



HHS Public Access

Author manuscript

Cancer Epidemiol Biomarkers Prev. Author manuscript; available in PMC 2020 February 01.

Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2019 February ; 28(2): 348–356. doi:
10.1158/1055-9965.EPI-17-0984.

Characterization of Squamous Cell Lung Cancers from Appalachian Kentucky

Jinpeng Liu^{#1}, Thilakam Murali^{#2}, Tianxin Yu^{#2}, Chunming Liu², Theru A. Sivakumaran³, Hunter N.B. Moseley², Igor B. Zhulin^{4,5}, Heidi L. Weiss¹, Eric B. Durbin^{1,6}, Sally R. Ellingson^{1,6}, Jinze Liu^{1,7}, Bin Huang^{1,8}, Brent J. Hallahan¹, Craig M. Horbinski⁹, Kurt Hodges¹⁰, Dana L. Napier¹, Thérèse Bocklage¹⁰, Joseph Mueller¹, Nathan L. Vanderford^{1,11}, David W. Fardo⁸, Chi Wang^{1,8,*}, and Susanne M. Arnold^{1,12,*}

¹Markey Cancer Center, University of Kentucky, Lexington, KY, US

²Department of Molecular & Cellular Biochemistry, University of Kentucky, Lexington, KY, US

³Department of Pathology and Laboratory Medicine, Phoenix Children's Hospital, Phoenix, AZ, US

⁴Department of Microbiology, The Ohio State University, Columbus, OH, US

⁵Center for Bioinformatics, Pavlov First Saint Petersburg State Medical University, Saint Petersburg, Russia

⁶Division of Biomedical Informatics, University of Kentucky, Lexington, KY, US

⁷Department of Computer Science, University of Kentucky, Lexington, KY, US

⁸Department of Biostatistics, University of Kentucky, University of Kentucky, Lexington, KY, US

⁹Department of Pathology and Neurosurgery, Northwestern University, Chicago, IL, US

¹⁰Department of Pathology and Laboratory Medicine, University of Kentucky, Lexington, KY, US

¹¹Department of Toxicology and Cancer Biology, University of Kentucky, University of Kentucky, Lexington, KY, US

¹²Department of Medicine, University of Kentucky, University of Kentucky, Lexington, KY, US

These authors contributed equally to this work.

Abstract

Background: Lung cancer is the leading cause of cancer mortality in the United States (US). Squamous cell carcinoma (SQCC) represents 22.6% of all lung cancers nationally, and 26.4% in Appalachian Kentucky (AppKY), where death from lung cancer is exceptionally high. The Cancer Genome Atlas (TCGA) characterized genetic alterations in lung SQCC, but this cohort did not focus on AppKY residents.

* **Correspondence to:** Susanne M. Arnold, MD, Markey Cancer Center, 800 Rose St., CC445, Lexington, KY 40536, Tel: 859-323-6522, susanne.arnold@uky.edu. Chi Wang, PhD, Markey Cancer Center, 800 Rose St., CC433, Lexington, KY 40536, Tel: 859-323-2045, chi.wang@uky.edu.

Conflict of interest statement: The authors declare no potential conflicts of interest.

Methods: Whole exome sequencing was performed on tumor and normal DNA samples from 51 lung SQCC subjects from AppKY. Somatic genomic alterations were compared between the AppKY and TCGA SQCC cohorts.

Results: From this AppKY cohort, we identified an average of 237 non-silent mutations per patient and, in comparison to TCGA, we found that *PCMTD1* (18%) and *IDH1* (12%) were more commonly altered in AppKY versus TCGA. Using *IDH1* as a starting point, we identified a mutually exclusive mutational pattern (*IDH1*, *KDM6A*, *KDM4E*, *JMJD1C*) involving functionally-related genes. We also found actionable mutations (10%) and/or intermediate or high tumor mutation burden (65%), indicating potential therapeutic targets in 65% of subjects.

Conclusions: This study has identified an increased percentage of *IDH1* and *PCMTD1* mutations in SQCC arising in the AppKY residents versus TCGA, with population-specific implications for the personalized treatment of this disease.

Impact: Our study is the first report to characterize genomic alterations in lung SQCC from AppKY. These findings suggest population differences in the genetics of lung SQCC between AppKY and US populations, highlighting the importance of the relevant population when developing personalized treatment approaches for this disease.

Keywords

genomic sequencing; squamous lung cancer

Introduction

Lung cancer is the leading cause of cancer death worldwide, with 160,000 deaths in the US annually¹. The state of Kentucky ranks highest in lung cancer incidence and mortality, and Appalachian region of Kentucky (AppKY) is the major driver of this extraordinary healthcare burden², see Supplementary Fig. S1 for lung cancer characteristics in Kentucky. We undertook the present study of lung squamous cell carcinomas (SQCC) from residents of AppKY to provide the first genomic characterization of lung cancers from this region and to test the hypothesis that genetic mutations in AppKY SQCC are distinct from the general population and may help explain the region's extremely high cancer incidence. Essential to this effort was the full sharing of the comprehensive genomic profile of lung SQCC in The Cancer Genome Atlas (TCGA)³, which provided the comparison of the initial 178 subjects from a US genomic profile that does not focus on Central Appalachians (the distribution of TCGA tissue source sites is provided in Supplementary Table S1).

Previous studies reported distinct genetic abnormalities in lung SQCCs: *TP53*, *NFE2L2*, *CDKN2A*, *PTEN* and *ALK1*³⁻⁷, and copy-number alterations in *SOX2*, *PDGFRA*, *FGFR1*, *WHSC1L1*, and *CDKN2A*^{3,8-10}, which are targets of therapeutic interventions. We present the results from whole-exome sequencing (WES) and analysis of 51 SQCC patients from AppKY, which includes an overview of somatic alterations and copy-number variations, explores unique mutational patterns, and provides a clinically actionable assessment of mutations in this population.

Materials and Methods

Patient cohort

This study was approved by the Institutional Review Board (IRB) of the University of Kentucky (protocol # 14–0071-P3H). Cancer, adjacent normal tissue and associated clinical data were obtained from 51 AppKY patients (2002–2013) with previously untreated stage I – III lung SQCC, excluding mixed histology. All specimens were collected from single surgical procedures prior to any cancer therapy and fresh frozen. All surgical samples were reviewed by a board-certified pathologist and confirmed to be SQCC by the Markey Cancer Center’s (MCC) Biospecimen Procurement and Translational Pathology (BPTP) Shared Resource Facility (SRF). The study was carried out in accordance with the Declaration of Helsinki and all subjects provided written informed consent in accordance with local IRB requirements.

Only residents of AppKY, as defined by the Appalachian Regional Commission’s authorizing legislation, were included in the study. All subjects were confirmed to be residents of AppKY by the Kentucky Cancer Registry (KCR) using multiple confirmatory identifiers, including: address, zip code, county, phone number, and were followed long-term by the SEER-KCR. MCC’s Cancer Research Informatics SRF provided all demographic and survival data for this project in a de-identified manner and served as the honest broker. The cohort includes 36 males, 15 females, median age of 65 years (range, 43 to 82 years) and 30 stage I, 12 stage II and 5 stage III patients. The median follow-up was 40 months and 35% of patients were alive at the time of submission. Reflective of AppKY, 98% of patients were of Caucasian race, and 76% were current smokers. Supplementary Table S2 lists summary demographic and clinical characteristics, compared with the TCGA SQCC cohort.

