



Published in final edited form as:

Science. 2014 October 10; 346(6206): 256–259. doi:10.1126/science.1256930.

Intra-tumor Heterogeneity in Localized Lung Adenocarcinomas Delineated by Multi-region Sequencing

Jianjun Zhang^{1,2}, Junya Fujimoto³, Jianhua Zhang⁴, David C. Wedge⁵, Xingzhi Song⁴, Jiexin Zhang⁶, Sahil Seth⁴, Chi-Wan Chow³, Yu Cao¹, Curtis Gumbs¹, Kathryn A. Gold², Neda Kalhor⁷, Latasha Little¹, Harshad Mahadeshwar⁴, Cesar Moran⁷, Alexei Protopopov⁴, Huangdong Sun⁴, Jiabin Tang^{4,¶}, Xifeng Wu⁸, Yuanqing Ye⁸, William N. William², Jack J. Lee⁹, John V. Heymach^{2,10}, Waun Ki Hong², Stephen Swisher¹¹, Ignacio I. Wistuba³, and P. Andrew Futreal^{1,12,*}

¹Department of Genomic Medicine, The University of Texas, MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX, 77030

²Department of Thoracic/Head and Neck Medical Oncology, The University of Texas, MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX, 77030

³Department of Translational Molecular Pathology, The University of Texas, MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX, 77030

⁴Applied Cancer Science Institute, The University of Texas, MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX, 77030

⁵Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK

⁶Department of Bioinformatics and Computational Biology, The University of Texas, MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX, 77030

⁷Department of Pathology, The University of Texas, MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX, 77030

⁸Department of Epidemiology, The University of Texas, MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX, 77030

⁹Department of Biostatistics, The University of Texas, MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX, 77030

¹⁰Department of Cancer Biology, The University of Texas, MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX, 77030

*Corresponding author. AFutreal@mdanderson.org (A.F.).

¶Current address: Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY, 10065.

List of supplementary materials

Patients and Methods

Supplementary Figure Legends

Figs. S1 to S6

Tables S1 to S4

References (31–44)

¹¹Department of Thoracic Surgery, The University of Texas, MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX, 77030

¹²Honorary Faculty, Wellcome Trust Sanger Institute, Hinxton, UK CB10 1SA

Abstract

Cancers are composed of populations of cells with distinct molecular and phenotypic features, a phenomenon termed intra-tumor heterogeneity (ITH). ITH in lung cancers has not been well studied. We applied multi-region whole exome sequencing (WES) on 11 localized lung adenocarcinomas. All tumors showed clear evidence of ITH. On average, 76% of all mutations and 20/21 known cancer gene mutations were identified in all regions of individual tumors suggesting single-region sequencing may be adequate to identify the majority of known cancer gene mutations in localized lung adenocarcinomas. With a median follow-up of 21 months post-surgery, 3 patients have relapsed and all 3 patients had significantly larger fractions of subclonal mutations in their primary tumors than patients without relapse. These data indicate larger subclonal mutation fraction may be associated with increased likelihood of postsurgical relapse in patients with localized lung adenocarcinomas.

Intra-tumor heterogeneity may have impacts on tumor biopsy strategy, characterization of actionable targets, treatment planning, and drug resistance (1–6). ITH has recently been elucidated in substantial detail in several cancer types using next-generation sequencing (NGS) approaches (7–14). Recent evidence supports a model of branched evolution leading to variable ITH in different tumors (9, 13, 15, 16). Studies in clear cell renal carcinoma (ccRCC) have demonstrated substantial ITH, with the majority of mutations in known cancer genes confined to spatially separated tumor regions except for VHL loss being a ubiquitous event (16, 17). These data suggest that a single biopsy may be inadequate for identifying all cancer gene mutations from a tumor, thus presenting an incomplete view of potential targets for therapy. Critically, the extent to which these observations in ccRCC apply to other solid tumors is currently not clear.

