Pan-cancer transcriptomic analysis associates long non-coding RNAs with key mutational driver events

Arghavan Ashouri1,*, Volkan I. Sayin2,*, Jimmy Van den Eynden1, Simranjit X. Singh2, Thales Papagiannakopoulos2 & Erik Larsson1

Thousands of long non-coding RNAs (lncRNAs) lie interspersed with coding genes across the genome, and a small subset has been implicated as downstream effectors in oncogenic pathways. Here we make use of transcriptome and exome sequencing data from thousands of tumours across 19 cancer types, to identify lncRNAs that are induced or repressed in relation to somatic mutations in key oncogenic driver genes. Our screen confirms known coding and non-coding effectors and also associates many new lncRNAs to relevant pathways. The associations are often highly reproducible across cancer types, and while many lncRNAs are co-expressed with their protein-coding hosts or neighbours, some are intergenic and independent. We highlight lncRNAs with possible functions downstream of the tumour suppressor TP53 and the master antioxidant transcription factor NFE2L2. Our study provides a comprehensive overview of lncRNA transcriptional alterations in relation to key driver mutational events in human cancers.

1 Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, The Sahlgrenska Academy, University of Gothenburg, SE-405 30 Gothenburg, Sweden. 2 Department of Pathology and Perlmutter Cancer Center, New York University School of Medicine, New York, New York 10016, USA. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to T.P. (email: thales.papagiannakopoulos@nyumc.org) or to E.L. (email: erik.larsson@gu.se).

DOI: 10.1038/ncomms13197
Transcriptomic studies have shown that mammalian genomes encode an abundance of messenger RNA (mRNA)-like transcripts that are often multiexonic and polyadenylated, but lack obvious protein-coding capacity\(^1\text{-}^3\). Increasingly, it is becoming clear that some of these long non-coding RNAs (lncRNAs) have important regulatory roles in cells, for example by controlling transcription through recruitment of histone-modifying complexes to chromatin\(^4\text{-}^5\), regulation of mRNA translation\(^6\text{-}^7\) or control of mRNA splicing\(^8\). Several studies also implicate dysregulation of lncRNAs as a contributing factor in human cancer, with early examples such as MALAT1 (ref. 9) now followed by many others\(^10\text{-}^14\).

Notably, an increasing number of studies describe lncRNAs that act as critical downstream effectors in established cancer-relevant pathways. Examples include the identification of lncRNA-p21 as a key target and mediator of the p53 DNA damage response\(^12\text{-}^{13}\), as well as other studies of p53-regulated long non-coding RNAs\(^14\text{-}^{16}\). Similarly, inhibition of Notch1 in T-cell acute lymphoblastic leukaemia identified lncRNAs acting as downstream targets and effectors of Notch signalling\(^17\), and both overexpression and siRNA silencing of MYC, followed by transcriptomic profiling, has been used to identify long non-coding RNAs acting downstream of c-Myc (refs 18,19). Another recent study compared the transcriptomes of melanocytes with and without expression of mutant BRAF(V600E), revealing differential expression of several lncRNAs including BANCR, shown to affect melanoma cell migration\(^20\).

While intriguing, these studies are still limited in number, and it is likely that lncRNAs have roles also in many other oncogenic programs. An attractive prospect is therefore to harness the increasing availability of mutational and transcriptomic data from tumours, to identify associations between coding oncogenic drivers and potential effector lncRNAs in a more systematic way. Recently, this idea was briefly explored in lung cancer using data from The Cancer Genome Atlas (TCGA), revealing associations between the mutational status of six key oncogenes and altered expression of lncRNAs (ref. 21). Although an obvious drawback to this approach is that it is correlative, that is, does not involve molecular perturbations, in return it has the potential to greatly facilitate the exploration of associations between lncRNAs and a large number of mutational driver events. Importantly, associations detected in this manner also represent direct observations from actual human tumours; a distinct advantage compared with data from cell culture systems. Still, the principle has not been systematically explored and evaluated in comprehensive human tumour materials.

Here we make use of molecular data generated by TCGA to determine expression levels and copy number amplitudes for GENCODE (refs 22,23) lncRNAs in more than 7,000 tumours across 19 types of human cancer. We establish a methodology to systematically search for associations between established key driver mutations and alterations in the expression of individual lncRNAs. The approach is evaluated in several ways, including silencing of the antioxidant transcription factor NFE2L2, which plays a critical role in lung tumorigenesis\(^24\), followed by transcriptome profiling to confirm predicted NFE2L2-responsive lncRNAs. We present a catalogue of associations, involving a broad repertoire of driver events, that are often consistent across many cancer types and that may serve as a reference and starting point for experimental studies. Additionally, our study provides an overview of lncRNA expression in human cancers.

**Results**

**LncRNA molecular profiles across 7,295 human tumours.** We established a framework whereby poly(A)-positive transcriptome sequencing (RNA-seq) and copy-number data from TCGA were used to derive lncRNAs expression and copy-number profiles across 7,295 human tumours. We additionally obtained coding mutation data for a subset of 4,698 samples (Fig. 1a; Table 1). The cohort spanned 19 cancer types, although breast cancer (BRCA) samples were subdivided into the four PAM50 (ref. 25) subtypes for a total of 22 types, each represented by between 49 and 517 tumours (Fig. 1b).

Briefly, expression levels were determined by realignment of RNA-seq libraries to the human genome and subsequent quantification of lncRNAs and coding genes in the GENCODE (ref. 22) annotation (Methods). We choose this lncRNA catalogue as it has been subjected to extensive manual curation and evaluation\(^23\). We considered 13,307 lncRNAs that were further subdivided into 7,885 ‘genic’ and 5,422 ‘intergenic’ loci, based on a 5 kb threshold with respect to the distance to the nearest coding gene (Fig. 1c). LncRNAs are often expressed at low levels, and high sequencing depth is therefore essential for proper quantification\(^26\). We consequently included only RNA-seq libraries generated on the Illumina HiSeq 2000/2500 platform that resulted in at least 25 million fragments mapping to...
GENCODE gene loci (average 54.8 million), in the end obtaining data from 7,295 samples that each represented a unique tumour and patient (Supplementary Data 1). Finally, IncRNAs and coding genes were assigned DNA copy number amplitudes by remapping of segmented data from the Affymetrix SNP6 platform, available for all 7,295 samples.

