HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2013 March 27.

Published in final edited form as:

Nature. 2012 September 27; 489(7417): 519–525. doi:10.1038/nature11404.

Comprehensive genomic characterization of squamous cell lung cancers

The Cancer Genome Atlas Research Network

Summary

Lung squamous cell carcinoma (lung SqCC) is a common type of lung cancer, causing approximately 400,000 deaths per year worldwide. Genomic alterations in lung SqCC have not been comprehensively characterized and no molecularly targeted agents have been developed specifically for its treatment. As part of The Cancer Genome Atlas (TCGA), we profiled 178 lung SqCCs to provide a comprehensive landscape of genomic and epigenomic alterations. Lung SqCC is characterized by complex genomic alterations, with a mean of 360 exonic mutations, 165 genomic rearrangements, and 323 segments of copy number alteration per tumor. We found statistically recurrent mutations in 18 genes in including mutation of *TP53* in nearly all specimens. Previously unreported loss-of-function mutations were seen in the *HLA-A* class I major histocompatibility gene. Significantly altered pathways included *NFE2L2/KEAP1* in 34%, squamous differentiation genes in 44%, *PI3K/AKT* in 47%, and *CDKN2A/RB1* in 72% of tumors. We identified a potential therapeutic target in the majority of tumors, offering new avenues of investigation for lung SqCC treatment.

Introduction

Lung cancer is the leading cause of cancer-related mortality worldwide, leading to an estimated 1.4 million deaths in 2010[1]. The discovery of recurrent mutations in the Epidermal Growth Factor Receptor (*EGFR*) kinase as well as fusions involving the Anaplastic Lymphoma Kinase (*ALK*), has led to a dramatic change in the treatment of

Correspondence to: Matthew Meyerson matthew_meyerson@dfci.harvard.edu.

[39]Oregon Health and Science University, Department of Molecular and Medical Genetics, Portland OR 97239 USA

Author Contributions: The TCGA 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 Centers, Cancer Genome Characterization Centers, and Genome Data Analysis Centers. All data were released through the Data Coordinating Center. Project activities were coordinated by the NCI and NHGRI Project Teams.

We specifically recognize the following investigators who made substantial contributions to the project.

Manuscript co-ordinators: P.S.H. and D.N.H. **Data co-ordinator:** M.D.W. **Analysis co-ordinators:** P.S.H. and N.S. **DNA sequence analysis:** P.S.H., M.S.L., A.S., B.H. and G.G. **mRNA sequence analysis:** M.D.W., J.L. and D.N.H. **DNA methylation analysis:** L.C., J.G.H. and L.D. **Copy number analysis:** A.C., G.S., N.H.P., R.K. and M.L. **Pathway analysis:** N.S., R.B., C.J.C., R.S., C.M., S.N., E.A.C., R.S., J.N.W. and C.S. **miRNA sequence analysis:** A.C. and G.R. **Pathology and clinical expertise:** W.D.T., B.E.J., D.A.W. and M.S.T. Project chairs: S.B.B., R.G. and M.M.

Author Information: The primary and processed data used to generate the analyses presented here can be downloaded by registered users at https://tcga-data.nci.nih.gov/tcga/tcgaDownload.jsp, https://cghub.ucsc.edu/ and https://tcga-data.nci.nih.gov/docs/publications/lusc_2012/

patients with lung adenocarcinoma, the most common type of lung cancer[2–5]. More recent data have suggested that targeting mutations in *BRAF*, *AKT1*, *ERBB2* and *PIK3CA* and fusions that involve *ROS1* and *RET* may also be successful[6,7]. Unfortunately, activating mutations in *EGFR* and *ALK* fusions are typically not present in the second most common type of lung cancer, lung squamous cell carcinoma (lung SqCC),[8] and targeted agents developed for lung adenocarcinoma are largely ineffective against lung SqCC.

Although no comprehensive genomic analysis of lung SqCCs has been reported, single-platform studies have identified regions of somatic copy number alteration in lung SqCCs, including amplification of *SOX2*, *PDGFRA* and *FGFR1/WHSC1L1* and deletion of *CDKN2A*[9,10]. DNA sequencing studies of lung SqCCs have reported recurrent mutations in a number of genes, including *TP53*, *NFE2L2*, *KEAP1*, *BAI3*, *FBXW7*, *GRM8*, *MUC16*, *RUNX1T1*, *STK11* and *ERBB4*[11,12]. *DDR2* mutations and *FGFR1* amplification have been nominated as therapeutic targets[13–15].

We have conducted a comprehensive study of lung SqCCs from a large cohort of patients as part of The Cancer Genome Atlas (TCGA) project. The twin aims are to characterize the genomic/epigenomic landscape of lung SqCC and to identify potential opportunities for therapy. We report an integrated analysis based on DNA copy number, somatic exonic mutations, mRNA sequencing (RNA-seq), mRNA expression, and promoter methylation for 178 histopathologically reviewed lung SqCCs in addition to whole-genome sequencing (WGS) of 19 samples and miRNA sequencing of 159 samples (Supplementary Table S1.1). Demographic and clinical data and results of the genomic analyses can be downloaded from the TCGA data portal (https://tcga-data.nci.nih.gov/tcga/tcgaDownload.jsp).

Samples and clinical data

Tumor samples were obtained from 178 patients with previously untreated stage I–IV lung SqCC. Germline DNA was obtained from adjacent, histologically normal tissues resected at the time of surgery (n=137) or peripheral blood (n=41). All patients provided written informed consent to conduct genomic studies in accordance with local Institutional Review Boards. The demographic characteristics are described in Supplementary Table S1.2. The median follow-up for the cohort was 15.8 months, and 60% of patients were alive at the time of last follow-up (data updated in November 2011). 96% of the patients had a history of tobacco use, similar to previous reports for North American lung SqCC patients[16]. DNA and RNA were extracted from patient specimens and were measured by several genomic assays, which included standard quality control assessments (Supplementary Methods, sections 2–8). A committee of experts in lung cancer pathology performed an additional review of all samples to confirm the histological subtype (Supplementary Figure 1.1 and Supplementary Methods, section 1).

Somatic DNA Alterations

The lung SqCCs analyzed in this study display a large number and variety of DNA alterations, with a mean 360 exonic mutations, 323 altered copy number segments and 165 genomic rearrangements per tumor.

Copy number alterations were analyzed using multiple platforms. Analysis of SNP 6.0 array data across the set of 178 lung SqCCs identified a high rate of copy number alteration (mean of 323 segments) when compared to other TCGA projects (as of February 1, 2012) including ovarian cancer (477 segments),[17] glioblastoma multiforme (282 segments),[18] colorectal carcinoma (213 segments), breast carcinoma (282 segments) and renal cell carcinoma (156 segments). These segments gave rise to regions of both focal and broad somatic copy number alterations (SCNAs) with a mean of 47 focal and 23 broad events per tumor (broad events defined as ≥50% of the length of the chromosome arm). There was strong concordance between the three independent copy number assays for all regions of SCNA (Supplementary Figures S2.1–2.4).

At the level of whole chromosome arm SCNAs, lung SqCCs exhibit many similarities to 205 cases of lung adenocarcinoma analyzed by TCGA (Supplementary Figure S2.1a). The most striking difference between these cancers is selective amplification of chromosome 3q in lung SqCC, as has been reported[9,19] ($p<1\times10^{-15}$ by Fisher's exact test). Using the SNP 6.0 array platform and GISTIC 2.0[20,21], we identified regions of significant copy number alteration (Supplementary Methods, section 2). There were 50 peaks of significant amplification or deletion ($q<0.05$), several of which included SCNAs previously seen in lung SqCCs including *SOX2*, *PDGFRA/KIT*, *EGFR*, *FGFR1/WHSC1L1*, *CCND1* and *CDKN2A*[9,10,19] (Supplementary Figure S2.1b and Supplementary Data S2.1 and S2.2). Other peaks defined regions of SCNA reported for the first time including amplifications of chromosomal segments containing *NFE2L2*, *MYC*, *CDK6*, *MDM2*, *BCL2L1* and *EYS* and deletions of *FOXP1*, *PTEN* and *NF1* (Supplementary Figure S2.1b).