To characterize ITH in localized lung adenocarcinomas, we applied multi-region WES on 48 tumor regions from 11 resected lung adenocarcinomas (8 stage I, 2 stage II and one stage III tumors, tumor size 2 – 4.6 cm), who had surgery with curative intent (Fig. 1 and Table S1). WES was conducted at mean depth of 277x. In total, 7,269 mutations were identified and 7,026 (97%) somatic mutations were validated by a separate bespoke capture sequencing experiment at mean depth of 863x (Table S2). The numbers of mutations varied substantially between tumors (Fig. S1), but no significant correlations were identified between mutation burden and age, gender, tumor size, lymph node status or smoking status.

A useful approach when considering ITH is to depict a given tumor as a tree structure with the trunk representing ubiquitous mutations present in all regions of the tumor, branches representing heterogeneous mutations present in only some regions of the tumor and private branches representing mutations that are present only in one region of the tumor – analogous to a phylogenetic tree. Placement of mutations on trunks versus branches reflects relative molecular time of acquisition, with branch mutations occurring, by definition, subsequent to trunk mutations. We applied this approach to multi-region sequencing data from these 11

lung adenocarcinomas. Evidence for ITH was found in each tumor studied. On average, 76% of all mutations were detected in all regions of the same tumors. However, the phylogenetic structure varied considerably between tumors (Fig. 1). We then characterized known cancer gene mutations, defined as nonsynonymous mutations identical to those previously reported in known cancer genes (18–23) or truncating mutations in known tumor suppressor genes, in the context of the derived phylogenetic tree structures. Thirteen of 14 known cancer gene point mutations were mapped to the trunks of the phylogenetic trees (Fig. 1, Table S3), indicating these mutations were acquired relatively early during evolution of these 11 tumors. In contrast to ccRCC, these data suggest that single-region sampling may be sufficient to identify the majority of known cancer gene mutations in localized lung adenocarcinomas.

We were also able to evaluate copy number changes relative to ITH. In contrast to ccRCC (16, 17, 24), we did not observe substantial difference in large-scale chromosome aberrations (Fig. S2A) and the log₂ ratio profiles were similar between different regions within the same tumors (Fig. S2B and Table S4). Further, amplification or deletion of known cancer genes (22) as well as their relative placement on the phylogenetic trees were delineated for these 11 lung adenocarcinomas. All of these events were mapped to the trunks of the phylogenetic trees (Fig. 1) suggesting that, like known cancer gene point mutations discussed above, amplification/deletion of known cancer genes were also early molecular events for these 11 tumors. Previous work in breast cancer also suggested that known cancer gene mutations were relatively early genetic events shared by all subclones of individual breast cancers (13). Taken together, these results indicate that different cancer types may have different relative timing of acquisition of cancer gene mutations. Further, the data would suggest in this subset of lung adenocarcinomas, there are likely mutations in non-canonical cancer genes that drive tumor development and subclonal divergence.

With a median follow-up of 21 months post surgery at the time of this report, 3 patients have had disease relapse. These 3 patients had significantly larger proportion of subclonal non-trunk mutations (branch plus private branch mutations) in their primary tumors than patients without relapse (average 40% in relapsed patients versus 17% in patients without relapse, $p=0.006$ by t test, Fig. 1). Although the sample size is small, these findings suggest the possibility that subclonal mutations may be important for cancer progression and that larger subclonal mutation fraction may be associated with an increased likelihood of post-surgical relapse in this subset of lung adenocarcinoma patients.

Analysis of NGS data relies heavily upon adequate sequencing depth to make high accuracy consensus base calls. We compared our WES data (average sequencing depth 277×) to deep sequencing data (average sequencing depth 863×) employed in validation with regards to known cancer gene mutation identification. In the tumor 499, a canonical *KRAS* p.G12C mutation was detected in only 1 of 4 tumor regions at exome depth but detected in all four tumor regions at increased sequencing depth (Table S2). Extending this analysis, we then compared deep sequencing data to WES data in defining ITH. The result showed many branch and private branch mutations defined by WES were detectable in all regions of individual tumors with increasing sequencing depth (Fig. 2). Taken together, these results

indicate that considerable depth of sequencing will be necessary to detect cancer gene mutations and accurately characterize ITH of lung adenocarcinomas.

