


Comprehensive molecular profiling of

View metadata, citation and similar papers at core.ac.uk

brought to you by  CORE

provided by Carolina Digital Repository

The Cancer Genome Atlas Research Network*

Adenocarcinoma of the lung is the leading cause of cancer death worldwide. Here we report molecular profiling of 230 resected lung adenocarcinomas using messenger RNA, microRNA and DNA sequencing integrated with copy number, methylation and proteomic analyses. High rates of somatic mutation were seen (mean 8.9 mutations per megabase). Eighteen genes were statistically significantly mutated, including *RIT1* activating mutations and newly described loss-of-function *MGA* mutations which are mutually exclusive with focal *MYC* amplification. *EGFR* mutations were more frequent in female patients, whereas mutations in *RBM10* were more common in males. Aberrations in *NF1*, *MET*, *ERBB2* and *RIT1* occurred in 13% of cases and were enriched in samples otherwise lacking an activated oncogene, suggesting a driver role for these events in certain tumours. DNA and mRNA sequence from the same tumour highlighted splicing alterations driven by somatic genomic changes, including exon 14 skipping in *MET* mRNA in 4% of cases. MAPK and PI(3)K pathway activity, when measured at the protein level, was explained by known mutations in only a fraction of cases, suggesting additional, unexplained mechanisms of pathway activation. These data establish a foundation for classification and further investigations of lung adenocarcinoma molecular pathogenesis.

Lung cancer is the most common cause of global cancer-related mortality, leading to over a million deaths each year and adenocarcinoma is its most common histological type. Smoking is the major cause of lung adenocarcinoma but, as smoking rates decrease, proportionally more cases occur in never-smokers (defined as less than 100 cigarettes in a lifetime). Recently, molecularly targeted therapies have dramatically improved treatment for patients whose tumours harbour somatically activated oncogenes such as mutant *EGFR*¹ or translocated *ALK*, *RET*, or *ROS1* (refs 2–4). Mutant *BRAF* and *ERBB2* (ref. 5) are also investigational targets. However, most lung adenocarcinomas either lack an identifiable driver oncogene, or harbour mutations in *KRAS* and are therefore still treated with conventional chemotherapy. Tumour suppressor gene abnormalities, such as those in *TP53* (ref. 6), *STK11* (ref. 7), *CDKN2A*⁸, *KEAP1* (ref. 9), and *SMARCA4* (ref. 10) are also common but are not currently clinically actionable. Finally, lung adenocarcinoma shows high rates of somatic mutation and genomic rearrangement, challenging identification of all but the most frequent driver gene alterations because of a large burden

of passenger events per tumour genome^{11–13}. Our efforts focused on comprehensive, multiplatform analysis of lung adenocarcinoma, with attention towards pathobiology and clinically actionable events.

Clinical samples and histopathologic data

We analysed tumour and matched normal material from 230 previously untreated lung adenocarcinoma patients who provided informed consent (Supplementary Table 1). All major histologic types of lung adenocarcinoma were represented: 5% lepidic, 33% acinar, 9% papillary, 14% micropapillary, 25% solid, 4% invasive mucinous, 0.4% colloid and 8% unclassifiable adenocarcinoma (Supplementary Fig. 1)¹⁴. Median follow-up was 19 months, and 163 patients were alive at the time of last follow-up. Eighty-one percent of patients reported past or present smoking. Supplementary Table 2 summarizes demographics. DNA, RNA and protein were extracted from specimens and quality-control assessments were performed as described previously¹⁵. Supplementary Table 3 summarizes molecular estimates of tumour cellularity¹⁶.

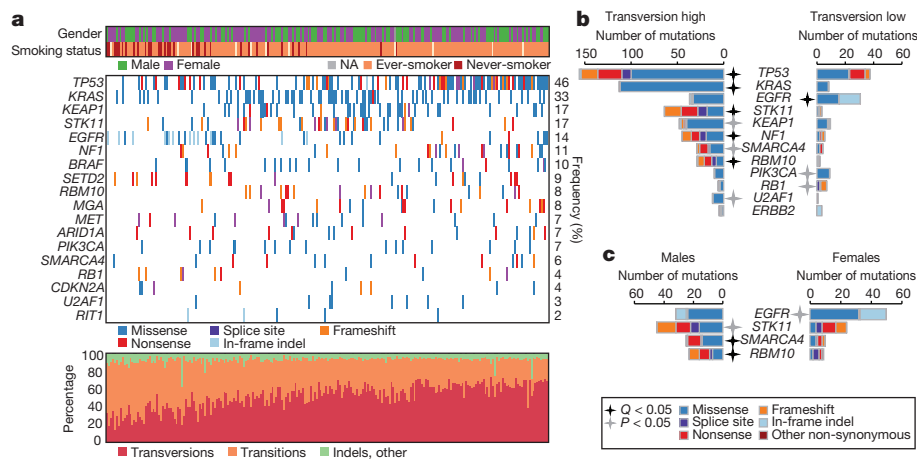


Figure 1 | Somatic mutations in lung adenocarcinoma. **a**, Co-mutation plot from whole exome sequencing of 230 lung adenocarcinomas. Data from TCGA samples were combined with previously published data¹² for statistical analysis. Co-mutation plot for all samples used in the statistical analysis ($n = 412$) can be found in Supplementary Fig. 2. Significant genes with a corrected P value less than 0.025 were identified using the MutSig2CV algorithm and are ranked in order of decreasing prevalence. **b**, **c**, The differential patterns of mutation between samples classified as transversion high and transversion low samples (**b**) or male and female patients (**c**) are shown for all samples used in the statistical analysis ($n = 412$). Stars indicate statistical significance using the Fisher's exact test (black stars: $q < 0.05$, grey stars: $P < 0.05$) and are adjacent to the sample set with the higher percentage of mutated samples.

Somatically acquired DNA alterations

We performed whole-exome sequencing (WES) on tumour and germline DNA, with a mean coverage of 97.6× and 95.8×, respectively, as performed previously¹⁷. The mean somatic mutation rate across the TCGA cohort was 8.87 mutations per megabase (Mb) of DNA (range: 0.5–48, median: 5.78). The non-synonymous mutation rate was 6.86 per Mb. MutSig2CV¹⁸ identified significantly mutated genes among our 230 cases along with 182 similarly-sequenced, previously reported lung adenocarcinomas¹². Analysis of these 412 tumour/normal pairs highlighted 18 statistically significant mutated genes (Fig. 1a shows co-mutation plot of TCGA samples ($n = 230$), Supplementary Fig. 2 shows co-mutation plot of all samples used in the statistical analysis ($n = 412$) and Supplementary Table 4 contains complete MutSig2CV results, which also appear on the TCGA Data Portal along with many associated data files (https://tcga-data.nci.nih.gov/docs/publications/luad_2014/). *TP53* was commonly mutated (46%). Mutations in *KRAS* (33%) were mutually exclusive with those in *EGFR* (14%). *BRAF* was also commonly mutated (10%), as were *PIK3CA* (7%), *MET* (7%) and the small GTPase gene, *RIT1* (2%). Mutations in tumour suppressor genes including *STK11* (17%), *KEAP1* (17%), *NF1* (11%), *RBI* (4%) and *CDKN2A* (4%) were observed. Mutations in chromatin modifying genes *SETD2* (9%), *ARID1A* (7%) and *SMARCA4* (6%) and the RNA splicing genes *RBM10* (8%) and *U2AF1* (3%) were also common. Recurrent mutations in the *MGA* gene (which encodes a Max-interacting protein on the *MYC* pathway¹⁹) occurred in 8% of samples. Loss-of-function (frameshift and nonsense) mutations in *MGA* were mutually exclusive with focal *MYC* amplification (Fisher's exact test $P = 0.04$), suggesting a hitherto unappreciated potential mechanism of *MYC* pathway activation. Coding single nucleotide variants and indel variants were verified by resequencing at a rate of 99% and 100%, respectively (Supplementary Fig. 3a, Supplementary Table 5). Tumour purity was not associated with the presence of false negatives identified in the validation data ($P = 0.31$; Supplementary Fig. 3b).