Overview of IncRNA expression in human cancer. We next sought to obtain an overview of the obtained IncRNA profiles, in part to ensure the overall technical validity. By comparison of the maximum length-normalized expression value (FPKM) across all samples we found that IncRNAs were expressed at considerably lower levels than coding genes in tumours (Fig. 2a), in agreement with several earlier observations. Notably, only 439,000 fragments were on average mapped to IncRNA loci in each sample, further reduced to 171,000 for intergenic IncRNAs and to be compared with 52.8 million for coding genes, again stressing the need for high sequencing depth in IncRNA transcriptomic studies.

While IncRNAs in general are weakly expressed, some are highly abundant and exceed even many coding mRNAs (ref. 30). We found that a small number of loci contributed a major fraction of all intergenic IncRNA-derived reads, notably with the top five constituting more than 50% in many cancers (for example, XIST (ref. 31) or HULC (ref. 32); Fig. 2b, Supplementary Table 1). The total fraction IncRNA-derived reads varied between cancers (P <1e-99, one-way ANOVA), to a large extent explained by abundant MALAT1 or NEAT1 in some types (Fig. 2b). This variability was largely mirrored in corresponding normal samples, and was not explained by MALAT1/NEAT1 copy number gain (Supplementary Fig. 1). Notably, these IncRNAs are both processed by RNase P from a poly(A)-positive into a poly(A)-negative transcript, and a shift in the ratio between forms or, alternatively, increased detection of poly(A)-negative RNA in some of the cancers, could conceivably contribute to the observed variability. Among top expressed loci in all cancers was also LINCC00657, previously shown to be conserved between human and mouse and highly expressed in endothelium.

We next investigated genome-wide correlations between DNA copy number amplitudes and expression levels. While these were generally much lower for IncRNAs compared with coding genes, this was largely explained by the overall low abundance of IncRNAs (Supplementary Fig. 2), and correlations approached that of coding genes when considering only highly expressed loci (Fig. 2c). The results support the overall validity of the data, as positive DNA-to-RNA correlations depend on proper IncRNA quantification.

Principal components analysis was applied to investigate whether cancer types were distinguishable based on their repertoires of expressed intergenic IncRNAs (Fig. 2d). We found that tumours from the same or related cancer types typically clustered closely together, supporting earlier results that many cancers have distinct IncRNA signatures. In the principal components analysis, glioblastoma and low-grade glioma samples were similar yet different from other cancers, suggesting a distinct IncRNA profile in neural tumours. Likewise, there was co-occurrence between breast (BRCA) and ovarian (OV) carcinomas, known to have several commonalities including copy number alterations. The results show that tumour types are clearly distinguishable by their IncRNA expression profiles, a likely reflection of the high degree of tissue specificity shown by many IncRNAs.

Associations between driver mutations and mRNA changes. We next explored the relationship between somatic driver mutations and transcriptional changes in tumours, with the ultimate goal of associating key mutational events with alterations in IncRNA expression. The principle has so far only seen limited use and the feasibility has not been evaluated, and we therefore wanted to first assess and establish our methodology on the protein-coding transcriptome.

Table 1 | Overview of included tumours and data types.

| Abbreviation | Description | Samples | Avg. fragments x 10⁶ (million) | Total fragments (million) | Copy number (Affy SNP6) | Mutations (exome) |
|--------------|------------|---------|-------------------------------|---------------------------|------------------------|-------------------|
| ACC          | Adrenocortical carcinoma | 73      | 46                            | 3,391                     | 73                     | 71                |
| BLCA         | Bladder carcinoma       | 390     | 48                            | 18,603                    | 390                    | 123               |
| BPCA         | Breast carcinoma (basal) | 189     | 58                            | 10,951                    | 189                    | 168               |
| BRCA Her2    | Breast carcinoma (Her2)  | 127     | 56                            | 7,161                     | 127                    | 115               |
| BRCA LumA    | Breast carcinoma (Luminal A) | 435   | 57                            | 24,805                    | 435                    | 391               |
| BRCA LumB    | Breast carcinoma (Luminal B) | 319   | 60                            | 19,120                    | 319                    | 281               |
| CESC         | Cervical carcinoma      | 286     | 53                            | 15,036                    | 286                    | 188               |
| GBM          | Glioblastoma multiforme | 134     | 42                            | 5,605                     | 134                    | 124               |
| HNSC         | Head & neck carcinoma  | 495     | 53                            | 26,029                    | 495                    | 278               |
| KICH         | Chromophobe renal carcinoma | 66     | 63                            | 4,170                     | 66                     | 66                |
| KIRC         | Clear cell renal carcinoma | 517   | 58                            | 30,095                    | 517                    | 408               |
| KIRP         | Papillary renal carcinoma | 277    | 51                            | 14,219                    | 277                    | 153               |
| LGG          | Low-grade glioma        | 511     | 55                            | 28,191                    | 511                    | 278               |
| LIHC         | Hepatocellular carcinoma | 358    | 49                            | 17,503                    | 358                    | 189               |
| LUAD         | Lung adenocarcinoma     | 467     | 49                            | 22,927                    | 467                    | 202               |
| LUSC         | Lung squamous cell carcinoma | 484  | 52                            | 25,389                    | 484                    | 171               |
| OV           | Serous ovarian carcinoma | 409    | 61                            | 24,835                    | 409                    | 190               |
| PRAD         | Prostate carcinoma      | 482     | 51                            | 24,793                    | 482                    | 414               |
| SKCM         | Melanoma                | 360     | 57                            | 20,369                    | 360                    | 272               |
| STAD         | Stomach carcinoma       | 364     | 63                            | 22,797                    | 364                    | 169               |
| THCA         | Thyroid carcinoma       | 503     | 62                            | 31,086                    | 503                    | 398               |
| UCS          | Uterine carcinosarcoma  | 49      | 48                            | 2,334                     | 49                     | 49                |
| Σ            |                       | 7295    | 55                            | 399,409                   | 7295                   | 4698              |

*Fragments (read pairs) mapped to GENCODE annotated genes.
We defined a set of 68 key driver genes commonly mutated in cancer, all established cancer genes with significant recurrence and in the present material (Supplementary Table 2, Methods). We next used non-parametric statistics (two-sided Wilcoxon rank sum test) to test for associations between mutations in these genes (exome data) and changes in mRNA levels, individually in each cancer type. As a first step, we applied mutations in these genes (exome data) and changes in mRNA levels, individually in each cancer type. As a first step, we applied 

\[ Z_{\text{score}} = \frac{\text{Estimate}}{\text{Standard Error}} \]

Notably, all NFE2L2 associations were positive, compatible with NFE2L2 being a transcriptional activator. Importantly, by comparison with available microarray data from A549 lung cancer cells treated with siRNA against NFE2L2 (ref. 44) we found that most NFE2L2-associated mRNAs detected in ≥2 cancers, and all detected in ≥4 cancers, were repressed 24 h post transfection (Fig. 3d).