Next, we analyzed the mutational spectrum of these 11 lung adenocarcinomas. Consistent with previous studies (18–20, 25), different mutation spectra were observed in smokers and non-smokers. Three never-smokers (292, 339 and 356) showed C>T-predominant mutation profiles. Three former smokers who quit more than 20 years ago (270, 472 and 4990) and one former smoker who had a 25 pack-year (pack year = number of packs per day x number of years) history of smoking and quit 6 years ago (283) also showed C>T-predominant mutation profiles, as in non-smokers. Two former smokers who had more than a 50 pack-year history of smoking and quit 5 years ago (317 and 499) and one former smoker, who had a 25 pack-year history of smoking and quit only 2.5 years ago (330) showed C>A-predominant mutation profiles consistent with the mutation profile of cigarette smoke exposure. The only current smoker, who had a 20 pack-year history of smoking, but had cut down to 2 cigarettes a day at the time of cancer diagnosis (324) showed an equivalent portion of C>T (26%) versus C>A (21%) substitutions in her tumor (Fig. 3A). These results indicate tumor mutation spectra in former smokers reflect not only quantity of smoking exposure, but also time since smoking cessation.

We next compared the mutational spectrum of trunk mutations versus non-trunk to explore the relative contribution of mutational processes over time. Significant differences in mutational spectrum were observed in 6 tumors indicating that specific mutational processes were likely operative at different times during development of these tumors (Fig. 3B and 3C). Of interest, two former smokers (317 and 330) and the current smoker (324) showed significant differences between trunk and non-trunk mutation spectrum with a shift from smoking-associated C>A transversions in trunk mutations to non-smoker-associated C>T transitions in non-trunk mutations.

Recent evidence has suggested APOBEC activity as a major source for C>T and C>G mutations (12, 26). We therefore investigated if there is evidence of an APOBEC mutational process in this subset of lung adenocarcinomas. On average, 28% of all mutations had a specific substitution pattern (C>T/G at TpCpW sites, where W is A or T), consistent with an APOBEC-mediated process (Fig. S3). APOBEC mutation signature enrichment was found to be more pronounced for non-trunk mutations compared to trunk mutations in 7 of the 11 patients, however this difference was statistically significant only for case 330 (Fig. 3D). These data suggest that an APOBEC-like process is contributing substantially to the mutations found in this subset of lung adenocarcinomas and that there is a trend towards this process being more pronounced in later, subclonal mutations – further highlighting the dynamic nature of mutational processes in play.

Substantial variation in the allele frequency of somatic mutations within each individual tumor region from a given tumor was observed in this set of lung adenocarcinomas (Fig. S4A). To more formally characterize subclonal fraction within each tumor region, we employed the ABSOLUTE algorithm (27). These analyses demonstrated that at least 29 out of 48 individual tumor regions showed evidence of intra-regional subclonal populations. The distribution of clonal and subclonal mutations was different among the sampled regions

within the same tumors in some patients (Fig. S4B) further suggesting single biopsy analysis would be inadequate to fully represent ITH in these tumors.

To explore the implications of how ITH assessment might be reduced to practice, we repeated the ABSOLUTE analysis on the combined sequencing data from all tumor regions of each patient to assess the global ITH on a per patient level, defined by the relative proportion of subclonal mutations. Similar to the phylogenetic analyses, all 3 patients with relapsed disease had larger subclonal fractions in their primary tumors (average 41% in patients with relapse versus 24% in patients without relapse, $p=0.045$ by t test, Fig. S5A). Using a complimentary Bayesian Dirichlet process (13) on the per patient combined data showed the same trend (average subclonal mutations 66% in patients with relapse versus 36% in patients without relapse, $p=0.035$ by t test, Fig. S5B). These results suggest that a measure of overall subclonal fraction may be of interest from a prognostic standpoint in this population of patients.