Past or present smoking associated with cytosine to adenine (C > A) nucleotide transversions as previously described both in individual genes and genome-wide^{12,13}. C > A nucleotide transversion fraction showed two peaks; this fraction correlated with total mutation count ($R^2 = 0.30$) and inversely correlated with cytosine to thymine (C > T) transition frequency ($R^2 = 0.75$) (Supplementary Fig. 4). We classified each sample (Supplementary Methods) into one of two groups named transversion-high (TH, $n = 269$), and transversion-low (TL, $n = 144$). The transversion-high group was strongly associated with past or present smoking ($P < 2.2 \times 10^{-16}$), consistent with previous reports¹³. The transversion-high and transversion-low patient cohorts harboured different gene mutations. Whereas *KRAS* mutations were significantly enriched in the transversion-high cohort ($P = 2.1 \times 10^{-13}$), *EGFR* mutations were significantly enriched in the transversion-low group ($P = 3.3 \times 10^{-6}$). *PIK3CA* and *RBI* mutations were likewise enriched in transversion-low tumours ($P < 0.05$). Additionally, the transversion-low tumours were specifically enriched for in-frame insertions in *EGFR* and *ERBB2* (ref. 5) and for frameshift indels in *RBI* (Fig. 1b). *RBI* is commonly mutated in small-cell lung carcinoma (SCLC). We found *RBI* mutations in transversion-low adenocarcinomas were enriched for frameshift indels versus single nucleotide substitutions compared to SCLC ($P < 0.05$)^{20,21} suggesting a mutational mechanism in transversion-low adenocarcinoma that is probably distinct from smoking in SCLC.

Gender is correlated with mutation patterns in lung adenocarcinoma²². Only a fraction of significantly mutated genes from the complete set reported in this study (Fig. 1a) were enriched in men or women (Fig. 1c). *EGFR* mutations were enriched in tumours from the female cohort ($P = 0.03$) whereas loss-of-function mutations within *RBM10*, an RNA-binding protein located on the X chromosome²³ were enriched in tumours from men ($P = 0.002$). When examining the transversion-high group, 16 out of 21 *RBM10* mutations were observed in males ($P = 0.003$, Fisher's exact test).

Somatic copy number alterations were very similar to those previously reported for lung adenocarcinoma²⁴ (Supplementary Fig. 5, Supplementary Table 6). Significant amplifications included *NKX2-1*, *TERT*,

MDM2, *KRAS*, *EGFR*, *MET*, *CCNE1*, *CCND1*, *TERC* and *MECOM* (Supplementary Table 6), as previously described²⁴, 8q24 near *MYC*, and a novel peak containing *CCND3* (Supplementary Table 6). The *CDKN2A* locus was the most significant deletion (Supplementary Table 6). Supplementary Table 7 summarizes molecular and clinical characteristics by sample. Low-pass whole-genome sequencing on a subset ($n = 93$) of the samples revealed an average of 36 gene-gene and gene-inter-gene

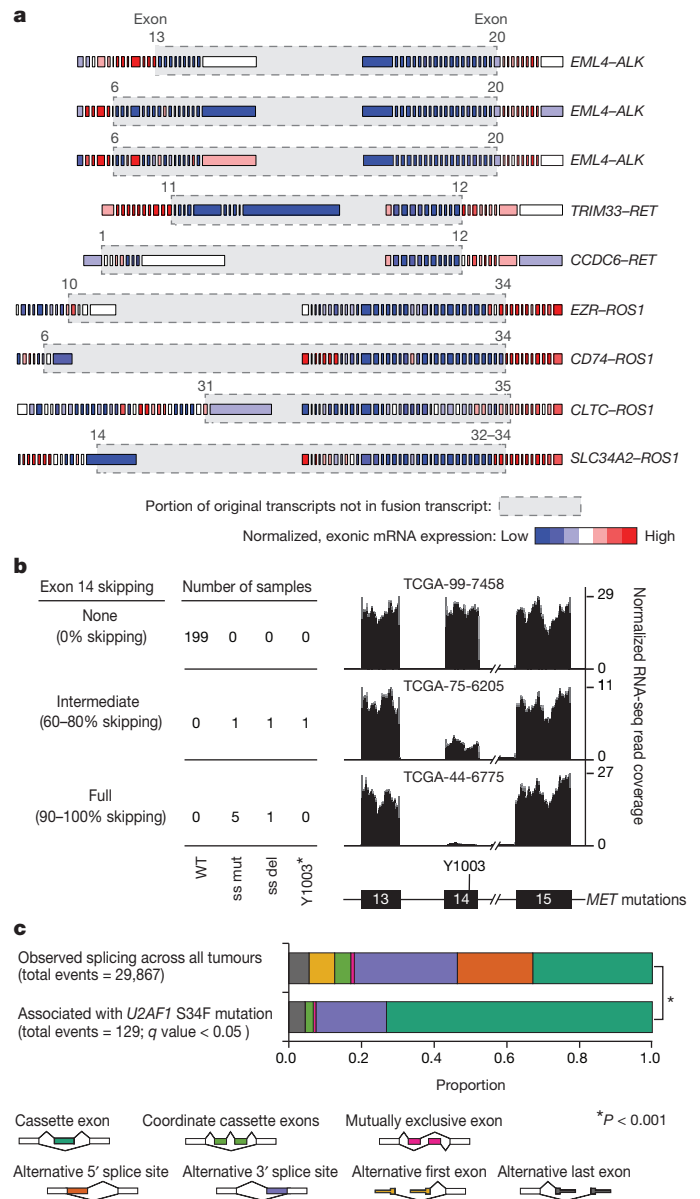


Figure 2 | Aberrant RNA transcripts in lung adenocarcinoma associated with somatic DNA translocation or mutation. a, Normalized exon level RNA expression across fusion gene partners. Grey boxes around genes mark the regions that are removed as a consequence of the fusion. Junction points of the fusion events are also listed in Supplementary Table 9. Exon numbers refer to reference transcripts listed in Supplementary Table 9. **b**, *MET* exon 14 skipping observed in the presence of exon 14 splice site mutation (ss mut), splice site deletion (ss del) or a Y1003* mutation. A total of 22 samples had insufficient coverage around exon 14 for quantification. The percentage skipping is (total expression minus exon 14 expression)/total expression. **c**, Significant differences in the frequency of 129 alternative splicing events in mRNA from tumours with *U2AF1* S34F compared to *U2AF1* WT tumours (q value < 0.05). Consistent with the function of *U2AF1* in 3' splice site recognition, most splicing differences involved cassette exon and alternative 3' splice site events (chi-squared test, $P < 0.001$).

rearrangements per tumour. Chromothripsis²⁵ occurred in six of the 93 samples (6%) (Supplementary Fig. 6, Supplementary Table 8). Low-pass whole genome sequencing-detected rearrangements appear in Supplementary Table 9.

Description of aberrant RNA transcripts

Gene fusions, splice site mutations or mutations in genes encoding splicing factors promote or sustain the malignant phenotype by generating aberrant RNA transcripts. Combining DNA with mRNA sequencing enabled us to catalogue aberrant RNA transcripts and, in many cases, to identify the DNA-encoded mechanism for the aberration. Seventy-five per cent of somatic mutations identified by WES were present in the RNA transcriptome when the locus in question was expressed (minimum 5×) (Supplementary Fig. 7a) similar to prior analyses¹⁵. Previously identified fusions involving *ALK* (3/230 cases), *ROS1* (4/230) and *RET* (2/230) (Fig. 2a, Supplementary Table 10), all occurred in transversion-low tumours ($P = 1.85 \times 10^{-4}$, Fisher's exact test).

MET activation can occur by exon 14 skipping, which results in a stabilized protein²⁶. Ten tumours had somatic *MET* DNA alterations with *MET* exon 14 skipping in RNA. In nine of these samples, a 5' or 3' splice site mutation or deletion was identified²⁷. *MET* exon 14 skipping was also found in the setting of a *MET* Y1003* stop codon mutation (Fig. 2b, Supplementary Fig. 8a). The codon affected by the Y1003* mutation is predicted to disrupt multiple splicing enhancer sequences, but the mechanism of skipping remains unknown in this case.

S34F mutations in *U2AF1* have recently been reported in lung adenocarcinoma¹² but their contribution to oncogenesis remains unknown. Eight samples harboured *U2AF1*^{S34F}. We identified 129 splicing events strongly associated with *U2AF1*^{S34F} mutation, consistent with the role of *U2AF1* in 3'-splice site selection²⁸. Cassette exons and alternative 3' splice sites were most commonly affected (Fig. 2c, Supplementary Table 11)²⁹. Among these events, alternative splicing of the *CTNNB1* proto-oncogene was strongly associated with *U2AF1* mutations (Supplementary Fig. 8b). Thus, concurrent analysis of DNA and RNA enabled delineation of both *cis* and *trans* mechanisms governing RNA processing in lung adenocarcinoma.

