Cell

Mapping the Hallmarks of Lung Adenocarcinoma with Massively Parallel Sequencing

Marcin Imielinski,[1,2,3,5,18] Alice H. Berger,[1,5,18] Peter S. Hammerman,[1,5,18] Bryan Hernandez,[1,18] Trevor J. Pugh,[1,5,18] Eran Hodis,[1] Jeonghee Cho,[6] James Suh,[7] Marzia Capelletti,[5] Andrey Sivachenko,[1] Carrie Sougnez,[1] Daniel Auclair,[1] Michael S. Lawrence,[1] Petar Stojanov,[1,5] Kristian Cibulskis,[1] Kyusam Choi,[6] Luc de Waal,[1,5] Tanaz Sharifnia,[1,5] Angela Brooks,[1,5] Heidi Greulich,[1,5] Shantanu Banerji,[1,5] Thomas Zander,[9,11] Danila Seidel,[9] Frauke Leenders,[9] Sascha Ansén,[9] Corinna Ludwig,[9] Walburga Engel-Riedel,[9] Erich Stoelben,[9] Jürgen Wolf,[9] Chandra Goparju,[8] Kristin Thompson,[1] Wendy Winckler,[1] David Kwiatkowski,[5] Bruce E. Johnson,[5] Pasi A. Jänne,[5] Vincent A. Miller,[12] William Pao,[14] William D. Travis,[13] Harvey I. Pass,[8] Stacey B. Gabriel,[1] Eric S. Lander,[1,4,15] Roman K. Thomas,[9,10,11,16,17] Levi A. Garraway,[1,5] Gad Getz,[1] and Matthew Meyerson[1,3,5,*]

[1]Broad Institute of Harvard and MIT, 7 Cambridge Center, Cambridge, MA 02142, USA
[2]Department of Pathology, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, USA
[3]Department of Pathology
[4]Department of Systems Biology
Harvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA
[5]Department of Medical Oncology, Dana Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02115, USA
[6]Samsung Research Institute, Samsung Medical Center, Seoul 135-967, Republic of Korea
[7]Department of Pathology
[8]Department of Cardiothoracic Surgery
Langone Medical Center, New York University, New York, NY 10016, USA
[9]Department of Internal Medicine and Center for Integrated Oncology
[10]Laboratory of Translational Cancer Genomics
Köln-Bonn, University of Cologne, 50924 Cologne, Germany
[11]Max Planck Institute for Neurological Research, 50924 Cologne, Germany
[12]Thoracic Oncology Service, Department of Medicine
[13]Department of Pathology
Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[14]Division of Hematology/Oncology, Vanderbilt-Ingram Cancer Center, Nashville, TN 37232, USA
[15]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA
[16]Department of Translational Genomics, University of Cologne, Weyertal 115b, 50931 Cologne, Germany
[17]Department of Pathology, University of Cologne, Kerpener Strasse 62, 50937 Cologne, Germany
[18]These authors contributed equally to this work
*Correspondence: matthew_meyerson@dfci.harvard.edu
http://dx.doi.org/10.1016/j.cell.2012.08.029

SUMMARY

Lung adenocarcinoma, the most common subtype of non-small cell lung cancer, is responsible for more than 500,000 deaths per year worldwide. Here, we report exome and genome sequences of 183 lung adenocarcinoma tumor/normal DNA pairs. These analyses revealed a mean exonic somatic mutation rate of 12.0 events/megabase and identified the majority of genes previously reported as significantly mutated in lung adenocarcinoma. In addition, we identified statistically recurrent somatic mutations in the splicing factor gene *U2AF1* and truncating mutations affecting *RBM10* and *ARID1A*. Analysis of nucleotide context-specific mutation signatures grouped the sample set into distinct clusters that correlated with smoking history and alterations of reported lung adenocarcinoma genes. Whole-genome sequence analysis revealed frequent structural rearrangements, including in-frame exonic alterations within *EGFR* and *SIK2* kinases. The candidate genes identified in this study are attractive targets for biological characterization and therapeutic targeting of lung adenocarcinoma.

INTRODUCTION

Lung cancer is a leading cause of death worldwide, resulting in more than 1.3 million deaths per year, of which more than 40% are lung adenocarcinomas (World Health Organization, 2012; Travis, 2002). Most often, tumors are discovered as locally advanced or metastatic disease, and despite improvements in molecular diagnosis and targeted therapies, the average 5 year

survival rate for lung adenocarcinoma is ~15% (Minna and Schiller, 2008).