By comparison with the molecular signatures database (MSigDB), we found that gene sets consistently associated with specific driver genes in our analysis were in several cases enriched for relevant pathways or direct responders identified by chemical or molecular perturbation (Supplementary Data 3). These included, for EGFR, genes repressed after treatment with the EGFR inhibitor CL-387785 (ref. 46) \( (q = 1.4e-2); \) NFE2L2, genes repressed after siRNA silencing of NFE2L2 (ref. 47) \( (q = 1.9e-19); \) RB1, genes induced on shRNA silencing of RB1 (ref. 48) \( (3.5e-11); \) KRAS, genes induced by NRAS overexpression \( (q = 1e-2), \) as well as various ERK/MAPK-related sets; TP53, genes responding differentially to DNA damage in TP53 null or wild type cells \( (q = 9.4e-7). \)

We noted that several canonical direct targets were among the detected genes. These included downstream effector and positive target of p53 CDKN1A (p21) (ref. 41), which showed a strong negative association with mutations in TP53, which are typically inactivating, in multiple cancers (Fig. 3b). Similarly, among genes consistently associated with activating mutations in NFE2L2, a master regulator of the oxidative stress response, were several well-known direct targets and downstream effectors including GCLC (Fig. 3c) and GCLM. Notably, all NFE2L2 associations were positive, compatible with NFE2L2 being a transcriptional activator.

Interestingly, by comparison with available microarray data from A549 lung cancer cells treated with siRNA against NFE2L2 (ref. 44) we found that most NFE2L2-associated mRNAs detected in ≥2 cancers, and all detected in ≥4 cancers, were repressed 24 h post transfection (Fig. 3d).

By comparison with the molecular signatures database (MSigDB), we found that gene sets consistently associated with specific driver genes in our analysis were in several cases enriched for relevant pathways or direct responders identified by chemical or molecular perturbation (Supplementary Data 3). These included, for EGFR, genes repressed after treatment with the EGFR inhibitor CL-387785 (ref. 46) \( (q = 1.4e-2); \) NFE2L2, genes repressed after siRNA silencing of NFE2L2 (ref. 47) \( (q = 1.9e-19); \) RB1, genes induced on shRNA silencing of RB1 (ref. 48) \( (3.5e-11); \) KRAS, genes induced by NRAS overexpression \( (q = 1e-2), \) as well as various ERK/MAPK-related sets; TP53, genes responding differentially to DNA damage in TP53 null or wild type cells \( (q = 9.4e-7). \)

We first noted that several canonical direct targets were among the detected genes. These included downstream effector and positive target of p53 CDKN1A (p21) (ref. 41), which showed a strong negative association with mutations in TP53, which are typically inactivating, in multiple cancers (Fig. 3b). Similarly, among genes consistently associated with activating mutations in NFE2L2, a master regulator of the oxidative stress response, were several well-known direct targets and downstream effectors including GCLC (Fig. 3c) and GCLM. Notably, all NFE2L2 associations were positive, compatible with NFE2L2 being a transcriptional activator.

Interestingly, by comparison with available microarray data from A549 lung cancer cells treated with siRNA against NFE2L2 (ref. 44) we found that most NFE2L2-associated mRNAs detected in ≥2 cancers, and all detected in ≥4 cancers, were repressed 24 h post transfection (Fig. 3d).

By comparison with the molecular signatures database (MSigDB), we found that gene sets consistently associated with specific driver genes in our analysis were in several cases enriched for relevant pathways or direct responders identified by chemical or molecular perturbation (Supplementary Data 3). These included, for EGFR, genes repressed after treatment with the EGFR inhibitor CL-387785 (ref. 46) \( (q = 1.4e-2); \) NFE2L2, genes repressed after siRNA silencing of NFE2L2 (ref. 47) \( (q = 1.9e-19); \) RB1, genes induced on shRNA silencing of RB1 (ref. 48) \( (3.5e-11); \) KRAS, genes induced by NRAS overexpression \( (q = 1e-2), \) as well as various ERK/MAPK-related sets; TP53, genes responding differentially to DNA damage in TP53 null or wild type cells \( (q = 9.4e-7). \)

We first noted that several canonical direct targets were among the detected genes. These included downstream effector and positive target of p53 CDKN1A (p21) (ref. 41), which showed a strong negative association with mutations in TP53, which are typically inactivating, in multiple cancers (Fig. 3b). Similarly, among genes consistently associated with activating mutations in NFE2L2, a master regulator of the oxidative stress response, were several well-known direct targets and downstream effectors including GCLC (Fig. 3c) and GCLM. Notably, all NFE2L2 associations were positive, compatible with NFE2L2 being a transcriptional activator.

Interestingly, by comparison with available microarray data from A549 lung cancer cells treated with siRNA against NFE2L2 (ref. 44) we found that most NFE2L2-associated mRNAs detected in ≥2 cancers, and all detected in ≥4 cancers, were repressed 24 h post transfection (Fig. 3d).

By comparison with the molecular signatures database (MSigDB), we found that gene sets consistently associated with specific driver genes in our analysis were in several cases enriched for relevant pathways or direct responders identified by chemical or molecular perturbation (Supplementary Data 3). These included, for EGFR, genes repressed after treatment with the EGFR inhibitor CL-387785 (ref. 46) \( (q = 1.4e-2); \) NFE2L2, genes repressed after siRNA silencing of NFE2L2 (ref. 47) \( (q = 1.9e-19); \) RB1, genes induced on shRNA silencing of RB1 (ref. 48) \( (3.5e-11); \) KRAS, genes induced by NRAS overexpression \( (q = 1e-2), \) as well as various ERK/MAPK-related sets; TP53, genes responding differentially to DNA damage in TP53 null or wild type cells \( (q = 9.4e-7). \)

We first noted that several canonical direct targets were among the detected genes. These included downstream effector and positive target of p53 CDKN1A (p21) (ref. 41), which showed a strong negative association with mutations in TP53, which are typically inactivating, in multiple cancers (Fig. 3b). Similarly, among genes consistently associated with activating mutations in NFE2L2, a master regulator of the oxidative stress response, were several well-known direct targets and downstream effectors including GCLC (Fig. 3c) and GCLM. Notably, all NFE2L2 associations were positive, compatible with NFE2L2 being a transcriptional activator.