Candidate driver genes

The receptor tyrosine kinase (RTK)/RAS/RAF pathway is frequently mutated in lung adenocarcinoma. Striking therapeutic responses are often achieved when mutant pathway components are successfully inhibited. Sixty-two per cent (143/230) of tumours harboured known activating mutations in known driver oncogenes, as defined by others³⁰. Cancer-associated mutations in *KRAS* (32%, $n = 74$), *EGFR* (11%, $n = 26$) and *BRAF* (7%, $n = 16$) were common. Additional, previously uncharacterized *KRAS*, *EGFR* and *BRAF* mutations were observed, but were not classified as driver oncogenes for the purposes of our analyses (see Supplementary Fig. 9a for depiction of all mutations of known and unknown significance); explaining the differing mutation frequencies in each gene between this analysis and the overall mutational analysis described above. We also identified known activating *ERBB2*-in-frame insertion and point mutations ($n = 5$)⁶, as well as mutations in *MAP2K1* ($n = 2$), *NRAS* and *HRAS* ($n = 1$ each). RNA sequencing revealed the aforementioned *MET* exon 14 skipping ($n = 10$) and fusions involving *ROS1* ($n = 4$), *ALK* ($n = 3$) and *RET* ($n = 2$). We considered these tumours collectively as oncogene-positive, as they harboured a known activating RTK/RAS/RAF pathway somatic event. DNA amplification events were not considered to be driver events before the comparisons described below.

We sought to nominate previously unrecognized genomic events that might activate this critical pathway in the 38% of samples without a RTK/RAS/RAF oncogene mutation. Tumour cellularity did not differ between oncogene-negative and oncogene-positive samples (Supplementary Fig. 9b). Analysis of copy number alterations using GISTIC³¹ identified unique focal *ERBB2* and *MET* amplifications in the oncogene-negative subset (Fig. 3a, Supplementary Table 6); amplifications in other wild-type proto-oncogenes, including *KRAS* and *EGFR*, were not significantly different between the two groups.

We next analysed WES data independently in the oncogene-negative and oncogene-positive subsets. We found that *TP53*, *KEAP1*, *NF1* and *RIT1* mutations were significantly enriched in oncogene-negative tumours ($P < 0.01$; Fig. 3b, Supplementary Table 12). *NF1* mutations have previously been reported in lung adenocarcinoma¹¹, but this is the first study, to our knowledge, capable of identifying all classes of loss-of-function

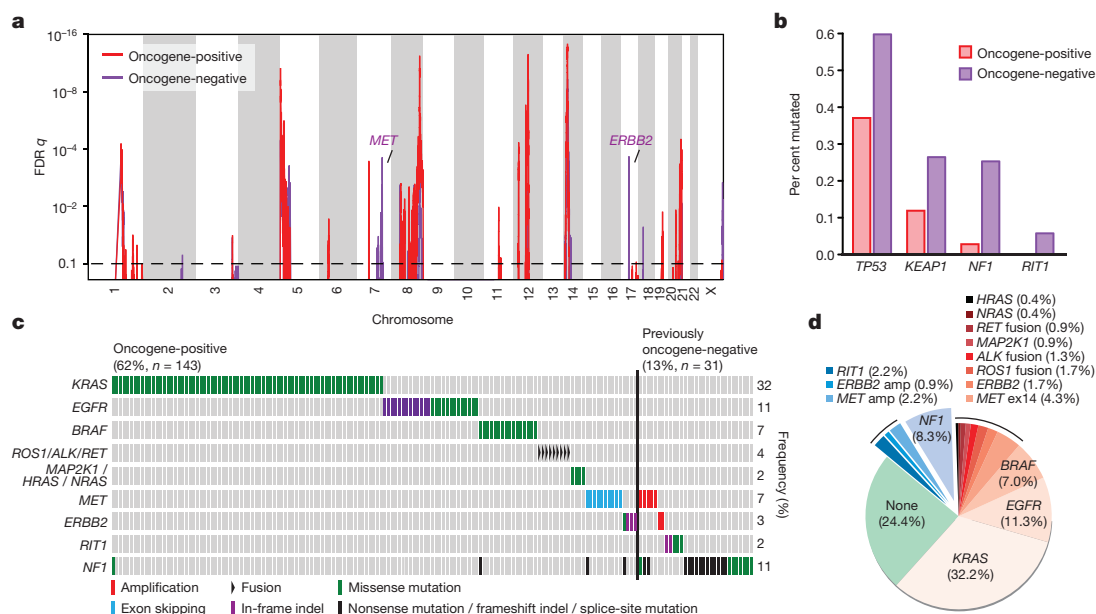
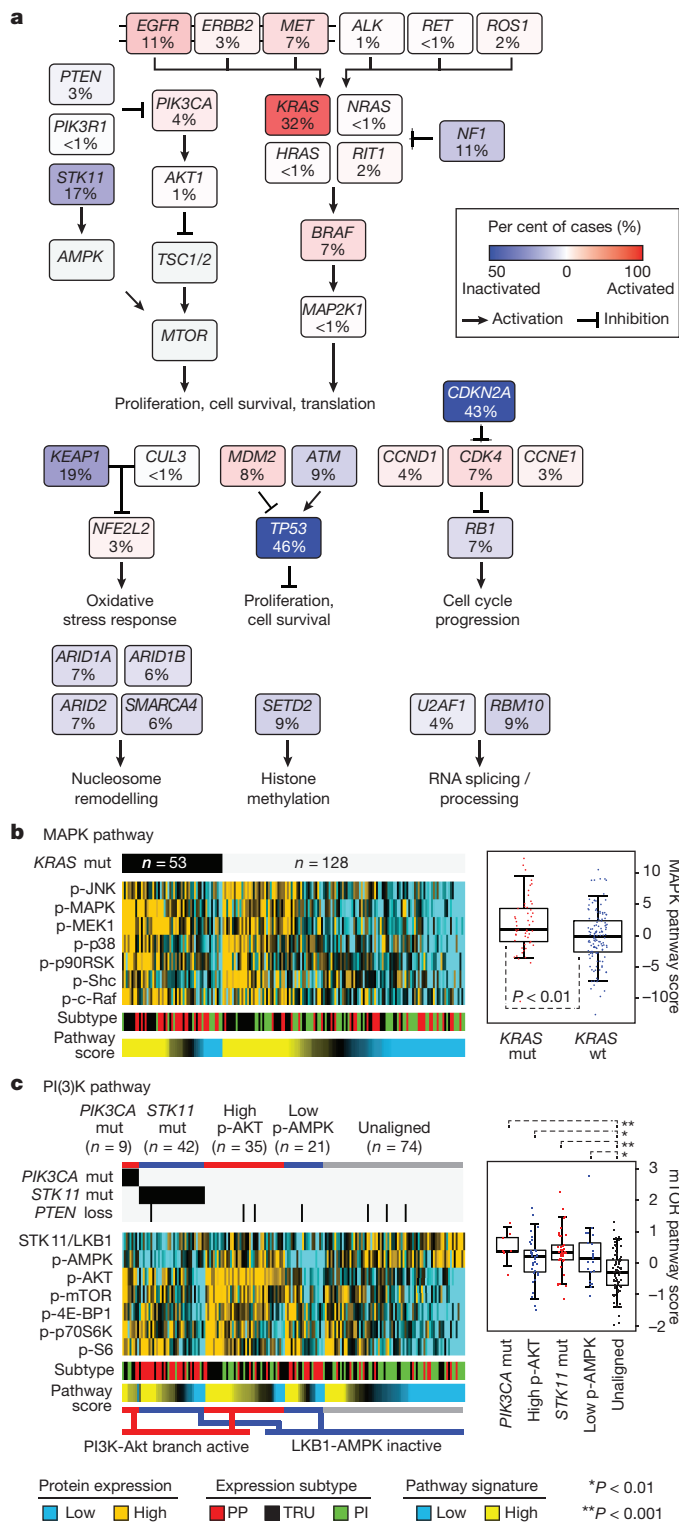


Figure 3 | Identification of novel candidate driver genes. **a**, GISTIC analysis of focal amplifications in oncogene-negative ($n = 87$) and oncogene-positive ($n = 143$) TCGA samples identifies focal gains of *MET* and *ERBB2* that are specific to the oncogene-negative set (purple). **b**, *TP53*, *KEAP1*, *NF1* and *RIT1* mutations are significantly enriched in samples otherwise lacking oncogene mutations (adjusted $P < 0.05$ by Fisher's exact test). **c**, Co-mutation plot of variants of known significance within the RTK/RAS/RAF pathway in lung

adenocarcinoma. Not shown are the 63 tumours lacking an identifiable driver lesion. Only canonical driver events, as defined in Supplementary Fig. 9, and proposed driver events, are shown; hence not every alteration found is displayed. **d**, New candidate driver oncogenes (blue: 13% of cases) and known somatically activated drivers events (red: 63%) that activate the RTK/RAS/RAF pathway can be found in the majority of the 230 lung adenocarcinomas.