Whole exome sequencing of 178 lung SqCCs and matched germline DNA, targeted 193,094 exons from 18,863 genes. Mean sequencing coverage across targeted bases in was 121X with 83% of target bases above 30X coverage. We identified a total of 48,690 non-silent mutations with a mean of 228 non-silent and 360 total exonic mutations per tumor, corresponding to a mean somatic mutation rate of 8.1 mutations per megabase (Mb) and median of 8.4/Mb. That rate is higher than rates observed in other TCGA projects including acute myelogenous leukemia (0.56/Mb), breast carcinoma (1.0/Mb), ovarian cancer[17] (2.1/Mb), glioblastoma multiforme[18] (2.3/Mb) and colorectal carcinoma (3.2/Mb) (data as of February 1, 2012, $p<2.2\times10^{-16}$ by t-test or Wilcoxon rank sum test for lung SqCC versus all others). In lung SqCC, CpG transitions and transversions were the most commonly observed mutation types, with mean rates of 9.9 and 10.7 per sequenced megabase of CpG context, respectively, for a total mutation rate of 20.6/Mb. At non-CpG sites, transversions at C:G sites were more common than transitions (7.3 vs. 2.9 per Mb; total = 10.2/Mb) and more common than transversions or transitions at A:T sites (1.5 vs. 1.3 per Mb; total = 2.8/Mb).

Significantly mutated genes were identified using a modified version of the MutSig algorithm (Supplementary Methods, section 3)[22,23]. We identified 10 genes with a false discovery rate (FDR) q value <0.1 (Supplementary Table S3.1): *TP53*, *CDKN2A*, *PTEN*, *PIK3CA*, *KEAP1*, *MLL2*, *HLA-A*, *NFE2L2*, *NOTCH1* and *RB1*, all of which demonstrated robust evidence of gene expression as defined by Reads Per Kilobase of exon model per Million mapped reads (RPKM) >1 (Figure 1). *TP53* mutation was observed in 81% of samples by automated analysis; visual review of sequencing reads identified an additional

9% of samples with potential mutations in regions of sub-optimal coverage or in samples with low purity. The majority of observed mutations in *NOTCH1* (8 of 17) were truncating alterations, suggesting loss-of-function, as has recently been reported for head and neck SCCs[22,24]. Mutations in *HLA-A* were also almost exclusively nonsense or splice site events (7 of 8).

To increase our statistical power to detect mutated genes in the setting of the observed high background mutation rate, we performed a secondary MutSig analysis only considering genes previously observed to be mutated in cancer according to the COSMIC database. This yielded 12 additional genes with FDR <0.1: *FAM123B (WTX), HRAS, FBXW7, SMARCA4, NF1, SMAD4, EGFR, APC, TSC1, BRAF, TNFAIP3* and *CREBBP* (Supplementary Table S3.1). Both the spectrum and the frequency of *EGFR* mutations differed from those seen in lung adenocarcinomas. The two most common alterations in lung adenocarcinoma, L858R and in-frame deletions in exon 19, were absent, whereas two L861Q mutations were detected in *EGFR*.

As described in Supplemental Figure S3.1 we verified somatic mutations by performing an independent hybrid-recapture of 76 genes in all samples. 1289 mutations were assayed and we achieved satisfactory coverage to have power to verify at 1283 positions. 1235 mutations were validated (96.2%) (Supplementary Figure S3.1 and Supplementary Methods, section 3). We also verified mutation calls using WGS and RNA-sequencing data with similar results (Supplementary Figures S3.1 and S4.3 and Supplementary Methods, sections 3 and 4).

Whole genome sequencing was performed for 19 tumor/normal pairs with a mean computed coverage of 54X. A mean of 165 somatic rearrangements was found per lung SqCC tumor pair (Supplementary Figure S3.2), a value in excess of that reported for WGS studies of other tumor types including colorectal carcinoma (75)[25], prostate carcinoma (108)[26], multiple myeloma (21)[23], and breast cancer (90)[27]. Although the majority of in-frame coding fusions detected in WGS were validated by RNA-seq, no recurrent rearrangements predicted to generate fusion proteins were identified (Supplementary Data S3.1 and S4.1).

Somatically altered pathways

Many of the somatic alterations we have identified in lung SqCCs appear to be drivers of pathways important to the initiation or progression of the cancer. Specifically, genes involved in oxidative stress response and squamous differentiation were frequently altered by mutation or SCNA. We observed mutations and copy number alterations of *NFE2L2* and *KEAP1* and/or deletion or mutation of *CUL3* in 34% of cases (Figure 2). *NFE2L2* and *KEAP1* code for proteins that bind each other, have been shown to regulate the cell's response to oxidative damage, chemo- and radiotherapy and are somatically altered in a variety of cancer types[28,29]. We found mutations in *NFE2L2* almost exclusively in one of two KEAP1 interaction motifs, DLG or ETGE. Mutations in *KEAP1* and *CUL3* showed a pattern consistent with loss-of-function and were mutually exclusive with mutations in *NFE2L2* (Figures 1c and 2). PARADIGM[30] analysis predicts that mutations in *NFE2L2* and

KEAP1 exert a significant functional impact (Supplementary Figure S7.C.1 and S.7.C.2 and Supplementary Methods, section 7).

We also found alterations in genes with known roles in squamous cell differentiation in 44% of samples, including overexpression and amplification of *SOX2* and *TP63*, loss-of-function mutations in *NOTCH1*, *NOTCH2*, and *ASCL4* and focal deletions in *FOXP1* (Figure 2). Although *NOTCH1* has been well characterized as an oncogene in hematologic cancers[31], *NOTCH1* and *NOTCH2* truncating mutations have been reported in cutaneous SCCs and lung SqCCs[32]. Truncating mutations in *ASCL4* are the first to be reported in human cancer and may have a lineage role given the requirement for *ASCL1* for survival of small cell lung cancer cells[33]. Alterations in *NOTCH1*, *NOTCH2* and *ASCL4* were mutually exclusive and exhibited minimal overlap with amplification of *TP63* and/or *SOX2* (Figure 2), suggesting that aberrations in those modulators of squamous cell differentiation have overlapping functional consequences.

mRNA expression profiling and subtype classification

Whole-transcriptome expression profiles were generated by RNA-sequencing for the entire cohort and by microarrays for a 121-sample subset. Of 20,502 genes analyzed, the mean RNA coverage indices were 19X and 6420 RPKM (Supplementary Figure S4.1 and Supplementary Methods, section 4). Previously reported lung SqCC gene expression-subtype signatures[34] were applied to both of the expression platforms, yielding four subtypes designated as classical (36%), basal (25%), secretory (24%) and primitive (15%). The concordance of subtypes between the two platforms was high (94% agreement) (Supplementary Figure S4.2). Significant correlations were found between the expression subtypes and genomic alterations in copy number, mutation and methylation (Figure 3). The classical subtype was characterized by alterations in *KEAP1*, *NFE2L2*, and *PTEN*, as well as pronounced hypermethylation and chromosomal instability. The 3q26 amplicon was present in all of the subtypes, but it was most characteristic of the classical subtype, which also showed the greatest overexpression of three known oncogenes on 3q: *SOX2*, *TP63*, and *PIK3CA*. RNA sequencing data suggested that high expression levels of *TP63*, in samples with and without amplification of *TP63*, was associated with dominant expression of the deltaN isoform (also called p40), which lacks the N-terminal transactivation domain, compared to the longer isoform, called tap63 (89% of tumors overexpressed deltaN compared to tap63; $p<2.2e-16$). That short deltaN isoform is thought to function as an oncogene[35,36], and its expression was most enriched in the classical subtype. In contrast, the primitive expression-subtype more commonly exhibited *RB1* and *PTEN* alterations, and the basal expression-subtype showed *NF1* alterations (Figure 3). Amplification of *FGFR1* and *WHSC1L1* was anti-correlated with the classical subtype and specifically with *NFE2L2* or *KEAP1* mutated samples. While *CDKN2A* alterations are common in lung SqCCs, they are not associated with any particular expression subtype (Figure 3).

Independent clustering of miRNA and methylation data indicated association with expression subtypes. The highest overall methylation was seen in the classical subtype (Figure 3, Supplementary Figures S5.1 and S6.1, Supplementary Methods, sections 5 and 6, Supplementary Data S6.1 and S6.2 and Supplementary Table 5.1). Integrative clustering

(iCluster)[37] of mRNA, miRNA, methylation, SCNA, and mutation data demonstrated concordance with the mRNA expression-subtypes and associated alterations. (Figure 3, Supplementary Figure S7. A.1 and Supplementary Methods, section 7). Independent correlation of somatic mutations, copy number alterations and gene expression signatures revealed significant subtype associations with alterations in the *TP53, PI3K, RB1* and *NFE2L2/KEAP1* pathways (Supplementary Figure S7. B.1 and Supplementary Methods, section 7).

Integrated analysis of the tumor suppressor locus for *CDKN2A*

Integrated multi-platform analyses showed that *CDKN2A*, a known tumor suppressor gene in lung SqCC[38] that encodes the INK4A/p16 and ARF/p14 proteins, is inactivated in 72% of cases of lung SqCC (Figure 4a and Supplementary Data S7.1)—by epigenetic silencing by methylation (21%), inactivating mutation (18%), exon 1β skipping (4%), and homozygous deletion (29%).

