The Prognostic Role of Long Non-Coding RNA in Hepatocellular Carcinoma

A report submitted as the examined component of
the project module SXL390

Daniel Beach

August 2021

(4984 words)

Abstract

Background: Hepatocellular carcinoma (HCC) is an aggressive cancer with poor 5-year survival rates. New prognostic biomarkers are needed to guide clinical treatment choices and improve patient outcomes. Long non-coding RNAs (lncRNAs) are RNA transcripts longer than 200 nucleotides with no protein-coding potential. They are dysregulated in cancer and detectable in tissues making them promising biomarker candidates. Using bioinformatic analysis, this study aimed to assess the prognostic value of lncRNAs in HCC and construct a lncRNA signature for prognostic evaluation of HCC patients.

Methods: Candidate HCC-related lncRNAs were downloaded from the LncRNADisease database. The Kaplan-Meier Plotter tool was used compare survival probabilities for patients with high and low expression of each candidate lncRNA, using data from a cohort of 364 HCC patients. The log-rank test was used to compare survival curves for high- and low-expression groups, producing p-values and hazard ratios with 95% confidence intervals. Individual lncRNAs with prognostic value were systematically tested in all possible combinations to produce an optimum prognostic lncRNA signature.

Results: The LncRNADisease database search yielded 5307 candidate lncRNAs. Survival analysis identified six lncRNAs associated with OS at $p<0.05$ (TP53TG1, TTTY15, MIAT, HCG18, XIST, NEAT1), and these were selected for construction of the lncRNA signature. Ten lncRNA combinations were significantly associated with OS after correction for multiple comparisons. The most prognostic signature was TP53TG1, TTTY15, MIAT ($p=0.00086$, HR=0.55, 95% CI=0.39-0.79). Subgroup analysis revealed a significant association with OS in patients with intermediate-stage HCC.

Conclusion: Bioinformatic analysis of online datasets has identified a novel prognostic signature which is associated with OS in a cohort of HCC patients. Our results support previous findings that lncRNAs TP53TG1, HCG18 and XIST are associated with survival in HCC. Additionally, we report novel evidence that the male-specific lncRNA TTTY15 is associated with OS in HCC.

(293 words)

List of Abbreviations

CDX	Cell-line derived xenograft
CI	Confidence interval
circRNA	Circular RNA
CP	Coding probability
CPAT	Coding Potential Assessment Tool
DEA	Differential expression analysis
HCC	Hepatocellular carcinoma
HR	Hazard ratio
KM-Plotter	Kaplan-Meier Plotter
lncRNA	long non-coding RNA
MIAT	Myocardial infarction associated transcript
miRNA	Micro RNA
ncRNA	Non-coding RNA
nt	Nucleotides
OS	Overall survival
RNA-seq	RNA sequence
TCGA	The Cancer Genome Atlas
TCGA-LIHC	The Cancer Genome Atlas - Liver Hepatocellular Carcinoma
TP53TG1	TP53 target gene 1
TTTY15	Testis specific transcript Y-linked 15

Table of Contents

List of Tables

List of Figures

1 Introduction

Hepatocellular carcinoma (HCC) is a globally important disease, accounting for a sixth of all cancer cases and a third of cancer deaths worldwide (Llovet et al., 2021). It is an aggressive cancer associated with high rates of metastasis and recurrence, and despite recent advances in diagnosis and treatment the prognosis is poor (Llovet et al., 2021).

HCC progresses from chronic liver disease and is associated with viral hepatitis infection, alcoholism, aflatoxin B exposure, diabetes, and obesity (Hu et al., 2018). These diverse aetiologies drive the development and progression of HCC via different mechanisms, resulting in a highly heterogeneous disease that is challenging to treat (Menyhárt et al., 2018).

Early HCC can be treated curatively with local ablation, surgical resection, or transplantation (Yang et al., 2021). However, patients with more advanced disease cannot currently be cured. Multiple-kinase inhibitors are the standard systemic chemotherapy for advanced HCC but provide only modest increases in survival (Yang et al., 2019).

New genomic sequencing techniques have revealed six distinct HCC subtypes, denoted G1-G6 (Table 1.1) (Yang et al., 2019). Subtypes vary considerably in terms of genetic mutations, clinical features, and prognostic outcomes (Yang et al., 2019). However, there is currently no stratification system to guide treatment on the basis of molecular subtype, and this significantly limits the effectiveness of current therapies (Yang et al., 2019). A biomarker-based prognostic approach is needed to predict patient risk more accurately and inform more personalised treatment choices (Yan et al., 2019). Alpha-fetoprotein is the only biomarker established for clinical use in HCC to date, and it has limited prognostic value (Yuan et al., 2021). Thus, new prognostic biomarkers for HCC are urgently needed.

Table 1.1 A summary of aetiologies and prognostic features associated with the main HCC classes and subtypes. Adapted from Llovet et al. (2021).

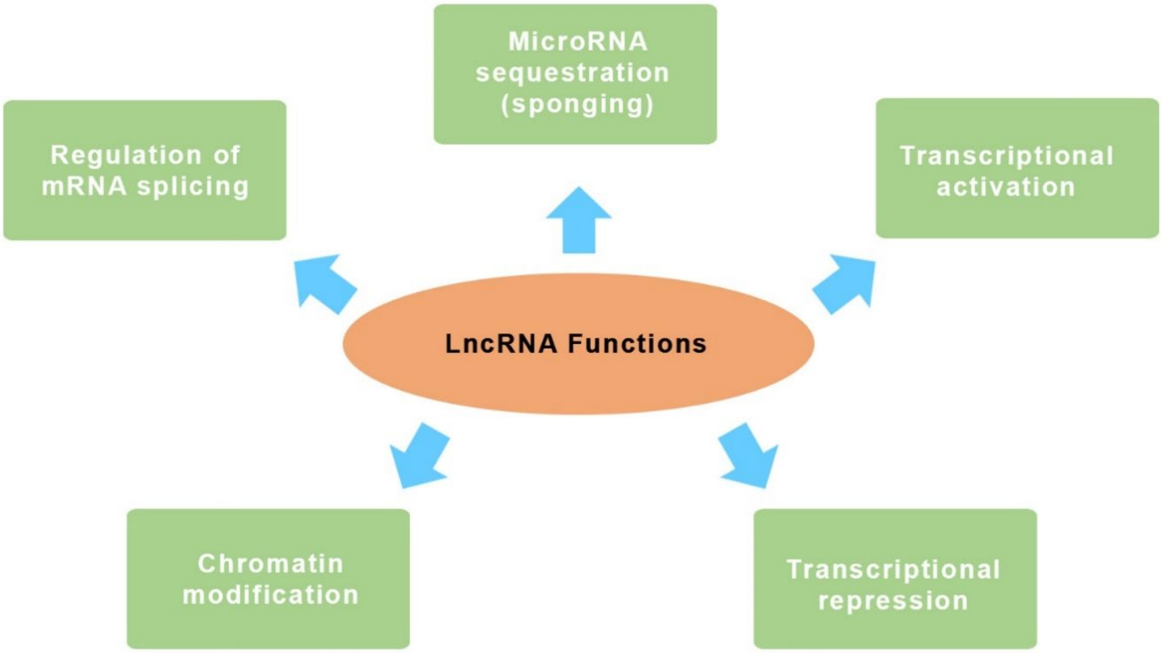
Class	Subtype	Main Aetiologies	Prognostic Features
Proliferative class	G1	Hepatitis B Virus	More aggressive tumours, high frequency of vascular invasion
	G2		
	G3		
Non-proliferative class	G4	Alcohol, Hepatitis C Virus, Non-alcoholic Fatty Liver Disease	Less aggressive tumours, low frequency of vascular invasion
	G5		
	G6		

The "central dogma" of molecular biology states that DNA encodes RNA which encodes protein: however, only 2% of the genome is protein-coding (Gulìa et al., 2017). The remaining 98% was previously dismissed as "junk DNA" but is now known to include several functional non-coding RNA (ncRNA) classes including circular RNAs (circRNAs), microRNAs (miRNAs) and long non-coding RNAs (lncRNAs) (Table 1.2) (Gulìa et al., 2017).

Table 1.2. Key non-coding RNA classes. Nt = nucleotides. Adapted from Gulìa et al. (2017).

Name	Acronym	Length (nt)	Description
Long non-coding RNAs	lncRNA	>200	Non-protein coding transcripts; heterogeneous class of RNAs
Circular RNA	circRNA	≈100–1600	Covalently closed RNA rings: some have coding functions; potential gene regulators and Micro-RNA "traps"
Small interfering RNA	siRNA	20–25	Double-stranded RNAs similar to Micro-RNA, operating through RNA interference
Micro-RNA	miRNA; miR	21–24	Function in RNA silencing and post-transcriptional regulation of gene expression
Small nucleolar RNAs	snoRNAs	60–300	Guide chemical modifications of other RNA

LncRNAs are RNA transcripts > 200 nucleotides in length with no protein coding ability (Hu et al., 2018). LncRNAs fold into complex 3-dimensional structures and perform wide-ranging functions including sequestering molecules, activating gene transcription, and modifying chromatin (Figure 1) (Kadali et al., 2018). LncRNAs have been shown to play a critical role in genetic and epigenetic regulation, directing cellular processes such as proliferation and the cell cycle (Yu et al., 2020). Furthermore, abnormal expression has been linked to the development and progression of cancers including HCC, in which they regulate metastasis, angiogenesis, and dedifferentiation (Hu et al., 2018). Emerging evidence suggests that lncRNAs are major drivers of carcinogenesis, and a deeper understanding of this ncRNA class could greatly improve the diagnosis, prognostic evaluation, and treatment of cancer (Yuan et al., 2021).