NF1 mutations and to statistically determine *NF1* mutations, as well as *KEAP1* and *TP53* mutations are enriched in the oncogene-negative subset of lung adenocarcinomas (Fig. 3c). All *RIT1* mutations occurred in the oncogene-negative subset and clustered around residue Q79 (homologous to Q61 in the switch II region of RAS genes). These mutations transform NIH3T3 cells and activate MAPK and PI(3)K signalling³², supporting a driver role for mutant *RIT1* in 2% of lung adenocarcinomas. This analysis increases the rate at which putative somatic lung adenocarcinoma driver events can be identified within the RTK/RAS/RAF pathway to 76% (Fig. 3d).



Recurrent aberrations in multiple key pathways and processes characterize lung adenocarcinoma (Fig. 4a). Among these were RTK/RAS/RAF pathway activation (76% of cases), PI(3)K-mTOR pathway activation (25%), p53 pathway alteration (63%), alteration of cell cycle regulators (64%, Supplementary Fig. 10), alteration of oxidative stress pathways (22%, Supplementary Fig. 11), and mutation of various chromatin and RNA splicing factors (49%).

We then examined the phenotypic sequelae of some key genomic events in the tumours in which they occurred. Reverse-phase protein arrays provided proteomic and phosphoproteomic phenotypic evidence of pathway activity. Antibodies on this platform are listed in Supplementary Table 13. This analysis suggested that DNA sequencing did not identify all samples with phosphoprotein evidence of activation of a given signalling pathway. For example, whereas *KRAS*-mutant lung adenocarcinomas had higher levels of phosphorylated MAPK than *KRAS* wild-type tumours had on average, many *KRAS* wild-type tumours displayed significant MAPK pathway activation (Fig. 4b, Supplementary Fig. 10). The multiple mechanisms by which lung adenocarcinomas achieve MAPK activation suggest additional, still undetected RTK/RAS/RAF pathway alterations. Similarly, we found significant activation of mTOR and its effectors (p70S6kinase, S6, 4E-BP1) in a substantial fraction of the tumours (Fig. 4c). Analysis of mutations in *PIK3CA* and *STK11*, *STK11* protein levels, and AMPK and AKT phosphorylation³³ led to the identification of three major mTOR patterns in lung adenocarcinoma: (1) tumours with minimal or basal mTOR pathway activation, (2) tumours showing higher mTOR activity accompanied by either *STK11*-inactivating mutation or combined low *STK11* expression and low AMPK activation and (3) tumours showing high mTOR activity accompanied by either phosphorylated AKT activation, *PIK3CA* mutation, or both. As with MAPK, many tumours lack an obvious underlying genomic alteration to explain their apparent mTOR activation.

Molecular subtypes of lung adenocarcinoma

Broad transcriptional and epigenetic profiling can reveal downstream consequences of driver mutations, provide clinically relevant classification and offer insight into tumours lacking clear drivers. Prior unsupervised analyses of lung adenocarcinoma gene expression have used varying nomenclature for transcriptional subtypes of the disease^{34–37}. To coordinate naming of the transcriptional subtypes with the histopathological³⁸, anatomic and mutational classifications of lung adenocarcinoma, we propose an updated nomenclature: the terminal respiratory unit (TRU, formerly bronchioid), the proximal-inflammatory (PI, formerly squamoid), and the proximal-proliferative (PP, formerly magnoid)³⁹ transcriptional subtypes (Fig. 5a). Previously reported associations of expression signatures with pathways and clinical outcomes^{34,36,39} were observed (Supplementary Fig. 7b) and integration with multi-analyte data revealed statistically significant genomic alterations associated with these transcriptional subtypes. The PP subtype was enriched for mutation of *KRAS*, along with inactivation of the *STK11* tumour suppressor gene by chromosomal loss, inactivating mutation, and reduced gene expression. In contrast, the PI subtype was characterized by solid histopathology and

Figure 4 | Pathway alterations in lung adenocarcinoma. **a**, Somatic alterations involving key pathway components for RTK signalling, mTOR signalling, oxidative stress response, proliferation and cell cycle progression, nucleosome remodelling, histone methylation, and RNA splicing/processing. **b, c**, Proteomic analysis by RPPA (*n* = 181) *P* values by two-sided *t*-test. Box plots represent 5%, 25%, 75%, median, and 95%. PP, proximal proliferative; TRU, terminal respiratory unit; PI, proximal inflammatory. **c**, mTOR signalling may be activated, by either Akt (for example, via PI(3)K) or inactivation of AMPK (for example, via *STK11* loss). Tumours were separated into three main groups: those with PI(3)K-AKT activation, through either *PIK3CA* activating mutation or unknown mechanism (high p-AKT); those with LKB1-AMPK inactivation, through either *STK11* mutation or unknown mechanism with low levels of LKB1 and p-AMPK; and those showing none of the above features.

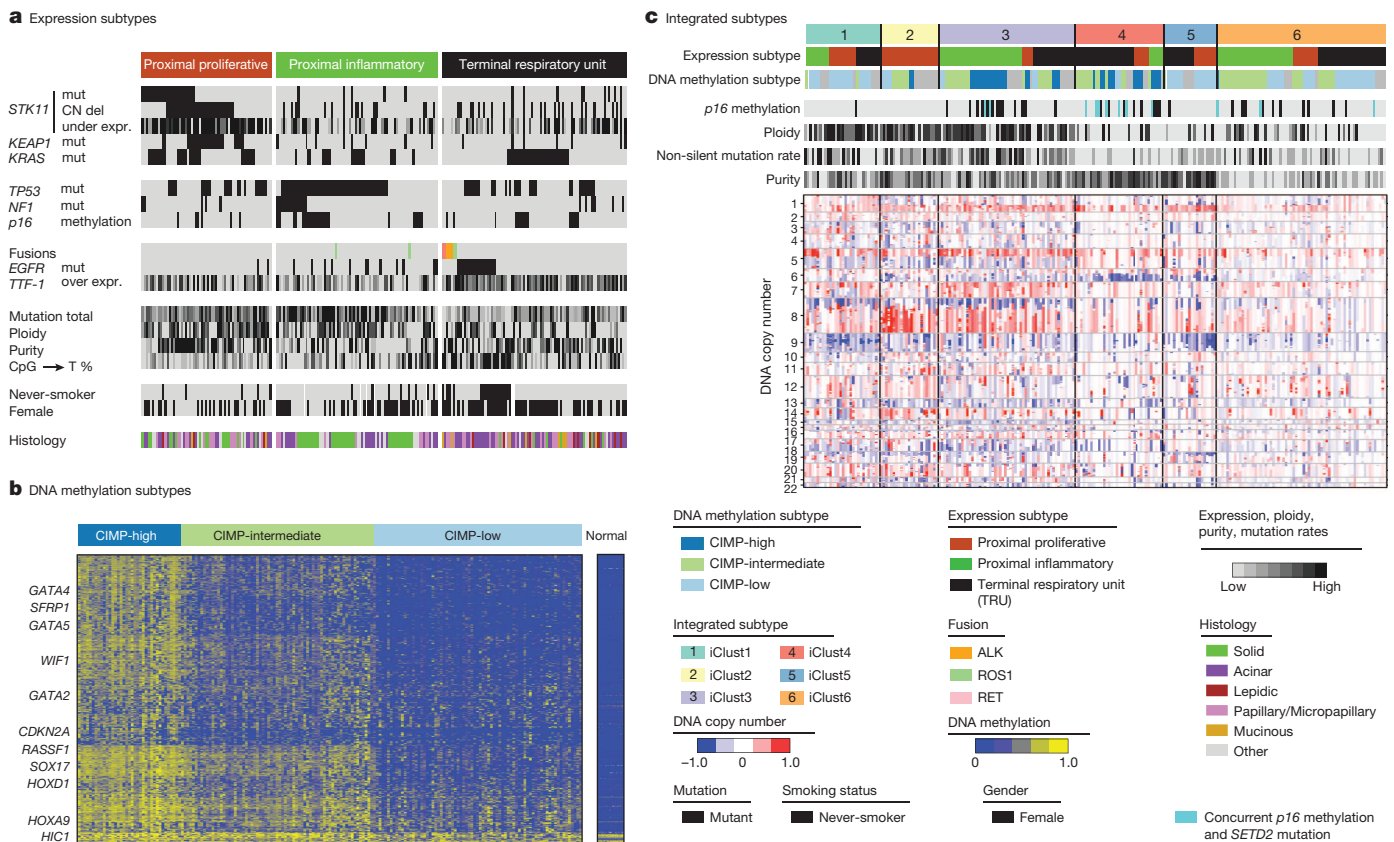


Figure 5 | Integrative analysis. a–c, Integrating unsupervised analyses of 230 lung adenocarcinomas reveals significant interactions between molecular subtypes. Tumours are displayed as columns, grouped by mRNA expression subtypes (a), DNA methylation subtypes (b), and integrated subtypes by

iCluster analysis (c). All displayed features are significantly associated with subtypes depicted. The CIMP phenotype is defined by the most variable CpG island and promoter probes.