Analysis of mRNA expression across the *CDKN2A* locus revealed four distinct patterns of *p16INK4* and *ARF* expression: complete absence of both *p16INK4* and *ARF* (33%); expression of high levels of both *p16INK4* and *ARF* (31%); high expression of *ARF* and absent *p16INK4* (31%); or expression of a transcript that represents a splicing of exon E1b from *ARF* with the shared E3 of *ARF* and *p16INK4*, generating a premature stop codon (4%) (Supplementary Figure S4.4). Almost all of the cases completely lacking *p16INK4* and *ARF* expression showed homozygous deletion (Figure 4b and Supplemental Data S7.1). In one case, *p16INK4* expression was detected but analysis of WGS data demonstrated an intergenic fusion event that resulted in detectable transcription between exon 1a of *p16INK4* and exon 18 of *KIAA1797* (Figures 4b, 4c). Interestingly, combined analysis of WGS and RNA-sequencing data identified tumor suppressor gene inactivation by intra- or interchromosomal rearrangement in *PTEN, NOTCH1, ARID1A, CTNNA2, VHL* and *NF1*, in 8 additional cases (Supplementary Data S3.1 and S4.1).

In addition to homozygous deletion, there are frequent mutational events in *CDKN2A* (Figure 4b and Supplementary Data 7.1). These account for 45% of the 56 cases with high *p16INK4* and *ARF* expression. Furthermore, methylation of the exon 1a promoter accounts for many other cases of *CDKN2A* inactivation (70% of lung SqCCs with *ARF* expression in the absence of detectable *p16INK4*). Seven additional tumors in the high-*ARF*/low-*INK4A* group had documented mutations of *INK4A*, primarily nonsense, suggesting nonsense-mediated decay as a mechanism. Of the 28% of tumors without *CDKN2A* alterations, *RB1* mutations were identified in eight cases and *CDK6* amplification in one case (Figure 4d).

Therapeutic targets

Molecularly targeted agents are now commonly used in patients with adenocarcinoma of the lung while no effective targeted agents have been developed specifically for lung SqCCs[13]. We analyzed our genomic data for evidence of the two common genomic alterations in adenocarcinomas of the lung: *EGFR* and *KRAS* mutation. Only one sample had a *KRAS* codon 61 mutation, and there were no exon 19 deletions or L858R mutations in *EGFR*.

However, amplifications of *EGFR* were found in 7% of cases as were two instances of the L861Q *EGFR* mutation, which confers sensitivity to erlotinib and gefitinib[39].

The presence of new potential therapeutic targets in lung SqCC was suggested by the observation that 96% of tumors (171 of 178) contain one or more mutations in tyrosine kinases, serine/threonine kinases, PI3K catalytic and regulatory subunits, nuclear hormone receptors, G protein-coupled receptors, proteases and tyrosine phosphatases (Supplementary Figure S7.D.1a and Supplementary Data S7.2 and S7.3). 50–77% of the mutations were predicted to have a medium or high functional impact as determined by Mutation Assessor score[40] (Supplementary Figure S7.D.1a) and 39% of tyrosine and 42% of serine/threonine kinase mutations were located in the kinase domain. Many of the alterations were in known oncogenes and tumor suppressors, as defined in the COSMIC database (Supplementary Data S7.3).

We selected potential therapeutic targets based on several features, including (i) availability of an FDA-approved targeted therapeutic agent or one under study in current clinical trials (Supplementary Data S7.2) (ii) confirmation of the altered allele in RNA-sequencing; and (iii) Mutation Assessor score[40]. Using those criteria, we identified 114 cases with somatic alteration of a potentially targetable gene (64%) (Supplementary Figure S7.D.1b and Supplementary Data S7.4). Among these we identified three families of tyrosine kinases, the ERBBs, FGFRs and JAKs, all of which were found to be mutated and/or amplified[41]. As discussed for *EGFR*, the mutational spectra in these potential therapeutic targets differed from those in lung adenocarcinoma (Supplementary Figure S7.D.2).[42]

To complement a gene-centered search for potential therapeutic targets, we analyzed core cellular pathways known to represent potential therapeutic vulnerabilities: PI3K/AKT, RTK and RAS. Analysis of the 178 lung SqCCs revealed alteration in at least one of those pathways in 69% of samples after restriction of the analysis to mutations confirmed by RNA-sequencing and to amplifications associated with overexpression of the target gene (Figure 5). Mutational events that have been curated in COSMIC are additionally shown in Supplementary Figure S7D.2 as is the distribution of mutations, amplifications and overexpression of the genes depicted in Figure 5C. Specifically, one of three components of the PI3K/Akt pathway (*PIK3CA*, *PTEN* or *AKT3*) was altered in 47% of tumors and receptor-tyrosine kinase (RTK) signaling likely altered by *EGFR* amplification, *BRAF* mutation or *FGFR* amplification or mutation in 26% of tumors. (Figures 5 and Supplementary Figure S7.D.3). Alterations in the PI3K/Akt pathway genes were mutually exclusive with *EGFR* alterations as identified by MEMo[43] (Supplementary Figure S7.D.4.). While the dependence of lung SqCC on many of these individual alterations remains to be defined functionally, this analysis suggests new areas for potential therapeutic development in this cancer.

Discussion

Lung SqCCs are characterized by a high overall mutation rate of 8.1 mutations/Mb and marked genomic complexity. Similar to high-grade serous ovarian carcinoma[17], almost all lung SqCCs display somatic mutation of *TP53*. There were also frequent alterations in the

CDKN2A/RB1, *NFE2L2/KEAP1/CUL3*, *PI3K/AKT* and *SOX2/TP63/NOTCH1* pathways, providing evidence of common dysfunction in cell cycle control, response to oxidative stress, apoptotic signaling and/or squamous cell differentiation. Pathway alterations clustered according to expression-subtype in many cases, suggesting that those subtypes have a biological basis.

A role for somatic mutation in the cancer hallmark of "avoiding immune destruction[44] is suggested by the presence of inactivating mutations in the *HLA-A* gene. Somatic loss-of-function alterations of *HLA-A* have not been reported previously in genomic studies of lung cancer. Given the recently reported efficacy of anti-Programmed Death 1 (PD1)[45] and anti-Cytotoxic T-Lymphocyte Antigen 4 (CTLA4) antibodies in NSCLC[46], these *HLA-A* mutations suggest a possible role for genotypic selection of patients for immunotherapies.

Targeted kinase inhibitors have been successfully used for treatment of lung adenocarcinoma but minimally so in lung SqCC. The observations reported here suggest that a detailed understanding of the possible targets in lung SqCCs may identify targeted therapeutic approaches. While *EGFR* and *KRAS* mutations, the two most common oncogenic aberrations in lung adenocarcinoma, are extremely rare in lung SqCC, alterations in the FGFR kinase family are common in lung SqCC. Lung SqCCs also share many alterations in common with head and neck squamous cell carcinomas without evidence of human papilloma virus (HPV) infection, including mutation in *PIK3CA*, *PTEN*, *TP53*, *CDKN2A*, *NOTCH1* and *HRAS*[22,24], suggesting that the biology of these two diseases may be similar.