MicroRNA sequestration (sponging)
Regulation of mRNA splicing
Transcriptional activation
LncRNA Functions
Chromatin modification
Transcriptional repression

LncRNAs are cancer and subtype-specific, and readily detectable in blood, urine and tissue samples making them promising biomarker candidates (Hu et al., 2018). Additionally, there is increasing evidence that lncRNA expression patterns may predict patient survival in HCC, demonstrating their potential prognostic value (Gu et al., 2018; Nie et al., 2020; Yan et al., 2019).

However, previous studies aiming to identify prognostic lncRNAs in HCC have a number of limitations. Single-centre clinical studies have been limited by small sample size and lack of ethnic and aetiological diversity among participants, reducing the reliability and generalisability of the results (Yu et al., 2020). In contrast, other recent studies have used bioinformatic analysis of large datasets to construct novel prognostic multi-lncRNA signatures for HCC, but results have varied considerably and there is little overlap between studies (Yang et al., 2020; Wang and Lei, 2021; Cao et al., 2021). Additionally, in some cases the statistical methods used may have overstated the significance of results (Yang et al., 2020; Cao et al., 2021). Naturally, this raises questions about the reliability of these findings, and whether they can be translated into clinically usable biomarkers.

While previous work provides a foundation of supporting evidence, the prognostic role of lncRNA in HCC is far from clear and further work is needed. We hypothesise that the expression of specific lncRNAs is significantly associated with patient survival in HCC. The current study aims to build on existing findings to establish a robust prognostic lncRNA signature for HCC, through bioinformatic dataset analysis.

2 Methods

2.1 Identifying HCC-related lncRNAs

Differential expression analysis (DEA) is commonly used to identify genes of interest in expression studies (Sonesona and Delorenzi, 2013). However, due to a lack of accessible bioinformatic tools, DEA of lncRNAs is beyond the scope of this study. Consequently, an alternative candidate gene approach was adopted.

LncRNADisease (v2.0) is a manually curated database of lncRNA-disease and circRNA-disease associations (Bao et al., 2019). The dataset includes more than 200,000 experimentally supported and algorithmically predicted lncRNA-disease associations. Data are collated from primary research papers identified via PubMed. Each entry includes a lncRNA symbol and associated disease, and details of the supporting evidence. LncRNADisease was selected for its size, scope, and robust inclusion criteria (Bao et al., 2019).

LncBook was identified as a second data source but was offline during the data collection phase and could not be used (Ma et al., 2019).

To identify HCC-related lncRNAs, LncRNADisease was searched using the terms: "disease = hepatocellular carcinoma". Results were downloaded and filtered to remove circRNAs and non-human results.

2.2 Shortlisting candidate lncRNAs

The Kaplan-Meier Plotter (KM-Plotter) is an online survival analysis tool for evaluating the effects of gene expression on survival in cancer patients (Menyhárt et al., 2018). 70,000 genes/ncRNAs are available for prognostic evaluation in 21 cancers, including HCC.

Search results from LncRNADisease were cross-referenced with the KM-Plotter to obtain a shortlist of lncRNAs available for survival analysis. Due to time constraints, only lncRNA symbols in the search results were cross-referenced. Alias symbols were not checked.

2.3 Verifying lncRNA classification

To ensure that only genuine lncRNAs were included in the analysis, the classification of shortlisted lncRNAs was checked using the Ensembl and GeneCards databases. Symbols with a classification other than "lncRNA" in both databases were excluded from further analysis.

2.4 Testing for protein coding potential

By definition, lncRNAs have no protein coding ability (Gulìa et al., 2017). However, 'bifunctional' lncRNAs with both coding and non-coding functions have recently been described (Huang et al., 2020). As our research question relates only to non-coding RNAs, including bifunctional RNAs with coding ability would be confounding. Accordingly, lncRNAs with documented or predicted protein coding ability were excluded.

The Coding Potential Assessment Tool (CPAT) is an online utility that can distinguish between coding and non-coding RNAs (Wang et al., 2013). Transcripts are analysed for hallmarks of coding potential and a coding probability (CP) score between 0 and 1 is calculated. CP scores ≥ 0.364 are suggestive of a protein coding sequence with 96.6% sensitivity and specificity (Wang et al., 2013).

Candidate lncRNA transcript sequences were downloaded from NCBI Gene and analysed with CPAT (Wang et al., 2013). LncRNAs with CP $\geq$0.364 were considered to be protein coding and excluded from further analysis.

2.5 Patient cohort

The KM-Plotter dataset includes 364 HCC patients from The Cancer Genome Atlas (TCGA) liver cancer project (TCGA-LIHC) (Menyhárt et al., 2018). The dataset comprises whole-transcriptome RNA sequence (RNA-seq) data from HCC tumour samples and associated clinical data (Menyhárt et al., 2018). Patients were enrolled at 24 sites across North America (Menyhárt et al., 2018). Table 2 summarises characteristics of the patient cohort.

Table 2. Characterisation of the patient sample used in this study. Adapted from Menyhart et al. (2018).

Cohort	Sample size
Total *n*	371
Gender	
Males	246
Females	118
Stage	
Stage I	170
Stage II	83
Stage III	83
Stage IV	4
Grade	
Grade 1	55
Grade 2	174
Grade 3	118
Grade 4	12
Race	
White/Caucasian	184
Black or African American	17
Asian	158
Vascular invasion	
Micro vascular invasion	90
Macro vascular invasion	16
None	203
Hepatitis Risk	
Yes	150
No	167
Alcohol risk	
Yes	115
No	202

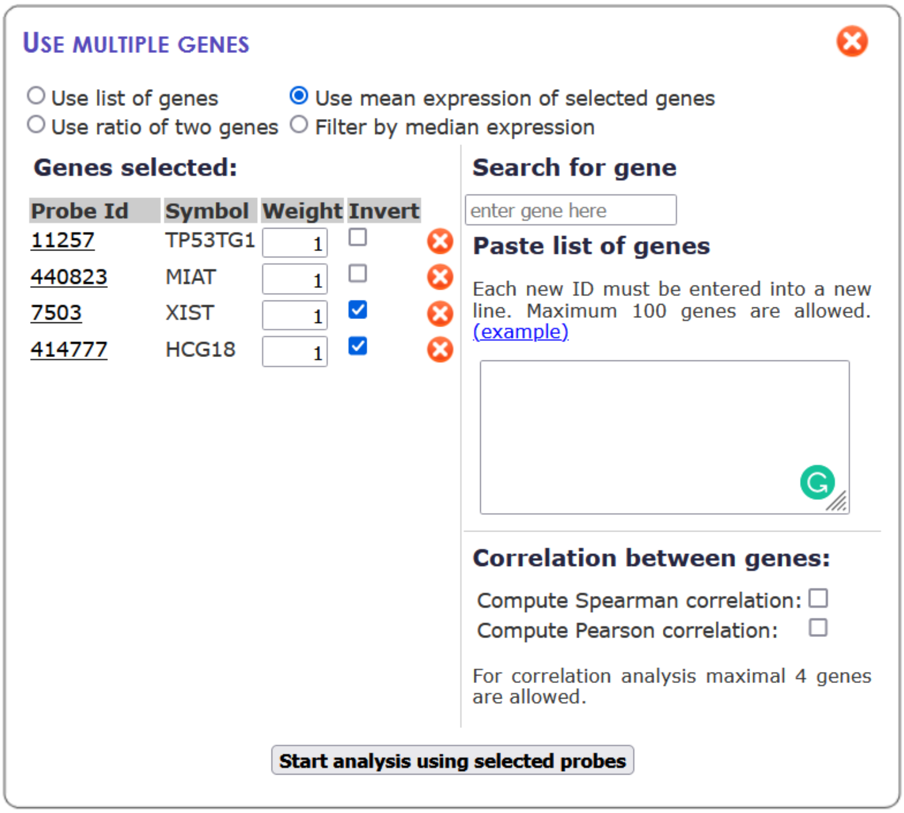
2.6 Survival analysis: single lncRNAs

The KM-Plotter was used to perform survival analyses for each shortlisted lncRNA. All analyses used the whole cohort (n=364). Patients were dichotomised into high- and low-expression groups on the basis of lncRNA expression data, using the median expression value as an objective cut-off point (Menyhárt et al., 2018). Overall survival (OS) was the selected endpoint in all analyses. The KM-Plotter was used to plot and compare survival curves for high- and low-expression groups, generating hazard ratios (HR) with 95% confidence intervals (CI) and log-rank p-values. The magnitude and direction of survival differences between groups was indicated by the HR. Log-rank p-values showed whether survival differences were statistically significant. A threshold of $p<0.05$ was used to select individual lncRNAs for inclusion in prognostic signature testing.