Interestingly, by comparison with available microarray data from A549 lung cancer cells treated with siRNA against NFE2L2 (ref. 44) we found that most NFE2L2-associated mRNAs detected in ≥2 cancers, and all detected in ≥4 cancers, were repressed 24 h post transfection (Fig. 3d).

By comparison with the molecular signatures database (MSigDB), we found that gene sets consistently associated with specific driver genes in our analysis were in several cases enriched for relevant pathways or direct responders identified by chemical or molecular perturbation (Supplementary Data 3). These included, for EGFR, genes repressed after treatment with the EGFR inhibitor CL-387785 (ref. 46) \( (q = 1.4e-2); \) NFE2L2, genes repressed after siRNA silencing of NFE2L2 (ref. 47) \( (q = 1.9e-19); \) RB1, genes induced on shRNA silencing of RB1 (ref. 48) \( (3.5e-11); \) KRAS, genes induced by NRAS overexpression \( (q = 1e-2), \) as well as various ERK/MAPK-related sets; TP53, genes responding differentially to DNA damage in TP53 null or wild type cells \( (q = 9.4e-7). \)
Figure 3 | Evaluation of methodology to associate driver mutations with transcriptional changes. (a) The feasibility of the approach was explored using coding mRNA data before application to LncRNAs. All coding genes were tested for associations with 68 mutational events, separately in each cancer type based on a subset of 4698 samples with available mutation data. The number of associated genes are shown for each event and cancer type (red shading) using a relatively inclusive threshold (P < 0.001 and absolute log_2 expression ratio > 1, Wilcoxon rank sum test, but only associations replicated in more than one tumour type were considered further (right part, blue shading). 1,121 such consistent associations (right box with arrow) were uncovered for 30 mutational events shown here, compared with data in two randomized cancers. The cancers with the highest mutation frequencies are indicated to the left for each event (descending order). (b) Reduced expression of CDKN1A/p21 in TP53 mutated compared with wild type tumours (cancers with P < 0.005 are shown, Wilcoxon rank sum test; colour codes from a). *** P < 0.001; ** P < 0.01; * P < 0.05. (c) Increased expression of GCLC in NFE2L2 (NRF2) mutated compared with wild type tumours. (d) Validation of coding mRNAs from a found to be associated with mutations in NFE2L2 in at least two cancers (n = 65) or four cancers (n = 10), by comparison with expression changes observed upon siRNA silencing of NFE2L2 in A549 cells 24 h post transfection55 (y-axis). The x-axis shows expression ratios in NFE2L2 mutated compared with wild type tumours for cancer types with association P-value less than 0.001 (vertical bars indicate, for each gene, the mean log_2 ratio across the relevant cancer types, while coloured dots show the individual cancers). Bars indicate the mean and error bars indicate s.e.m.

appears that by considering signals replicated in multiple cancers we enrich for transcripts that are canonical downstream responders or direct targets.

LncRNAs are altered in relation to key mutational events. Having established our methodology on the coding transcriptome, we next used the same principle to uncover associations between key mutational driver events and altered LncRNA expression (complete unfiltered result in Supplementary Data 4). We focus here on a subset of LncRNAs showing clear and consistent associations in multiple cancers, although additional relevant effects might be present at lower stringencies. Our analysis identified 189 consistent associations (replicated in ≥2 cancers as established above for mRNAs) involving 21 mutational events and 169 unique LncRNAs (Fig. 4a; Supplementary Data 5), while none were found using randomized data (Supplementary Fig. 4).

LncRNAs may be co-expressed with neighbouring or overlapping coding genes3, suggesting that associations may arise secondary to mRNA expression changes. Hence, for each of the 189 associations we compared its strength (lowest P-value) with wild type tumours for cancer types with association P-value less than 0.001 (vertical bars indicate, for each gene, the mean log_2 ratio across the relevant cancer types, while coloured dots show the individual cancers). Bars indicate the mean and error bars indicate s.e.m.
Figure 4 | Associations between driver mutations and lncRNA expression levels. (a) 13,307 lncRNAs were tested for associations with key mutational events, as described in Fig. 3 for coding genes. Only associations replicated in more than one cancer type were further considered. 189 such consistent associations (right boxes with arrows) were found for 21 mutational events shown here, compared with 0 in randomized data. (b) Dependencies between associated lncRNAs and proximal coding genes (closest neighbours on either side as well as any coding gene within 100 kb). Dots represent the 189 lncRNAs identified in a, with the x-axis showing the association P value (lowest amongst all cancer types, Wilcoxon rank sum test) and the y-axis indicating the P value for the best proximal coding gene in the same cancer type. Dot colour indicates the cancer type, while dot size shows the expression correlation (Pearson’s r) between the lncRNA and the relevant coding gene in this cancer type. LncRNAs below the horizontal dotted line lack strongly associated proximal coding genes (P < 0.01). (c) Expression of the β-catenin target NKD1 and its intronic antisense transcript RP11-401P9.6 in CTNNB1 mutated compared with wild type tumours (cancers with P < 0.05 are shown, Wilcoxon rank sum test; colour codes from a). The NKD1 genomic context is shown, with annotated lncRNAs in blue. *P < 0.05, **P < 0.001, ***P < 1e−4. (d) Reduced expression of LINC01021 and RP3-510D11.2 in TP53 mutated compared with wild type tumours. (e) p53-dependent expression of RP11-115D19.1 in TCGA tumours (left) as well as in HCT116 colon carcinoma cells (TP53 +/+ or −/−) treated with the p53-activating agent Nutlin-3 (ref. 55) (right; GEO accession GSE53966). The genomic location of the primary expressed isoform ENST00000513572.1 is indicated. Bars indicate the mean and error bars indicate s.e.m.