The current study has identified a potentially targetable gene or pathway alteration in the majority of the lung SqCC samples studied. The data presented here can help organize efforts to analyze lung SqCC clinical tumor specimens for a panel of specific, actionable mutations to select patients for appropriately targeted clinical trials. These data could thereby help to facilitate effective personalized therapy 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 commercial technology for capture and sequencing of exomes from tumor DNA and normal DNA 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 lesions. GISTIC[23,24] analysis of the circular-binary-segmented Affymetrix SNP 6.0 copy number data was used to identify recurrent amplification and deletion peaks. Consensus clustering approaches were used to analyze mRNA, miRNA and methylation subtypes using previous approaches[20,21,34,38,41,44].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. [Accessed February, 2012] WHO Statistics. <http://www.who.int/mediacentre/factsheets/fs297/en/>
2. Soda M, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007; 448:561–566. nature05945 [pii]. 10.1038/nature05945 [PubMed: 17625570]
3. Paez JG, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science. 2004; 304:1497–1500. [pii]. 10.1126/science.10993141099314 [PubMed: 15118125]
4. Lynch TJ, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med. 2004; 350:2129–2139. NEJMoa040938 [pii]. 10.1056/NEJMoa040938 [PubMed: 15118073]
5. Pao W, et al. EGF receptor gene mutations are common in lung cancers from “never smokers” and are associated with sensitivity of tumors to gefitinib and erlotinib. Proc Natl Acad Sci U S A. 2004; 101:13306–13311. [pii]. 10.1073/pnas.0405220101040522010 1 [PubMed: 15329413]
6. Felip E, Gridelli C, Baas P, Rosell R, Stahel R. Metastatic non-small-cell lung cancer: consensus on pathology and molecular tests, first-line, second-line, and third-line therapy: 1st ESMO Consensus Conference in Lung Cancer; Lugano 2010. Ann Oncol. 2011; 22:1507–1519. mdr150 [pii]. 10.1093/annonc/mdr150 [PubMed: 21536661]
7. Ju YS, et al. A transforming KIF5B and RET gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. Genome Res. 2012 gr.133645.111 [pii]. 10.1101/gr.133645.111
8. Rekhtman N, et al. Clarifying the spectrum of driver oncogene mutations in biomarker-verified squamous carcinoma of lung: lack of EGFR/KRAS and presence of PIK3CA/AKT1 mutations. Clin Cancer Res. 2012 1078-0432.CCR-11-2109 [pii]. 10.1158/1078-0432.CCR-11-2109
9. Bass AJ, et al. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. Nat Genet. 2009; 41:1238–1242. ng.465 [pii]. 10.1038/ng.465 [PubMed: 19801978]
10. Ramos AH, et al. Amplification of chromosomal segment 4q12 in non-small cell lung cancer. Cancer Biol Ther. 2009; 8:2042–2050. 9764 [pii]. [PubMed: 19755855]
11. Shibata T, et al. Cancer related mutations in NRF2 impair its recognition by Keap1-Cul3 E3 ligase and promote malignancy. Proc Natl Acad Sci U S A. 2008; 105:13568–13573. 0806268105 [pii]. 10.1073/pnas.0806268105 [PubMed: 18757741]
12. Kan Z, et al. Diverse somatic mutation patterns and pathway alterations in human cancers. Nature. 2010; 466:869–873. nature09208 [pii]. 10.1038/nature09208 [PubMed: 20668451]
13. Hammerman PS, Sos ML, Ramos A, Xu C. Mutations in the DDR2 kinase gene identify a novel therapeutic target in squamous cell lung cancer. Cancer Discovery. 2011; 1
14. Weiss J, et al. Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. Sci Transl Med. 2010; 2:62ra93. 2/62/62ra93 [pii]. 10.1126/scitranslmed.3001451
15. Dutt A, et al. Inhibitor-Sensitive FGFR1 Amplification in Human Non-Small Cell Lung Cancer. PLoS One. 2011; 6:e20351. PONE-D-11-00027 [pii]. 10.1371/journal.pone.0020351 [PubMed: 21666749]
16. Kenfield SA, Wei EK, Stampfer MJ, Rosner BA, Colditz GA. Comparison of aspects of smoking among the four histological types of lung cancer. Tob Control. 2008; 17:198–204. tc.2007.022582 [pii]. 10.1136/tc.2007.022582 [PubMed: 18390646]
17. Integrated genomic analyses of ovarian carcinoma. Nature. 2011; 474:609–615. nature10166 [pii]. 10.1038/nature10166 [PubMed: 21720365]
18. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008; 455:1061–1068. nature07385 [pii]. 10.1038/nature07385 [PubMed: 18772890]
19. Tonon G, et al. High-resolution genomic profiles of human lung cancer. Proc Natl Acad Sci U S A. 2005; 102:9625–9630. 0504126102 [pii]. 10.1073/pnas.0504126102 [PubMed: 15983384]
20. Beroukhim R, et al. The landscape of somatic copy-number alteration across human cancers. Nature. 2010; 463:899–905. nature08822 [pii]. 10.1038/nature08822 [PubMed: 20164920]

21. Mermel CH, et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 2011; 12:R41. gb-2011-12-4-r41 [pii]. 10.1186/gb-2011-12-4-r41 [PubMed: 21527027]
22. Stransky N, et al. The mutational landscape of head and neck squamous cell carcinoma. Science. 2011; 333:1157–1160. science.1208130 [pii]. 10.1126/science.1208130 [PubMed: 21798893]
23. Chapman MA, et al. Initial genome sequencing and analysis of multiple myeloma. Nature. 2011; 471:467–472. nature09837 [pii]. 10.1038/nature09837 [PubMed: 21430775]
24. Agrawal N, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science. 2011; 333:1154–1157. science.1206923 [pii]. 10.1126/science.1206923 [PubMed: 21798897]
25. Bass AJ, et al. Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion. Nat Genet. 2011; 43:964–968. ng.936 [pii]. 10.1038/ng.936 [PubMed: 21892161]
26. Berger MF, et al. The genomic complexity of primary human prostate cancer. Nature. 2011; 470:214–220. nature09744 [pii]. 10.1038/nature09744 [PubMed: 21307934]
27. Stephens PJ, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature. 2009; 462:1005–1010. nature08645 [pii]. 10.1038/nature08645 [PubMed: 20033038]
28. Singh A, et al. Dysfunctional KEAP1-NRF2 interaction in non-small-cell lung cancer. PLoS Med. 2006; 3:e420. 06-PLME-RA-0322R2 [pii]. 10.1371/journal.pmed.0030420 [PubMed: 17020408]
29. Singh A, Bodas M, Wakabayashi N, Bunz F, Biswal S. Gain of Nrf2 function in non-small-cell lung cancer cells confers radioresistance. Antioxid Redox Signal. 2010; 13:1627–1637.10.1089/ars.2010.3219 [PubMed: 20446773]
30. Vaske CJ, et al. Inference of patient-specific pathway activities from multidimensional cancer genomics data using PARADIGM. Bioinformatics. 2010; 26:i237–245. btq182 [pii]. 10.1093/bioinformatics/btq182 [PubMed: 20529912]
31. Aster JC, Blacklow SC, Pear WS. Notch signalling in T-cell lymphoblastic leukaemia/lymphoma and other haematological malignancies. J Pathol. 2011; 223:262–273.10.1002/path.2789 [PubMed: 20967796]
32. Wang NJ, et al. Loss-of-function mutations in Notch receptors in cutaneous and lung squamous cell carcinoma. Proc Natl Acad Sci U S A. 2011; 108:17761–17766. 1114669108 [pii]. 10.1073/pnas.1114669108 [PubMed: 22006338]
33. Osada H, Tatematsu Y, Yatabe Y, Horio Y, Takahashi T. ASH1 gene is a specific therapeutic target for lung cancers with neuroendocrine features. Cancer Res. 2005; 65:10680–10685. 65/23/10680 [pii]. 10.1158/0008-5472.CAN-05-1404 [PubMed: 16322211]
34. Wilkerson MD, et al. Lung squamous cell carcinoma mRNA expression subtypes are reproducible, clinically important, and correspond to normal cell types. Clin Cancer Res. 2010; 16:4864–4875. 1078-0432.CCR-10-0199 [pii]. 10.1158/1078-0432.CCR-10-0199 [PubMed: 20643781]
35. Bishop JA, et al. p40 (DeltaNp63) is superior to p63 for the diagnosis of pulmonary squamous cell carcinoma. Mod Pathol. 2011 modpathol2011173 [pii]. 10.1038/modpathol.2011.173
36. Massion PP, et al. Significance of p63 amplification and overexpression in lung cancer development and prognosis. Cancer Res. 2003; 63:7113–7121. [PubMed: 14612504]
37. Shen R, Olshen AB, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. Bioinformatics. 2009; 25:2906–2912. btp543 [pii]. 10.1093/bioinformatics/btp543 [PubMed: 19759197]
38. Wikman H, Kettunen E. Regulation of the G1/S phase of the cell cycle and alterations in the RB pathway in human lung cancer. Expert Rev Anticancer Ther. 2006; 6:515–530.10.1586/14737140.6.4.515 [PubMed: 16613540]
39. Kancha RK, Peschel C, Duyster J. The epidermal growth factor receptor-L861Q mutation increases kinase activity without leading to enhanced sensitivity toward epidermal growth factor receptor kinase inhibitors. J Thorac Oncol. 2011; 6:387–392. 01243894-201102000-00020 [pii]. 10.1097/JTO.0b013e3182021f3e [PubMed: 21252719]
40. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 2011; 39:e118. gkr407 [pii]. 10.1093/nar/gkr407 [PubMed: 21727090]

41. Govindan R. Summary of the proceedings from the 10th annual meeting of molecularly targeted therapy in non-small cell lung cancer. J Thorac Oncol. 2010; 5:S433. 01243894-201012002-00001 [pii]. 10.1097/JTO.0b013e3181fd6fe1 [PubMed: 21102234]

42. Ding L, et al. Somatic mutations affect key pathways in lung adenocarcinoma. Nature. 2008; 455:1069–1075. nature07423 [pii]. 10.1038/nature07423 [PubMed: 18948947]

43. Ciriello G, Cerami E, Sander C, Schultz N. Mutual exclusivity analysis identifies oncogenic network modules. Genome Res. 2012; 22:398–406. gr.125567.111 [pii]. 10.1101/gr.125567.111 [PubMed: 21908773]

44. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011; 144:646–674. S0092-8674(11)00127-9 [pii]. 10.1016/j.cell.2011.02.013 [PubMed: 21376230]

45. Brahmer JR, et al. Phase I study of single-agent anti-programmed death-1 (MDX-1106) in refractory solid tumors: safety, clinical activity, pharmacodynamics, and immunologic correlates. J Clin Oncol. 2010; 28:3167–3175. JCO.2009.26.7609 [pii]. 10.1200/JCO.2009.26.7609 [PubMed: 20516446]

46. Lynch TJ, Bondarenko IN, Luft A, Serwatowski P, Barlesi F, Chacko RT, Sebastian M, Siegel J, Cuillerot J, Reck M. Phase II trial of ipilimumab (IPI) and paclitaxel/carboplatin (P/C) in first-line stage IIIb/IV non-small cell lung cancer (NSCLC). J Clin Oncol. 2010; 28:7531.