Resectable localized disease accounts for 30–50% of all non-small cell lung cancers with increasing prevalence as screening is more widely implemented (28–30). Given that this subset of patients has tumors surgically resected as standard of care, there is an opportunity to confirm these preliminary observations by deep sequencing multi-region samples obtained from resected tumors. The question of whether sequencing targeted cancer gene panels versus whole exome will yield sufficient mutation data for meaningful analyses and the most appropriate algorithms for analyses will need to be addressed in order to fully test if the clinical correlation suggested in these data is born out in larger patient cohorts.

Evidence of marked regional ITH in ccRCC suggested substantial challenges to personalized oncology based on single tumor biopsy to portray the mutational landscape. This study, however, provides evidence that ITH patterns may be different between cancer types. With the caveat of limited sample size fully acknowledged, these data suggest that whilst multi-region sampling is needed to fully assess ITH complexity, single biopsy analysis at appropriate depth might be sufficient to identify the majority of known cancer gene mutations in this subset of lung adenocarcinomas. Studies in much larger cohorts, ideally with comprehensive clinical annotation and repeat biopsy at relapse are needed to fully understand the clinical impact of ITH and insights afforded by these types of analyses. Furthermore, extension to epigenetic and phenotypic assessment through regional DNA methylation, chromatin state and RNA/protein expression studies over time and under treatment are needed to fully understand the impact of ITH on the biology of the cancer itself and its impact on the clinical phenotype of cancer patients.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by the Cancer Prevention and Research Institute of Texas (R120501), UT Systems Stars Award (PS100149), the Welch Foundation Robert A. Welch Distinguished University Chair Award (G-0040), Department of Defense PROSPECT grant (W81XWH-07-1-0306), the UT Lung Specialized Programs of Research Excellence grant (P50CA70907), the MD Anderson Cancer Center Support Grant (CA016672), NIH T32 Research

Training in Academic Medical Oncology Grant (CA-009666) and the A. Lavoy Moore Endowment Fund. The authors would like to thank Drs. Lynda Chin and Roeland Verhaak for constructive discussions. The DNA sequencing data were submitted to dbGaP (Accession number is pending).

References and Notes

1. Yap TA, Gerlinger M, Futreal PA, Pusztai L, Swanton C. Intratumor heterogeneity: seeing the wood for the trees. *Sci Transl Med.* Mar 28.2012 4:127ps10.
2. Swanton C. Intratumor heterogeneity: evolution through space and time. *Cancer Res.* Oct 1.2012 72:4875. [PubMed: 23002210]
3. Swanton C, et al. Predictive biomarker discovery through the parallel integration of clinical trial and functional genomics datasets. *Genome Med.* 2010; 2:53. [PubMed: 20701793]
4. Horswell S, Matthews N, Swanton C. Cancer heterogeneity and “The Struggle for Existence”: Diagnostic and analytical challenges. *Cancer Lett.* Nov 8.2012
5. Fisher R, Pusztai L, Swanton C. Cancer heterogeneity: implications for targeted therapeutics. *Br J Cancer.* Feb 19.2013 108:479. [PubMed: 23299535]
6. Saunders NA, et al. Role of intratumoural heterogeneity in cancer drug resistance: molecular and clinical perspectives. *EMBO Mol Med.* Aug.2012 4:675. [PubMed: 22733553]
7. Navin N, et al. Inferring tumor progression from genomic heterogeneity. *Genome research.* Jan. 2010 20:68. [PubMed: 19903760]
8. Navin N, et al. Tumour evolution inferred by single-cell sequencing. *Nature.* Apr 7.2011 472:90. [PubMed: 21399628]
9. Anderson K, et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature.* Jan 20.2011 469:356. [PubMed: 21160474]
10. Sottoriva A, et al. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proceedings of the National Academy of Sciences of the United States of America.* Mar 5.2013 110:4009. [PubMed: 23412337]
11. Campbell PJ, et al. Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proceedings of the National Academy of Sciences of the United States of America.* Sep 2.2008 105:13081. [PubMed: 18723673]
12. Nik-Zainal S, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell.* May 25.2012 149:979. [PubMed: 22608084]
13. Nik-Zainal S, et al. The life history of 21 breast cancers. *Cell.* May 25.2012 149:994. [PubMed: 22608083]
14. Yachida S, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature.* Oct 28.2010 467:1114. [PubMed: 20981102]
15. Shah SP, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature.* Oct 8.2009 461:809. [PubMed: 19812674]
16. Gerlinger M, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *The New England journal of medicine.* Mar 8.2012 366:883. [PubMed: 22397650]
17. Gerlinger M, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nature genetics.* Feb 2.2014
18. Imielinski M, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell.* Sep 14.2012 150:1107. [PubMed: 22980975]
19. Ding L, et al. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature.* Oct 23.2008 455:1069. [PubMed: 18948947]
20. Govindan R, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell.* Sep 14.2012 150:1121. [PubMed: 22980976]
21. Watson IR, Takahashi K, Futreal PA, Chin L. Emerging patterns of somatic mutations in cancer. *Nature reviews Genetics.* Oct.2013 14:703.
22. Vogelstein B, et al. Cancer genome landscapes. *Science.* Mar 29.2013 339:1546. [PubMed: 23539594]