co-mutation of *NF1* and *TP53*. Finally, the TRU subtype harboured the majority of the *EGFR*-mutated tumours as well as the kinase fusion expressing tumours. TRU subtype membership was prognostically favourable, as seen previously³⁴ (Supplementary Fig. 7c). Finally, the subtypes exhibited different mutation rates, transition frequencies, genomic ploidy profiles, patterns of large-scale aberration, and differed in their association with smoking history (Fig. 5a). Unsupervised clustering of miRNA sequencing-derived or reverse phase protein array (RPPA)-derived data also revealed significant heterogeneity, partially overlapping with the mRNA-based subtypes, as demonstrated in Supplementary Figs 12 and 13.

Mutations in chromatin-modifying genes (for example, *SMARCA4*, *ARID1A* and *SETD2*) suggest a major role for chromatin maintenance in lung adenocarcinoma. To examine chromatin states in an unbiased manner, we selected the most variable DNA methylation-specific probes in CpG island promoter regions and clustered them by methylation intensity (Supplementary Table 14). This analysis divided samples into two distinct subsets: a significantly altered CpG island methylator phenotype-high (CIMP-H(igh)) cluster and a more normal-like CIMP-L(ow) group, with a third set of samples occupying an intermediate level of methylation at CIMP sites (Fig. 5b). Our results confirm a prior report⁴⁰ and provide additional insights into this epigenetic program. CIMP-H tumours often showed DNA hypermethylation of several key genes: *CDKN2A*, *GATA2*, *GATA4*, *GATA5*, *HIC1*, *HOXA9*, *HOXD13*, *RASSF1*, *SFRP1*, *SOX17* and *WIF1* among others (Supplementary Fig. 14). WNT pathway genes are significantly over-represented in this list (P value = 0.0015) suggesting that this is a key pathway with an important driving role within this subtype. *MYC* overexpression was significantly associated with the CIMP-H phenotype as well (P = 0.003).

Although we did not find significant correlations between global DNA methylation patterns and individual mutations in chromatin remodeling genes, there was an intriguing association between *SETD2* mutation

and *CDKN2A* methylation. Tumours with low *CDKN2A* expression due to methylation (rather than due to mutation or deletion) had lower ploidy, fewer overall mutations (Fig. 5c) and were significantly enriched for *SETD2* mutation, suggesting an important role for this chromatin-modifying gene in the development of certain tumours.

Integrative clustering⁴¹ of copy number, DNA methylation and mRNA expression data found six clusters (Fig. 5c). Tumour ploidy and mutation rate are higher in clusters 1–3 than in clusters 4–6. Clusters 1–3 frequently harbour *TP53* mutations and are enriched for the two proximal transcriptional subtypes. Fisher's combined probability tests revealed significant copy number associated gene expression changes on 3q in cluster one, 8q in cluster two, and chromosome 7 and 15q in cluster three (Supplementary Fig. 15). The low ploidy and low mutation rate clusters four and five contain many TRU samples, whereas tumours in cluster 6 have comparatively lower tumour cellularity, and few other distinguishing molecular features. Significant copy number-associated gene expression changes are observed on 6q in cluster four and 19p in cluster five. The CIMP-H tumours divided into a high ploidy, high mutation rate, proximal-inflammatory CIMP-H group (cluster 3) and a low ploidy, low mutation rate, TRU-associated CIMP-H group (cluster 4), suggesting that the CIMP phenotype in lung adenocarcinoma can occur in markedly different genomic and transcriptional contexts. Furthermore, cluster four is enriched for *CDKN2A* methylation and *SETD2* mutations, suggesting an interaction between somatic mutation of *SETD2* and deregulated chromatin maintenance in this subtype. Finally, cluster membership was significantly associated with mutations in *TP53*, *EGFR* and *STK11* (Supplementary Fig. 15, Supplementary Table 6).

Conclusions

We assessed the mutation profiles, structural rearrangements, copy number alterations, DNA methylation, mRNA, miRNA and protein expression

of 230 lung adenocarcinomas. In recent years, the treatment of lung adenocarcinoma has been advanced by the development of multiple therapies targeted against alterations in the RTK/RAS/RAF pathway. We nominate amplifications in *MET* and *ERBB2* as well as mutations of *NFI* and *RIT1* as driver events specifically in otherwise oncogene-negative lung adenocarcinomas. This analysis increases the fraction of lung adenocarcinoma cases with somatic evidence of RTK/RAS/RAF activation from 62% to 76%. While all lung adenocarcinomas may activate this pathway by some mechanism, only a subset show tonic pathway activation at the protein level, suggesting both diversity between tumours with seemingly similar activating events and as yet undescribed mechanisms of pathway activation. Therefore, the current study expands the range of possible targetable alterations within the RTK/RAS/RAF pathway in general and suggests increased implementation of *MET* and *ERBB2/HER2* inhibitors in particular. Our discovery of inactivating mutations of *MGA* further underscores the importance of the *MYC* pathway in lung adenocarcinoma.

This study further implicates both chromatin modifications and splicing alterations in lung adenocarcinoma through the integration of DNA, transcriptome and methylome analysis. We identified alternative splicing due to both splicing factor mutations in *trans* and mutation of splice sites in *cis*, the latter leading to activation of the *MET* gene by exon 14 skipping. Cluster analysis separated tumours based on single-gene driver events as well as large-scale aberrations, emphasizing lung adenocarcinoma's molecular heterogeneity and combinatorial alterations, including the identification of coincident *SETD2* mutations and *CDKN2A* methylation in a subset of CIMP-H tumours, providing evidence of a somatic event associated with a genome-wide methylation phenotype. These studies provide new knowledge by illuminating modes of genomic alteration, highlighting previously unappreciated altered genes, and enabling further refinement in sub-classification for the improved personalization of treatment for this deadly disease.

METHODS SUMMARY

All specimens were obtained from patients with appropriate consent from the relevant institutional review board. DNA and RNA were collected from samples using the Allprep kit (Qiagen). We used standard approaches for capture and sequencing of exomes from tumour DNA and normal DNA¹⁵ and whole-genome shotgun sequencing. Significantly mutated genes were identified by comparing them with expectation models based on the exact measured rates of specific sequence lesions⁴². GISTIC analysis of the circular-binary-segmented Affymetrix SNP 6.0 copy number data was used to identify recurrent amplification and deletion peaks³¹. Consensus clustering approaches were used to analyse mRNA, miRNA and methylation subtypes using previous approaches¹⁵. The publication web page is (https://tcga-data.nci.nih.gov/docs/publications/luad_2014/). Sequence files are in CGHub (<https://cghub.ucsc.edu/>).

Received 11 June 2013; accepted 22 April 2014.

Published online 9 July 2014.

- Paez, J. G. *et al.* *EGFR* mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304**, 1497–1500 (2004).
- Kwak, E. L. *et al.* Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N. Engl. J. Med.* **363**, 1693–1703 (2010).
- Bergthron, K. *et al.* ROS1 rearrangements define a unique molecular class of lung cancers. *J. Clin. Oncol.* **30**, 863–870 (2012).
- Drilon, A. *et al.* Response to cabozantinib in patients with RET fusion-positive lung adenocarcinomas. *Cancer Discov.* **3**, 630–635 (2013).
- Stephens, P. *et al.* Lung cancer: intragenic ERBB2 kinase mutations in tumours. *Nature* **431**, 525–526 (2004).
- Takahashi, T. *et al.* p53: a frequent target for genetic abnormalities in lung cancer. *Science* **246**, 491–494 (1989).
- Sanchez-Cespedes, M. *et al.* Inactivation of LKB1/STK11 is a common event in adenocarcinomas of the lung. *Cancer Res.* **62**, 3659–3662 (2002).
- Shapiro, G. I. *et al.* Reciprocal Rb inactivation and p16^{INK4} expression in primary lung cancers and cell lines. *Cancer Res.* **55**, 505–509 (1995).
- Singh, A. *et al.* Dysfunctional KEAP1-NRF2 interaction in non-small-cell lung cancer. *PLoS Med.* **3**, e420 (2006).
- Medina, P. P. *et al.* Frequent *BRG1/SMARCA4*-inactivating mutations in human lung cancer cell lines. *Hum. Mutat.* **29**, 617–622 (2008).
- Ding, L. *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455**, 1069–1075 (2008).
- Imielinski, M. *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150**, 1107–1120 (2012).