TCGA Network Authors and Affiliations

Genome Sequencing Centers

Broad Institute – Peter S Hammerman([1,5]), Michael S. Lawrence([1]), Douglas Voet([1]), Rui Jing([1]), Kristian Cibulskis([1]), Andrey Sivachenko([1]), Petar Stojanov([1]), Aaron McKenna([1]), Eric S Lander([1,2,3]), Stacey Gabriel([4]), Gad Getz([1,4]), Carrie Sougnez([4]), Marcin Imielinski ([1]), Elena Helman ([1]), Bryan Hernandez ([1]), Nam H. Pho ([1]), Matthew Meyerson([1,5,6])

Genome Characterization Centers

BC Cancer Agency – Andy Chu([7]), Hye-Jung E. Chun([7]), Andrew J. Mungall([7]), Erin Pleasance([7]), A. Gordon Robertson([7]), Payal Sipahimalani([7]), Dominik Stoll([7]), Miruna Balasundaram([7]), Inanc Birol([7]), Yaron S.N. Butterfield([7]), Eric Chuah([7]), Robin J.N. Coope([7]), Richard Corbett([7]), Noreen Dhalla([7]), Ranabir Guin([7]), An He([7]), Carrie Hirst([7]), Martin Hirst([7]), Robert A. Holt([7]), Darlene Lee([7]), Haiyan I. Li([7]), Michael Mayo([7]), Richard A. Moore([7]), Karen Mungall([7]), Ka Ming Nip([7]), Adam Olshen([8]), Jacqueline E. Schein([7]), Jared R. Slobodan([7]), Angela Tam([7]), Nina Thiessen([7]), Richard Varhol([7]), Thomas Zeng([7]), Yongjun Zhao([7]), Steven J.M. Jones([7]), Marco A. Marra([7]) *Broad Institute* – Gordon Saksena([1]), Andrew D. Cherniack([1]), Stephen E. Schumacher([1,5]), Barbara Tabak([1,5]), Scott L. Carter([1]), Nam H. Pho([1]), Huy Nguyen([1]), Robert C. Onofrio([4]), Andrew Crenshaw([1]), Kristin Ardlie([4]), Rameen Beroukhim([1,5]), Wendy Winckler([1,4]), Peter S Hammerman ([1,5]) Gad Getz([1,4]), Matthew Meyerson([1,5,6]), *Brigham*

[1]The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University Cambridge, MA 02142 USA
[5]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215 USA
[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142 USA
[3]Department of Systems Biology, Harvard University, Boston, MA 02115 USA
[4]Genetic Analysis Platform, The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02142 USA
[6]Department of Pathology, Harvard Medical School, Boston, MA 02115 USA
[7]Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, BC V5Z, Canada
[8]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA 94143 USA

and Women's Hospital and Harvard Medical School – Alexei Protopopov ([9,91]), Jianhua Zhang ([9,91]), Angela Hadjipanayis ([10,11]), Semin Lee ([12]), Ruibin Xi ([12]), Lixing Yang([12]), Xiaojia Ren([9,10,11]), Hailei Zhang ([1,9]), Sachet Shukla ([1,9]), Peng-Chieh Chen ([10,11]), Psalm Haseley ([11,12]), Eunjung Lee ([11,12]), Lynda Chin ([1,5,9,14,91]), Peter J. Park ([11,12,13]), Raju Kucherlapati ([10,11]), *Memorial Sloan-Kettering Cancer Center (TCGA Pilot Phase only)* Nicholas D. Socci ([27]), Yupu Liang ([27]), Nikolaus Schultz ([27]), Laetitia Borsu ([27]), Alex E. Lash ([27]), Agnes Viale ([27]), Chris Sander ([27]), Marc Ladanyi ([28, 30]), *University of North Carolina, Chapel Hill*: J. Todd Auman([15,16]), Katherine A. Hoadley([17,18,19]), Matthew D. Wilkerson([19]), Yan Shi([19]), Christina Liquori ([19]), Shaowu Meng([19]), Ling Li([19]), Yidi J. Turman([19]), Michael D. Topal([18,19]), Donghui Tan([20]), Scot Waring([19]), Elizabeth Buda([19]), Jesse Walsh([19]), Corbin D. Jones([21]), Piotr A. Mieczkowski([17]), Darshan Singh([24]), Junyuan Wu([19]), Anisha Gulabani([19]), Peter Dolina([19]), Tom Bodenheimer([19]), Alan P. Hoyle([19]), Janae V. Simons([19]), Matthew G. Soloway([19]), Lisle E. Mose([18]), Stuart R. Jefferys([18]), Saianand Balu([19]), Brian D. O'Connor([19]), Jan F. Prins([22]), Jinze Liu ([23]), Derek Y. Chiang([17,19]), D. Neil Hayes([19,24]), Charles M. Perou([17,18,19]), *University of Southern California / Johns Hopkins* – Leslie Cope ([26]), Ludmila Danilova ([26]), Daniel J. Weisenberger([25]), Dennis T. Maglinte([25]), Fei Pan([25]), David J. Van Den Berg([25]), Timothy Triche Jr([25]), Stephen B. Baylin([26]), & Peter W. Laird([25])

Genome Data Analysis Centers

Broad Institute – Gad Getz([1]), Michael Noble([1]), Doug Voet([1]), Gordon Saksena([1]), Nils Gehlenborg([1,12]), Daniel DiCara([1]), Jinhua Zhang([9, 91]), Hailei Zhang([1]), Chang-Jiun Wu([5,91]), Spring Yingchun Liu([1]), Michael S. Lawrence([1]), Lihua Zou([1]), Andrey

[9]Belfer Institute for Applied Cancer Science, Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115 USA
[91]These authors are currently affiliated with the Institute for Applied Cancer Science, Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX 77030 USA
[10]Department of Genetics, Harvard Medical School, Boston, MA 02115 USA
[11]Division of Genetics, Brigham and Women's Hospital, Boston, MA 02115 USA
[12]The Center for Biomedical Informatics, Harvard Medical School, Boston, MA 02115 USA
[14]Department of Dermatology, Harvard Medical School, Boston, MA 02115 USA
[13]Informatics Program, Children's Hospital, Boston, MA 02115 USA
[27]Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 USA
[28]Department of Molecular Oncology, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 USA
[30]Department of Pathology and Human Oncology & Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 USA
[15]Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA
[16]Institute for Pharmacogenetics and Individualized Therapy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA
[17]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA
[18]Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, Chapel Hill, NC 27599 USA
[19]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA
[20]Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA
[21]Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA
[24]Department of Internal Medicine, Division of Medical Oncology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA
[22]Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA
[23]Department of Computer Science, University of Kentucky, Lexington, KY.40506.
[26]Cancer Biology Division, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, Baltimore, MD 21231 USA
[25]University of Southern California Epigenome Center, University of Southern California, Los Angeles, CA 90033 USA