Sample preparation and whole-exome sequencing

High molecular weight genomic DNA was extracted from matched tumor and adjacent normal tissues in all 51 cases using the DNAeasy Blood & Tissue Kit or the MagAttract HMW DNA Kit (both from QIAGEN). Tumor contents of samples are provided in Supplementary Table S3. DNA was quantified using the Qubit dsDNA HS Assay Kit (Invitrogen), and quality was assessed using E-gel, 0.8% (Invitrogen). Exome capture and sequencing were performed using Illumina Nextera Rapid Capture Exome v1.2 targeting 212,158 exonic regions. Paired-end sequencing ($2 \times 110\text{bp}$) was carried out using HiSeq2500 sequencing platform at the University of Illinois at Urbana-Champaign.

Somatic mutation analysis

Sequencing reads were trimmed and filtered using Cutadapt (v1.4.1)¹¹, then aligned to human reference genome b37/hg19 using BWA (v.0.7.9a)¹². PCR duplicates were removed using Picard (<http://broadinstitute.github.io/picard/>, v1.115). The Genome Analysis Toolkit (GATK v3.1–1)¹³ was used for local indels realignment and base quality recalibration. Somatic point mutations and indels were detected using MuTect (v1.1.4)¹⁴ and SomaticIndelDetector (GATK v2.3–9), respectively, with default settings. Mutations were annotated using Oncotator (v1.4.1.0)¹⁵. Significantly mutated genes were identified using MutSigCV (v1.4)¹⁶. Gene alteration rates between the AppKY cohort and the TCGA were

compared using Fisher's exact tests with Benjamini–Hochberg procedure to calculate the false discovery rate (FDR). For genes that showed a significant difference in alteration rates between the two cohorts, the exact logistic regression was used to further evaluate the difference between the two cohorts after adjusting for clinical and demographic variables including age, gender, stage and smoking. The analyses were performed using R (v3.3.1) and SAS (v9.3).

Somatic copy-number alterations (SCNA) analysis was conducted using ExomeCNV¹⁷, an R statistical package. Exonic CNAs were inferred based on the depth-of-coverage ratio between matched tumor and normal samples. Then, CNAs calls were combined into larger segments using circular binary segmentation in DNACopy¹⁸. Gistic2.0¹⁹ with a confidence level of 0.95 was used on the copy ratio profiles to identify significantly amplified/deleted regions.

To evaluate the clinical relevance of the somatic genomic alterations identified in our cohort, we downloaded the OncoKB database²⁰ (accessed in December 2017) to identify FDA approved drugs for the FDA-recognized and standard care biomarkers.

The mutual exclusivity of IDH1-related gene sets was examined by MEGSA²¹, which implements a likelihood ratio test. The p-value was calculated based on a mixture distribution with 0.5 probability at point mass zero and 0.5 probability as a chi-square distribution of 1 degree of freedom, as described in the MEGSA paper²¹.

Identification of IDH1 homologs

Human IDH1 protein sequence (NP_005887.2) was queried using BLASTP against the RefSeq database. The top 500 hits, including representatives from all major eukaryotic supergroups (except Rhizaria), were aligned using ClustalX²² with default parameters. The neighbor joining tree was built from this alignment. From the tree, a clade containing IDH1 and IDH2 homologs was identified and sequences from representative genomes of distantly related organisms were retrieved and realigned. The maximum likelihood phylogenetic tree was built from this alignment using the MEGA6 package²³ with default parameters. IDH1 and IDH2 orthologs were identified in individual clades that had the same topology as the eukaryotic ribosomal tree²⁴ and checked for consistency with RefSeq annotations.

Mutagenesis

pCSC-SP-PW-GFP (aka:pBOB-GFP) was a gift from Inder Verma (Addgene plasmid #12337). pCSC-Sp-pw-IDH1-GFP and pCSC-Sp-pw-IDH1 R132H-GFP were a gift from Hai Yan, which encode for wildtype human IDH1 (GenBank:CAG38738.1) and mutant IDH1 R132H (encoding for human IDH1 protein that has mutant aa132 from Arg to His), respectively. Both plasmids were digested with EcoRI and NOT1, and ligated into pcDNA3.1 to make pcDNA3.1-IDH1-GFP. Then pcDNA3.1-IDH1-GFP used for mutagenesis to make pcDNA3.1-IDH1 R132H-GFP, pcDNA3.1-IDH1 V178A-GFP, pcDNA3.1-IDH1 A307S-GFP and pcDNA3.1-IDH1 L352P-GFP (shortened as WT, R132H, V178A, A307S and L352P hereafter). Primers used for mutagenesis are listed in Supplementary Method S2. Mutagenesis was performed with Fushion E in a 50µl reaction volume. PCR reactions consisted of 19 cycles as follows: 95°C for 50 sec, 60°C for 50 sec,

68°C for 3 min. The PCR reaction product was digested with DpnI at 37 °C for 1 h, and transformed into DH5a competent cells. Colonies were selected for sequencing to verify the mutations.

Isocitrate dehydrogenase activity assays

HEK293T (ATCC) cells (see Supplementary Method S3 for the authentication) were transfected with the above plasmids using Calcium Phosphate transfection. Each well of cells from the 6-well plate were transfected with 1µg of plasmids. Cells were then suspended in 0.02% Triton-X100 PBS, homogenized 48 h post-transfection, and sonicated with three 20s pulses. 10 µg cell lysate (in 5 µl of 0.02% Triton-X100 PBS) was added to the reaction mix (including 33 mM Tris-Cl pH7.5, 2 mM MnCl₂, 107 µM NADP⁺) in total volume of 180 µl at room temperature. 20 µl of 800 µM isocitrate was added to start the reaction. Absorbance at 340 nm was monitored for 1 h at 10 min intervals. NADPH production was calculated as: µmole of NADPH produced/ml sample/min = DA₃₄₀×3×1000/6220. The enzyme activity assay was based on previous reports²⁵.

Immunohistochemistry (IHC) staining to confirm IDH1 R132H mutation

Tissue samples were stained with anti-IDH1 R132H antibody (Dianova DIA-H09, 1:20) using Target Retrieval Solution (Dako) according to manufacturer's instructions. Known IDH1 R132H positive controls and negative mouse IgG controls were run simultaneously. A detailed protocol is in Supplementary Method S4.

Pathway construction via systems biology analysis

Gene families that are grouped based on sequence or function were downloaded from the HGNC database²⁶ and gene families with similar or complementary functions to significantly mutated genes in our AppKY cohort were extracted. Likewise, protein-protein interactions involving the gene-products of significantly mutated genes and their related gene families were extracted from protein interaction databases^{27,28} to find interaction partners. Both sets of genes were used to identify and expand mutually exclusive mutational patterns in the AppKY mutational dataset. The resulting mutually exclusive patterns were then used to construct pathways around *IDH1* based on known interactions, guided by the principle that similar functions may be modified for cancer-related processes.