23. Forbes SA, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic acids research*. Jan.2011 39:D945. [PubMed: 20952405]
24. Martinez P, et al. Parallel evolution of tumour subclones mimics diversity between tumours. *The Journal of pathology*. Aug.2013 230:356. [PubMed: 23716380]
25. Hainaut P, Pfeifer GP. Patterns of p53 G→T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke. *Carcinogenesis*. Mar.2001 22:367. [PubMed: 11238174]
26. Alexandrov LB, et al. Signatures of mutational processes in human cancer. *Nature*. Aug 22.2013 500:415. [PubMed: 23945592]
27. Carter SL, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nature biotechnology*. May.2012 30:413.
28. Bulzebruck H, et al. New aspects in the staging of lung cancer. Prospective validation of the International Union Against Cancer TNM classification. *Cancer*. Sep 1.1992 70:1102. [PubMed: 1515985]
29. Mahadevia PJ, et al. Lung cancer screening with helical computed tomography in older adult smokers: a decision and cost-effectiveness analysis. *JAMA : the journal of the American Medical Association*. Jan 15.2003 289:313.
30. Church TR, et al. Results of initial low-dose computed tomographic screening for lung cancer. *The New England journal of medicine*. May 23.2013 368:1980. [PubMed: 23697514]
31. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. Mar 1.2010 26:589. [PubMed: 20080505]
32. Cibulskis K, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature biotechnology*. Mar.2013 31:213.
33. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. Nov 1.2009 25:2865. [PubMed: 19561018]
34. Meacham F, et al. Identification and correction of systematic error in high-throughput sequence data. *BMC bioinformatics*. 2011; 12:451. [PubMed: 22099972]
35. Loman NJ, et al. Performance comparison of benchtop high-throughput sequencing platforms. *Nature biotechnology*. May.2012 30:434.
36. Kluge AG, Farris JS. Quantitative Phyletics and Evolution of Anurans. *Syst Zool*. 1969; 18:1.
37. Eck, R.; Dayhoff, M. Atlas of Protein Sequence and Structure 1966. National Biomedical Research Foundation, Silver Spring; Maryland: 1966.
38. Felsenstein J. PHYLIP – Phylogeny Inference Package (Version 3.2). *Cladistics*. 1989; 5:164.
39. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. Mar 15.2010 26:841. [PubMed: 20110278]
40. Koboldt DC, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research*. Mar.2012 22:568. [PubMed: 22300766]
41. Roberts SA, et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nature genetics*. Sep.2013 45:970. [PubMed: 23852170]
42. Bolli N, et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nature communications*. 2014; 5:2997.
43. Van Loo P, et al. Allele-specific copy number analysis of tumors. *Proceedings of the National Academy of Sciences of the United States of America*. Sep 28.2010 107:16910. [PubMed: 20837533]
44. Stephens PJ, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature*. Jun 21.2012 486:400. [PubMed: 22722201]