Sivachenko([1]), Pei Lin([1]), Petar Stojanov([1]), Rui Jing([1]), Juok Cho([1]), Marc-Danie Nazaire([1]), Jim Robinson([1]), Helga Thorvaldsdottir([1]), Jill Mesirov([1]), Peter J. Park([11,12,13]), Lynda Chin([1,5,9,14,91]) *Memorial Sloan-Kettering Cancer Center* – Nikolaus Schultz([27]), Rileen Sinha([27]), Giovanni Ciriello([27]), Ethan Cerami([27]), Benjamin Gross([27]), Anders Jacobsen([27]), Jianjiong Gao([27]), B. Arman Aksoy([27]), Nils Weinhold([27]), Ricardo Ramirez([27]), Barry S. Taylor([27]), Yevgeniy Antipin([27]), Boris Reva([27]), Ronglai Shen([29]), Qianxing Mo([29]), Venkatraman Seshan([29]), Paul K Paik ([31]), Marc Ladanyi([28, 30]), Chris Sander([27]) *The University of Texas MD Anderson Cancer Center* – Rehan Akbani([32]), Nianxiang Zhang([32]), Bradley M. Broom([32]), Tod Casasent([32]), Anna Unruh([32]), Chris Wakefield([32]), R. Craig Cason ([33]), Keith A. Baggerly([32]), John N. Weinstein([32,34]), *University of California, Santa Cruz / Buck Institute* – David Haussler([35,36]), Christopher C. Benz([37]), Joshua M. Stuart([35]), Jingchun Zhu([35]), Christopher Szeto([35]), Gary K. Scott([37]), & Christina Yau([37]). Sam Ng([35]), Ted Goldstein([35]), Peter Waltman([35]), Artem Sokolov([35]), Kyle Ellrott([35]), Eric A. Collisson([38]), Daniel Zerbino([35]), Christopher Wilks([35]), Singer Ma([35]), Brian Craft([35]), *University of North Carolina, Chapel Hill*: Matthew D. Wilkerson([19]), J. Todd Auman([15,16]), Katherine A. Hoadley([17,18,19]), Ying Du([19]), Christopher Cabanski ([19]), Vonn Walter ([19]), Darshan Singh ([19]) Junyuan Wu([19]), Anisha Gulabani([19]), Tom Bodenheimer([19]), Alan P. Hoyle([19]), Janae V. Simons([19]), Matthew G. Soloway([19]), Lisle E. Mose([18]), Stuart R. Jefferys([18]), Saianand Balu([19]), J.S. Marron ([94]), Yufeng Liu ([95]), Kai Wang ([23]), Jinze Liu ([23]), Jan F. Prins([19]), D. Neil Hayes([19,24]), Charles M. Perou([17,18,19]), *Baylor College of Medicine* - Chad J Creighton([40]), Yiqun Zhang([40])

Pathology Committee

William D. Travis ([41]), Natasha Rekhtman([41]), Joanne Yi([42]), Marie C. Aubry([42]), Richard Cheney([43]), Sanja Dacic([44]), Douglas Flieder([45]), William Funkhouser([46]), Peter Illei([47]), Jerome Myers ([48]), Ming Sound Tsao([49])

[29]Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 USA
[31]Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 USA
[32]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030 USA
[33]Division of Pathology and Laboratory Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX 77030 USA
[34]Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030 USA
[35]Department of Biomolecular Engineering and Center for Biomolecular Science and Engineering, University of California Santa Cruz, Santa Cruz, CA 95064 USA
[36]Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, CA 95064 USA
[37]Buck Institute for Age Research, Novato, CA 94945 USA
[38]Division of Hematology/Oncology, University of California San Francisco, San Francisco, CA 94143 USA
[94]Department of Statistics and Operations Research, University of North Carolina Medical Center, Chapel Hill, NC 27599 USA
[95]Carolina Center for Genome Sciences, University of North Carolina Medical Center, Chapel Hill, NC 27599 USA
[40]Human Genome Sequencing Center and Dan L. Duncan Cancer Center Division of Biostatistics, Baylor College of Medicine, Houston, TX 77030 USA
[41]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY 10065 USA
[42]Department of Pathology, Mayo Clinic, Rochester, MN, 55905 USA
[43]Department of Pathology, Roswell Park Cancer Institute, Buffalo, NY 14263 USA
[44]Department of Pathology, University of Pittsburgh Cancer Center, Pittsburgh, PA 15213 USA
[45]Department of Pathology, Fox Chase Cancer Center, Philadelphia, PA 19111 USA
[46]Department of Pathology, University of North Carolina Medical Center, Chapel Hill, NC 27599 USA
[47]Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD 21287 USA
[48]Department of Pathology, Penrose-St. Francis Health System, Colorado Springs, CO 80907 USA
[49]Department of Pathology and Medical Biophysics, Ontario Cancer Institute and Princess Margaret Hospital, Toronto, ON Canada

Biospecimen Core Resources

International Genomics Consortium – Robert Penny([50]), David Mallery ([50]), Troy Shelton([50]), Martha Hatfield([50]), Scott Morris([50]), Peggy Yena([50]), Candace Shelton([50]), Mark Sherman([50]),& Joseph Paulauskis([50])

Disease working group

Matthew Meyerson([1,5,6]), Stephen B. Baylin([26]), Ramaswamy Govindan([51]), Rehan Akbani([32]), Ijeoma Azodo([60]), David Beer([52]), Ron Bose([51]), Lauren A. Byers ([54]) David Carbone([53]), Li-Wei Chang ([51]), Derek Chiang([17,19]), Andy Chu([7]), Elizabeth Chun([7]), Eric Collisson([38]), Leslie Cope([26]), Chad J Creighton([40]), Ludmila Danilova([26]), Li Ding([51]), Gad Getz([1,4]), Peter S Hammerman([1,5]), D Neil Hayes([19,23]), Bryan Hernandez([1]), James Herman([26]), John Heymach([54]), Cristiane Ida([42]), Marcin Imielinski([1,6]), Bruce Johnson([5]), Igor Jurisica([55]), Jacob Kaufman([53]), Farhad Kosari([60]), Raju Kucherlapati([10,11]), David Kwiatkowski([5]), Marc Ladanyi([28, 30]), Michael Lawrence([1]), Christopher A Maher([51]), Andy Mungall([7]), Sam Ng([35]), William Pao([53]), Martin Peifer([56, 92]), Robert Penny([50]), Gordon Robertson([7]), Valerie Rusch([57]) Chris Sander([27]), Nikolaus Schultz([27]), Ronglai Shen([29]), Jill Siegfried([58]), Rileen Sinha([27]), Andrey Sivachenko([1]), Carrie Sougnez([4]), Dominik Stoll([7]), Joshua Stuart([35]), Roman K Thomas([56, 92, 93]), Sandra Tomaszek([60]), Ming-Sound Tsao([49]), William D Travis([41]), Charles Vaske([35]), John N Weinstein([32,34]), Daniel Weisenberger([25]), David Wheeler([59]), Dennis A Wigle([60]), Matthew D Wilkerson([19]), Christopher Wilks([25]), Ping Yang([60]), Jianjua (John) Zhang([9])

Data Coordination Center

Mark A. Jensen([61]), Robert Sfeir([61]), Ari B. Kahn([61]), Anna L. Chu([61]), Prachi Kothiyal([61]), Zhining Wang([61]), Eric E. Snyder([61]), Joan Pontius([61]), Todd D. Pihl([61]), Brenda Ayala([61]), Mark Backus([61]), Jessica Walton([61]), Julien Baboud([61]), Dominique L. Berton([61]), Matthew C. Nicholls([61]), Deepak Srinivasan([61]), Rohini Raman([61]), Stanley Girshik([61]), Peter A. Kigonya([61]), Shelley Alonso([61]), Rashmi N. Sanbhadti([61]), Sean P. Barletta([61]), John M. Greene([61]) & David A. Pot([61])

[50]International Genomics Consortium, Phoenix, AZ 85004 USA
[51]Division of Oncology, Department of Medicine and The Genome Institute, Washington University School of Medicine, St. Louis, MO 63110 USA
[60]Mayo Clinic College of Medicine, Rochester, MN 55905 USA
[52]Department of Surgery, University of Michigan, Ann Arbor, MI 48109 USA
[54]The University of Texas MD Anderson Cancer Center, Houston, TX 77030 USA
[53]Departments of Hematology/Oncology and Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN 37232 USA
[55]Ontario Cancer Institute IBM Life Sciences Discovery Centre, Toronto, ON Canada
[56]Department of Translational Genomics, University of Cologne, Cologne, Germany
[92] Max Planck Institute for Neurological Research, Cologne, Germany
[57]Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY 10065 USA
[58]Department of Pharmacology and Chemical Biology, University of Pittsburgh Medical Center, Pittsburgh, PA 15232 USA
[93] Department of Translational Cancer Genomics, Center of Integrated Oncology, University of Cologne, Cologne, Germany
[59]Human Genome Sequencing Center, Baylor College of Medcine, Houston, TX 77030 USA
[61]SRA International, Fairfax, VA 22033 USA