(Supplementary Data 5). We found that the genic lncRNAs, which constituted roughly half (98/189) of the associations (Fig. 4a), were typically hosted by coding genes that likewise were strongly associated with the same driver events (Fig. 4b). Expectedly, these lncRNAs-mRNA pairs were normally co-expressed (indicated by dot size in Fig. 4b). For example, two out of three genic lncRNAs associated here with mutations in CTNNB1 (β-catenin) co-localize with NKD1, an established transcripational target of β-catenin complexes51 and a marker of aberrant Wnt/β-catenin signalling22 (Fig. 4c shows RP11-401P9.6, antisense intronic to NKD1). The intergenic associated lncRNAs (91/189) were generally more independent and showed less
co-expression with nearby coding genes (Fig. 4b). Interestingly, co-expressed coding neighbours would in many cases show weaker associations, compatible with lncRNAs being primary targets rather than the opposite.

**Known and novel TP53-dependent lncRNAs.** We note that, out of three intergenic lncRNAs highlighted in a recent study as being the strongest lncRNA responders to DNA damage and direct targets of p53 (ref. 14), two were among the TP53-associated lncRNAs identified in our screen, in both cases due to consistent reduction in TP53 mutated tumours (LINC01021, also known as LOC643401, and RP3-510D11.2; Fig. 4d). The third, LINCO0086 (also known as LED), as well as the recently identified p53 effector MLKL-1-ASI (also known as PINT)15, showed similar patterns but did not meet our stringent criteria (Supplementary Fig. 5). LINC01021 was recently confirmed as a p53 target and a negative regulator of proliferation in SW480 colorectal cancer cells23. As some TP53 mutations can have a gain-of-function effect in addition to disrupting normal TP53 action, we additionally performed our association screen using this subset. This identified one intergenic lncRNAs, RP16-46F2.2, which was activated in relation to TP53 gain-of-function mutations independently from its coding neighbours (Supplementary Fig. 6).

By re-analysis of nuclear run-on data (GRO-seq) from HCT116 colon cancer cells treated with the p53-activating agent Nutlin-3 (ref. 55), we found that lncRNAs negatively associated with TP53 mutations in our screen typically showed reduced transcription in TP53 null compared with wild type cells (Supplementary Fig. 7), supporting the overall validity of the results. The strongest reduction (139-fold) was seen for RP11-115D19.1, which was 2.9-fold transcriptionally induced 1 h after Nutlin-3 treatment in wild type cells while being almost completely blocked in TP53 null cells (Fig. 4e). Prominent reduction in TP53 mutated compared with wild type tumours was seen in multiple cancers for this lncRNA, with the strongest association in the BRCA luminal B subtype (P = 4.1e-21, Wilcoxon rank sum test; Fig. 4e). Confirmatory analyses using additional available data15,56 showed that RP11-115D19.1 was induced in human fibroblasts treated with the DNA damaging agent doxorubicin and reduced in response to TP53 siRNA in breast cancer cells exposed to ionizing radiation, although levels were low in these cells (Supplementary Fig. 8). While classified here as genic due to its antisense localization near the SNCA 3' UTR (ref. 57), we found that its primary expressed isoform resided 30 kb downstream of SNCA and 370 kb upstream of GPRIN3 (P = 0.59 and 0.29, respectively, in BRCA luminal B).

**Activation of lncRNAs downstream of NFE2L2.** Having obtained an overview, we next focused on NFE2L2 (encoding the transcription factor NRF2), a master activator of genes involved in the cellular antioxidant response. Activating mutations in NFE2L2 can facilitate tumour progression and protect cells from chemotherapy, and occur frequently in several cancers including squamous cell lung carcinomas (LUSC)42. Although knowledge about lncRNAs in the antioxidant response is limited, a recent study showed that NFE2L2 mediates activation of SCAL1 IncRNA in response to cigarette smoke58, and SCAL1 and other IncRNAs were subsequently shown to correlate with NFE2L2 mutations in lung cancer data from TCGA (ref. 21). Additionally, NFE2L2 inhibits the pluripotency IncRNAs ROR in embryonic stem cells59.

Fifteen lncRNAs (10 genic and 5 intergenic) were identified here as consistently induced in tumours with NFE2L2 gain-of-function mutations (Fig. 4a; Supplementary Data 5). To confirm NFE2L2 responsiveness of these lncRNAs, we silenced NFE2L2 in A549 lung cancer cells that have constitutively high levels of the NRF2 protein due to a loss-of-function mutation in KEAP1, the negative regulator of NRF2 (ref. 60). We performed high coverage total RNA sequencing (> 50 million reads per library) to compare three samples transfected with NFE2L2 siRNAs to four control transfections (Supplementary Data 6). This revealed a strong transcriptional response (56 IncRNAs and 1069 coding genes at q < 0.01 using DESeq2 (ref. 61)) dominated by genes whose expression was reduced, notably with 5 of the 10 top repressed lncRNAs being predicted targets from our screen (Fig. 5a, left panel). Furthermore, out of 11 predicted NFE2L2 activated lncRNAs that were also expressed in A549 cells, 8 were repressed more than two-fold (Fig. 5a, right panel). Three repressed IncRNAs were further evaluated with RT-qPCR: LINCO0942, RP11-284F21.7 and RP11-345L23.1. In all cases, this confirmed responsiveness to NFE2L2 siRNA silencing in A549, as well as one additional lung cancer cell line, H838, which similarly harbours a loss-of-function mutation in KEAP1 (Fig. 5b).

These results demonstrate a high degree of accuracy in identifying NFE2L2-responsive lncRNAs from the tumour data.

We next focused on the three NFE2L2-dependent lncRNAs confirmed by RT-qPCR. These showed increased expression in relation to NFE2L2 activating mutations primarily in bladder (BLCA), cervical (CESC), head and neck (HNSC) and LUSC carcinomas (Fig. 5c). LINCO0942 resides 4.5 kb downstream of ERC1 (Fig. 5c), but notably was not co-expressed with ERC1 or other nearby genes (Supplementary Data 5). RP11-284F21.9/10.7 is a cluster of three annotated lncRNAs antisense to BCAN (Fig. 5c), all showing positive NFE2L2 associations most notably in LUSC, that based on RNA-seq data appears to be part of the same transcript (RP11-284F21.7 was preferred for RT-qPCR since it is spliced). BCAN is a proteoglycan linked to invasiveness in glioma62, which lacked expression in A549/H838 but interestingly showed an inverse association with NFE2L2 mutations in LUSC (P = 4.5e-4, Wilcoxon rank sum test). RP11-345L23.1 (also called LINCO1564) is intergenic but 11 kb upstream of GCC1 (83 kb from the closest RefSeq isoform) in the antioxidant orientation (Fig. 5c).