Molecular genotyping is now routinely used to guide clinical care of lung adenocarcinoma patients, largely due to clinical trials that demonstrated superior efficacy of targeted kinase inhibitors as compared to standard chemotherapy for patients with *EGFR* mutations or *ALK* fusions (Kwak et al., 2010; Pao and Chmielecki, 2010). In addition to *EGFR* and *ALK* alterations found in ~15% of U.S. cases, lung adenocarcinomas frequently harbor activating mutations in *KRAS*, *BRAF*, *ERBB2*, and *PIK3CA* or translocations in *RET* and *ROS1* (Pao and Hutchinson, 2012), all of which are being pursued as targets in ongoing clinical trials (http://clinicaltrials.gov/). Lung adenocarcinomas also often harbor loss-of-function mutations and deletions in tumor suppressor genes *TP53*, *STK11*, *RB1*, *NF1*, *CDKN2A*, *SMARCA4*, and *KEAP1* (Ding et al., 2008; Kan et al., 2010; Sanchez-Cespedes et al., 2002). Unfortunately, such alterations are difficult to exploit therapeutically. Therefore, knowledge of additional genes altered in lung adenocarcinoma is needed to further guide diagnosis and treatment.

Previous efforts in lung adenocarcinoma genome characterization include array-based profiling of copy number changes (Tanaka et al., 2007; Weir et al., 2007), targeted sequencing of candidate protein-coding genes (Ding et al., 2008; Kan et al., 2010), and whole-genome sequencing of a single tumor/normal pair (Ju et al., 2012; Lee et al., 2010). These studies identified somatic focal amplifications of *NKX2-1*, substitutions and copy number alterations in known oncogenes and tumor suppressor genes, and recurrent in-frame fusions of *KIF5B* and *RET*. These studies have also nominated several putative cancer genes with somatic mutations (*EPHA* family, *NTRK* family, *TLR4*, *LPHN3*, *GRM1*, and *GLI*), but the functional consequence of many alterations is unknown. A recent study describing whole-exome sequencing of 16 lung adenocarcinomas (Liu et al., 2012) enumerated several mutated genes but did not identify genes undergoing positive selection for mutation in the studied tumors.

In this study, we used next-generation sequencing to sequence the exomes and/or genomes of DNA from 183 lung adenocarcinomas and matched normal adjacent tissue pairs. In addition to verifying genes with frequent somatic alteration in previous studies of lung adenocarcinoma, we identified novel mutated genes with statistical evidence of selection and that likely contribute to pathogenesis. Together, these data represent a significant advance toward a comprehensive annotation of somatic alterations in lung adenocarcinoma.

RESULTS

Patient Cohort Description

We sequenced DNA from 183 lung adenocarcinomas and matched normal tissues by using paired-end massively parallel sequencing technology (Bentley et al., 2008). The cohort included 27 never-smokers, 17 light smokers (defined by less than ten pack years of tobacco use), 118 heavy smokers (more than ten pack years), and 21 patients of unknown smoking status (Table 1). The cohort included 90 stage I, 36 stage II, 22 stage III, and 10 stage IV lung adenocarcinoma cases, as well as 25 patients with unknown stage. All tumors were chemotherapy-naive, primary resection specimens except for one case with whole-genome sequence data (LU-A08-43) that was a postchemotherapy metastatic tumor from a never-smoker. Sample acquisition details are provided in Extended Experimental Procedures. Additional clinical descriptors of the cohort are provided in Table 1. Comprehensive clinical and histopathological annotations, sequence characteristics, and major variants for each patient in the study are provided in Table S1 (available online).

Table 1. Summary of Clinical Features

Age at Surgery (Median; Range)	66 (36–87)
Gender	
male	95
female	88
Smoking Status (AJCC 7th Edition)	
never-smoker	27
smoker >10 years	118
smoker ≤10 years	17
n/a	21
pack years (median; range)	30 (0–128)
Survival	
follow-up available	135
follow-up unavailable	48
PFS in months (median; range)	9 (0–63)
Tumor Stage	
I	90
II	36
III	22
IV	10
n/a	25

Distribution of selected clinical variables from 183 lung adenocarcinoma cases.

Mutation Detection and Validation

We examined 183 lung adenocarcinoma tumor/normal pairs with a combination of whole-exome sequencing (WES) or whole-genome sequencing (WGS): 159 WES, 23 WES and WGS, and 1 WGS only. Exomes were sequenced to a median fold coverage of 92 (range: 51–201) on 36.6 Mb of target sequence (Fisher et al., 2011). Genomes were sequenced to a median coverage of 69 (range: 25–103) in the tumor and 36 (range: 28–55) in the normal, with the higher tumor coverage to adjust for stromal contamination. Complementary SNP array analysis of 183 pairs was used to detect genome-wide somatic copy number alterations. See Table 2 and Extended Experimental Procedures for more details.

We identified somatic substitutions and small insertions and deletions (indels) through statistical comparison of paired tumor/normal sequence data by using algorithms calibrated for stromally contaminated cancer tissues (Banerji et al., 2012; Stransky et al., 2011) (www.broadinstitute.org/cancer/cga; Extended Experimental Procedures). Exonic regions of the 183 cases contained 77,736 somatic variants corresponding to