2.7 Survival analysis: lncRNA signatures

lncRNAs that were prognostic at $p<0.05$ were combined in all possible permutations to generate a list of potential prognostic "signatures". The KM-Plotter was used to generate survival plots for each signature using the whole cohort (n=364). Patients were split into high- and low-expression groups based on the mean expression of all lncRNAs in the signature (Figure 2.1). The median expression value was used as the cut-off point for dichotomisation and the endpoint was OS. LncRNAs that reduced OS in the individual analysis (HR>1) were 'inverted' when tested alongside lncRNAs that increased OS (HR<1), using the KM-Plotter 'invert' feature (Figure 2.1). This allowed lncRNAs with opposing prognostic effects to be tested in combination without the conflicting effects on survival cancelling each other out. The KM-Plotter calculated log-rank p-values and HRs with 95% CIs for each signature.

Figure 2.1 Screenshot of KM-Plotter “use multiple genes” feature (Menyhárt et al., 2018).



USE MULTIPLE GENES

○ Use list of genes ● Use mean expression of selected genes
○ Use ratio of two genes ○ Filter by median expression

Genes selected:

Probe Id	Symbol	Weight	Invert
11257	TP53TG1	1	☐
440823	MIAT	1	☐
7503	XIST	1	☑
414777	HCG18	1	☑

Search for gene

enter gene here

Paste list of genes

Each new ID must be entered into a new line. Maximum 100 genes are allowed. (example)

Correlation between genes:

Compute Spearman correlation: ☐
Compute Pearson correlation: ☐

For correlation analysis maximal 4 genes are allowed.

2.8 Subgroup analysis

Subgroup analysis was performed on the most prognostic signature to identify subtype-specific associations with stage, grade, vascular invasion, gender, race, alcohol consumption and hepatitis infection.

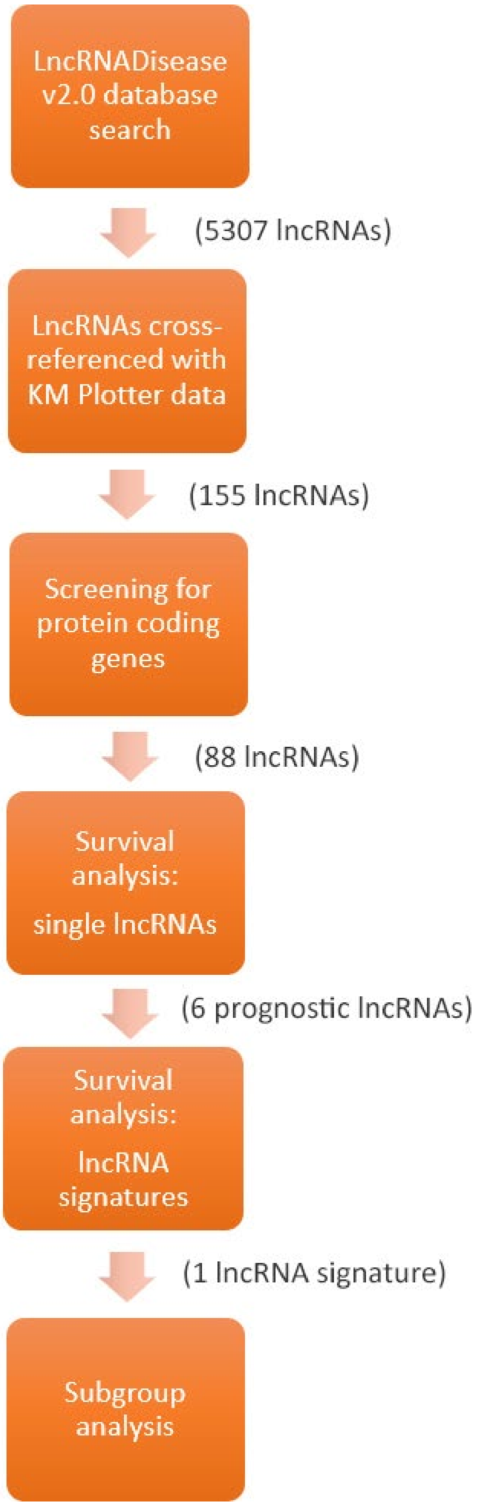
Survival analyses were performed as per the method for lncRNA signatures, but discrete patient subgroups were tested in isolation instead of the whole cohort.

2.9 Correction for multiple hypothesis testing

The KM-Plotter “Multiple Testing Correction” tool was used to calculate a corrected p-value threshold for statistical significance, to account for the multiple tests performed. The Hochberg “step-up” procedure was identified as the most appropriate correction method for the study design and number of tests performed (Gyorffy et al., 2005). The correction included all log-rank p-values and specified a significance threshold of p=0.05.

The study design is summarised in Figure 2.2.

Figure 2.2 A summary of the study design.



3 Results

3.1 Candidate lncRNAs

The LncRNADisease database search returned 5307 results (supplementary file 1). After exclusion of circRNAs and non-human results, 5269 HCC-related lncRNAs remained. 237 lncRNAs had an experimentally validated association with HCC, and 5302 had a predicted association.

Cross-referencing candidate lncRNAs with the KM-Plotter database produced a shortlist of 155 lncRNAs available for survival analysis (supplementary file 2).

3.2 Verified lncRNAs

Checks against the Ensembl and GeneCards databases revealed that the shortlist of candidate lncRNAs contained 42 mislabelled protein coding genes, two small nucleolar RNAs (Table 3.1), and four pseudogenes (non-functional 'copies' of coding genes) (Tutar, 2012). In total, 52 genes had a non-lncRNA classification and were excluded from further analysis.

The remaining 107 candidate genes had a confirmed "lncRNA" classification.

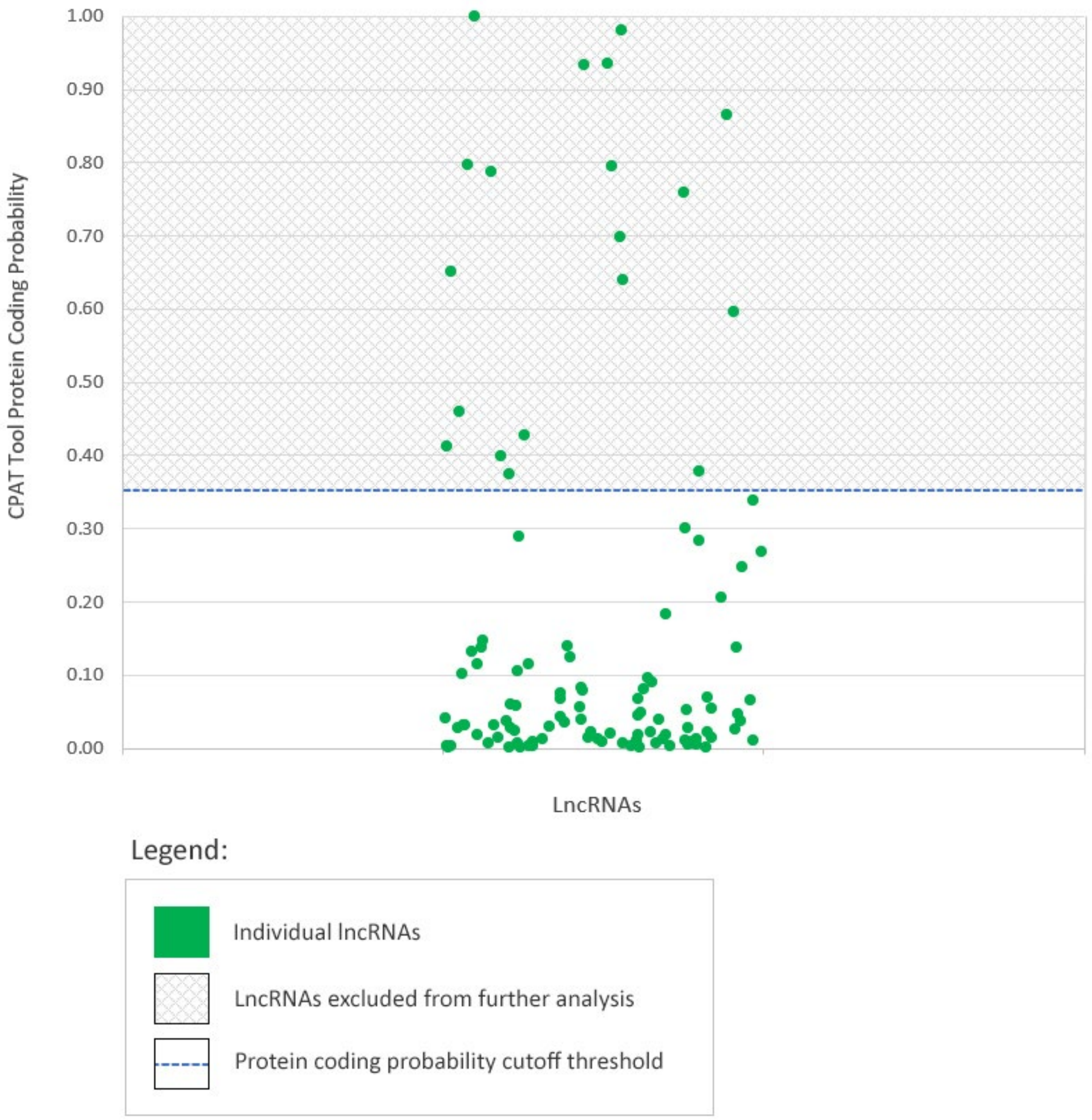
Table 3.1. Ensembl/GeneCards classification for 155 candidate genes from the LncRNADisease database.