Results

Somatic alterations identified from the AppKY cohort

Overview of somatic alterations—The mean coverage of WES across the targeted regions was 104× with 92% of targeted bases being covered at 30× (Supplementary Table S4). Raw sequencing data are available at dbGaP (Accession: phs001651.v1.p1). We identified 16,005 somatic single-nucleotide variants and 217 somatic insertions or deletions (indels) across 51 matched tumor and normal pairs in the protein coding regions (Supplementary Table S5). Of those mutations, 12,117 were predicted to be non-silent mutations resulting in an amino acid change. The mean mutation rate in our cohort was 237 non-silent mutations per patient, corresponding to 8.5 mutations per megabases (Mb). Among non-silent mutations, transitions and transversions at CpG sites were the most

commonly observed mutation types, with rates of 11.5 per Mb and 15.5 per Mb, respectively. For non-CpG sites, transitions were more frequently observed at C:G sites (3.2 per Mb) than at A:T sites (1.8 per Mb). Similarly, transversions were more frequently observed at C:G sites (8.0 per Mb) than at A:T sites (2.0 per Mb). Mutation rate for each sample is provided in Supplementary Table S6.

Significantly mutated genes—We identified 3 genes that were significantly mutated (i.e., non-silent mutation rates higher than background mutation rates) in the AppKY cohort with an FDR < 0.2 using MutSigCV¹⁶: *TP53*, *PCMTD1* and *IDHI*. To increase the statistical power of our analysis, we followed the approach of the TCGA SQCC report³ and performed a secondary MutSigCV¹⁶ analysis to only consider genes causally implicated in cancer according to the COSMIC database²⁹. This approach enabled us to identify 11 additional genes that were significantly mutated with an FDR < 0.2: *PIK3CA*, *RNF43*, *MLLT10*, *STK11*, *NFE2L2*, *DEK*, *POT1*, *ATP2B3*, *HRAS*, *HOXA11* and *HOXA13* (Fig. 1). The description of each gene symbol used in this study is provided in Supplementary Table S7.

Copy number variation analysis—SCNAs were analyzed using WES data. We identified regions with significant SCNAs using Gistic2.0¹⁹. There were 18 peaks of significant amplification and 34 peaks of significant deletions (FDR<0.25). Significantly amplified regions were 3q27 (*MCF2L2*), 8p11 (*FGFR1*, *TACCI*, *WHSC1L1*, *LETM2*, *RNF5PI*), 11q13 (*CCND1*-oncogene), 7q21.2 (*CDK6*), 19q13, 13q34, 5p15, 8q24 (*MYC*-oncogene) and deleted regions were 9p21 (*CDKN2A*-tumor suppressor, *CDKN2B*), 8p23, 10q23 (*PTEN*, *CFL1PI*, *KLLN*), 17p13, 4q28.2 (*VEGFC*), 22q13.2 (*CHEK2*). Consistent amplification patterns were seen in certain related sets of genes, such as stem cell renewal genes. Detailed SCNA results are provided in Supplementary Tables S8-S9, S10-S11 and Supplementary Fig. S2-S3.

Comparative mutational analysis with other cohorts—We first compared somatic mutations and SCNAs of AppKY lung SQCC to TCGA cohort^{3,30-32}. We focused our comparison on significantly mutated genes in at least one cohort by the MutSigCV¹⁶ analysis. Our comparative analysis presented here (Table 1) included somatic mutations (point mutations and indels) only in the calculation of gene alteration rate. The comparison including both somatic mutations and SCNAs is provided in Supplementary Table S12 with similar conclusions. Both cohorts showed similar rates of alterations for *TP53* (68.6% AppKY, 80.9% TCGA, FDR q-value=1.000), *PIK3CA* (11.8% AppKY, 15.7% TCGA, FDR q-value=1.000), *NOTCH1* (11.8% AppKY, 8.4% TCGA, FDR q-value=1.000) and *PTEN* (5.9% AppKY, 7.9% TCGA, FDR q-value=1.000).

Significant differences in mutation rates between the AppKY and TCGA cohorts were observed. The *IDHI* mutations were observed in 11.8% of patients in the AppKY cohort. In contrast, only 1.1% of patients in the TCGA cohort had *IDHI* mutations (FDR q-value=0.039). Similarly, the AppKY cohort also showed a higher rate of mutations in *PCMTD1* (17.6% AppKY vs. 3.9% TCGA, FDR q-value=0.045). Even after adjusting for age, gender, stage, and smoking via exact logistic regression, mutation frequencies are still

significantly different between the AppKY and TCGA cohorts for *IDH1* (p-value=0.0024) and *PCMTD1* (p-value=0.019).

We also compared the somatic mutations and SCNAs of AppKY to three other lung SQCC cohorts^{31–33}, see Supplementary Table S12. Mutation rates of *IDH1* and *PCMTD1* were significantly higher in AppKY than all other cohorts.

Because *IDH1* and *PCMTD1* showed significantly higher alteration rates in the AppKY cohort, we performed an in-depth analysis of alterations seen in these two genes and in other genes based on previously observed mutually exclusive mutational patterns and/or either known or reasonably hypothesized interactions.

Clinically actionable mutations assessment—We investigated the somatic mutations/SCNAs observed in our cohort in association with FDA approved agents or published or ongoing clinical trials for non-small-cell lung carcinoma (NSCLC) or other tumor types. 5 subjects (10%) had actionable mutations, defined as FDA approved drugs (either for this indication or another cancer type), with a total of 8 somatic mutations/SCNAs events found in these 5 individuals. Additionally, we found that 33 out of 51 subjects (65%) had high (>20 mut/MB) or intermediate (6–20 mut/MB) tumor mutation burden (TMB), indicating an additional group of therapeutic choices for this population using checkpoint inhibitors. Overall, 65% of subjects had actionable mutations with FDA approved drugs and/or TMB that was high or intermediate. Many others had mutations that are under clinical investigation (Supplementary Table S13).

Analysis of alterations in *IDH1* and associated pathways

Prediction of the effect of *IDH1* mutations—Mutations in *IDH1* and its homolog *IDH2* coding for cytosolic and mitochondrial isocitrate dehydrogenases, correspondingly, are common in gliomas³³ and myeloid neoplasms³⁴, but rare in lung cancer. We observed multiple *IDH1* variants: R132H, V178A, A307S and L352P, and the R132H variant was confirmed by immunohistochemistry (Supplementary Fig. S4). The *IDH1* variant R132H (Supplementary Fig. S5A) is reported in a variety of cancers and the role of various R132 missense substitutions has been studied extensively. These mutations are generally heterozygous, suggesting a gain-of-function by the enzyme, and supported by mechanistic studies demonstrating that the R132H variant protein has an aberrant enzymatic activity, converting α -ketoglutarate (2OG) to (R)-2-hydroxyglutarate (2HG)³⁵. This enantiomer of 2HG acts as an oncometabolite and interferes with cell differentiation³⁶.

