-Vinkovic, D., Murrell, A. & Balasubramanian, S. Binding interactions between long noncoding RNA HOTAIR and PRC2 proteins. *Biochemistry* **52**, 9519–9527 (2013).
165. Vikesaa, J. et al. RNA-binding IMPs promote cell adhesion and invadopodia formation. *EMBO J.* **25**, 1456–1468 (2006).
166. Tessier, C. R., Doyle, G. A., Clark, B. A., Pitot, H. C. & Ross, J. Mammary tumor induction in transgenic mice expressing an RNA-binding protein. *Cancer Res* **64**, 209–214 (2004).
167. Mohamadkhani, A. Long Noncoding RNAs in Interaction With RNA Binding Proteins in Hepatocellular Carcinoma. *Hepat. Mon.* **14**, e18794 (2014).
168. Schaeffer, D. F. et al. Insulin-like growth factor 2 mRNA binding protein 3 (IGF2BP3) overexpression in pancreatic ductal adenocarcinoma correlates with poor survival. *BMC Cancer* **10**, 59 (2010).
169. Lederer, M., Bley, N., Schleifer, C. & Huttmelmaier, S. The role of the oncofetal IGF2 mRNA-binding protein 3 (IGF2BP3) in cancer. *Semin Cancer Biol.* **29**, 3–12 (2014).
170. Rabbitts, T. H., Forster, A., Larson, R. & Nathan, P. Fusion of the dominant negative transcription regulator CHOP with a novel gene FUS by translocation t(12;16) in malignant liposarcoma. *Nat. Genet.* **4**, 175–180 (1993).
171. Crozat, A., Aman, P., Mandahl, N. & Ron, D. Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. *Nature* **363**, 640–644 (1993).
172. Han, D. et al. Long noncoding RNA H19 indicates a poor prognosis of colorectal cancer and promotes tumor growth by recruiting and binding to eIF4A3. *Oncotarget* **7**, 22159–22173 (2016).
173. Yao, N., Yu, L., Zhu, B., Gan, H. Y. & Guo, B. Q. LncRNA GIHCG promotes development of ovarian cancer by regulating microRNA-429. *Eur. Rev. Med. Pharm. Sci.* **22**, 8127–8134 (2018).
174. Jiang, X., Li, Q., Zhang, S., Song, C. & Zheng, P. Long noncoding RNA GIHCG induces cancer progression and chemoresistance and indicates poor prognosis in colorectal cancer. *Onco Targets Ther.* **12**, 1059–1070 (2019).
175. Liu, G., Jiang, Z., Qiao, M. & Wang, F. Lnc-GIHCG promotes cell proliferation and migration in gastric cancer through miR-1281 adsorption. *Mol. Genet. Genom. Med.* **7**, e711 (2019).
176. Yang, J. P., Yang, X. J., Xiao, L. & Wang, Y. Long noncoding RNA PVT1 as a novel serum biomarker for detection of cervical cancer. *Eur. Rev. Med. Pharm. Sci.* **20**, 3980–3986 (2016).
177. Chen, X. et al. Long Noncoding RNA PVT1 as a Novel Diagnostic Biomarker and Therapeutic Target for Melanoma. *Biomed. Res Int* **2017**, 7038579 (2017).
178. Tao, K. et al. Clinical significance of urothelial carcinoma associated 1 in colon cancer. *Int J. Clin. Exp. Med* **8**, 21854–21860 (2015).
179. Wen, J. J., Ma, Y. D., Yang, G. S. & Wang, G. M. Analysis of circulating long non-coding RNA UCA1 as potential biomarkers for diagnosis and prognosis of osteosarcoma. *Eur. Rev. Med. Pharm. Sci.* **21**, 498–503 (2017).
180. Qi, P., Zhou, X. Y. & Du, X. Circulating long non-coding RNAs in cancer: current status and future perspectives. *Mol. Cancer* **15**, 39 (2016).
181. Yan, H., Chai, H. & Zhao, H. Detecting lncRNA-Cancer Associations by Combining miRNAs, Genes, and Prognosis With Matrix Factorization. *Front Genet* **12**, 639872 (2021).
182. Sanchez-Salcedo, R., Miranda-Castro, R., de-Los-Santos-Alvarez, N. & Lobo-Castan, M. J. Dual electrochemical genosensor for early diagnosis of prostate cancer through lncRNAs detection. *Biosens. Bioelectron.* **192**, 113520 (2021).
183. Zhou, Z., Shen, Y., Khan, M. R. & Li, A. LncReg: a reference resource for lncRNA-associated regulatory networks. *Database (Oxford)* **2015**, bav083 (2015).
184. Bao, Z. et al. LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res* **47**, D1034–D1037 (2019).
185. Chou, C. H. et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res* **46**, D296–D302 (2018).
186. Cui, T. et al. MNDIR v2.0: an updated resource of ncRNA-disease associations in mammals. *Nucleic Acids Res* **46**, D371–D374 (2018).