Verified Gene Classification	Number in raw data
Long non-coding RNA (LncRNA)	107
Protein coding genes	42
Pseudogenes	4
Small nucleolar RNAs (snoRNAs)	2
Total	155

3.3 LncRNAs with protein coding potential

CPAT CP scores were obtained for 107 verified lncRNAs (Figure 3.1). 88 lncRNAs were confirmed as having no protein coding ability (CP < 0.364). 19 lncRNAs were shown to have significant protein coding potential (CP > 0.364) and were excluded from further analysis.

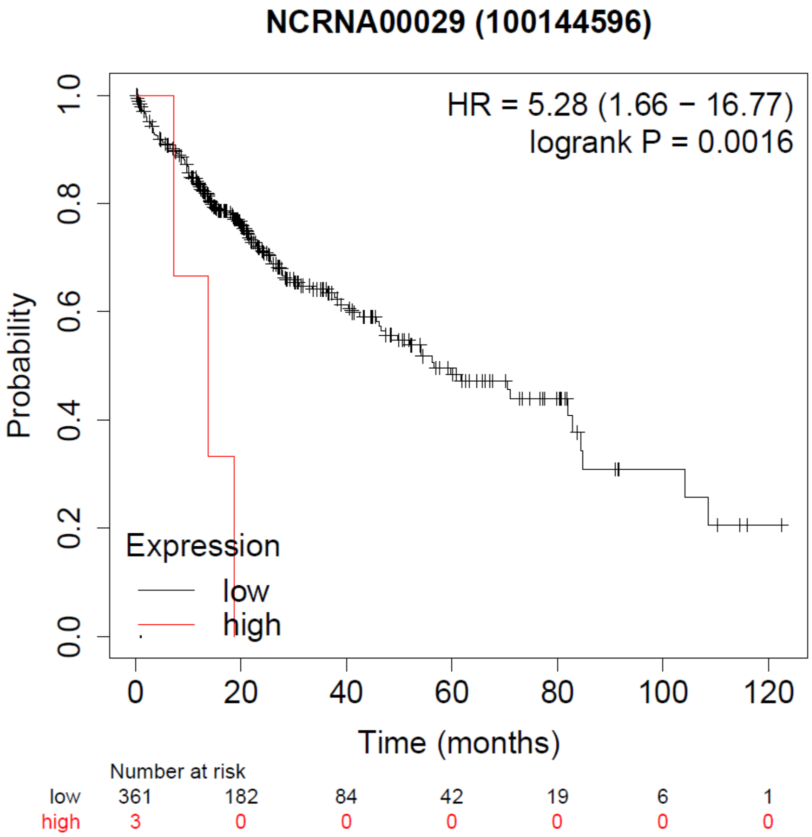
Figure 3.1 Coding probability (CP) scores for 107 HCC-related lncRNAs. Y-axis position indicates the CP score. Dashed line indicates the CP threshold of 0.364. CP≥0.364 indicates high protein coding potential. LncRNAs with CP<0.364 were selected for further analysis.



3.4 Identification of prognostic lncRNAs

Kaplan-Meier analysis was completed for 88 lncRNAs. Forty-four lncRNAs produced survival plots that were highly unbalanced, with many more patients in one group than the other (Figure 3.2). Unbalanced dichotomisation of patients reduces the reliability of survival analysis (Hsieh, 1992). Accordingly, 44 lncRNAs that produced highly unbalanced patient splits (defined as >10% difference between groups) were excluded from further analysis (supplementary file 4).

Figure 3.2. Survival plot for the lncRNA "NCRNA00029" demonstrating highly unbalanced patient dichotomisation. At month zero there are 361 patients in the low-expression group and 3 patients in the high-expression group.



Six lncRNAs (TP53TG1, TTTY15, MIAT, NEAT1, HCG18 and XIST) were associated with OS at $p<0.05$ (Figure 3.3).

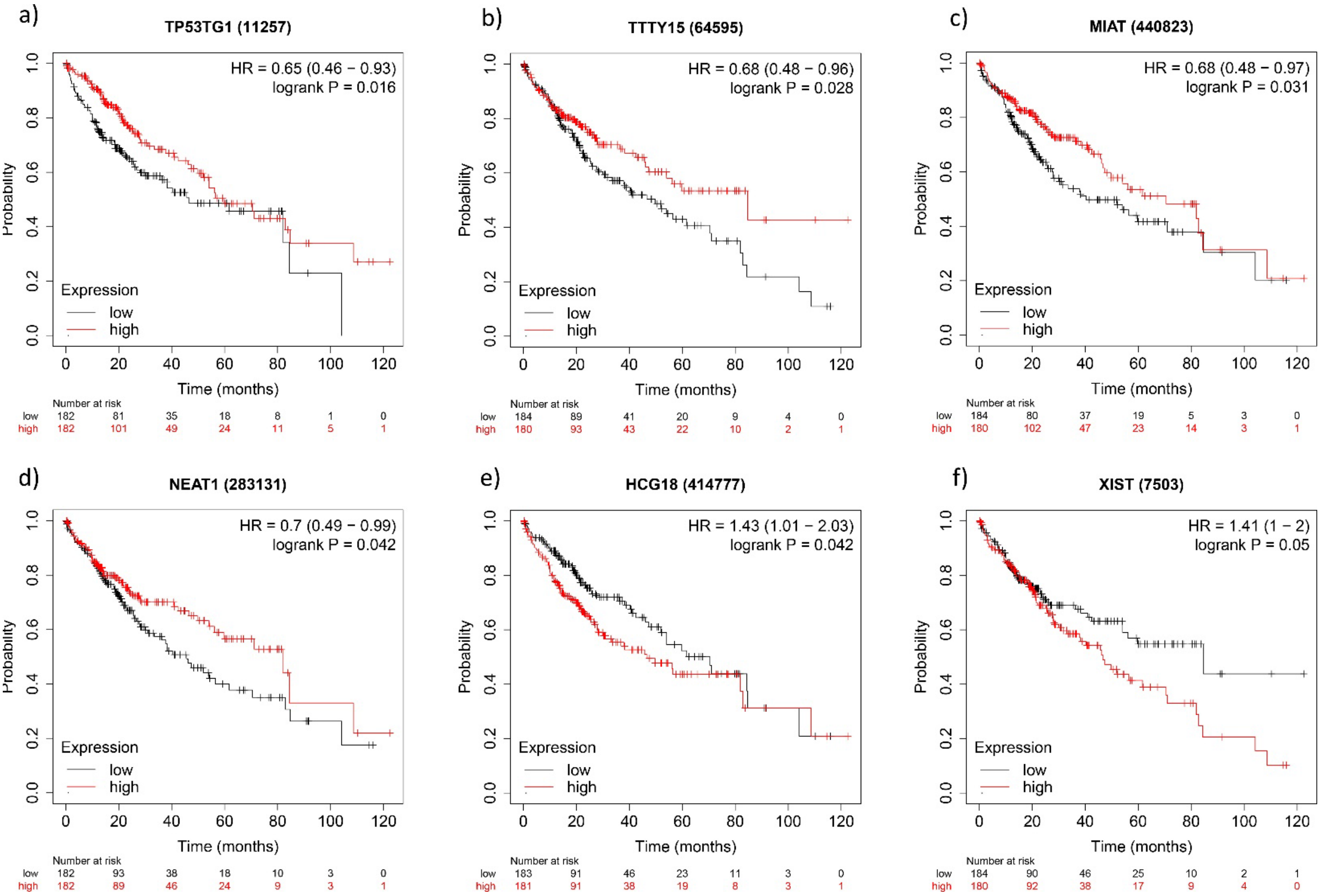
Survival plots for TP53TG1, TTTY15, MIAT and NEAT1 show that patients in the high-expression group had consistently higher survival probabilities over time compared to the low-expression group (Figure 3.3). HR values <1 show that high expression lowered the risk of death by 35% for TP53TG1, 32% of TTTY15, 32% for MIAT and 30% for NEAT1 (Figure 3.3).

Survival plots for HCG18 and XIST show that patients in the high expression group had consistently lower survival probabilities compared to the low expression group. HR values >1 show that high expression increased the risk of death by 43% for HCG18 and 41% for XIST (Figure 3.3).

After correction for multiple comparisons, the p-value significance threshold was adjusted to $p\leq0.003$. No individual lncRNAs remained significantly associated with OS at this adjusted threshold.

The six lncRNAs associated with OS at $p<0.05$ were selected for construction of a prognostic signature.