To understand potential consequences of the other detected *IDH1* variants (V178A, A307S, and L352P) (Supplementary Fig. S5A), we applied a recently developed evolutionary approach³⁷, based on the principle that most deleterious, and hence potentially disease-promoting mutations, result in reduced evolutionary fitness and thus are selected against during evolution. Homologous genes derive from a common ancestor gene, while orthologous genes diverge after a speciation event in two different species; paralogous genes occur within a single species and diverge after a duplication event. Unlike orthologous genes, a paralogous gene evolves new function(s), making the distinction between the roles of orthologous and paralogous genes in disease critical for estimating disease risk using

molecular conservation³⁷. We have identified both IDH1 and IDH2 orthologs in representative genomes from all major eukaryotic supergroups and built a maximum-likelihood phylogenetic tree (Supplementary Fig. S6) from their multiple sequence alignment (Supplementary Fig. S7). Satisfactorily, we found that position corresponding to R132 in the human IDH1 protein is absolutely invariant, not only in orthologous sequences, but in all IDH homologs (Supplementary Fig. S6), which is consistent with deleterious effects of its substitution. Similar to R132, both A307 and L352 are also invariant residues in all IDH1 and IDH2 orthologs and all other IDH1 homologs with uncertain evolutionary history from all major eukaryotic supergroups (Fig. 2A and Supplementary Fig. S7). Because no substitutions in these positions occurred since the last eukaryotic common ancestor, any changes in these positions were predicted to be disease-promoting. While position V178 is not invariable among all homologs, the only allowable substitutions are V178I (occasionally found in both IDH1 and IDH2) and V178C (occasionally found only in IDH2) (Fig. 2A and Supplementary Fig. S7). No V178A substitution was ever detected in any IDH homologs, including the most distant ones, and might be cancer-promoting. We therefore tested the activity of these mutations using an enzymatic activity assay.

Effect of IDH1 mutations on enzyme activity—To test the function of IDH1 and the effect of different variants on IDH1 functions, we constructed plasmids with wildtype (WT) *IDH1* and mutant *IDH1* genes (pcDNA3.1-IDH1-A307S; pcDNA3.1-IDH1-R132H; pcDNA3.1-IDH1-V178A; and pcDNA3.1-IDH1-L352P). We tested the enzymatic activity of the WT and each IDH1 variant by analysis of isocitrate dehydrogenase activity that directly tests NADPH production. We found that R132H and L352P mutations significantly attenuated net NADPH production of IDH1 (Fig. 2B), while A307S and V178 mutations had no significant effect. In the context of other R132 IDH1 studies, attenuation of net NADPH production by the R132H variant enzyme implies that production of 2HG in the oncogenic reaction consumes NADPH. These results suggest that R132H is a point mutation that disables or attenuates some enzymatic activity of IDH1.

Placing IDH1 within a functional pathway context—As previously mentioned, certain variants of IDH1 are known to produce the oncometabolite 2HG,^{38,39} which showed inhibitory effects on 2OG-dependent enzymes, with the histone demethylases (KDM) most sensitive to inhibition⁴⁰. There are two classes of KDMs: 2OG-dependent and FAD-dependent. The biochemical function of both classes of KDMs is to demethylate specific lysine residues in histones, leading to regulation of gene expression⁴¹. KDMs may also regulate gene expression via demethylation of other residues on histones⁴². Based on this information and our discovery of mutually exclusive mutational patterns between certain histone demethylases and methyl transferases, we proceeded to ask if mutations in *IDH1* share a mutually exclusive pattern with 2OG-dependent enzymes in this lung SQCC population. We found that mutations in 2OG-dependent KDMs are mutually exclusive with *IDH1* (Fig. 3), suggesting that mutations in either *IDH1* or the 2OG-dependent KDMs lead to a common inhibition of histone demethylation. The mutually exclusive mutational pattern involving *IDH1* is statistically significant ($P=0.018$ based on the MEGSA²¹ method). This mutual exclusion is a novel observation in lung SQCC, which has not previously been reported. More than 35% of AppKY patients have mutations in 2OG-dependent protein

demethylases, the vast majority of them in KDMs (Supplementary Fig. S8). Furthermore, when all lysine demethylases are included in the analyses, only one FAD-dependent, *KDM1A*, is found to be mutated in one case. These data suggest that *IDH1* mutations may regulate gene expression via inhibition of 2OG-dependent KDMs. We further evaluated the mutations in the KDMs to see if they had functional consequences and found mutations possibly affecting a variety of specific regions in each of the different KDMs. The mutations in the KDMs are not localized to a specific region, are highly dispersive across each gene, and functionally affect protein-protein interactions, post-translational modification sites, and metal-binding (Supplementary Table S14), suggesting a general loss-of-function. This loss-of-function interpretation is further strengthened by the fact that *IDH1* mutations responsible for the production of 2HG, which is inhibitory to KDMs⁴⁰, are mutually exclusive with mutations in the above mentioned KDMs (Fig. 3). The mutational patterns observed between *IDH1* and KDMs suggest that restoring the KDM 2HG-inhibited function in cases with certain *IDH1* mutations may prevent cancer signaling through IDH1⁴³.

Analysis of alterations in *PCMTD1* and associated pathways

Localization of *PCMTD1* mutations—*PCMTD1* has an N-terminal canonical iso-aspartate methyl transferase (PCMT) domain, which in another protein has been shown to methylate iso-aspartate and aspartate residues on proteins including histone H4, and suggests a role in protein repair or turnover^{44,45}. *PCMTD1*'s C-terminal domain is not well characterized, and the cellular function(s) of the gene-product are not known. In the AppKY dataset, mutations in *PCMTD1* were always observed in the C-terminus coding region of the protein and never in the N-terminus region. These results are similar to other cancer studies including pancreatic cancer, melanoma, aggressive rhabdomyosarcoma and others (Supplementary Fig. 5B, Table 2 and Supplementary Table S15). Therefore, the C-terminus coding region of *PCMTD1* appears to be a mutation hotspot.

Smoking signature in *TP53* gene and possible relationship to *PCMTD1*

mutations—A recent report indicates that lysine methyltransferases (KMTs), *KMT2A* and *KMT2D*, are upregulated by gain-of-function *TP53* mutations (mutations in the DNA binding domain)⁴⁶. *PCMTD1* is also a methyltransferase (MT). As mentioned earlier, isoaspartate residues of *TP53* have been shown to be methylated, and this in turn has been shown to regulate levels of *TP53* as well as its function during DNA damage⁴⁷. *CUL5*, a *PCMTD1* interacting protein is recruited to target the *TP53* protein for proteasomal degradation⁴⁸. We explored the connections between *PCMTD1* and *TP53*, the most frequently mutated gene in the AppKY dataset (69%). *TP53* mutations in this cohort showed a strong signature for a smoking-associated mutational pattern, with frequent mutations in the protein regions 157–159 and 192–193⁴⁹. We also found that the mutations within the smoking signature, specifically the 157–159 region frequently co-occur with mutations in *PCMTD1* (Supplementary Table S16).

Discussion

From our analyses and other studies, there is growing evidence that numerous pathways converge on protein modification enzymes, including MTs and protein demethylases, that

function via direct protein modification, and in the regulation of gene expression via chromatin modification. Therefore, regulation of protein MTs and demethylases affects the methylation status of histones and other substrates such as signaling proteins⁵⁰. For example, mutations in PI3K/AKT signaling regulate H3K4 methylation through KDM5A⁵⁰, and PIK3CA and AKT phosphorylate KDMs and KMTs, which alters their functions and renders them oncogenic^{50,51}. Thus, these methyltransferases and demethylases may be promising targets in cancer therapy.

The observation of a smoking-associated mutational signature in *TP53* is not surprising⁵² given the high rate of smoking in AppKY, and this signature appears to frequently co-occur with mutations in *PCMTD1*. We hypothesize that PCMTD1 could function as a regulator of TP53, although further study will be needed to examine this hypothesis. In the AppKY population, concentrations of arsenic, chromium and nickel are higher than the US national levels⁵³. The toxicity of carcinogenic metals has been shown to be mediated by altering histone methylation via 2OG-dependent enzymes^{54,55}. In addition to the known link to tobacco exposure, we hypothesize that environmental exposures relevant to AppKY may be contributing to the development of this (R)-2-hydroxyglutarate-specific cancer mechanism in our cohort. This could help explain the *IDH1* and 2OG-dependent KDMs mutually exclusive pattern seen only in the AppKY cohort.