Genes downstream of NFE2L2 should in theory be responsive also to mutations in KEAP1, which exert their oncogenic activity by stabilizing the NRF2 protein42. KEAP1 was not considered in our initial screen due to absence in the Cancer Gene Census38 (Methods), providing an opportunity to further test NFE2L2 responsiveness. We found that all three IncRNAs were induced in KEAP1-mutated compared with wild type tumours in several cancers, despite NFE2L2 and KEAP1 mutations being near-mutually exclusive (co-occurrence in three tumours) (Fig. 5d). This further reinforces NFE2L2-dependent expression of these IncRNAs.

**LINCO0942 is a direct target of NFE2L2.** One of the validated NFE2L2-dependent lncRNAs, LINCO0942, was selected for further characterization. BLAST search against known proteins in UniProt and analysis using the coding potential calculator tool63 supported its annotated non-coding status. An analysis of 1019 samples in the Cancer Cell Line Encyclopaedia64 using cBioPortal65 showed that many canonical NFE2L2 targets were found among top transcripts co-expressed with LINCO0942, including among others GCLM at rank 6 and TXNRD1 at rank 14, further establishing NFE2L2-dependent expression of LINCO0942. The annotated transcription initiation site was supported by RNA polymerase II chromatin immunoprecipitation (ChIP) data from ENCODE, and an upstream regulatory region was further indicated by DNase I hypersensitivity and FAIRE signals (Fig. 6a). Notably, we found that these signals coincided with an element that closely matches the consensus sequence of the NRF2-binding anti-oxidant
Figure 5 | Validation of lncRNAs induced in relation to activating mutations in NFE2L2. (a) Total RNA sequencing of A549 lung cancer cells treated with NFE2L2 siRNAs (n = 3 transfections using three unique sequences) or control siRNAs (n = 4 transfections using two unique sequences) during 48 h. Left panel: volcano plot showing 56 lncRNAs (black dots) altered at q < 0.01 (DESeq2), with aggregation of predicted NFE2L2-responsive lncRNAs (red circles) among top repressed transcripts. Right panel: predicted NFE2L2-responsive lncRNAs are repressed 48 h post NFE2L2 inhibition. The x-axis shows expression ratios in NFE2L2 mutated compared with wild type tumours for cancer types with association P < 0.001 (Wilcoxon rank sum test). Vertical bars indicate, for each gene, the mean log2 ratio across the relevant cancer types and coloured dots show the individual cancers using colours from Fig. 1b. The plot includes 11/15 predicted lncRNAs detectable in A549 cells (>10 reads in one sample). (b) Validation of select transcripts by RT-qPCR, comparing cells treated with NFE2L2 siRNAs or control siRNA as described in a, with additional results from H838 lung cancer cells. Values were normalized to ACTB, and are shown relative to the controls. P-values were determined using Student’s t-test. (c) Induction of LINCO0942, RPI1-284F21.7 and RPI1-345L23.1 in NFE2L2 mutated compared with wild type tumours (cancers with P < 0.05 are shown, Wilcoxon rank sum test). Genomic contexts are shown; blue, lncRNAs; green, coding genes. *P < 0.05; **P < 0.01; ***P < 1e-4. (d) Induction of LINCO0942, RPI1-284F21.7 and RPI1-345L23.1 in KEAP1 mutated compared with wild type tumours. Error bars indicate s.e.m.

response element (ARE)\(^{65}\), at ~120 bp upstream of the annotated transcription start site (Fig. 6a).

To assess binding of NRF2 to the LINCO0942 promoter, we performed ChIP analysis in A549 and H838 cells using two different NRF2 antibodies or control IgG. We evaluated the predicted ARE site as well as two positive and three negative control regions using qPCR. This revealed a positive enrichment relative to IgG with both antibodies in both A549 and H838 cells (P < 0.02 for all comparisons, Student’s t-test), with a stronger signal in H838 (16-fold enrichment relative IgG), which contain basally higher levels of LINCO0942 (Fig. 6b). These data confirm direct binding of NRF2 to an ARE element at the LINCO0942 promoter.

To further assess the ability of NRF2 to regulate the expression of LINCO0942 by binding to the ARE, we cloned a 1,000 bp region encompassing the LINCO0942 transcription start and the binding site into a promoter-less luciferase reporter vector. Luciferase activity was detected 48 h post transfection in both A549 and H838 cells, furthering the support of the promoter region (Fig. 6c). Co-transfections with siRNAs targeting NFE2L2 reduced the reporter activity in both cell lines (P < 1.8e-3, Student’s t-test), and point mutations in the ARE promoter further reduced the signal (P < 5.2e-5) while largely abolishing the response to NFE2L2 silencing (Fig. 6c). The luciferase reporter assays in conjunction with ChIP data demonstrate that NRF2 can activate LINCO0942 transcriptionally by directly binding to its promoter. This provides mechanistic insight into our finding that human tumours with NFE2L2 gain-of-function mutations display increased expression of LINCO0942.

To assess the role of LINCO0942 in cells with NRF2 activation, we inhibited it in A549 cells using locked nucleic acid (LNA) antisense oligonucleotides (ASOs). Of three different ASOs targeting LINCO0942, two were highly efficient (83% and 89% reduction, respectively; P < 6.9e-5 using Student’s t-test), while one was less potent (27% reduction; P = 0.015) (Fig. 7a). We next assayed the expression of three canonical coding components of the antioxidant programme orchestrated by NFE2L2 to suppress reactive oxygen species (ROS) accumulation in cells, all of which were detected as NFE2L2 activated in the tumour data (Supplementary Data 2): GCLC, crucial for synthesis of the antioxidant glutathione; NQO1, which prevents production of ROS from quinones and TXNRD1, which catalyses reduction of the antioxidant thioredoxin from the oxidized to the reduced form\(^{67}\).