Figure 3.3. Kaplan-Meier plots for 6 lncRNAs associated with OS at $p<0.05$ in a cohort of HCC patients. n=364. a) TP53TG1, b) TTTY15, c) MIAT, d) NEAT1, e) HCG18, f) XIST. Red survival curves represent high-expression groups. Black curves represent low-expression groups. The number of patients in low/high groups at each time interval is displayed below the plot. Log-rank p-values and HR with 95% CI are shown.



3.5 Identification of a prognostic lncRNA signature

Combining TP53TG1, TTTY15, MIAT, NEAT1, HCG18 and XIST in all possible permutations produced 57 potential prognostic signatures. Forty-eight signatures were significantly associated with OS: 24 at p<0.05, 21 at p<0.01, and 2 at p<0.001. In all cases, HR<1 indicated that high expression of the signature was associated with longer OS.

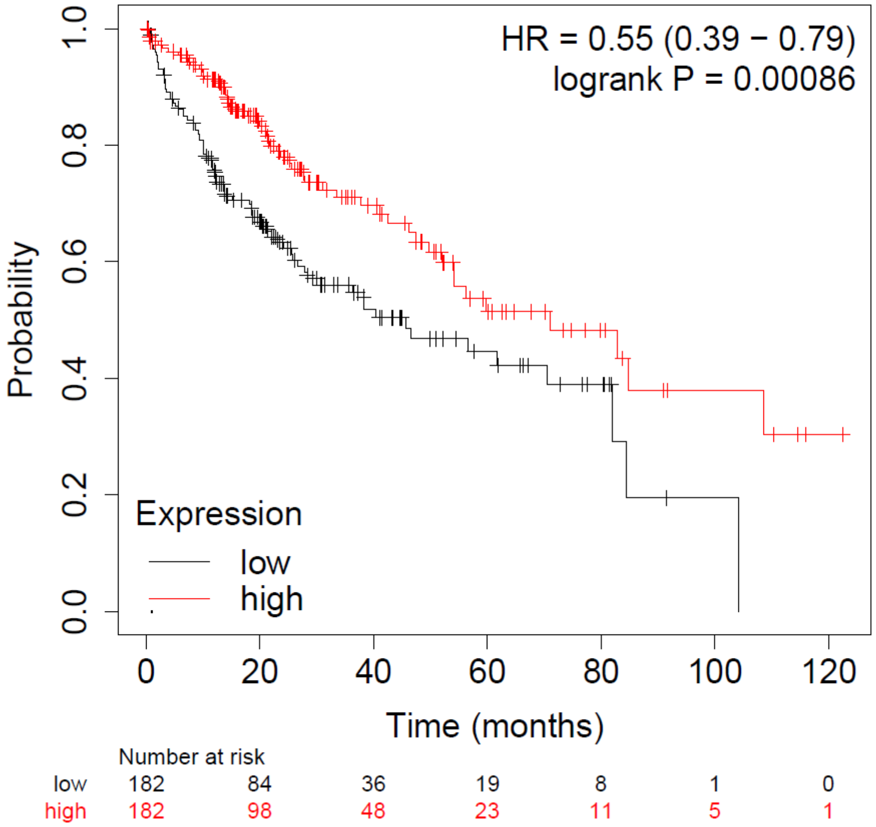
After correction for multiple comparisons, 10 signatures remained significantly associated with OS at p≤0.003 (Table 3.2).

Table 3.2. P-values and HR with 95% CI for 10 lncRNA signatures significantly associated with OS at p≤0.003. lncRNAs associated with increased OS are blue. lncRNAs associated with reduced OS are red.

LncRNA Signature	P	HR (95% CI)
TP53TG1_TTTY15_MIAT	0.00086	0.55 (0.39 - 0.79)
TP53TG1_MIAT_HCG18_XIST	0.00095	0.56 (0.39 - 0.79)
TP53TG1_MIAT_XIST	0.001	0.56 (0.39 - 0.79)
TP53TG1_TTTY15_MIAT_XIST	0.001	0.56 (0.39 - 0.79)
TP53TG1_TTTY15_MIAT_HCG18_XIST	0.0012	0.56 (0.39 - 0.8)
TP53TG1_MIAT	0.0023	0.58 (0.41 - 0.83)
TTTY15_MIAT	0.0026	0.59 (0.41 - 0.83)
TP53TG1_HCG18_XIST	0.0029	0.59 (0.41 - 0.84)
TP53TG1_MIAT_NEAT1_HCG18	0.003	0.58 (0.41 - 0.84)
TP53TG1_TTTY15_MIAT_NEAT_HCG18	0.003	0.59 (0.41 - 0.84)

The signature TP53TG1_TTTY15_MIAT showed the strongest association with OS (Table 3.2). The survival plot shows that patients in the high-expression group had a consistently higher probability of survival compared to the low-expression group (Figure 3.4). Overall, the risk of death was 45% higher for patients in the low-expression group (HR=0.55).

Figure 3.4. Survival analysis based on expression of the 3-lncRNA signature TP53TG1_TTTY15_MIAT in a group of 364 HCC patients. Black/red curves represent low/high-expression groups respectively. Log-rank p-value and HR with 95% CI are shown.



3.6 Subgroup analysis

Subgroup analysis revealed that high TP53TG1_TTTY15_MIAT expression was associated with increased OS in male patients (p=0.0075), Asian patients (p=0.0054) and patients with hepatitis virus infection (p=0.0263) (Table 3.3). The signature was prognostic for patients with and without alcohol risk (p=0.04, p=0.0056) (Table 3.3). The protective effect appeared to apply to intermediate stages and tumour grades only, with notable associations at Stage 2 (p=0.0281), Stages 2+3 combined (p=0.001) and Grade 2 (p=0.0092).

After correction for multiple comparisons, only the Stage 2+3 subgroup association remained significant.

Table 3.3. P-values and hazard ratios with 95% CI for subgroup analysis of the 3-lncRNA signature: TP53TG1_TTTY15_MIAT. Highlighted = significant after Hochberg correction.

Subgroup		P-value	HR (95% CI)
Stage	1 (n=170)	0.7533	0.91 (0.49 - 1.67)
	1+2 (n=257)	0.0607	0.63 (0.39 - 1.03)
	2 (n=83)	0.0281	0.4 (0.17 - 0.93)
	2+3 (n=166)	0.001	0.45 (0.28 - 0.73)
	3 (n=83)	0.4249	0.79 (0.44 - 1.42)
	3+4 (n=87)	0.5196	0.83 (0.47 - 1.47)
	4 (n=4)	(Sample too small to analyse)	
Grade	Grade 1 (n=55)	0.105	0.45 (0.17 - 1.21)
	Grade 2 (n=174)	0.0092	0.5 (0.3 - 0.85)
	Grade 3 (n=118)	0.0507	0.55 (0.3 - 1.01)
	Grade 4 (n=12)	(Sample too small to analyse)	
Vascular Invasion	None (n=203)	0.1028	0.65 (0.39 - 1.09)
	Micro (n=90)	0.4265	0.73 (0.34 - 1.58)
	Macro (n=16)	(Sample too small to analyse)	
Gender	Male (n=246)	0.0075	0.55 (0.35 - 0.86)
	Female (n=118)	0.5284	0.84 (0.48 - 1.46)
Race	White (n=181)	0.69	0.69 (0.43 - 1.09)
	Black/African American (n=17)	(Sample too small to analyse)	
	Asian (n=155)	0.0054	0.42 (0.23 - 0.79)
Sorafenib treatment	Treated (n=29)	0.7004	1.24 (0.41 - 3.73)
Alcohol consumption	Yes (n=117)	0.04	0.51 (0.27 - 0.98)
	None (n=202)	0.0056	0.52 (0.32 - 0.83)
Hepatitis virus	Yes (n=150)	0.0263	0.47 (0.24 - 0.93)
	None (n=167)	0.2859	0.78 (0.5 - 1.23)

3.7 Hochberg corrected significance threshold

Following Hochberg correction for multiple hypothesis testing, the threshold for statistical significance was adjusted to $p \leq 0.003$.

4 Discussion

Survival analyses based on the expression of 88 candidate lncRNAs in a cohort of 364 patients has identified 10 novel lncRNA signatures that are significantly associated with OS in HCC ($p \leq 0.003$) (Table 3.2). The 3-lncRNA signature TP53TG1_TTTY15_MIAT showed the strongest association with survival. Six individual lncRNAs were associated with OS at $p<0.05$, but no significant associations remained after correction for multiple comparisons (Figure 3.3). However, all 6 lncRNAs were included in at least one significant prognostic signature, suggesting that each has a role in predicting patient survival. Collectively, these results support the hypothesis that the expression of specific lncRNAs is associated with patient survival in HCC.

4.1 Novel findings

4.1.1 A 3-lncRNA prognostic signature for HCC

We have identified a 3-lncRNA signature which could serve as a clinical prognostic biomarker in HCC patients. Of the 10 signatures identified, TP53TG1_TTTY15_MIAT showed the strongest association with survival, and high expression of the signature was associated with significantly longer OS (Figure 3.4).