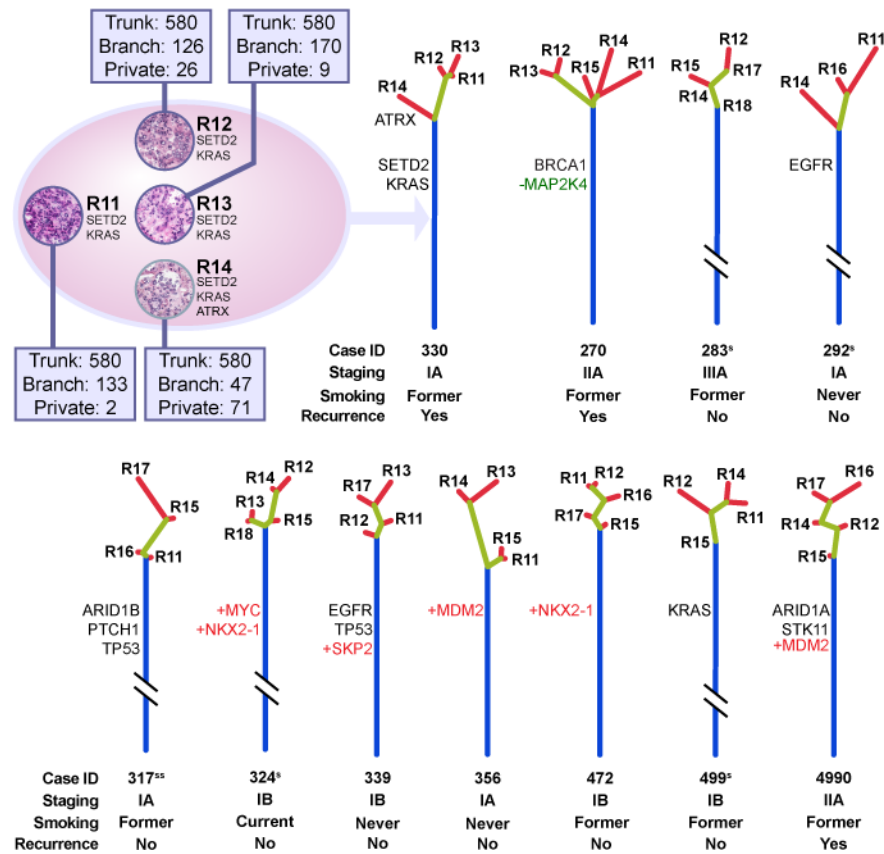


Fig. 1.

Assessment of ITH of 11 lung adenocarcinomas by multi-region sequencing. An example of regional mutation distribution (case 330) of resected tumors (represented by the ellipsis) is shown at the up left corner. Mutated cancer genes are indicated to the right of representative H&E image of each sequenced tumor region. Numbers of trunk, branch and private branch mutations for each region are indicated in associated windows. A phylogenetic tree was generated from all validated mutations utilizing Wagner parsimony method in PHYLIP. Blue, yellow and red lines represent trunk, branch and private branches respectively. Trees are anchored at germline DNA sequence obtained from peripheral blood of the relevant patients. Known cancer gene mutations are mapped to the trunks and branches as indicated. Point mutations, amplifications and deletions of known cancer genes are presented as black, +red and –green respectively. Trunk and branch lengths are proportional to the numbers of mutations acquired on the corresponding trunk or branch. Note: 5 tumors have their trunk lengths reduced to 10%^S or 2%^{SS} of original length for visualization purposes.