- Govindan, R. *et al.* Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* **150**, 1121–1134 (2012).
- Travis, W. D., Brambilla, E. & Riely, G. J. New pathologic classification of lung cancer: relevance for clinical practice and clinical trials. *J. Clin. Oncol.* **31**, 992–1001 (2013).
- The Cancer Genome Atlas Research Network Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
- Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nature Biotechnol.* **30**, 413–421 (2012).
- Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnol.* **31**, 213–219 (2013).
- Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
- Hurlin, P. J., Steingrimsson, E., Copeland, N. G., Jenkins, N. A. & Eisenman, R. N. Mga, a dual-specificity transcription factor that interacts with Max and contains a T-domain DNA-binding motif. *EMBO J.* **18**, 7019–7028 (1999).
- Peifer, M. *et al.* Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nature Genet.* **44**, 1104–1110 (2012).
- Rudin, C. M. *et al.* Comprehensive genomic analysis identifies *SOX2* as a frequently amplified gene in small-cell lung cancer. *Nature Genet.* **44**, 1111–1116 (2012).
- Tokumo, M. *et al.* The relationship between epidermal growth factor receptor mutations and clinicopathologic features in non-small cell lung cancers. *Clin. Cancer Res.* **11**, 1167–1173 (2005).
- Coleman, M. P. *et al.* A novel gene, DXS8237E, lies within 20 kb upstream of UBE1 in Xp11.23 and has a different X inactivation status. *Genomics* **31**, 135–138 (1996).
- Weir, B. A. *et al.* Characterizing the cancer genome in lung adenocarcinoma. *Nature* **450**, 893–898 (2007).
- Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
- Kong-Beltran, M. *et al.* Somatic mutations lead to an oncogenic deletion of Met in lung cancer. *Cancer Res.* **66**, 283–289 (2006).
- Seo, J. S. *et al.* The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res.* **22**, 2109–2119 (2012).
- Wu, S., Romfo, C. M., Nilsen, T. W. & Green, M. R. Functional recognition of the 3' splice site AG by the splicing factor U2AF³⁵. *Nature* **402**, 832–835 (1999).
- Brooks, A. N. *et al.* A pan-cancer analysis of transcriptome changes associated with somatic mutations in *U2AF1* reveals commonly altered splicing events. *PLoS ONE* **9**, e87361 (2014).
- Pao, W. & Hutchinson, K. E. Chipping away at the lung cancer genome. *Nature Med.* **18**, 349–351 (2012).
- Beroukhim, R. *et al.* Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc. Natl Acad. Sci. USA* **104**, 20007–20012 (2007).
- Berger, A. H. *et al.* Oncogenic *RIT1* mutations in lung adenocarcinoma. *Oncogene* <http://dx.doi.org/10.1038/onc.2013.581> (2014).
- Creighton, C. J. *et al.* Proteomic and transcriptomic profiling reveals a link between the PI3K pathway and lower estrogen-receptor (ER) levels and activity in ER⁺ breast cancer. *Breast Cancer Res.* **12**, R40 (2010).
- Wilkerson, M. D. *et al.* Differential pathogenesis of lung adenocarcinoma subtypes involving sequence mutations, copy number, chromosomal instability, and methylation. *PLoS ONE* **7**, e36530 (2012).
- Beer, D. G. *et al.* Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nature Med.* **8**, 816–824 (2002).
- Hayes, D. N. *et al.* Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. *J. Clin. Oncol.* **24**, 5079–5090 (2006).
- Bhattacharjee, A. *et al.* Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl Acad. Sci. USA* **98**, 13790–13795 (2001).
- Travis, W. D. *et al.* International association for the study of lung cancer/American Thoracic Society/European Respiratory Society international multidisciplinary classification of lung adenocarcinoma. *J. Thoracic Oncol.* **6**, 244–285 (2011).
- Yatabe, Y., Mitsudomi, T. & Takahashi, T. TTF-1 expression in pulmonary adenocarcinomas. *Am. J. Surg. Pathol.* **26**, 767–773 (2002).
- Shinjo, K. *et al.* Integrated analysis of genetic and epigenetic alterations reveals CpG island methylator phenotype associated with distinct clinical characters of lung adenocarcinoma. *Carcinogenesis* **33**, 1277–1285 (2012).
- Mo, Q. *et al.* Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proc. Natl Acad. Sci. USA* **110**, 4245–4250 (2013).
- Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements This study was supported by NIH grants: U24 CA126561, U24 CA126551, U24 CA126554, U24 CA126543, U24 CA126546, U24 CA137153, U24 CA126563, U24 CA126544, U24 CA143845, U24 CA143858, U24 CA144025, U24 CA143882, U24 CA143866, U24 CA143867, U24 CA143848, U24 CA143840, U24 CA143835, U24 CA143799, U24 CA143883, U24 CA143843, U54 HG003067, U54 HG003079 and U54 HG003273. We thank K. Guebert and L. Gaffney for assistance and C. Gunter for review.

Author Contributions The Cancer Genome Atlas Research Network contributed collectively to this study. Biospecimens were provided by the tissue source sites and processed by the biospecimen core resource. Data generation and analyses were performed by the genome sequencing centres, cancer genome characterization centres and genome data analysis centres. All data were released through the data coordinating centre. The National Cancer Institute and National Human Genome Research Institute project teams coordinated project activities. We also acknowledge the following TCGA investigators who made substantial contributions to the project: E. A. Collisson (manuscript coordinator); J. D. Campbell, J. Chmielecki, (analysis coordinators); C. Sougnez (data coordinator); J. D. Campbell, M. Rosenberg, W. Lee, J. Chmielecki, M. Ladanyi, and G. Getz (DNA sequence analysis); M. D. Wilkerson, A. N. Brooks, and D. N. Hayes (mRNA sequence analysis); L. Danilova and L. Cope (DNA methylation analysis); A. D. Cherniack (copy number analysis); M. D. Wilkerson and A. Hadjipanayis (translocations); N. Schultz, W. Lee, E. A. Collisson, A. H. Berger, J. Chmielecki, C. J. Creighton, L. A. Byers and M. Ladanyi (pathway analysis); A. Chu and A. G. Robertson (miRNA sequence analysis); W. Travis and D. A. Wigle (pathology and clinical expertise); L. A. Byers and G. B. Mills (reverse phase protein arrays); S. B. Baylin, R. Govindan and M. Meyerson (project chairs).

Author Information The primary and processed data used to generate the analyses presented here can be downloaded by registered users from The Cancer Genome Atlas at (<https://tcga-data.nci.nih.gov/tcga/Download.jsp>). All of the primary sequence files are deposited in cGHub and all other data are deposited at the Data Coordinating Center (DCC) for public access (<http://cancergenome.nih.gov/>), (<https://cghub.ucsc.edu/>) and (https://tcga-data.nci.nih.gov/docs/publications/luad_2014/). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.M. (matthew_meyerson@dfci.harvard.edu).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

The Cancer Genome Atlas Research Network

Disease analysis working group Eric A. Collisson¹, Joshua D. Campbell², Angela N. Brooks^{2,3}, Alice H. Berger², William Lee⁴, Juliann Chmielecki², David G. Beer⁵, Leslie Cope⁶, Chad J. Creighton⁷, Ludmila Danilova⁶, Li Ding⁸, Gad Getz^{2,9,10}, Peter S. Hammerman², D. Neil Hayes¹¹, Bryan Hernandez², James G. Herman⁶, John V. Heymach¹², Igor Jurisica¹³, Raju Kucherlapati⁹, David Kwiatkowski¹⁴, Marc Ladanyi⁴, Gordon Robertson¹⁵, Nikolaus Schultz⁴, Ronglai Shen⁴, Rileen Sinha⁴, Carrie Sougnez², Ming-Sound Tsao¹³, William D. Travis⁴, John N. Weinstein¹², Dennis A. Wigle¹⁶, Matthew D. Wilkerson¹¹, Andy Chu¹⁵, Andrew D. Cherniack², Angela Hadjipanayis⁹, Mara Rosenberg², Daniel J. Weisenberger¹⁷, Peter W. Laird¹⁷, Amie Radenbaugh¹⁸, Singer Ma¹⁸, Joshua M. Stuart¹⁸, Lauren Averett Byers¹², Stephen B. Baylin⁶, Ramaswamy Govindan⁸, Matthew Meyerson^{2,3}