HCC subtypes involving distinct genetic mutations and signalling pathways have been linked to different aetiologies, and the prevalence of different aetiologies varies globally (Llovet et al., 2021). It was therefore anticipated that lncRNAs associated with OS could vary according to patient risk factors and ethnicity, as demonstrated previously (Menyhárt et al., 2018).

In line with this hypothesis, subgroup effects were noted in male patients, Asian patients, hepatitis patients, and those with intermediate disease (Stage 2, Stage 2+3 combined, Grade 2) (Table 3.3). However, only the association with Stage 2+3 patients combined remained significant after correcting for multiple comparisons, suggesting that any prognostic value may be limited to patients in intermediate stages of the disease.

It is possible that small sample sizes at the subgroup level lacked the statistical power to detect some genuine subgroup associations, especially at the adjusted significance threshold of $p \leq 0.003$. Subgroups including “Stage 4, “Grade 4”, and “Black/African American” were too small to analyse. To fully elucidate any subgroup-specific effects, the analysis should be repeated in a larger dataset.

4.1.2 Testis specific transcript Y-linked 15 (TTTY15)

TTTY15 has not previously been associated with HCC. We found that increased expression was associated with improved OS, suggesting a tumour-suppressive role. TTTY15 is Y-linked and thus only expressed in males (Stelzer et al., 2016). This could explain why the signature TP53TG1_TTTY15_MIAT was only prognostic in male patients and suggests this subtype effect was genuine, despite not reaching statistical significance. TTTY15 expression has previously been linked to poor OS in prostate cancer and improved OS in non-small-cell lung cancer in men (Xiao et al., 2019; Lai et al., 2019). LncRNAs are known to perform both ubiquitous and tissue specific functions, and this is consistent with the varying effects observed in different tumour types (Jiang et al., 2016). TTTY15 was found to influence gene expression by functioning as a miRNA sponge in these cancers, suggesting a possible mechanism for involvement in HCC (Xiao et al., 2019; Lai et al., 2019). Its Y-linked expression pattern and established associations with male-specific cancers suggests that TTTY15 could play a broader role in male cancer susceptibility, and further studies could investigate this connection. Functional studies are also needed to identify the mechanisms underlying its association with HCC.

4.2 Confirmation of previous work

4.2.1 TP53 target gene 1 (TP53TG1)

Our results confirm the recent findings of Chen et al. (2021), who demonstrated that reduced expression of TP53TG1 is predictive of poor OS and aggressive phenotypes in HCC patients. *In vitro* analysis has shown that TP53TG1 functions as a tumour-suppressor through ubiquitin-mediated degradation of the peroxisomal protein PRDX4 (Chen et al., 2021). This in turn inactivates the Wnt/β-catenin signalling pathway, halting the cell-cycle and proliferation (Chen et al., 2021). TP53TG1 is induced by the oncogene TP53 which is frequently mutated in aggressive HCC subtypes, suggesting that it could be an effective marker for aggressive disease (Llovet et al., 2021). In the current study, TP53TG1 was the most prognostic individual lncRNA. It also features in 9 of the 10 the statistically significant signatures identified, demonstrating a robust association with OS.

4.2.2 HLA Complex Group 18 (HCG18)

Our results show that high HCG18 expression reduces OS in HCC. In line with this, upregulation of HCG18 promotes the proliferation and migration of HCC cells *in vitro*, and downregulation has been shown to inhibit HCC growth in a cell-line derived xenograft (CDX) mouse model (Zou et al., 2020). The proposed mechanism for this oncogenic effect is sponging of the tumour-suppressive miRNA miR-214-3p, and upregulation of Centromere Protein M; a protein involved in chromosome segregation during mitosis (Zou et al., 2020). Our finding that high HCG18 expression reduces OS in HCC patients adds to the existing evidence for a functional oncogenic role.

4.2.3 X Inactive Specific Transcript (XIST)

In line with the findings of two small clinical studies, we found that increased expression of XIST was associated with poor OS in HCC (Mo et al., 2017; Liu et al., 2021). Mo et al. (2017) found that XIST was upregulated in HCC tumours and associated with increased tumour size and shorter disease-free survival. Inhibition of XIST *in vivo* and *in vitro* decreased proliferation and increased apoptosis, suggesting a mechanistic role in disease progression. Functional investigation revealed that XIST may sponge the tumour-suppressive microRNA miR-139-5p resulting in activation of the PI3K-Akt signalling pathway, which is frequently dysregulated in aggressive G1/G2 subtypes (Llovet et al., 2021). These results were recently replicated by Liu et al. (2021) who found that increased XIST was associated with poor OS in a cohort of 42 HCC patients.

These corroborating results confirm that TP53TG1, XIST and HCG18 are promising prognostic biomarkers that should be further investigated for clinical use. In particular, TP53TG1 and XIST could be used to detect aggressive HCC phenotypes. As mechanistic relationships have also been proposed for all three, these lncRNAs could be novel treatment targets.

4.3 Conflicting results

4.3.1 Myocardial Infarction Associated Transcript (MIAT)

Our data show that increased expression of MIAT is associated with improved OS in HCC (Figure 3.3). In contrast, evidence from cell-line experiments suggests that MIAT plays a fundamental role in HCC development and progression. (Da et al., 2020). Huang et al. (2018) found that upregulation of MIAT increased the proliferation and invasiveness of HCC cells in a CDX mouse model. MIAT exerted this effect by sponging the miRNA mir-214 *in vitro*, resulting in increased expression of the polycomb protein EZH2 and activation of the Wnt/β-catenin signalling pathway (Huang et al., 2018). Furthermore, Zhao et al. (2019) found that downregulating MIAT *in vitro* inhibited the proliferation of HCC cells by promoting apoptosis. Whilst these results are puzzling in the context of the current study, they are not directly related to the survival of HCC patients, and results from cell-line models cannot be directly extrapolated to humans (Nandwani et al., 2021). In contrast, the current study used a large patient cohort and thus presents robust evidence of a positive correlation between MIAT expression and OS. However, previous findings cannot be dismissed, and further work is needed to address these conflicting results.

4.3.2 Nuclear Paraspeckle Assembly Transcript 1 (NEAT1)

In contrast with the findings of a retrospective study by Liu et al (2017), we found that increased NEAT1 expression was associated with longer OS. This study included 88 Asian patients from a single centre in China. The reliability of the findings is therefore limited by the small sample size, lack of diversity, and retrospective study design. The results of the present study are more reliable and generalisable in comparison due to the large and diverse patient sample used (Table 2). However, all patients in the current study were recruited in North America, so it is possible that an unidentified factor, such as the prevalence of different HCC subtypes, underlies the different prognostic effects in these two locations. Subgroup analyses stratified on the prevalent risk factors and disease aetiologies in these two locations would help to determine whether this is the case.

4.4 Alternative signatures

Three prognostic lncRNA signatures have been suggested for HCC recently (Wang and Lei, 2021; Cao et al., 2021; Yang et al., 2020). There is no overlap between the findings of these studies and the signatures identified in the present study (Table 4). This disparity can be explained by the different methodologies used.

Table 4. A summary of results, methods, and data sources from three recent HCC lncRNA signature studies (Wang and Lei, 2021; Cao et al., 2021; Yang et al., 2020). Highlighted = protein coding/high coding probability (Stelzer et al., 2016)

Study	LncRNAs in the prognostic signature	Data source	Methods	Sample size
Wang et al., 2021	LINC01060, LINC01136, EGLN3-AS1, RP11-20J15.2, AC025580.1, HOXC-AS2, AC114912.1, LINC01517, AL592043.1, AC089983.1, DDX11-AS1	The Cancer Genome Atlas (TCGA)	• Differential expression analysis • Univariate cox regression Multivariate cox regression • Kaplan-Meier survival analysis	342
Cao et al., 2021	LINC01649, LINC01060, LINC00462, LINC01559, LINC00632, LINC00200, LINC01224, LOC100996671, LINC01508, LINC00668, LINC00942, LINC01970, LINC02202, SMIM32, FLJ36000, LRRC77P, DLX2_DT, MIR137HG	The Cancer Genome Atlas (TCGA)	• Differential expression analysis • Univariate cox regression • Multivariate cox regression • Kaplan-Meier survival analysis	310
Yang et al., 2020	MAPKAPK5-AS1, MUC20-OT1, DGCR5, RPL23A pseudogene	Gene Expression Omnibus (GEO)	• Differential expression analysis • Univariate cox regression • Multivariate cox regression • Kaplan-Meier survival analysis	225

All three recent studies used differential expression analysis (DEA) to identify lncRNAs of interest (Table 4). Wang and Lei (2021) and Cao et al. (2021) performed DEA on the same TGCA-LIHC dataset used in the present study. Yang et al. (2020) analysed microarray data from the GEO database. DEA is an unbiased genome-wide analytical method that can identify previously unannotated lncRNAs (Soneson and Delorenzi, 2013). In contrast, by necessity the present study used a candidate gene approach and was limited to investigating lncRNAs included in the LncRNADisease database. This is the primary limitation of the present study.