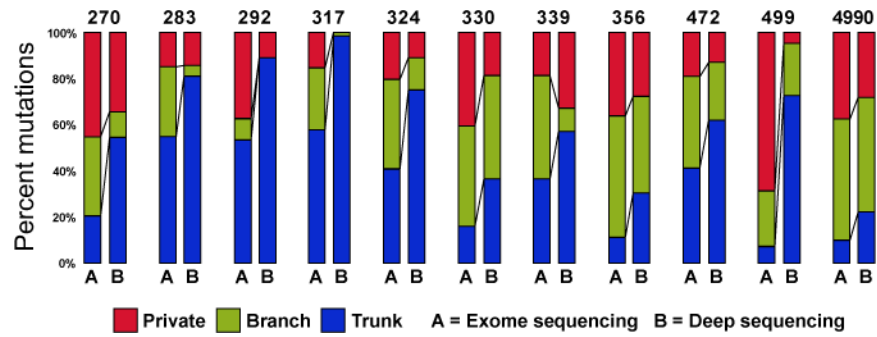


Fig. 2. Distribution of trunk, branch and private mutations defined by exome sequencing (average sequencing depth of 277 \times) versus deep sequencing (average sequencing depth of 863 \times). Only validated mutations meeting the following criteria are included: total counts in tumor DNA ≥ 100 ; total counts in germ line DNA ≥ 50 ; variant allele frequency (VAF) $\geq 5\%$ in tumor DNA and VAF=0 in germ line DNA.

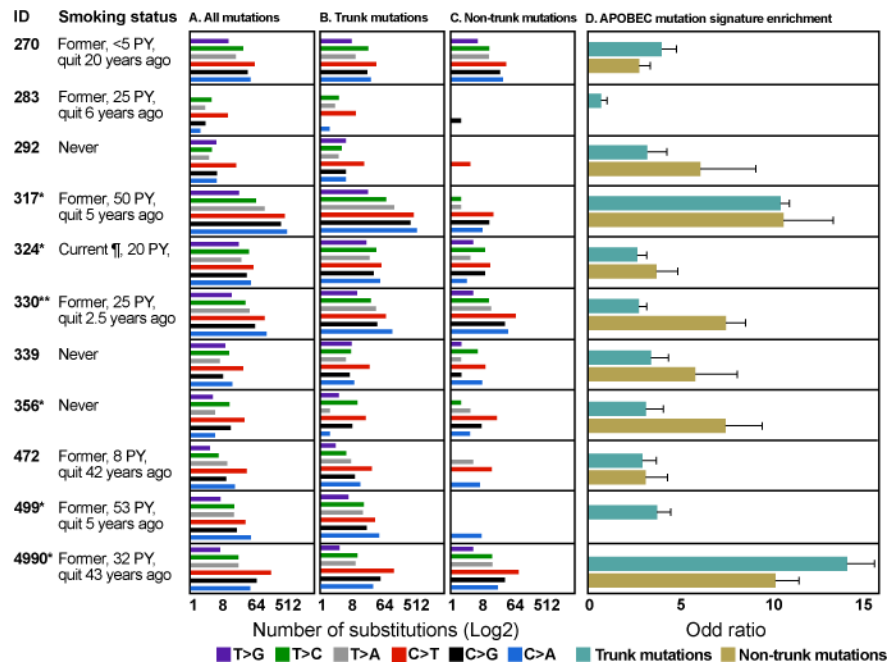


Fig. 3. Mutation spectrum of the 11 lung adenocarcinomas. (A) Mutation spectrum of all validated mutations. (B) Mutation spectrum of trunk mutations. (C) Mutation spectrum of non-trunk mutations. The difference of mutation spectrum between trunk and non-trunk mutations in each patient was evaluated with Fisher's exact test and significant p values are shown as *($P < 0.05$) and **($P < 0.01$). (D) APOBEC mutation signature enrichment odds ratio for trunk and non-trunk mutations. 95% confidence intervals for Fisher's exact test are indicated. PY: pack year. ¶: Patient had cut down to 2 cigarettes a day at the time of cancer diagnosis.