187. Pinero, J. et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* **45**, D833–D839 (2017).
188. Lee, B. T. et al. The UCSC Genome Browser database: 2022 update. *Nucleic Acids Res* **50**, D1115–D1122 (2022).
189. Kent, W. J. et al. The human genome browser at UCSC. *Genome Res* **12**, 996–1006 (2002).
190. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res* **12**, 656–664 (2002).
191. Karolchik, D. et al. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* **32**, D493–D496 (2004).
192. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).
193. Raney, B. J. et al. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* **30**, 1003–1005 (2014).
194. Kent, W. J. et al. Exploring relationships and mining data with the UCSC Gene Sorter. *Genome Res* **15**, 737–741 (2005).
195. Rosenbloom, K. R. et al. ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res* **41**, D56–D63 (2013).
196. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
197. Kent, W. J. & Haussler, D. Assembly of the working draft of the human genome with GigAssembler. *Genome Res* **11**, 1541–1548 (2001).
198. Kroh, E. M., Parkin, R. K., Mitchell, P. S. & Tewari, M. Analysis of circulating microRNA biomarkers in plasma and serum using quantitative reverse transcription-PCR (qRT-PCR). *Methods* **50**, 298–301 (2010).
199. Chen, G., Wang, J. & Cui, Q. Could circulating miRNAs contribute to cancer therapy? *Trends Mol. Med* **19**, 71–73 (2013).
200. Dobosy, J. R. et al. A methyl-deficient diet modifies histone methylation and alters Igf2 and H19 repression in. *Prostate Prostate* **68**, 1187–1195 (2008).
201. Solanas, M. et al. Differential expression of H19 and vitamin D3 upregulated protein 1 as a mechanism of the modulatory effects of high virgin olive oil and high corn oil diets on experimental mammary tumours. *Eur. J. Cancer Prev.* **18**, 153–161 (2009).
202. Pritchard, C. C. et al. Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. *Cancer Prev. Res (Philos.)* **5**, 492–497 (2012).
203. Gresner, P., Gromadzinska, J. & Wasowicz, W. Reference genes for gene expression studies on non-small cell lung cancer. *Acta Biochim Pol.* **56**, 307–316 (2009).
204. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet* **17**, 333–351 (2016).
205. Logsdon, G. A., Vollger, M. R. & Eichler, E. E. Long-read human genome sequencing and its applications. *Nat. Rev. Genet* **21**, 597–614 (2020).
206. Grillone, K. et al. Non-coding RNAs in cancer: platforms and strategies for investigating the genomic “dark matter”. *J. Exp. Clin. Cancer Res* **39**, 117 (2020).
207. Mercer, T. R. et al. Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. *Nat. Protoc.* **9**, 989–1009 (2014).
208. Kreil, D. P., Russell, R. R. & Russell, S. Microarray oligonucleotide probes. *Methods Enzymol.* **410**, 73–98 (2006).
209. de Sena Brandine, G. & Smith, A. D. Falco: high-speed FastQC emulation for quality control of sequencing data. *F1000Res* **8**, 1874 (2019).
210. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).
211. Chen, S. et al. AfterQC: automatic filtering, trimming, error removing and quality control for fastq data. *BMC Bioinforma.* **18**, 80 (2017).
212. Volders, P. J. et al. LNCipedia 5: towards a reference set of human long non-coding RNAs. *Nucleic Acids Res* **47**, D135–D139 (2019).
213. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
214. Langmead, B., Wilks, C., Antonescu, V. & Charles, R. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35**, 421–432 (2019).
215. Kovaka, S. et al. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**, 278 (2019).
216. Zhao, S., Ye, Z. & Stanton, R. Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA* **26**, 903–909 (2020).
217. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
218. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
219. Choi, S. W., Mak, T. S. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* **15**, 2759–2772 (2020).
220. Gene Ontology, C. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* **49**, D325–D334 (2021).
221. Kanehisa, M., Sato, Y. & Kawashima, M. KEGG mapping tools for uncovering hidden features in biological data. *Protein Sci* **31**, 47–53 (2021).
222. Huang, S., Yee, C., Ching, T., Yu, H. & Garmire, L. X. A novel model to combine clinical and pathway-based transcriptomic information for the prognosis prediction of breast cancer. *PLoS Comput Biol.* **10**, e1003851 (2014).
223. Huang, S. et al. Novel personalized pathway-based metabolomics models reveal key metabolic pathways for breast cancer diagnosis. *Genome Med* **8**, 34 (2016).

## ACKNOWLEDGEMENTS

This research was supported by grants K01ES025434 awarded by NIEHS through funds provided by the trans-NIH Big Data to Knowledge (BD2K) initiative ([www.bd2k.nih.gov](http://www.bd2k.nih.gov)), P20 COBRE GM103457 awarded by NIH/NIGMS, R01 LM012373 and R01 5R01LM012907 awarded by NLM, and R01 HD084633 awarded by NICHD to L.X. Garmire.

## AUTHOR CONTRIBUTIONS

L.X.G. envisioned the project. C.B. wrote the manuscript with the help of B.H. and L.X.G. All authors have read, revised and approved the final manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41698-022-00283-7>.

**Correspondence** and requests for materials should be addressed to Cedric Badowski or Lana X. Garmire.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022