Recent studies have used multivariate analyses such as cox regression to test prognostic signatures (Wang and Lei, 2021; Cao et al., 2021). This type of analysis can identify and adjust for confounding variables in models and assess the interaction between variables in subgroup analyses (Schober and Vetter, 2021). As a univariable analysis, the Kaplan-Meier method used in the present study evaluates the impact of one variable at a time (Schober and

Vetter, 2021). Consequently, other potentially prognostic factors such as age were not controlled for in the present study, and this may be a source of bias. Furthermore, it has not been possible to determine whether the lncRNAs and signatures identified are independent prognostic factors.

Methods used in recent lncRNA signature studies may limit the reliability of results. In the present study, the median lncRNA/signature expression score was used to divide patients into high- and low-expression groups for survival analysis. In contrast, Cao et al. (2021) tested every possible cut-off and selected an "optimal" value to maximise survival differences between groups. This introduces a multiple comparisons problem that was not adjusted for in the study. Similarly, in Yang et al. (2020), multiple analyses were performed on the validation datasets, but no adjustment of the significance threshold was made to account for this. As such, the significance of the results presented in these two studies may be overstated.

In the current study, stringent checks on gene classification and protein coding ability were performed to ensure only genuine lncRNAs were included in the analysis. In contrast, the signatures reported by Cao et al. (2021) and Yang et al. (2020) include protein coding genes or lncRNAs with high coding potential (Table 4) (Stelzer et al., 2016). Whilst this may not affect the prognostic ability of the signatures, it limits their usefulness in terms of lncRNA research.

Recent studies have used a split-sample design in which patients are randomly divided into "training" and "testing" datasets (Wang and Lei, 2021; Cao et al., 2021). Prognostic signatures are developed in the "training" dataset and validated in the "testing" dataset, demonstrating reproducibility. External validation has not been performed on the results of the present study, and replication in an independent dataset is required.

4.5 Strengths and limitations

Strengths of the current study include the novel bioinformatic approach which allowed the analysis of a large number of lncRNAs, the large sample size, and the stringent elimination of protein coding genes. Additionally, the integrity of significant results has been maintained through rigorous Hochberg correction to account for multiple comparisons.

There are also a number of limitations. The candidate gene approach limited analysis to lncRNAs that are already known. Furthermore, only one primary database was accessible during the data collection phase. The quality and completeness of data obtained from the LncRNADisease database was questionable, as evidenced by the number of protein coding genes and other RNA classes filtered out at the "verification" stage.

Analysis was also limited to lncRNAs that were present in the KM-Plotter database. Only 155 of 5307 lncRNAs were available in the KM-Plotter, limiting the number of analyses that could be conducted. LncRNA nomenclature was another complicating factor. One lncRNA may have several aliases, but due to time constraints alias symbols were not checked. The implication of both is that prognostically important lncRNAs may have been missed, and evidence suggests this may be the case. For example, the lncRNAs ANRIL and HOTTIP have previously been linked to survival in HCC patients but were not present in the KM-Plotter (Ghafouri-Fard et al., 2021). Had they been available for analysis in the present study, the resulting prognostic signatures could have been very different.

Survival analyses for some lncRNAs were affected by a technical issue. Dichotomizing patients at the median cut-off point resulted in a highly uneven patient split for 44 lncRNAs. The reason for this issue is unknown. However, unbalanced dichotomisation is known to affect the reliability of survival analysis, and accordingly these 44 lncRNA were excluded from the results (Hsieh, 1992).

Despite a large sample size, some subgroups were too small to analyse and low patient numbers in others may have masked genuine subgroup-specific effects. Furthermore, options to stratify patients by known prognostic risk factors such as age and smoking were not provided, nor was it not possible to adjust for these as potentially confounding factors. These issues limit the conclusions that can be drawn from the subtype analysis.

In summary, the methods used were sufficient to provide a reasonably robust answer to the research question. However, the limiting factors noted introduce some uncertainty which should be considered when interpreting the results. As the results have not yet been validated in an independent dataset, the evidence presented is preliminary at this stage.

5 Conclusion

Bioinformatic analysis of online datasets has identified a novel 3-lncRNA signature that could be an independent prognostic factor in Stage 2+3 HCC patients. Nine additional prognostic signatures are presented for future evaluation. Additionally, we report novel evidence that the male-specific lncRNA TTTY15 is associated with OS in HCC. Our results support previous findings that TP53TG1, HCG18, XIST are prognostically important lncRNAs in HCC. Future studies could further investigate these lncRNAs for clinical use and investigate their potential as novel treatment targets. Conflicting results on the effects of NEAT1 and MIAT expression require further investigation.

References

Bao, Z., Yang, Z., Huang, Z., Zhou, Y., Cui, Q., Dong, D. (2019) 'LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases' *Nucleic Acids Research*, vol. 47, no. D1, pp. D1034–D1037. Available at https://doi.org/10.1093/nar/gky905

Cao, J., Wu, L., Lei, X., Shi, K., Shi, L., Shi, Y. (2021) 'Long non-coding RNA-based signature for predicting prognosis of hepatocellular carcinoma' *Bioengineered*, vol. 12, no. 1, pp. 673–681. Available at https://doi.org/10.1080/21655979.2021.1878763

Chen, B., Lan, J., Xiao, Y., Liu, P., Guo, D., Gu, Y., Song, Y., Zhong, Q., Ma, D., Lei, P., Liu, Q. (2021) 'Long noncoding RNA TP53TG1 suppresses the growth and metastasis of hepatocellular carcinoma by regulating the PRDX4/β-catenin pathway' *Cancer Letters*, vol. 513, pp.75–89. Available at https://doi.org/10.1016/j.canlet.2021.04.022

Da, C., Gong, C.-Y., Nan, W., Zhou, K.-S., Wu, Z.-L., Zhang, H.-H. (2020) 'The role of long non-coding RNA MIAT in cancers' *Biomedicine & Pharmacotherapy*, vol. 129, pp. 110359. Available at https://doi.org/10.1016/j.biopha.2020.110359

Ghafouri-Fard, S., Gholipour, M., Hussen, B.M., Taheri, M. (2021) 'The Impact of Long Non-Coding RNAs in the Pathogenesis of Hepatocellular Carcinoma' *Frontiers in Oncology*, vol. 11, pp. 649107-649107. Available at https://doi.org/10.3389/fonc.2021.649107

Gu, J., Zhang, X., Miao, R., Ma, X., Xiang, X., Fu, Y., Liu, C., Niu, W., Qu, K. (2018) 'A three-long non-coding RNA-expression-based risk score system can better predict both overall and recurrence-free survival in patients with small hepatocellular carcinoma' *Aging*, vol. 10, no. 7, pp. 1627–1639. Available at https://doi.org/10.18632/aging.101497

Gulìa, C., Baldassarra, S., Signore, F., Rigon, G., Pizzuti, V., Gaffi, M., Briganti, V., Porrello, A., Piergentili, R. (2017) 'Role of Non-Coding in the Etiology of Bladder Cancer' *Genes*, vol. 8, no. 11, pp. 339. Available at https://doi.org/10.3390/genes8110339

Gyorffy, B., Gyorffy, A., Tulassay, Z. (2005) '[The problem of multiple testing and solutions for genome-wide studies]' *Orvosi Hetilap1*, vol. 146, no. 12, pp. 559–563. Available at https://pubmed.ncbi.nlm.nih.gov/15853065/

Hsieh, F.Y. (1992) 'Comparing sample size formulae for trials with unbalanced allocation using the logrank test' *Statistics in Medicine*, vol. 11, no. 8, pp. 1091–1098. Available at https://doi.org/10.1002/sim.4780110810

Hu, X., Jiang, J., Xu, Q., Ni, C., Yang, L., Huang, D. (2018) 'A Systematic Review of Long Noncoding RNAs in Hepatocellular Carcinoma: Molecular Mechanism and Clinical Implications' *BioMed Research International*, vol. 2018, pp. 8126208–8126208. Available at https://doi.org/10.1155/2018/8126208

Huang, X., Gao, Y., Qin, J., Lu, S. (2018) 'lncRNA MIAT promotes proliferation and invasion of HCC cells via sponging miR-214' *American Journal of Physiology:*

Gastrointestinal and Liver Physiology, vol. 314, no. 5, pp. G559–G565. Available at https://doi.org/10.1152/ajpgi.00242.2017

Huang, Z., Zhou, J.-K., Peng, Y., He, W., Huang, C. (2020) 'The role of long noncoding RNAs in hepatocellular carcinoma' *Molecular Cancer*, vol. 19, no. 1, pp. 77. Available at https://doi.org/10.1186/s12943-020-01188-4

Jiang, C., Li, Y., Zhao, Z., Lu, J., Chen, H., Ding, N., Wang, G., Xu, J., Li, X. (2016) 'Identifying and functionally characterizing tissue-specific and ubiquitously expressed human lncRNAs' *Oncotarget*, vol. 7, no. 6, pp. 7120–7133. Available at https://doi.org/10.18632/oncotarget.6859

Kadali, V., Chandran, S., Murthy, S. (2018) 'Long Non-Coding RNAs and their "Orchestration" in Cancers' *Journal of Applied Biology & Biotechnology*, vol. 6, no.5. pp. 57–60. Available at https://doi.org/10.7324/JABB.2018.60509