Notably, all three LINCO0942 ASOs caused a significant reduction in GCLC mRNA (P < 3.2e-3), importantly with the two more potent molecules causing stronger mRNA reductions (53% and 70%, respectively; Fig. 7b), as well as a decrease in GCLC protein levels (Fig. 7c). No consistent mRNA changes were seen...
Fig. 9). Consistent with a reduction in GCLC posttranscriptional mechanism (Supplementary Fig. 10). The effect was due to reduced transcription rather than a glutathione, was reduced (Fig. 7d; Supplementary Fig. 10). Increase in ROS (Fig. 7e). LINC00942 expression in the antioxidant response downstream of NFE2L2. We observed that the antioxidant glutathione might lead to accumulation of cellular ROS levels are similar to what has been observed during cancer. We first found that a naïve approach, based on association with the more potent ASOs. These effects on glutathione and ASOs led to a notable 2.6e-3), in particular in cells treated with the anticipated site. POL2, polymerase 2 ChIP (K562 cells, HudsonAlpha data); DNase, DNase I hypersensitivity (K562 cells, UNC). (b) ChIP analysis using two different NFE2L2 antibodies (Ab 1, Abcam EP1808Y; Ab 2, Diagenode C15410242) or IgG (n = 3 IPs for each antibody). A 132 bp segment centred on the predicted response element (a), as well as two positive control sites (in NQO1 and HMOX1) and three negative control regions were assayed using qPCR. Signals are shown relative to IgG (fold enrichment) and s.e.m. is indicated. (c) A 1 kb segment encompassing the transcription start and the NRF2 site, either wild type or mutated, was cloned into a promoter-free Gaussia luciferase (GLuc) reporter plasmid (pEZX-PG04). The plasmid was co-transfected with two different NFE2L2 siRNAs or two control siRNAs (n = 3 transfections each) in A549 and H838 cells. The signal is shown normalized to secreted alkaline phosphatase (SEAP), expressed as a control reporter from the same vector. P-values from Student’s t-test combining the six transfections for control or treatment siRNAs. Error bars indicate s.e.m.

for the other two genes or NFE2L2 itself (Fig. 7b). Similar reductions in GCLC mRNA were observed in H838 lung cancer cells upon treatment with LINC00942 ASOs (Supplementary Fig. 9). GCLC mRNA stability was not affected, showing that the effect was due to reduced transcription rather than a posttranscriptional mechanism (Supplementary Fig. 10). Consistent with a reduction in GCLC, its enzymatic product, glutathione, was reduced (Fig. 7d; P < 1.3e-3). Reduced levels of the antioxidant glutathione might lead to accumulation of cellular ROS. We observed that LINC00942 ASOs led to a notable increase in ROS (Fig. 7e; P < 2.6e-3), in particular in cells treated with the more potent ASOs. These effects on glutathione and ROS levels are similar to what has been observed during NFE2L2 silencing68. Collectively, our results suggest a role for LINC00942 in the antioxidant response downstream of NFE2L2.

Discussion

By systematically investigating patterns of altered lncRNA expression in relation to key mutational events, enabled by a wealth of molecular data provided by TCGA, we provide a comprehensive catalogue of candidate non-coding RNAs that may play a functional role as part of oncogenic programs in cancer. We first found that a naïve approach, based on association tests in individual cancer types, often produces an excessive number of mRNA and lncRNAs hits of questionable relevance, likely explained by strong expression subtypes that may differ with respect to driver mutation frequencies35. Instead, by focusing on effects that are replicated across cancer types, more informative results are obtained, often enriched for known downstream responder mRNAs. Applying this strategy to lncRNAs recapitulated known biology in relation to TP53 and also provided numerous new predictions, some of which were further highlighted. It should be noted that we focus here on strong consistent associations, using a stringent false discovery rate cut off, and that many additional relevant effects are likely identifiable by applying less strict filtering criteria to the full results (provided as Supplementary Data 4). These results should form a valuable complement to studies that directly manipulate factors of interest in cell culture systems, in that they describe effects clearly visible in human tumour materials.

TP53-dependent regulation of lncRNAs has been the focus of several studies12,14–16, but the role of lncRNAs in most other cancer-related pathways remains elusive. Here, we highlight lncRNAs consistently induced in tumours with activation of the transcription factor NFE2L2 (encoding NRF2), which is critical for the oxidative stress response. We demonstrate by RNA-seq that predicted NFE2L2-dependent lncRNAs typically show immediate responsiveness to NFE2L2 silencing in A549 cells, where NRF2 is constitutively active80. Although some detected associations may be due to indirect effects, these results suggest that many represent immediate downstream consequences. In particular, for LINC00942, we confirmed that NRF2 directly binds to an upstream regulatory element in its promoter to
induce its transcription. Additionally, transfection with ASOs complementary to LINC00942 suggested a functional role in the NFE2L2 antioxidant pathway by altering levels of GCLC, which is required for synthesis of the key antioxidant glutathione and a known target of NRF2. Compatible with a reduction in GCLC, which is required for glutathione synthesis. By using orthogonal methods such as CRISPR/Cas9 targeted mutagenesis and CRISPR/Cas9-based transcriptional repression/activation, future studies will be able to establish the function and mechanism of action for LINC00942 or other lncRNAs downstream of NFE2L2.

In summary, we have systematically investigated alterations in lncRNA expression in relation to key mutational driver events in human cancers. Although earlier studies have been able to tie lncRNAs to individual oncogenic pathways, we provide a broader catalogue that may serve as a reference and starting point for future experimental studies.

Methods

LncRNA annotation and RNA-seq-based quantification. Unaligned RNA-seq data were downloaded from the cglHub repository (detailed sample information is available in Supplementary Data 1). Libraries from non-embargoed tumours available on 28 April 2015, that met several additional criteria, were included. For consistency, we only considered libraries produced on Illumina HiSeq 2000/2500 machines and excluded all FFPE samples. Only primary tumours (TCGA type 1) were included, except for SKCM where metastasis samples (type 6) were instead considered. In cases where multiple libraries were available for a given tumour, we included the largest. Finally, only tumours with available copy number data were included. Sequences were aligned to the human hg19 assembly (excluding alternative haplotype regions) using HiSat (ref. 69) (options --known-splicesite-infile and --no-mixed --no-discordant --no-unal). Sequences were aligned to the human hg19 assembly (excluding alternative haplotype regions) using HiSat (ref. 69) (options --known-splicesite-infile and --no-mixed --no-discordant --no-unal).

