Figure legends:

**Fig 1. Characterization of lncRNA expression across tumours and normal tissues. (A)** t-SNE dimensionality reduction visualization of lncRNA expression profiles across 28 tumour types from TCGA labelled by original cancer type. **(B)** Summary of differentially ranked lncRNAs between TCGA patients and GTEx samples from matching primary organ sites. **(C)** Summary of prognostic lncRNAs (adjusted Wald *P* < 0.1)across TCGA cancer types. **(D)** Top 50 significantly co-expressed lncRNA-lncRNA pairs in adrenocortical carcinoma (ACC) and low-grade glioma (LGG). Spearman correlation values are shown and white blocks represent non-significant correlations.

**Fig 2. Identification of highly prognostic lncRNA candidates. (A)** Workflow for integrating machine learning methods to identify the most clinically relevant lncRNAs using cross-validation and elastic-net regularization. **(B)** Summary of c-indices obtained for models trained and evaluated using cross-validations. Results shown for cancer types with significant differences between clinical, lncRNA and combined models (left). Summary of the number of lncRNA candidates selected in more than 50% of cross-validation for each cancer type (right). **(C)** Evaluation of individual lncRNA candidate-based Cox-PH models compared to clinical variable based models in each cancer type (top). Evaluation of lncRNA candidate-based survival models compared to models fit using randomly sampled lncRNAs.

**(D)** lncRNA candidates split by their location relative to protein-coding genes, a common method to classify lncRNAs. **(E)** Performance of lncRNA candidate-based survival models relative to their neighboring protein-coding genes, n=126 pairs compared. lncRNAs are labelled by genomic classification.

**Fig 3. External validation of lncRNA prognostic candidates. (A).** Pearson correlation between Hazard-Ratios obtained in lncRNA-based survival models in TCGA cohorts versus PCAWG cohorts. Gene names shown for five lncRNA candidates that were significantly associated with survival in both cohorts, Wald *P* < 0.05). Kaplan-Meier plots representing patient cohorts split by lncRNA expression with log rank p-value for (B) LIHC, (C) KIRC, (D) HNSC, (E) LUAD & (F) LIHC. Dashed line represents time at which survival probability is 50%.

**Fig 4. Copy number aberrations and epigenetic differences in lncRNA defined risk groups. (A)** lncRNA candidates that overlapped copy number segments separated by risk type showing the percentage or risk patients that had an amplification (red bars) or deletion (blue bars). (**B)** *RP13-1032I1.7* lncRNA transcript abundance in cervical cancer stratified by levels of copy number: deletion (segment mean < -0.3), other (-0.3 < segment mean < 0.3) and amplification (segment mean > 0.3). (**C**) lncRNA candidates that overlapped CpG probes whose methylation was significantly associated with transcript abundance (Kruskal Wallis adjusted *P* < 0.05 and Spearman *P* < 0.05). Risk with methylation indicates patients whose lncRNA defined hazard matched expected methylation direction. For example, for a hazardous lncRNA, risk is high expression and expected methylation levels would indicate a negative correlation where patients with low methylation would have high lncRNA abundance. These would be risk patients with methylation. Patients with low expression and high methylation are indicated as non-risk group with methylation signal. Patients with no correlation between methylation and transcript abundance are labeled as others. (**D**) *CTC-297N7.5* lncRNA transcript abundance in liver cancer stratified by levels of methylation: unmethylated (beta values < 0.25), other (0.25 < beta values < 0.75) and methylation (beta values > 0.75).

**Fig 5. Clinical and molecular features associated with lncRNA expression. (A)** Overview of lncRNAs associated with different clinical or molecular features using a chi-square test. Boxes marked by a “v” indicate that survival models fit using both the clinical/molecular feature and lncRNA candidate outperform clinical/molecular based model. (**B**) Combined survival models (using lncRNA candidate and its associated clinical/molecular feature) concordance values versus clinical/molecular feature only based models. Line drawn through 0,0. **(C)** Kaplan-Meier plot showing patient stratification in low grade glioma using both IDH1 mutation status and two lncRNA candidate expression profiles *HOXB-AS2* (top) and *HOXA10-AS* (bottom).

**Fig 6. Identification of protein-coding gene signatures associated with lncRNA expression. (A)** Summary of differential expression analysis comparing the expression of protein-coding genes between each lncRNA candidate defined patient risk groups (low versus high lncRNA expressing patients). The number of cancer gene census genes within these signatures is shown in addition to the number of pathways enriched by these genes using gProfiler. (**B**) Pathway enrichment map showing pathways associated with lncRNA candidates in low grade glioma. Each node represents a pathway (GO term and Reactome terms). The nodes are colored by the lncRNA candidates they are associated with as defined by ActivePathways. Pathway clusters were manually annotated. (**C**) Transcript abundance of lncRNA candidates and protein-coding genes identified in the brain development pathway were used to cluster LGG and GBM patients. Relative risk was calculated for LGG patients using multivariate models accounting for *HOXB-AS2* and *HOXA10-AS*. IDH mutation status is labeled for all patients.