Tissue Source Sites

Ming Tsao([49]), Bizhan Bandarchi-Chamkhaleh([49]), Jeff Boyd([45]), JoEllen Weaver([45]), Dennis A. Wigle ([60]), Ijeoma A. Azodo([60]), Sandra C. Tomaszek([60]), Marie Christine Aubry([62]), Christiane M. Ida([62]), Ping Yang([63]), Farhad Kosari([64]), Malcolm V. Brock([65]), Kristen Rogers([65]), Marian Rutledge([66]), Travis Brown([65]), Beverly Lee([66]), James Shin([67]), Dante Trusty([67]), Rajiv Dhir([68]), Jill M. Siegfried([69]), Olga Potapova([70]), Konstantin V. Fedosenko([71]), Elena Nemirovich-Danchenko([70]), Valerie Rusch([57]), Maureen Zakowski([72]), Mary V. Iacocca ([73]), Jennifer Brown([73]), Brenda Rabeno([73]), Christine Czerwinski([73]), Nicholas Petrelli([73]), Zhen Fan([74]), Nicole Todaro([74]), John Eckman([48]), Jerome Myers([48]), W. Kimryn Rathmell([19]), Leigh B. Thorne([75]), Mei Huang([75]), Lori Boice([75]), Ashley Hill([19]), Robert Penny([50]), David Mallery([50]), Erin Curley([50]), Candace Shelton([50]), Peggy Yena([50]), Carl Morrison([43]), Carmelo Gaudioso([43]), John M.S. Bartlett([76]), Sugy Kodeeswaran([76]), Brent Zanke([76]), Harman Sekhon([77]), Kerstin David([78]), Hartmut Juhl([79]), Xuan Van Le([80]), Bernard Kohl([80]), Richard Thorp([80]), Nguyen Viet Tien([81]), Nguyen Van Bang([82]), Howard Sussman([83]), Bui Duc Phu([82]), Richard Hajek([84]), Nguyen Phi Hung([85]) & Khurram Z. Khan([86]), Thomas Muley ([96])

Project Team

National Cancer Institute – Kenna R. Mills Shaw([87]), Margi Sheth ([87]), Liming Yang ([87]), Ken Buetow([88]), Tanja Davidsen([88]), John A. Demchok([87]), Greg Eley([88]), Martin Ferguson([89]), Laura A. L. Dillon([87]) & Carl Schaefer([88]) *National Human Genome*

[62]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester MN 55905 USA
[63]Department of Health Sciences Research, Mayo Clinic, Rochester MN 55905 USA
[64]Center for Individualized Medicine, Mayo Clinic, Rochester MN 55905 USA
[65]Department of Surgery, 600 North Wolfe Street, Baltimore, MD 21287 USA
[66]Department of Oncology, 600 North Wolfe Street, Baltimore, MD 21287 USA
[67]Department of Pathology, 600 North Wolfe Street, Baltimore, MD 21287 USA
[68]University of Pittsburgh, Pittsburgh, PA, 15213 USA
[69]Department of Pharmacology and Chemical Biology, University of Pittsburgh Cancer Institute, Pittsburgh, PA, 15213 USA
[70]Cureline, Inc., South San Francisco, CA 94080 USA
[71]City Clinical Oncology Dispensary, St. Petersburg, Russia
[72]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY 10065 USA
[73]Helen F. Graham Cancer Center, Newark, Delaware 19713 USA
[74]St. Joseph Medical Center, Towson, MD 21204 USA
[75]UNC Tissue Procurement Facility, Department of Pathology, UNC Lineberger Cancer Center, Chapel Hill, NC 27599, USA
[76]Ontario Tumour Bank, Ontario Institute for Cancer Research, Toronto, Ontario, M5G 0A3, Canada
[77]Ontario Tumour Bank – Ottawa site, The Ottawa Hospital, Ottawa, Ontario, K1H 8L6, Canada
[78]Indivumed GmbH, Hamburg, Germany
[79]Indivumed Inc, Kensington, MD 20895, USA
[80]ILSBio, LLC, Chestertown, MD 21620, USA
[81]Ministry of Health, Hanoi, Vietnam
[82]Hue Central Hospital, Hue City, Vietnam
[83]Stanford University Medical Center, Stanford, CA 94305 USA
[84]Center for Minority Health Research, University of Texas, M.D. Anderson Cancer Center, Houston, TX 77030 USA
[85]National Cancer Institute, Hanoi, Vietnam
[86]ILSBio LLC, Karachi, Pakistan
[96]ThoraxKlinik, Heidelberg University Hospital, Heidelberg, Germany.
[87]The Cancer Genome Atlas Program Office, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892 USA
[88]Center for Biomedical Informatics and Information Technology (CBIIT), National Cancer Institute, National Institutes of Health, Rockville, MD 20852 USA
[89]MLF Consulting, Arlington, MA 02474 USA

Research Institute: Mark S. Guyer([90]), Bradley A. Ozenberger([90]), Jacqueline D. Palchik([90]), Jane Peterson([90]), Heidi J. Sofia([90]), & Elizabeth Thomson([90])

Writing Committee

Peter S Hammerman([1,5]), D Neil Hayes([19,23]), Matthew D Wilkerson([19]), Nikolaus Schultz([27]), Ron Bose([51]), Andy Chu([7]), Eric A Collisson([38]), Leslie Cope([26]), Chad J Creighton([40]), Gad Getz([1,4]), James Herman([26]), Bruce E Johnson([5]), Raju Kucherlapati([10,11]), Marc Ladanyi([28, 30]), Christopher A Maher([51]), Gordon Robertson([7]), Chris Sander([27]), Ronglai Shen([27]), Rileen Sinha([27]), Andrey Sivachenko([1]), Roman K Thomas([56, 92, 93]), William D Travis([41]), Ming-Sound Tsao([49]), John N Weinstein([32,34]), Dennis A Wigle([60]), Stephen B Baylin([26]), Ramaswamy Govindan([51]), and Matthew Meyerson([1,5,6])

[90]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892 USA

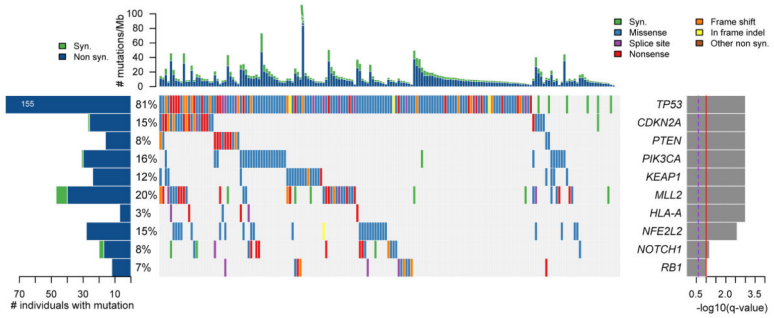

Figure 1. Significantly mutated genes in lung SqCC

Significantly mutated genes (q-value <0.1) identified by exome sequencing are listed vertically by q-value. The percentage of lung SqCC samples with a mutation detected by automated calling is noted at the left. Samples displayed as columns, with the overall number of mutations plotted at the top and samples arranged to emphasize mutual exclusivity among mutations.

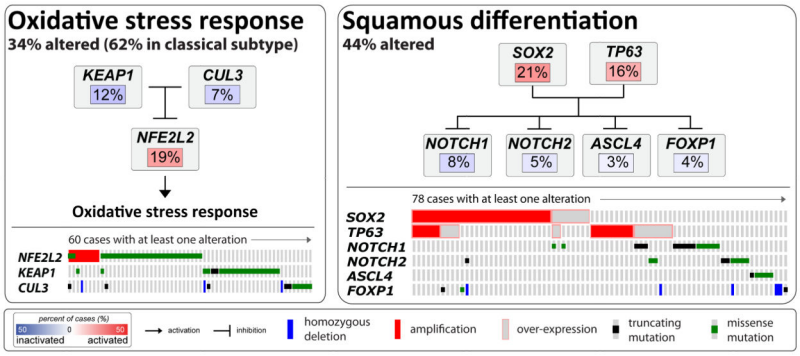

Figure 2. Somatically altered pathways in squamous cell lung cancer

Left, Alterations in oxidative response pathway genes by somatic mutation as defined by somatic mutation, copy number alteration or up- or down-regulation. Frequencies of alteration are expressed as a percentage of all cases, with background in red for activated genes and blue for inactivated genes. **Right,** Alterations in genes that regulate squamous differentiation, as defined in the left panel.