Genome sequencing centres: The Eli & Edythe L. Broad Institute Mara Rosenberg², Stacey B. Gabriel², Kristian Cibulskis², Carrie Sougnez², Jaegil Kim², Chip Stewart², Lee Lichtenstein², Eric S. Lander^{2,19}, Michael S. Lawrence², Getz^{2,9,10}, **Washington University in St. Louis** Cyriac Kandoth⁵, Robert Fulton⁵, Lucinda L. Fulton⁵, Michael D. McLellan⁵, Richard K. Wilson⁵, Kai Ye⁵, Catrina C. Fronick⁵, Christopher A. Maher⁵, Christopher A. Miller⁵, Michael C. Wendt⁵, Christopher Cabanski⁵, Li Ding⁸, Elaine Mardis⁵, Ramaswamy Govindan⁸, **Baylor College of Medicine** Chad J. Creighton⁷, David Wheeler⁷

Genome characterization centres: Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency Miruna Balasundaram¹⁵, Yaron S. N. Butterfield¹⁵, Rebecca Carlsen¹⁵, Andy Chu¹⁵, Eric Chuah¹⁵, Noreen Dhalla¹⁵, Ranabir Guin¹⁵, Carrie Hirst¹⁵, Darlene Lee¹⁵, Haiyan I. Li¹⁵, Michael Mayo¹⁵, Richard A. Moore¹⁵, Andrew J. Mungall¹⁵, Jacqueline E. Schein¹⁵, Payal Sipahimalani¹⁵, Angela Tam¹⁵, Richard Varhol¹⁵, A. Gordon Robertson¹⁵, Natasha Wye¹⁵, Nina Thiessen¹⁵, Robert A. Holt¹², Steven J. M. Jones¹⁵, Marco A. Marra¹⁵, **The Eli & Edythe L. Broad Institute** Joshua D. Campbell², Angela N. Brooks^{2,3}, Juliann Chmielecki², Marcini Imielinski^{2,9,10}, Robert C. Onofrio², Eran Hodis⁹, Travis Zack², Carrie Sougnez², Elena Helman², Chandra Sekhar Pedamallu², Jill Mesirov², Andrew D. Cherniack², Gordon Saksena², Steven E. Schumacher², Scott L. Carter², Bryan Hernandez², Levi Garraway^{2,3,9}, Rameen Beroukhi^{2,3,9}, Stacey B. Gabriel², Gad Getz^{2,9,10}, Matthew Meyerson^{2,3,9}, **Harvard Medical School/Brigham & Women's Hospital/MD Anderson Cancer Center** Angela Hadjipanayis^{9,14}, Semin Lee^{9,14}, Harshad S. Mahadeshwar¹², Angeliki Pantazi^{9,14}, Alexei Protopopov¹², Xiaojia Ren⁹, Sahil Sethi¹², Xingzhi Song¹², Jiabin Tang¹², Lixing Yang⁹, Jianhua Zhang¹², Peng-Chieh Chen⁹, Michael Parfenov^{9,14}, Andrew Wei Xu^{9,14}, Netty Santos^{9,14}, Lynda Chin¹², Peter J. Park^{9,14} & Raju Kucherlapati^{9,14}, **University of North Carolina, Chapel Hill** Katherine A. Hoadley¹¹, J. Todd Auman¹¹, Shaowu Meng¹¹, Yan Shi¹¹, Elizabeth Buda¹¹, Scot Waring¹¹, Umadevi Veluvolu¹¹, Donghui Tan¹¹, Piotr A. Mieczkowski¹¹, Corbin D. Jones¹¹, Janae V. Simons¹¹, Matthew G. Soloway¹¹, Tom Bodenheimer¹¹, Stuart R. Jefferys¹¹, Jeffrey Roach¹¹, Alan P. Hoyle¹¹, Junyuan Wu¹¹, Saianand Balu¹¹, Darshan Singh¹¹, Jan F.

Prins¹¹, J.S. Marron¹¹, Joel S. Parker¹¹, D. Neil Hayes¹¹, Charles M. Perou¹¹, **University of Kentucky** Jinze Liu²⁰, **The USC/JHU Epigenome Characterization Center** Leslie Cope⁶, Ludmila Danilova⁶, Daniel J. Weisenberger¹⁷, Dennis T. Maglinte¹⁷, Philip H. Lai¹⁷, Moiz S. Bootwalla¹⁷, David J. Van Den Berg¹⁷, Timothy Triche Jr¹⁷, Stephen B. Baylin⁶, Peter W. Laird¹⁷

Genome data analysis centres: The Eli & Edythe L. Broad Institute Mara Rosenberg², Lynda Chin¹², Jianhua Zhang¹², Juok Cho², Daniel DiCara², David Heiman², Pei Lin², William Mallard², Douglas Voet², Hailei Zhang², Lihua Zou², Michael S. Noble², Michael S. Lawrence², Gordon Saksena², Nils Gehlenborg², Helga Thorvaldsdottir², Jill Mesirov², Marc-Danie Nazaire², Jim Robinson², Gad Getz^{2,9,10}, **Memorial Sloan-Kettering Cancer Center** William Lee⁴, B. Arman Aksoy⁴, Giovanni Ciriello⁴, Barry S. Taylor⁴, Gideon Dresdner⁴, Jianjing Gao⁴, Benjamin Gross⁴, Venkatraman E. Seshan⁴, Marc Ladanyi⁴, Boris Reva⁴, Rileen Sinha⁴, S. Onur Sumer⁴, Nils Weinhold⁴, Nikolaus Schultz⁴, Ronglai Shen⁴, Chris Sander⁴; **University of California, Santa Cruz/BUck Institute** Sam Ng¹⁸, Singer Ma¹⁸, Jingchun Zhu¹⁸, Amie Radenbaugh¹⁸, Joshua M. Stuart¹⁸, Christopher C. Benz²¹, Christina Yau²¹ & David Haussler^{18,22}, **Oregon Health & Sciences University** Paul T. Spellman²³, **University of North Carolina, Chapel Hill** Matthew D. Wilkerson¹¹, Joel S. Parker¹¹, Katherine A. Hoadley¹¹, Patrick K. Kimes¹¹, D. Neil Hayes¹¹, Charles M. Perou¹¹, **The University of Texas MD Anderson Cancer Center** Bradley M. Broom¹², Jing Wang¹², Yiling Lu¹², Patrick Kwok Shing Ng¹², Lixia Diao¹², Lauren Averett Byers¹², Wenbin Liu¹², John V. Heymach¹², Christopher I. Amos¹², John N. Weinstein¹², Rehan Akbani¹², Gordon B. Mills¹²

Biospecimen core resource: International Genomics Consortium Erin Curley²⁴, Joseph Paulauskis²⁴, Kevin Lau²⁴, Scott Morris²⁴, Troy Shelton²⁴, David Mallery²⁴, Johanna Gardner²⁴, Robert Penny²⁴