Lai, I.-L., Chang, Y.-S., Chan, W.-L., Lee, Y.-T., Yen, J.-C., Yang, C.-A., Hung, S.-Y., Chang, J.-G. (2019) 'Male-Specific Long Noncoding RNA TTTY15 Inhibits Non-Small Cell Lung Cancer Proliferation and Metastasis via TBX4' *International Journal of Molecular Sciences*, vol. 20, no. 14, pp. 3473. Available at https://doi.org/10.3390/ijms20143473

Liu, L., Jiang, H., Pan, H., Zhu, X. (2021) 'LncRNA XIST promotes liver cancer progression by acting as a molecular sponge of miR-200b-3p to regulate ZEB1/2 expression' *The Journal of International Medical Research*, vol. 49, no. 5, pp. 03000605211016211. Available at https://doi.org/10.1177/03000605211016211

Liu, Z., Chang, Q., Yang, F., Liu, B., Yao, H.-W., Bai, Z.-G., Pu, C.-S., Ma, X.-M., Yang, Y., Wang, T.-T., Guo, W., Zhou, X.-N., Zhang, Z.-T. (2017) 'Long non-coding RNA NEAT1 overexpression is associated with unfavorable prognosis in patients with hepatocellular carcinoma after hepatectomy: A Chinese population-based study' *European Journal of Surgical Oncology*, vol. 43, no. 9, pp. 1697–1703. Available at https://doi.org/10.1016/j.ejso.2017.06.013

Llovet, J.M., Kelley, R.K., Villanueva, A., Singal, A.G., Pikarsky, E., Roayaie, S., Lencioni, R., Koike, K., Zucman-Rossi, J., Finn, R.S. (2021) 'Hepatocellular carcinoma' *Nature Review Disease Primers*, vol. 7, no.1, pp. 1–28. Available at https://doi.org/10.1038/s41572-020-00240-3

Ma, L., Cao, J., Liu, L., Du, Q., Li, Z., Zou, D., Bajic, V.B., Zhang, Z. (2019) 'LncBook: a curated knowledgebase of human long non-coding RNAs' *Nucleic Acids Research*, vol. 47, no. D1, pp. D128–D134. Available at https://doi.org/10.1093/nar/gky960

Menyhárt, O., Nagy, Á., Győrffy, B. (2018) 'Determining consistent prognostic biomarkers of overall survival and vascular invasion in hepatocellular carcinoma' *Royal Society Open Science*, vol. 5, no. 12, pp. 181006-181006. Available at https://doi.org/10.1098/rsos.181006

Mo, Y., Lu, Y., Wang, P., Huang, S., He, L., Li, D., Li, F., Huang, J., Lin, X., Li, X., Che, S., Chen, Q. (2017) 'Long non-coding RNA XIST promotes cell growth by regulating miR-139-

5p/PDK1/AKT axis in hepatocellular carcinoma' *Tumor Biology*, vol. 39, no. 2, pp. 101042831769099. Available at https://doi.org/10.1177/1010428317690999

Nandwani, A., Rathore, S., Datta, M. (2021) 'LncRNAs in cancer: Regulatory and therapeutic implications' *Cancer Letters*, vol. 501, pp. 162–171. Available at https://doi.org/10.1016/j.canlet.2020.11.048

Nie, Y., Li, Y., Xu, Y., Jiao, Y., Li, W. (2020) 'Long non-coding RNA BACE1-AS is an independent unfavorable prognostic factor in liver cancer' *Oncology Letters*, vol. 20, no. 5, pp. 1-9. Available at https://doi.org/10.3892/ol.2020.12065

Schober, P., Vetter, T.R. (2021) 'Kaplan-Meier Curves, Log-Rank Tests, and Cox Regression for Time-to-Event Data' *Anesthesia & Analgesia*, vol. 132, no. 4, pp. 969–970. Available at https://doi.org/10.1213/ANE.0000000000005358

Soneson, C., Delorenzi, M. (2013) 'A comparison of methods for differential expression analysis of RNA-seq data' *BMC Bioinformatics*, vol. 14, no.1, pp. 91. Available at https://doi.org/10.1186/1471-2105-14-91

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T.I., Nudel, R., Lieder, I., Mazor, Y., Kaplan, S., Dahary, D., Warshawsky, D., Guan-Golan, Y., Kohn, A., Rappaport, N., Safran, M., Lancet, D. (2016) 'The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses' *Current Protocols in Bioinformatics*, vol. 54, no. 1, pp. 1.30.1-1.30.33. Available at https://doi.org/10.1002/cpbi.5

Tutar, Y. (2012) 'Pseudogenes' *Comparative and Functional Genomics*, vol. 2012, pp. 424526-4. Available at https://doi.org/10.1155/2012/424526

Wang, A., Lei, J. (2021) 'Identification of an 11-lncRNA signature with high performance for predicting the prognosis of hepatocellular carcinoma using bioinformatics analysis' *Medicine (Baltimore)*, vol. 100, no. 5, pp. e23749- e23749. Available at https://doi.org/10.1097/MD.0000000000023749

Wang, L., Park, H. J., Dasari, S., Wang, S., Kocher, J.-P., & Li, W. (2013) 'CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model' *Nucleic Acids Research*, vol. 41, no. 6, pp. e74-e47. Available at https://doi.org/10.1093/nar/gkt006

Xiao, G., Yao, J., Kong, D., Ye, C., Chen, R., Li, L., Zeng, T., Wang, L., Zhang, W., Shi, X., Zhou, T., Li, J., Wang, Y., Xu, C.L., Jiang, J., Sun, Y. (2019) 'The Long Noncoding RNA TTTY15, Which Is Located on the Y Chromosome, Promotes Prostate Cancer Progression by Sponging let-7' *European Urology*, vol. 76, no.3, pp. 315–326. Available at https://doi.org/10.1016/j.eururo.2018.11.012

Yan, J., Zhou, C., Guo, K., Li, Q., Wang, Z. (2019) 'A novel seven-lncRNA signature for prognosis prediction in hepatocellular carcinoma' *Journal of Cellular Biochemistry*, vol. 120, no. 1, pp. 213–223. https://doi.org/10.1002/jcb.27321

Yang, J.D., Hainaut, P., Gores, G.J., Amadou, A., Plymoth, A., Roberts, L.R. (2019) 'A global view of hepatocellular carcinoma: trends, risk, prevention and management' *Nature*

Reviews Gastroenterology & Hepatology, vol. 16, no. 10, pp. 589–604. Available at https://doi.org/10.1038/s41575-019-0186-y

Yang, Z., Yang, Y., Zhou, G., Luo, Y., Yang, W., Zhou, Y., Yang, J. (2020) 'The Prediction of Survival in Hepatocellular Carcinoma Based on A Four Long Non-coding RNAs Expression Signature' *Journal of Cancer*, vol. 11, no. 14, pp. 4132–4144. Available at https://doi.org/10.7150/jca.40621

Yu, X., Zhang, J., Yang, R., Li, C. (2020) 'Identification of Long Noncoding RNA Biomarkers for Hepatocellular Carcinoma Using Single-Sample Networks' *BioMed Research International*, vol. 2020, pp. 8579651-8579651. Available at https://doi.org/10.1155/2020/8579651

Yuan, D., Chen, Y., Li, X., Li, J., Zhao, Y., Shen, J., Du, F., Kaboli, P.J., Li, M., Wu, X., Ji, H., Cho, C.H., Wen, Q., Li, W., Xiao, Z., Chen, B. (2021) 'Long Non-Coding RNAs: Potential Biomarkers and Targets for Hepatocellular Carcinoma Therapy and Diagnosis' *International Journal of Biological Sciences*, vol. 17, no. 1, pp. 220–235. Available at https://doi.org/10.7150/ijbs.50730

Zhao, L., Hu, K., Cao, J., Wang, P., Li, J., Zeng, K., He, X., Tu, P.-F., Tong, T., Han, L. (2019) 'lncRNA miat functions as a ceRNA to upregulate sirt1 by sponging miR-22-3p in HCC cellular senescence' *Aging*, vol. 11, no. 17, pp. 7098–7122. Available at https://doi.org/10.18632/aging.102240

Zou, Y., Sun, Z., Sun, S. (2020) 'LncRNA HCG18 contributes to the progression of hepatocellular carcinoma via miR-214-3p/CENPM axis' *The Journal of Biochemistry*, vol. 168, no. 5, pp. 535–546. Available at https://doi.org/10.1093/jb/mvaa073

Acknowledgements

I would like to acknowledge and offer my warmest thanks to my tutor, Dr Susan Cliffe, and mentor Dr Francesco Crea, for their invaluable support and mentorship.

I would also like to thank Priyadarsini Nambiar and Azuma Kalu for their patience, encouragement, and expert advice.

Finally - I would like to thank my family for helping me to find the time and focus needed to complete my degree. Without your support, this project would not have been possible.

Appendices

Supplementary file 1: Search results from the LncRNADisease database

Supplementary file 2: Candidate LncRNAs available in the KM-Plotter

Supplementary file 3: Results from gene classification checks and CPAT testing

Supplementary file 4: Results from survival analyses of all individual lncRNAs

Supplementary file 5: Results from survival analyses of all lncRNA signatures