This study is the first characterization of the genomic alterations in lung SQCC from AppKY residents. Our data shares several findings with the TCGA, namely high rates of *TP53*, *NOTCH1*, *PTEN* and *PI3KCA*, the complexity of genomic patterns, and well-recognized pathways upregulated in SQCC lung cancer. However, the AppKY SQCC has a specific genetic signature characterized by an increased number of *IDH1* and *PCMTD1* mutations, as compared to the TCGA. The findings in this study have important mechanistic implications for how SQCC lung cancers develop in AppKY residents and provide insights into treatment. The 10% potentially actionable mutations/SCNAs observed in our AppKY cohort (based on FDA-approved drugs) coupled with 65% of subjects with high or intermediate mutation burden indicates that a majority of these patients have potential molecular targets for treatment (Supplementary Table S13) including *ERBB2* amplification with FDA approved monoclonal antibodies and tyrosine kinase inhibitors; *PDGFRA*, and *TSC2* where targeted agents are approved in other tumor types; as well as other mutations with targeted therapies under active investigation (*HRAS*, *KRAS*, *PTEN*, *NOTCH1*, *NF1*, *BRAF*). The current study adds to the body of literature that supports drug development based on mutations in lung SQCC and highlights genomic population differences that are relevant. By utilizing therapies specific to actionable mutations that are common in our AppKY population, we can provide a more personalized approach through directed drug discovery targeting highly mutated genes, such as *IDH1* and *PCMTD1*.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors wish to thank Drs. Jill M. Kolesar and Andrew N. Lane for helpful comments, the MCC Research Communications Office for its assistance in manuscript editing, and Dr. Youngwook Kim for providing the somatic mutation and copy number variation data from their paper³⁰.

Grant Support

T. Murali, H.N.B. Moseley and C. Wang were supported by NCI grant R21 CA205778. I.B. Zhulin was supported by National Institute of General Medical Sciences grant R01 GM072285. H.L. Weiss, J. Liu, C. Wang and S.M. Arnold were supported by National Institute of Health grant UL1 TR001998. This work was also supported by NCI grant P30 CA177558 which supports the Biostatistics and Bioinformatics SRF, the Biospecimen Procurement and Translation Pathology SRF, and the Cancer Research Informatics SRF of the University of Kentucky MCC.

References

- Howlader N et al. SEER Cancer Statistics Review, 1975–2012, <<https://seer.cancer.gov/index.html>> (2015).
- Friedell GH, Linville LH & Hullet S Cancer control in rural Appalachia Cancer 83, 1868–1871, doi: 10.1002/(SICI)1097-0142(19981015)83:8+<1868::AID-CNCR40>3.0.CO;2-G (1998).
- The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. Nature 489, 519–525, doi:10.1038/nature11404 (2012). [PubMed: 22960745]
- Singh A et al. Dysfunctional KEAP1-NRF2 interaction in non-small-cell lung cancer. PLoS Med 3, e420, doi:10.1371/journal.pmed.0030420 (2006). [PubMed: 17020408]
- Andujar P et al. p16INK4A inactivation mechanisms in non-small-cell lung cancer patients occupationally exposed to asbestos. Lung Cancer 67, 23–30, doi:10.1016/j.lungcan.2009.03.018 (2010). [PubMed: 19375815]
- Jin G et al. PTEN mutations and relationship to EGFR, ERBB2, KRAS, and TP53 mutations in non-small cell lung cancers. Lung Cancer 69, 279–283, doi:10.1016/j.lungcan.2009.11.012 (2010). [PubMed: 20018398]
- Engelman JA Targeting PI3K signalling in cancer: opportunities, challenges and limitations. Nat Rev Cancer 9, 550–562, doi:10.1038/nrc2664 (2009). [PubMed: 19629070]
- Ramos AH et al. Amplification of chromosomal segment 4q12 in non-small cell lung cancer. Cancer Biol Ther 8, 2042–2050 (2009). [PubMed: 19755855]
- Weiss J et al. Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. Sci Transl Med 2, 62ra93, doi:10.1126/scitranslmed.3001451 (2010).
- Tonon G et al. High-resolution genomic profiles of human lung cancer. Proc Natl Acad Sci U S A 102, 9625–9630, doi:10.1073/pnas.0504126102 (2005). [PubMed: 15983384]
- Martin M Cutadapt removes adapter sequences from high-throughput sequencing reads. 17, doi: 10.14806/ej.17.1.200 (2011).
- Li H & Durbin R Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England) 26, 589–595, doi:10.1093/bioinformatics/btp698 (2010).
- McKenna A et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297–1303, doi:10.1101/gr.107524.110 (2010). [PubMed: 20644199]
- Cibulskis K et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 31, 213–219, doi:10.1038/nbt.2514 (2013). [PubMed: 23396013]
- Ramos AH et al. Oncotator: cancer variant annotation tool. Hum Mutat 36, E2423–2429, doi: 10.1002/humu.22771 (2015). [PubMed: 25703262]
- Lawrence MS et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 499, 214–218, doi:10.1038/nature12213 (2013). [PubMed: 23770567]

17. Sathirapongsasuti JF et al. Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics (Oxford, England)* 27, 2648–2654, doi: 10.1093/bioinformatics/btr462 (2011).
18. Seshan Venkatraman E., A. O. DNACopy: DNA copy number data analysis. R package version 1.44.0.
19. 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* 12, R41, doi:10.1186/gb-2011-12-4-r41 (2011). [PubMed: 21527027]
20. Chakravarty D et al. OncoKB: A Precision Oncology Knowledge Base. *JCO precision oncology* 2017, doi:10.1200/PO.17.00011 (2017).
21. Hua X et al. MEGSA: A Powerful and Flexible Framework for Analyzing Mutual Exclusivity of Tumor Mutations. *Am J Hum Genet* 98, 442–455, doi:10.1016/j.ajhg.2015.12.021 (2016). [PubMed: 26899600]
22. Larkin MA et al. Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)* 23, 2947–2948, doi:10.1093/bioinformatics/btm404 (2007).
23. Tamura K, Stecher G, Peterson D, Filipski A & Kumar S MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30, 2725–2729, doi:10.1093/molbev/mst197 (2013). [PubMed: 24132122]
24. Letunic I & Bork P Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic acids research* 39, W475–478, doi:10.1093/nar/gkr201 (2011). [PubMed: 21470960]
25. Jin G et al. 2-hydroxyglutarate production, but not dominant negative function, is conferred by glioma-derived NADP-dependent isocitrate dehydrogenase mutations. *PLoS One* 6, e16812, doi: 10.1371/journal.pone.0016812 (2011). [PubMed: 21326614]
26. Gray KA, Yates B, Seal RL, Wright MW & Bruford EA Genenames.org: the HGNC resources in 2015. *Nucleic acids research* 43, D1079–1085, doi:10.1093/nar/gku1071 (2015). [PubMed: 25361968]
27. Orchard S et al. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic acids research* 42, D358–363, doi:10.1093/nar/gkt1115 (2014). [PubMed: 24234451]
28. Stark C et al. BioGRID: a general repository for interaction datasets. *Nucleic acids research* 34, D535–539, doi:10.1093/nar/gkj109 (2006). [PubMed: 16381927]
29. Futreal PA et al. A census of human cancer genes. *Nat Rev Cancer* 4, 177–183, doi:10.1038/nrc1299 (2004). [PubMed: 14993899]
30. Kim Y et al. Integrative and Comparative Genomic Analysis of Lung Squamous Cell Carcinomas in East Asian Patients. *Journal of Clinical Oncology* 32, 121–128, doi:10.1200/JCO.2013.50.8556 (2014). [PubMed: 24323028]
31. Campbell JD et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nature genetics* 48, 607–616, doi:10.1038/ng.3564 (2016). [PubMed: 27158780]
32. Li C et al. Whole Exome Sequencing Identifies Frequent Somatic Mutations in Cell-Cell Adhesion Genes in Chinese Patients with Lung Squamous Cell Carcinoma. *Scientific reports* 5, 14237, doi: 10.1038/srep14237 (2015). [PubMed: 26503331]
33. Yan H et al. IDH1 and IDH2 mutations in gliomas. *The New England journal of medicine* 360, 765–773, doi:10.1056/NEJMoa0808710 (2009). [PubMed: 19228619]
34. Molenaar RJ et al. Clinical and biological implications of ancestral and non-ancestral IDH1 and IDH2 mutations in myeloid neoplasms. *Leukemia* 29, 2134–2142, doi:10.1038/leu.2015.91 (2015). [PubMed: 25836588]
35. Dang L et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* 462, 739–744, doi:10.1038/nature08617 (2009). [PubMed: 19935646]
36. Lu C et al. IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* 483, 474–478, doi:10.1038/nature10860 (2012). [PubMed: 22343901]
37. Adebali O, Reznik AO, Ory DS & Zhulin IB Establishing the precise evolutionary history of a gene improves prediction of disease-causing missense mutations. *Genetics in medicine : official*