Tissue source sites: Analytical Biological Service, Inc. Charles Saller²⁵, Katherine Tarvin²⁵, **Brigham & Women's Hospital** William G. Richards¹⁴; **University of Alabama at Birmingham** Robert Cerfolio²⁶, Ayesha Bryant²⁶, **Cleveland Clinic:** Daniel P. Raymond²⁷, Nathan A. Pennell²⁷, Carol Farver²⁷; **Christiana Care** Christine Czerwinski²⁸, Lori Huelsenbeck-Dill²⁸, Mary Iacocca²⁸, Nicholas Petrelli²⁸, Brenda Rabenc²⁸, Jennifer Brown²⁸, Thomas Bauer²⁸; **Cureline** Oleg Dolzhanisky²⁹, Olga Potapova²⁹, Daniil Rotin²⁹, Olga Voronina²⁹, Elena Nemirovich-Danchenko²⁹, Konstantin V. Fedosenko²⁹; **Emory University** Anthony Gai³⁰, Madhusmita Behera³⁰, Suresh S. Ramalingam³⁰, Gabriel Sica³⁰; **Fox Chase Cancer Center** Douglas Flieder³¹, Jeff Boyd³¹, JoEllen Weaver³¹; **ILSbio** Bernard Kohli³², Dang Huy Quoc Thinh³²; **Indiana University** George Sandusky³³, **Indivumed** Hartmut Juhl³⁴, **John Flynn Hospital** Edwina Duhig^{35,36}; **Johns Hopkins University** Peter Illei⁶, Edward Gabrielson⁶, James Shin⁶, Beverly Lee⁶, Kristen Rodgers⁶, Dante Trusty⁶, Malcolm V. Brock⁶; **Lahey Hospital & Medical Center** Christina Williamson³⁷, Eric Burks³⁷, Kimberly Rieger-Christ³⁷, Antonia Holway³⁷, Travis Sullivan³⁷; **Mayo Clinic** Dennis A. Wigle¹⁶, Michael K. Asiedu¹⁶, Farhad Kosari¹⁶; **Memorial Sloan-Kettering Cancer Center** William D. Travis⁴, Natasha Rekhman⁴, Maureen Zakowski⁴, Valerie W. Rusch⁴; **NYU Langone Medical Center** Paul Zippile³⁸, James Suh³⁸, Harvey Pass³⁸, Chandra Goparaju³⁸, Yvonne Owusu-Sarpong³⁸, **Ontario Tumour Bank** John M. S. Bartlett³⁹, Sugy Kodeeswaran³⁹, Jeremy Parfitt³⁹, Harmanjatinder Sekhon³⁹, Monique Albert³⁹; **Penrose St. Francis Health Services** John Eckman⁴⁰, Jerome B. Myers⁴⁰; **Roswell Park Cancer Institute** Richard Cheney⁴¹, Carl Morrison⁴¹, Carmelo Gaudio⁴¹; **Rush University Medical Center** Jeffrey A. Borgia⁴², Philip Bonomi⁴², Mark Pool⁴², Michael J. Liptay⁴²; **St. Petersburg Academic University** Fedor Moiseenko⁴³, Irina Zaytseva⁴³; **Thoraxklinik am Universitätsklinikum Heidelberg, Member of Biomaterial Bank Heidelberg (BMBH) & Biobank Platform of the German Centre for Lung Research (DZL)** Hendrik Dienemann⁴⁴, Michael Meister⁴⁴, Philipp A. Schnabel⁴⁵, Thomas R. Mulez⁴⁴; **University of Cologne** Martin Peifer⁴⁶; **University of Miami** Carmen Gomez-Fernandez⁴⁷, Lynn Herbert⁴⁷, Sophie Egea⁴⁷; **University of North Carolina** Mei Huang¹¹, Leigh B. Thorne¹¹, Lori Boice¹¹, Ashley Hill Salazar¹¹, William K. Funkhouser¹¹, W. Kimryn Rathmell¹¹; **University of Pittsburgh** Rajiv Dhir⁴⁸, Samuel A. Yousem⁴⁸, Sanja Dacic⁴⁸, Frank Schneider⁴⁸, Jill M. Siegfried⁴⁸; **The University of Texas MD Anderson Cancer Center** Richard Hajek¹²; **Washington University School of Medicine** Mark A. Watson⁸, Sandra McDonald⁸, Bryan Meyers⁸; **Queensland Thoracic Research Center** Belinda Clarke³⁵, Ian A. Yang³⁵, Kwun M. Fong³⁵, Lindy Hunter³⁵, Morgan Windsor³⁵, Rayleen V. Bowman³⁵; **Center Hospitalier Universitaire Vaudois** Solange Peters⁴⁹, Igor Letovanec⁴⁹; **Ziauddin University Hospital** Khurram Z. Khan⁵⁰

Data Coordination Centre Mark A. Jensen⁵¹, Eric E. Snyder⁵¹, Deepak Srinivasan⁵¹, Ari B. Kahn⁵¹, Julien Baboud⁵¹, David A. Pot⁵¹

Project team: National Cancer Institute Kenna R. Mills Shaw⁵², Margi Sheth⁵², Tanja Davidsen⁵², John A. Demchok⁵², Liming Yang⁵², Zhining Wang⁵², Roy Tarnuzzer⁵², Jean Claude Zenklusen⁵²; **National Human Genome Research Institute** Bradley A. Ozenberger⁵³, Heidi J. Sofia⁵³

Expert pathology panel William D. Travis⁴, Richard Cheney⁴¹, Belinda Clarke³⁵, Sanja Dacic⁴⁸, Edwina Duhig^{36,35}, William K. Funkhouser¹¹, Peter Illei⁶, Carol Farver²⁷, Natasha Rekhman⁴, Gabriel Sica³⁰, James Suh³⁸ & Ming-Sound Tsao¹³

¹University of California San Francisco, San Francisco, California 94158, USA. ²The Eli and Edythe L. Broad Institute, Cambridge, Massachusetts 02142, USA. ³Dana Farber Cancer Institute, Boston, Massachusetts 02115, USA. ⁴Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ⁵University of Michigan, Ann Arbor, Michigan 48109, USA. ⁶Johns Hopkins University, Baltimore, Maryland 21287, USA. ⁷Baylor College of

Medicine, Houston, Texas 77030, USA. ⁸Washington University, St. Louis, Missouri 63108, USA. ⁹Harvard Medical School, Boston, Massachusetts 02115, USA. ¹⁰Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ¹¹University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ¹²University of Texas MD Anderson Cancer Center, Houston, Texas 77054, USA. ¹³Princess Margaret Cancer Centre, Toronto, Ontario M5G 2M9, Canada. ¹⁴Brigham and Women's Hospital Boston, Massachusetts 02115, USA. ¹⁵BC Cancer Agency, Vancouver, British Columbia V5Z 4S6, Canada. ¹⁶Mayo Clinic, Rochester, Minnesota 55905, USA. ¹⁷University of Southern California, Los Angeles, California 90033, USA. ¹⁸University of California Santa Cruz, Santa Cruz, California 95064, USA. ¹⁹Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. ²⁰University of Kentucky, Lexington, Kentucky 40515, USA. ²¹Buck Institute for Age Research, Novato, California 94945, USA. ²²Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, California 95064, USA. ²³Oregon Health and Science University, Portland, Oregon 97239, USA. ²⁴International Genomics Consortium, Phoenix, Arizona 85004, USA. ²⁵Analytical Biological Services, Inc., Wilmington, Delaware 19801, USA. ²⁶University of Alabama at Birmingham, Birmingham, Alabama 35294, USA. ²⁷Cleveland Clinic, Cleveland, Ohio 44195, USA. ²⁸Christiana Care, Newark, Delaware 19713, USA. ²⁹Cureline, Inc., South San Francisco, California 94080, USA. ³⁰Emory University, Atlanta, Georgia 30322, USA. ³¹Fox Chase Cancer Center, Philadelphia, Philadelphia 19111, USA. ³²ILSbio, Chestertown, Maryland 21620, USA. ³³Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. ³⁴Individumed, Silver Spring, Maryland 20910, USA. ³⁵The Prince Charles Hospital and the University of Queensland Thoracic Research Center, Brisbane, 4032, Australia. ³⁶Sullivan Nicolaidis Pathology & John Flynn Hospital, Tugun 4680, Australia. ³⁷Lahey Hospital and Medical Center, Burlington, Massachusetts 01805, USA. ³⁸NYU Langone Medical Center, New York, New York 10016, USA. ³⁹Ontario Tumour Bank, Ontario Institute for Cancer Research, Toronto, Ontario M5G 0A3, Canada. ⁴⁰Penrose St. Francis Health Services, Colorado Springs, Colorado 80907, USA. ⁴¹Roswell Park Cancer Center, Buffalo, New York 14263, USA. ⁴²Rush University Medical Center, Chicago, Illinois 60612, USA. ⁴³St. Petersburg Academic University, St Petersburg 199034, Russia. ⁴⁴Thoraxklinik am Universitätsklinikum Heidelberg, 69126 Heidelberg, Germany. ⁴⁵University Heidelberg, 69120 Heidelberg, Germany. ⁴⁶University of Cologne, 50931 Cologne, Germany. ⁴⁷University of Miami, Sylvester Comprehensive Cancer Center, Miami, Florida 33136, USA. ⁴⁸University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA. ⁴⁹Center Hospitalier Universitaire Vaudois, Lausanne and European Thoracic Oncology Platform, CH-1011 Lausanne, Switzerland. ⁵⁰Ziauddin University Hospital, Karachi, 75300, Pakistan. ⁵¹SRA International, Inc., Fairfax, Virginia 22033, USA. ⁵²National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. ⁵³National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.