Copy number analyses. Copy number amplitudes for GENCODE genes were determined from segmented copy-number data (Affymetrix SNP6 platform, minus othervariant) and an ‘intergenic’ set (n = 5,422) based on proximity to the nearest coding gene, here also including coding RefSeq genes in addition to GENCODE. LncRNAs and coding genes were quantified using HTSeq-count (option --intersection-strict and FPKM expression values were determined while normalizing each library using robust size factors as described). LncRNA expression profiles are available as Supplementary Data 7.
Inclusion criteria for mutational events. A set of 88 mutational events were defined, representing relatively frequently mutated and established cancer genes (Supplementary Table 1). Genes were required to be detected as significantly mutated in at least one cancer type at q < 0.05 by either MutSigCV (ref. 39) or SomaCast (ref. 40) and mutated at a frequency above 5% in at least one cancer type, in addition to being listed as known cancer genes in the Cancer Gene Census list of known cancer genes18.

Associations between mutational subtypes and RNA levels. RNA levels for IncRNAs and coding genes were evaluated for associations with driver mutations subtypes, separately in each cancer type. RNA samples were subselected based on similarity (Pearson’s r) to the mean expression profile of each of the PAM50 subtypes, determined from samples available from PAM50 classifications from TCGA. Exome data in bam format were obtained from the Broad Institute (latest release available on 24 February 2015). Genes were considered as mutated when having either in-frame indels or nonsynonymous single nucleotide substitutions in the coding mRNA or spliced alternative isoforms. The Wilcoxon rank sum test (as implemented in Matlab, Mathworks Inc.; approximate method) was used to assess differential expression in mutated compared with non-mutated samples, while requiring genes/IncRNAs to be detectable in at least 10% of samples in both groups (scripts available from the authors on request). A pseudo value of 0.1 FPKM was added to genes/lncRNAs to be detectable in at least 10% of samples in both groups (scripts available from the authors on request). A pseudo value of 0.1 FPKM was added to genes/lncRNAs to be detectable in at least 10% of samples in both groups.

 Luciferase assays. A 1,000 bp region encompassing the NRF2 binding site and the annotated transcription start of LINC00942 (–800 to +200 bp, chr12:6108, 855–1,609,854) was cloned into a Gaussia luciferase (GLuc) reporter vector (pEZ-PX-PG04, Genepcooia), which also contains a reference reporter gene, Secreted Alkaline Phosphatase (SEAP). Additionally, a construct containing a mutated version of the NRF2 binding site was generated (ATGACCTGCGCA > ACCGACTGCGCA). A549 and H838 cells were cultured in a 12-well plate 24 h prior to transfection, and at a confluence of approximately 80% cells were transfected with the wild type or mutant plasmid together with one of two different control siRNAs or two siRNAs targeting LINC0092. Each siRNA/plasmid combination was transfected in triplicates, using Lipofectamin 2,000 (Life technologies, cat. no. 11668027). The activities of GLuc, adjacent to LINC00942 promoter, and SEAP were quantified with the Secrete-Pair Dual Luminescence Assay Kit (Genepcooia, cat. no. SPDA-D010) 48 h after transfection.

 Data availability. The NFE2L2 siRNA RNA-seq data have been deposited in GEO using the accession number GSE75452 (http://www.ncbi.nlm.nih.gov/geo/query/ acc.cgi?acc=GSE75452). All remaining data are contained within the Article and Supplementary Information files or is available from the authors upon request.

 References

1. Gutmann, M. et al. Ab initio reconstruction of cell-type specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat. Biotechnol. 28, 503–510 (2010).

2. Carninci, P. et al. The transcriptional landscape of the mammalian genome. Science 309, 1559–1563 (2005).

3. Djebali, S. et al. Landscape of transcription in human cells. Nature 489, 101–108 (2012).

4. Plath, K. et al. Role of histone H3 lysine 27 methylation in X inactivation. Science 300, 131–135 (2003).

5. Gupta, R. A. et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464, 1071–1076 (2010).

6. Carrieri, C. et al. Long non-coding antisense RNA controls Uchl1 translation. Mol. Cell. 47, 648–655 (2012).

7. Tripipati, V. et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol. Cell. 39, 925–938 (2010).

8. 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).

9. Prensner, J. R. & Chinnaiyan, A. M. The emergence of lncRNAs in cancer biology. Cancer Discov. 1, 391–407 (2011).

10. Esteller, M. Non-coding RNAs in human disease. Nat. Rev. Genet. 12, 861–874 (2011).

11. Menard, S. et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464, 1071-1076 (2010).
Acknowledgements

The results published here are in whole or part based upon data generated by The Cancer Genome Atlas pilot project established by the NCI and NHGRI. Information about TCGA and the investigators and institutions who constitute the TCGA research network can be found at http://cancergenome.nih.gov. We are most grateful to the patients, investigators, clinicians, technical personnel and funding bodies who contributed to TCGA, thereby making this study possible. This work was supported by grants from the Knut and Alice Wallenberg Foundation (E.L.), the Swedish Foundation for Strategic Research (E.L.), the Swedish Medical Research Council (E.L. and V.I.S.), the Swedish Cancer Society (E.L.), the Åke Wiberg Foundation (E.L.), the Lars Erik Lundberg Foundation for Research and Education (E.L.) and the Assar Gabrielsson Foundation (V.I.S.). V.I.S. is a recipient of EMBO Long Term Fellowship ALTF 1451–2015 co-funded by the European Commission (LTCO-FUND2013, GA-2013-609409) with support from Marie Curie Actions. The computations were in part performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under project b2012108.

Author contributions

A.A., J.V.E. and E.L. performed bioinformatics analyses; A.A. and V.I.S. designed, performed and analysed experiments; S.X.S. performed experiments; E.L. wrote the paper with contributions from T.P., A.A. and V.I.S.; T.P. supervised and designed experiments; E.L. conceived the study.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Ashouri, A. et al. Pan-cancer transcriptomic analysis associates long non-coding RNAs with key mutational driver events. Nat. Commun. 7, 13197 doi: 10.1038/ncomms13197 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2016