- journal of the American College of Medical Genetics 18, 1029–1036, doi:10.1038/gim.2015.208 (2016). [PubMed: 26890452]
38. Cairns RA & Mak TW Oncogenic isocitrate dehydrogenase mutations: mechanisms, models, and clinical opportunities. *Cancer Discov* 3, 730–741, doi:10.1158/2159-8290.CD-13-0083 (2013). [PubMed: 23796461]
 39. Ward PS et al. Identification of additional IDH mutations associated with oncometabolite R(-)-2-hydroxyglutarate production. *Oncogene* 31, 2491–2498, doi:10.1038/onc.2011.416 (2012). [PubMed: 21996744]
 40. Joberty G et al. Interrogating the Druggability of the 2-Oxoglutarate-Dependent Dioxygenase Target Class by Chemical Proteomics. *ACS Chem Biol* 11, 2002–2010, doi:10.1021/acschembio.6b00080 (2016). [PubMed: 27197014]
 41. Labbe RM, Holowatyj A & Yang ZQ Histone lysine demethylase (KDM) subfamily 4: structures, functions and therapeutic potential. *Am J Transl Res* 6, 1–15 (2013). [PubMed: 24349617]
 42. Walport LJ et al. Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. *Nature communications* 7, 11974, doi:10.1038/ncomms11974 (2016).
 43. Mondesir J, Willekens C, Touat M & de Botton S IDH1 and IDH2 mutations as novel therapeutic targets: current perspectives. *Journal of blood medicine* 7, 171–180, doi:10.2147/JBM.S70716 (2016). [PubMed: 27621679]
 44. Biterge B, Richter F, Mittler G & Schneider R Methylation of histone H4 at aspartate 24 by protein L-isoaspartate O-methyltransferase (PCMT1) links histone modifications with protein homeostasis. *Sci Rep* 4, 6674, doi:10.1038/srep06674 (2014). [PubMed: 25327473]
 45. McFadden PN & Clarke S Methylation at D-aspartyl residues in erythrocytes: possible step in the repair of aged membrane proteins. *Proc Natl Acad Sci U S A* 79, 2460–2464 (1982). [PubMed: 6123997]
 46. Zhu J et al. Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. *Nature* 525, 206–211, doi:10.1038/nature15251 (2015). [PubMed: 26331536]
 47. Lee JC et al. Protein L-isoaspartyl methyltransferase regulates p53 activity. *Nature communications* 3, 927, doi:10.1038/ncomms1933 (2012).
 48. Okumura F, Joo-Okumura A, Nakatsukasa K & Kamura T The role of cullin 5-containing ubiquitin ligases. *Cell division* 11, 1, doi:10.1186/s13008-016-0016-3 (2016). [PubMed: 27030794]
 49. Halvorsen AR et al. TP53 Mutation Spectrum in Smokers and Never Smoking Lung Cancer Patients. *Frontiers in genetics* 7, 85, doi:10.3389/fgene.2016.00085 (2016). [PubMed: 27242894]
 50. Hamamoto R, Saloura V & Nakamura Y Critical roles of non-histone protein lysine methylation in human tumorigenesis. *Nat Rev Cancer* 15, 110–124, doi:10.1038/nrc3884 (2015). [PubMed: 25614009]
 51. Xu K et al. EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. *Science* 338, 1465–1469, doi:10.1126/science.1227604 (2012). [PubMed: 23239736]
 52. Schoenberg NE, Huang B, Seshadri S & Tucker TC Trends in cigarette smoking and obesity in Appalachian Kentucky. *Southern medical journal* 108, 170–177, doi:10.14423/SMJ.000000000000245 (2015). [PubMed: 25772051]
 53. Johnson N et al. Concentrations of arsenic, chromium, and nickel in toenail samples from Appalachian Kentucky residents. *Journal of environmental pathology, toxicology and oncology : official organ of the International Society for Environmental Toxicology and Cancer* 30, 213–223 (2011).
 54. Arita A et al. The effect of exposure to carcinogenic metals on histone tail modifications and gene expression in human subjects. *J Trace Elem Med Biol* 26, 174–178, doi:10.1016/j.jtemb.2012.03.012 (2012). [PubMed: 22633395]
 55. Chervona Y, Arita A & Costa M Carcinogenic metals and the epigenome: understanding the effect of nickel, arsenic, and chromium. *Metallomics* 4, 619–627, doi:10.1039/c2mt20033c (2012). [PubMed: 22473328]