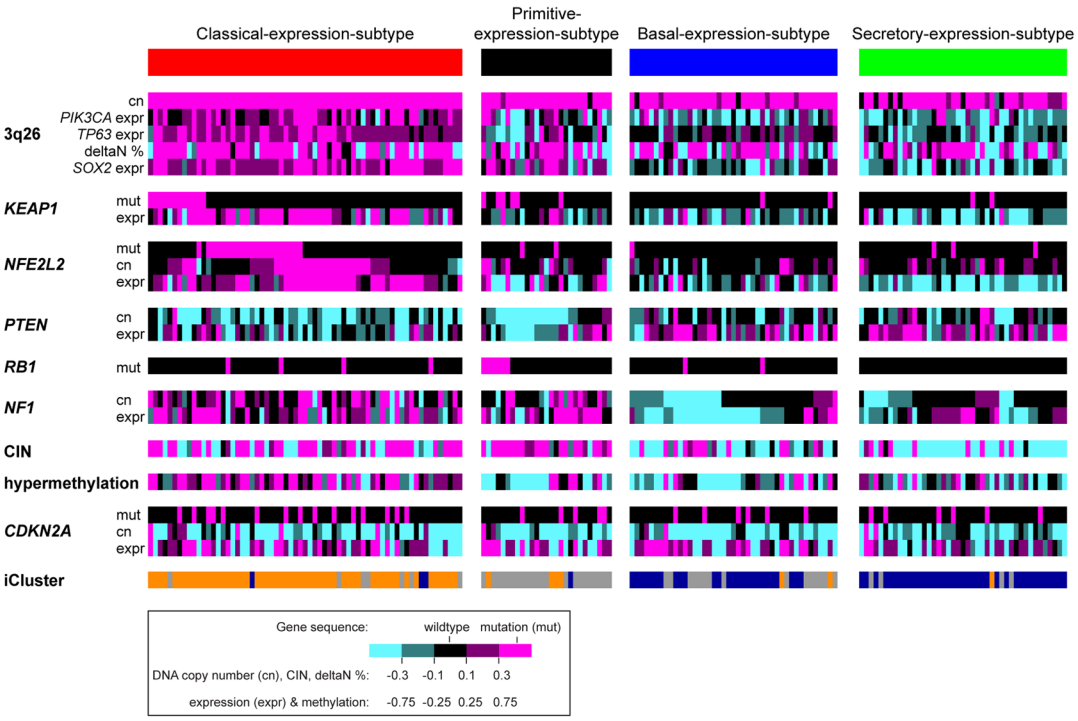

Figure 3. Gene expression subtypes integrated with genomic alterations

Tumors are displayed as columns, grouped by gene expression subtype. Subtypes were compared by Kruskal-Wallis tests for continuous features and by Fisher's exact tests for categorical features. Displayed features displayed showed significant association with gene expression subtype ($P<0.05$), except for *CDKN2A* alterations. deltaN expression percentage represents transcript isoform usage between the *TP63* isoforms, deltaN and tap63, as determined by RNA-sequencing. Chromosomal instability (CIN) is defined by the mean of the absolute values of chromosome arm copy numbers from the GISTIC[23,24] output. Absolute values are used so that amplification and deletion alterations are counted equally. Hypermethylation scores and iCluster assignments are described in Supplementary Figure S6.1 and S7.A1, respectively. CIN, methylation, gene expression, and deltaN values were standardized for display using z-score transformation.

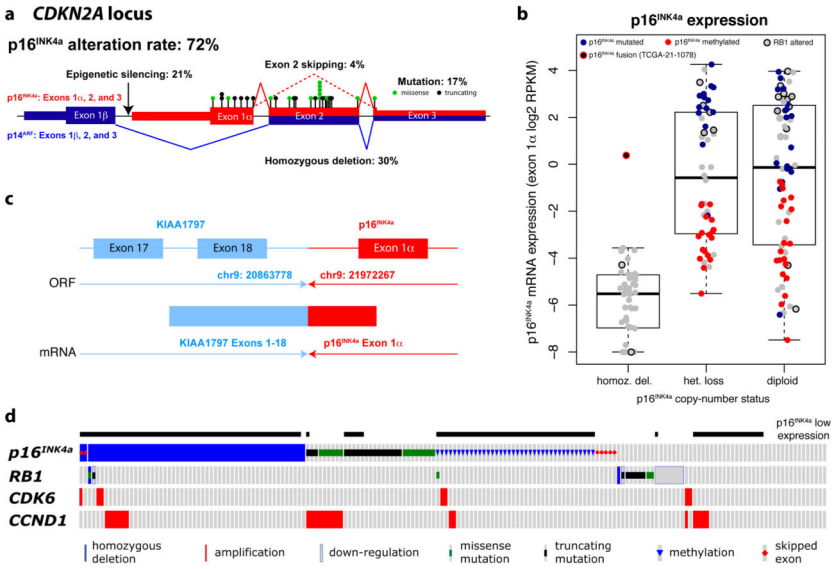

Figure 4. Multi-faceted characterization of mechanisms of *CDKN2A* loss

a, Schematic view of the exon structure of *CDKN2A* demonstrating the types of alterations identified in the study. The locations of point mutation are denoted by black and green circles. **b,** *CDKN2A* expression (y-axis) versus *CDKN2A* copy number (x-axis). Samples are represented by circles and colored-coded by specific type of *CDKN2A* alteration. **c,** Diagram of the *KIAA1797-CDKN2A* fusion identified by whole genome sequencing. **d,** *CDKN2A* alterations and expression levels (binary) in each sample.

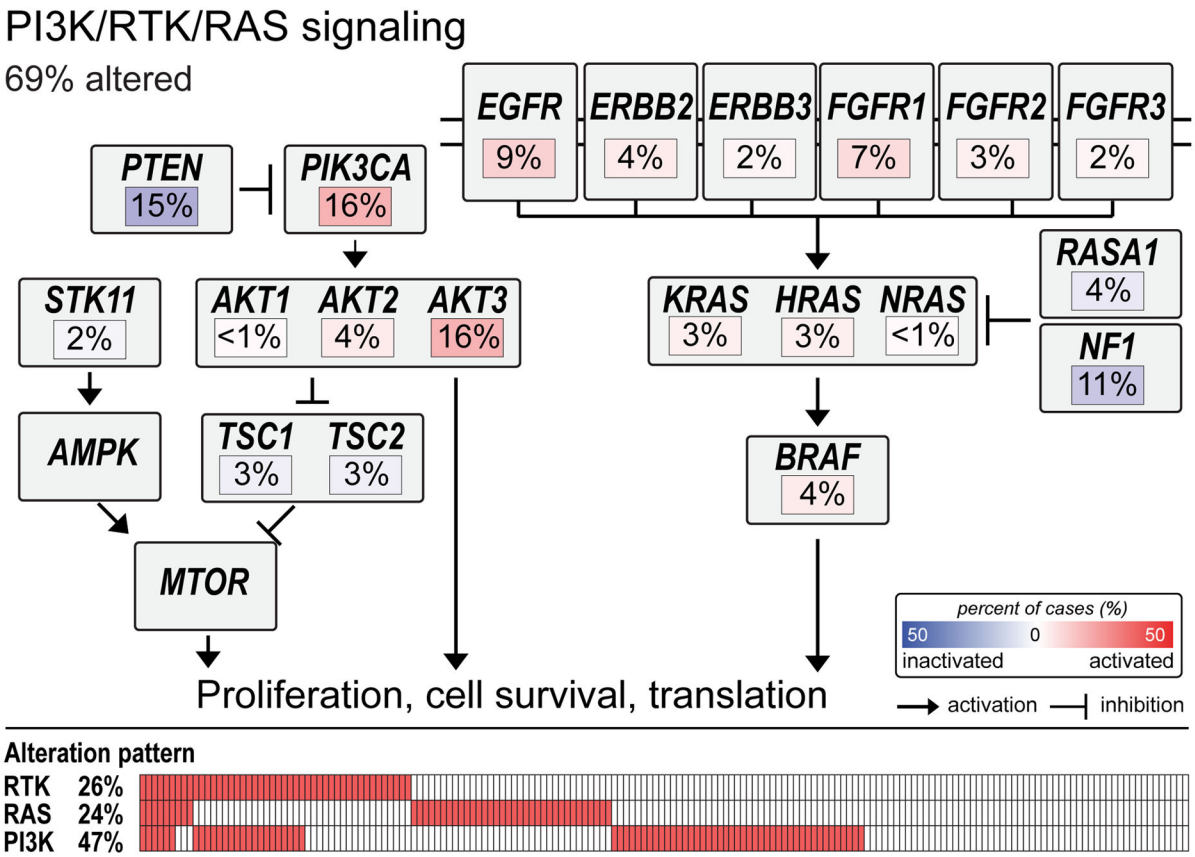

Figure 5. Alterations in targetable oncogenic pathways in lung SqCCs

Pathway diagram showing the percentage of samples with alterations in the PI3K/RTK/RAS pathways. Alterations are defined by somatic mutations, homozygous deletions, high-level, focal amplifications, and, in some cases, by significant up- or down-regulation of gene expression (*AKT3, FGFR1, PTEN*).