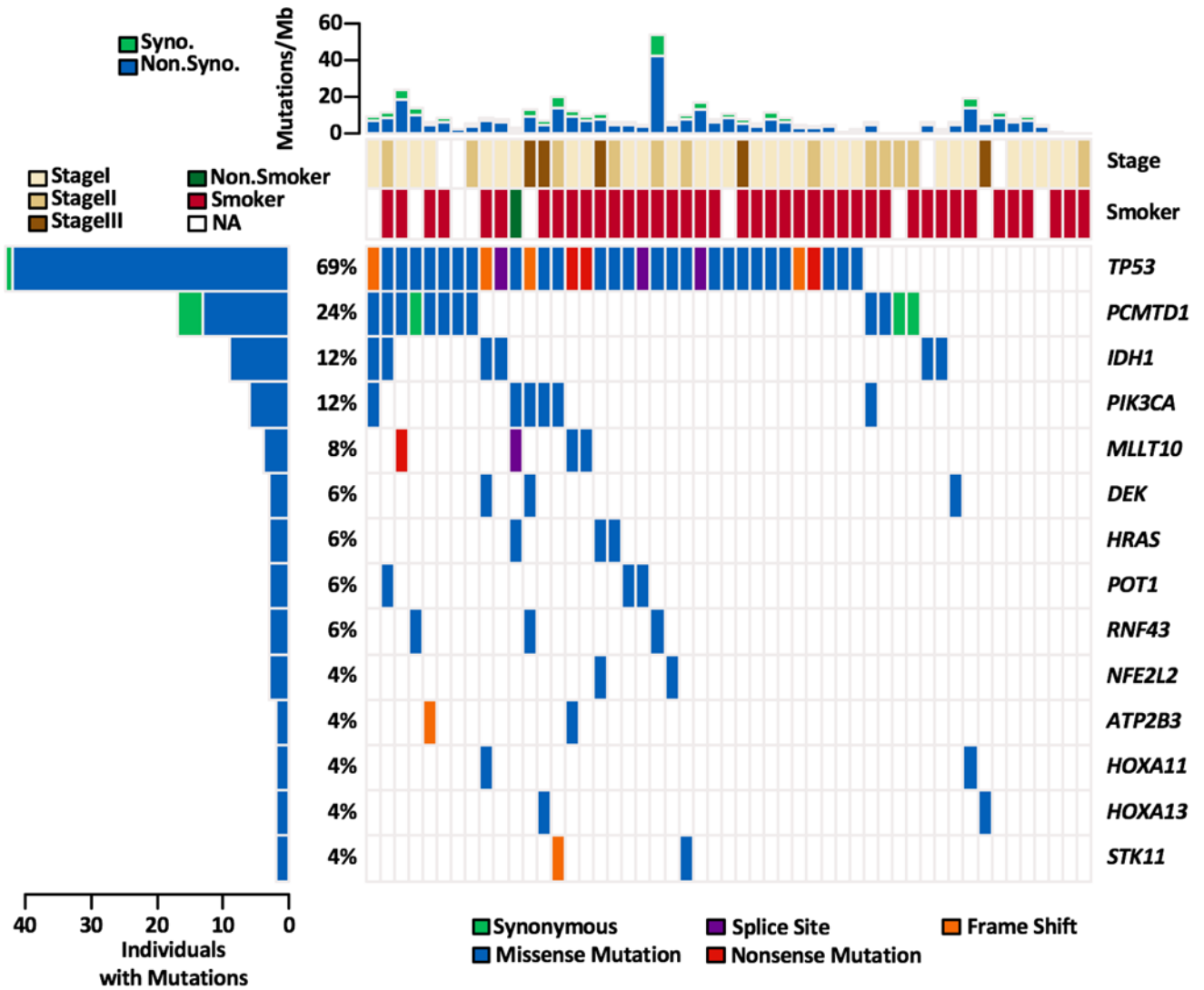


Figure 1. Significantly mutated genes in lung SQCC.

Significantly mutated genes (FDR<0.2) from whole-exome sequencing of 51 samples from Appalachian Kentucky patients. The number and percentage of samples with mutations in each gene are shown on the left. Samples are displayed as columns, with the overall number of mutations, smoking status, and tumor stage plotted at the top.

A

	132	178	307	352
Human IDH1	I I G R H A Y	G G G V A M G	E A E A A H G	N K E L A F F
Frog IDH1	I I G R H A Y	C G G V A L G	E A E A A H G	N Q E L K N F
Fish IDH1	I I G R H A H	T G G V A M G	E S E A A H G	N A E L K T F
Nematode IDH1	I I G R H A H	G P G V S L S	E A E A A H G	N S A L E T F
Worm IDH1	V I G R H A H	S G G V A L G	E A E A A H G	N D A L A R F
Lancelet IDH1	V I G R H A F	G G G V A L G	E S E A A H G	N E E L K T F
Human IDH2	T I G R H A H	A G G V G M G	E A E A A H G	N Q D L I R F
Frog IDH2	T I G R H A H	A G G V G M G	E A E A A H G	N Q D L I N F
Fish IDH2	T I G R H A F	A G G C G M G	E A E A A H G	N D D L I K F
Nematode IDH2	T I G R H A F	S G G V G L A	E A E A A H G	N E A L K T F
Worm IDH2	I I G R H A H	S G G C G M G	E S E A A H G	N K E L L K F
Lancelet IDH2	V I G R H A H	G G G C G M G	E A E A A H G	N Q D L V K F

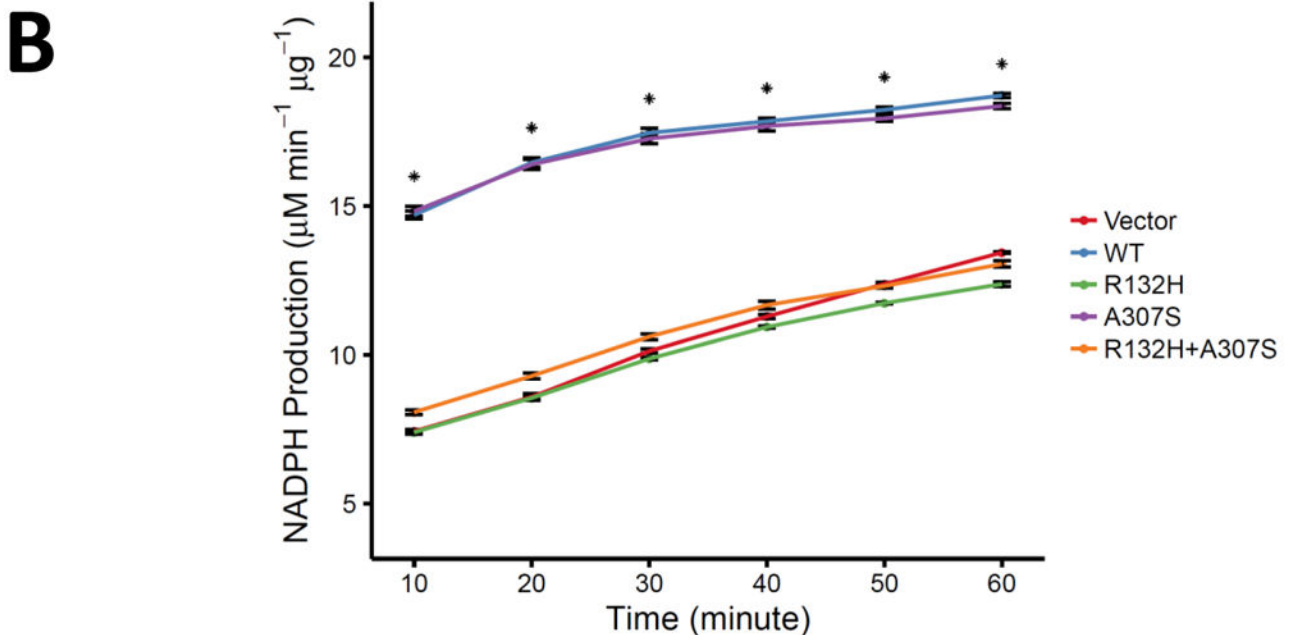


Figure 2. Functional analysis of IDH1 variants.

(A) Segments of multiple sequence alignment for representative IDH1 (upper set) and IDH2 (lower set) orthologs, showing conservation of Arg132, Val178, Ala207, and Leu352.

Numbers are provided for a human IDH1 protein. A complete alignment and sequence accession numbers are shown in Supplementary Fig. S7. Positions 132, 178, 307, and 352 are marked and highlighted in yellow, whereas substitutions in these positions are highlighted in blue. For all other positions, residues that are identical to those in the human IDH1 are highlighted in gray. Human, *Homo sapiens*; Frog, *Xenopus tropicalis*; Fish, *Takifugu rubripes*; Nematode, *Caenorhabditis elegans*; Worm, *Saccoglossus kowalevskii*;

Lancelet, *Branchiostoma floridae*. **(B)** Effect of IDH1 variants on enzyme activity. Left: effect of R132H and A307S mutants; Right: effect of V178A and L352P mutants. The two-sample t-test was performed to compare each IDH1 mutant versus the wild type and the Bonferroni correction was used for multiple comparison adjustment. • Statistically significant reductions of NADPH production comparing IDH1 R132H versus wild type; ♦ Statistically significant reductions of NADPH production comparing IDH1 L352P versus wild type.

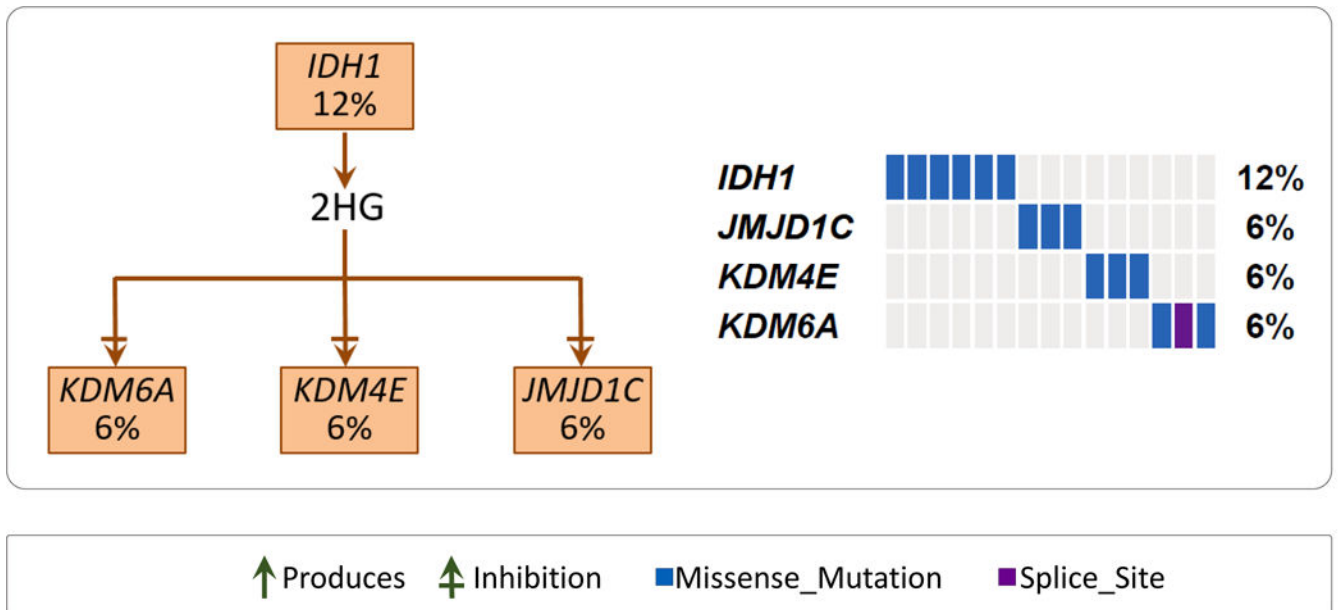


Figure 3. IDH1 mutations and IDH1 associated pathway analysis.

Variant IDH1 may produce the oncometabolite 2HG that inhibits 2OG-dependent dioxygenases; the 2OG-dependent dioxygenases are highly sensitive to inhibition by 2HG. Mutations in *IDH1* and 2OG dependent enzymes are mutually exclusive. The number and percentage of samples with mutations in each gene are shown on the left. Samples are displayed as columns.

Table 1.

Somatic alteration rate comparison between AppKY and TCGA of Lung SQCC. The comparison focuses on genes that were identified as significantly mutated based on the MutSigCV analysis in at least one of the two cohorts.

Hugo Symbol*	AppKY (%)	TCGA (%)	p-value [§]	q-value [¥]
<i>IDH1</i> ⊗	11.80%	1.10%	0.002	0.039
<i>PCMTD1</i> ⊗	17.60%	3.90%	0.002	0.045
<i>DEK</i>	5.90%	0.00%	0.011	0.200
<i>NFE2L2</i> ⊕	3.90%	15.20%	0.032	0.584
<i>CDKN2A</i> 鋼	3.90%	14.60%	0.050	0.830
<i>HOXA11</i>	3.90%	0.00%	0.049	0.830
<i>TP53</i> ⊕	68.60%	80.90%	0.082	1.000
<i>PTEN</i> 鋼	5.90%	7.90%	0.770	1.000
<i>PIK3CA</i> ⊕	11.80%	15.70%	0.655	1.000
<i>KEAP1</i> 鋼	9.80%	12.40%	0.806	1.000
<i>KMT2D</i> 鋼	9.80%	19.70%	0.142	1.000
<i>HLA-A</i> 鋼	7.80%	3.40%	0.236	1.000
<i>NOTCH1</i> 鋼	11.80%	8.40%	0.424	1.000
<i>RBI</i> 鋼	2.00%	6.70%	0.307	1.000
<i>RNF43</i>	5.90%	1.70%	0.126	1.000
<i>MLLT10</i>	7.80%	3.90%	0.269	1.000
<i>STK11</i>	3.90%	1.70%	0.309	1.000
<i>POT1</i>	5.90%	2.20%	0.186	1.000
<i>ATP2B3</i>	3.90%	2.20%	0.617	1.000
<i>HRAS</i>	5.90%	2.80%	0.381	1.000
<i>HOXA13</i>	3.90%	0.60%	0.125	1.000

* ⊗: significantly mutated in AppKY only; ⊕: significantly mutated in TCGA only; ⊕: significantly mutated in both cohorts

[§]The p-value was based on the Fisher's exact test to compare percentages of samples that had somatic alterations (somatic mutations or SCNAs) in the two cohorts.

[¥]The q-value was based on the Benjamini–Hochberg procedure. Genes with significant differences (FDR<0.2) in the alteration rate are shown in bold.

Table 2.

PCMTD1 mutations. The *PCMTD1* mutations reported in the literature are in the C-terminal SOCS Box. *PCMTD1* mutations in cancers are rarely found in the PCMT domain. The vast majority of mutations (except 1 case in TCGA Lung SQCC and 1 case in Glioblastoma) occur in the SOCS Box.

Study PMID	Cancer	PCMT (1–239)	SOCS Box (240–356)			% of cases
			BC (~16)	Spacer (~82)	Cul5 Box (~15)	
22960745	Lung SQCC	Yes	No	Yes	Yes	4%
24793135	Aggressive Rhabdomyosarcoma	No	No	Yes	No	65%
22622578	Melanoma	No	Yes	Yes	Yes	28%
22610119	Prostrate	No	No	No	Yes	1%
24816255	Gastric Carcinoma	No	No	Yes	No	7%
25855536	Pancreatic Cancer	No	Yes	Yes	No	7%
24120142	Glioblastoma	Yes	No	Yes	No	1%
AppKY	Lung SQCC	No	No	Yes	Yes	18%